

STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF THE KRAB-DOMAIN ASSOCIATED PROTEIN 1

Thèse N° 7154

Présentée le 18 octobre 2019

à la Faculté des sciences de la vie

Unité du Prof. Dal Peraro

Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Giulia FONTI

Acceptée sur proposition du jury

Prof. D. Trono, président du jury

Prof. M. Dal Peraro, directeur de thèse

Prof. V. Zoete, rapporteur

Prof. B. Luisi, rapporteur

Prof. P. Barth, rapporteur

2019

Acknowledgments

I would like to sincerely thank my thesis supervisor, Prof. Matteo Dal Peraro, for welcoming me within his team of scientists and for supporting my research during these years.

A special thanks goes to all the LBM members for the great support, the nice discussions and the friendly environment. In particular, thank you Alessio, Chan and Deniz for all the crazy and nice moments we spent together, all the laughs and the support you show me especially in the difficult moments. During these years we have established a friendship that will last in the future, for which I am deeply grateful.

I would also like to thank all the amazing people working with us here in EPFL. Among them I want to sincerely thank Julia Prébandier for the outstanding and excellent work but especially for her fairness and friendship.

Finally, I want to thank my family for unconditionally supporting and for believing in me. I would like to dedicate this thesis to my mother, who has always encouraging me through my studies with love and patience. I want also to thank my partner for helping me become a better version of myself through tremendous support and patience.

Content

CHAPTER 1 INTRODUCTION.....	11
1.1 GENE REGULATION.....	11
1.2 KRAB ZINC FINGER TRANSCRIPTION FACTORS.....	13
1.3 KAP1: A MASTER REGULATOR OF THE GENOME.....	16
1.4 INTEGRATIVE MODELLING TO SOLVE FLEXIBLE LARGE COMPLEXES.....	24
1.5 VIRTUAL SCREENING TO ACCELERATE THE FINDING OF HIGH AFFINITY BINDERS.....	26
1.6 THESIS OBJECTIVES.....	28
1.6.1 Aim 1: KAP1 structural and functional characterization.....	28
1.6.2 Aim 2: Discovery of small-molecule inhibitors for KAP1 Bromodomain.....	28
CHAPTER 2 METHODS.....	30
2.1 EXPRESSION AND PURIFICATION OF RECOMBINANT PROTEINS.....	30
2.1.1 Cloning.....	30
2.2 EXPRESSION, SOLUBILITY AND MINI-PURIFICATION TESTS.....	31
2.3 LARGE SCALE PURIFICATION.....	32
2.4 SIZE-EXCLUSION CHROMATOGRAPHY COUPLED TO MULTI-ANGLE LIGHT SCATTERING.....	34
2.4.1 Theoretical basis.....	34
2.4.2 Practical application.....	36
2.5 CIRCULAR DICHROISM (CD) SPECTRA.....	37
2.5.1 Theoretical basis.....	37
2.5.2 Practical application.....	38
2.6 ANALYTICAL ULTRACENTRIFUGATION (AUC).....	38
2.6.1 Theoretical basis.....	38
2.6.2 Practical application.....	39
2.7 HIGH RESOLUTION TECHNIQUES.....	40
2.7.1 Crystallization trials.....	40
2.7.2 Electron Microscopy experiments.....	41
2.8 SMALL-ANGLE X-RAY SCATTERING DATA COLLECTION AND ANALYSIS.....	42
2.8.1 Theoretical basis.....	42
2.8.2 Practical application.....	48
2.9 MOLECULAR MODELING AND FLEXIBLE SAXS DATA FITTING.....	49
2.9.1 Theoretical basis.....	49
2.9.2 Practical application.....	53
2.10 KAP1 AND THE BINDING PARTNERS.....	54
2.10.1 Isothermal titration calorimetry (ITC).....	54
2.10.2 Single molecule experiments.....	55
2.11 IN VITRO SUMOYLATION ASSAY.....	58
2.12 VIRTUAL SCREENING.....	58
2.12.1 Theoretical basis.....	58
2.12.2 Practical application.....	62
2.13 DIFFERENTIAL SCANNING FLUORIMETRY.....	63
CHAPTER 3 BIOPHYSICAL AND STRUCTURAL CHARACTERIZATION OF KAP1.....	65
3.1 EXPRESSION AND PURIFICATION OF RECOMBINANT PROTEINS.....	66
3.1.1 The RBCC domain.....	66
3.1.2 KAP1 FL constructs (Δ KAP1, KAP1 FL).....	67
3.2 BIOPHYSICAL CHARACTERIZATION.....	69
3.2.1 KAP1 is a homodimer in solution.....	69
3.2.2 KAP1 is well folded and show an α -helical structure.....	70
3.3 CRYSTALLIZATION TRIALS.....	73
3.3.1 The RBCC domain formed crystals but they diffracted at low resolution.....	73
3.3.2 RBCC-T4 Lysozyme.....	73

3.3.3	<i>New constructs for RBCC</i>	75
3.3.4	<i>KAP1 small binders selection for crystallization trials</i>	77
3.4	SMALL-ANGLE X-RAY SCATTERING	79
3.4.1	<i>Data collection and preliminary analysis</i>	79
3.5	NEGATIVE STAINING AND CRYO-ELECTRON MICROSCOPY.....	84
3.6	MOLECULAR MODELING AND FLEXIBLE SAXS DATA FITTING	85
3.7	FUNCTIONAL STUDIES: KAP1 ARCHITECTURE ALLOWS RING-DEPENDENT AUTOSUMOYLATION ...	94
3.8	BINDING PARTNERS: KAP1 ASYMMETRY IS FUNCTIONAL FOR RECRUITING HP1 α	97
3.9	CONCLUSIONS	105
CHAPTER 4 DISCOVERING SMALL MOLECULE COMPOUNDS TARGETING THE KAP1 BROMODOMAIN		110
4.1	BROMODOMAINS: STRUCTURE AND FUNCTIONS.....	110
4.2	BROMODOMAINS INHIBITORS	113
4.2.1	<i>Triazolothienodiazepine and Benzodiazepine templates</i>	115
4.2.2	<i>Isoxazole-azepine templates</i>	116
4.2.3	<i>Dimethyl Isoxazole templates</i>	117
4.2.4	<i>Quinoline, Quinazoline and Quinazolone templates</i>	119
4.2.5	<i>Other templates</i>	120
4.2.6	<i>Azobenzene templates</i>	120
4.2.7	<i>Benzoimidazole templates</i>	121
4.2.8	<i>Other templates</i>	122
4.3	KAP1 PHD-BROMODOMAIN MODULE.....	123
4.3.1	<i>KAP1 PHD-Br domain: function and structure</i>	123
4.4	VIRTUAL SCREENING: PROTEIN AND LIGANDS PREPARATION	126
4.5	EXPERIMENTAL VALIDATION	137
4.6	CONCLUSIONS	143
CHAPTER 5 CONCLUSIONS AND FUTURE PERSPECTIVES		146
CHAPTER 6 APPENDIX		154
6.1	APPENDIX 1: GELS GALLERY.....	154
6.2	APPENDIX 2: CRYSTAL GALLERY	155
6.3	APPENDIX 3: CRYO EM GALLERY	157
6.4	APPENDIX 4: <i>AB INITIO</i> DAMMIN MODELS.....	158
6.5	APPENDIX 5: C651A MUTANT.....	162
6.6	APPENDIX 6: UNCROPPED WESTERN BLOT IMAGES RELATED TO FIG. 41	163
6.7	APPENDIX 7: 18 COMMON INHIBITORS FOUND IN THE OPEN AND CLOSE CONFORMATIONS	164
6.8	APPENDIX 8: POSITIVE COMPOUNDS RETRIEVED FROM THERMAL SHIFT ASSAY.....	165
6.9	APPENDIX 9: TABLE 4-SAXS DATA.....	166
6.10	REFERENCES AND BIBLIOGRAPHY	167

List of figures

Figure 1 Schematic representation of proteins involved in gene transcription.....	12
Figure 2 KRAB ZFP _s families.....	14
Figure 3 KRAB-ZFP _s KAP1 repressor complex.....	15
Figure 4 TRIM family members classification.....	16
Figure 5 RBCC domain structures.....	19
Figure 6 KAP1 sequence architecture.....	19
Figure 7 SUMOylation and phosphorylation regulate different KAP1 function.....	21
Figure 8 HP1 structure.....	23
Figure 9 Workflow for structure-based virtual screening (SBVS).....	26
Figure 10 Detection of the scattered intensity from a biological sample.....	34
Figure 11 Schematic representation of a triple detector SEC-MALS setting.....	34
Figure 12 Circular Dichroism spectra of representative secondary structures.....	36
Figure 13 Schematic diagram of the sample cell used in AUC experiment.....	37
Figure 14 Schematic representation of a SAXS experiment.....	41
Figure 15 SAXS scattering curve.....	42
Figure 16 Guinier plots of different samples.....	43
Figure 17 Pair distance distribution functions representative of different sample shapes.....	44
Figure 18 Schematic representation of typical Kratky plots.....	45
Figure 19 Summary of the different plots analyzed in this thesis obtained from SAXS data.....	45
Figure 20 Example of Dammin models.....	46
Figure 21 Flowchart of the non-linear Cartesian normal mode method for SAXS data fitting.....	51
Figure 22 Schematic representation of a thermal shift assay.....	63
Figure 23 Cloning of RBCC construct.....	65
Figure 24 RBCC mini-purification.....	66
Figure 25 KAP1 FL constructs.....	67
Figure 26 KAP1 FL constructs mini-purification test.....	68
Figure 27 Biophysical characterization of KAP1 FL constructs.....	69
Figure 28 Circular Dichroism (CD) analysis of the KAP1 constructs.....	70
Figure 29 RBCC variable temperature CD spectra.....	71
Figure 30 KAP1 FL variable temperature CD spectra.....	71
Figure 31 RBCC-T4L sequence and purification.....	73
Figure 32 Limited proteolysis experiment.....	74
Figure 33 New RBCC fragments.....	75
Figure 34 KAP1 AviTag.....	77
Figure 35 KAP1 AviTag in vitro biotinylation test.....	78
Figure 36 SAXS data collection and Guinier analysis.....	80
Figure 37 Kratky plot of the KAP1 constructs.....	81
Figure 38 RBCC and KAP1 FL P(r) function and Ab initio models.....	83
Figure 39 KAP1 FL TEM gallery.....	84
Figure 40 Overall architecture of the RBCC domain.....	85
Figure 41 Flowchart of the NOLB NMA for SAXS data fitting.....	86
Figure 42 Flexible SAXS data fitting of KAP1 FL models.....	87
Figure 43 Native asymmetry of KAP1.....	88
Figure 44 Asymmetric domain arrangement in KAP1 FL.....	89
Figure 45 Analysis of the KAP1 FL models in the first density level.....	90
Figure 46 Comparison of P(r) functions.....	91
Figure 47 Integrative modelling analysis of Δ KAP1.....	92
Figure 48 KAP1 constructs for the autoSUMOylation activity.....	94
Figure 49 Role of the RING domain in the auto-SUMOylation of KAP1.....	96
Figure 50 HP1 α FL construct A. HP1 α FL cartoon model.....	98
Figure 51 KAP1 dimer binds only one HP1 dimer.....	99
Figure 52 SAXS data analysis of KAP1-HP1 complex.....	100
Figure 53 KAP1 stabilizes HP1 α -chromatin interactions.....	103
Figure 54 Bromodomains classification.....	110

Figure 55 Bromodomain structure.....	111
Figure 56 Bromodomains interaction with acetylated lysine.....	111
Figure 57 Bromodomains inhibitors.....	113
Figure 58 Triazolothienodiazepine templates.....	115
Figure 59 Isoxazole-azepine templates.....	116
Figure 60 Isoxazole derivates.....	117
Figure 61 Quinoline, Quinazoline, Quinazolone derivates.....	118
Figure 62 Triazolophthalazine templates.....	119
Figure 63 Azobenzene templates.....	120
Figure 64 Benzoimidazole templates.....	121
Figure 65 KAP1 and BPTG PHD finger domains comparison.....	124
Figure 66 KAP1 and PCAF Bromodomains comparison.....	125
Figure 67 ZA loop of KAP1 compared with other Bromodomains.....	126
Figure 68 Hydrophobicity analysis.....	128
Figure 69 KAP1 αZ helix.....	129
Figure 70 Bromodomains comparison.....	130
Figure 71 Grid box preparation.....	131
Figure 72 Binding mode of KAP1 inhibitors.....	132
Figure 73 Analysis of KAP1 Bromodomain in the “close” conformation.....	133
Figure 74 Scaffold Hunter analysis of KAP1 Bromodomain in the “open” conformation.....	135
Figure 75 PHD-Br domain purification.....	137
Figure 76 Boltzmann fit of the melt curve data.....	139
Figure 77 F2031-0348 binding poses.....	141
Figure 78 Fitting of the RBCC X-ray structures to the SAXS data.....	147
Figure 79 Schematic representation of the KAP1-HP1-di-nucleosome complex.....	149
Figure 80 Schematic representation of a transcriptional repression assay.....	151
Table 1 List of constructs used in this thesis.....	31
Table 2 smTIRF values.....	103
Table 3 List of the compounds that bind KAP1 PHD-Br domain.....	138
Table 4 SAXS data.....	165

Abstract

While conventional structural biology techniques still struggle to provide high-resolution structures of large biological assemblies characterized by great flexibility, integrative modeling (IM) approaches are today emerging as a powerful strategy to define the architecture of several macromolecular complexes. IM methods aim to provide a structural description of such dynamic systems starting from high-resolution structures of the single subunits and integrating computational methods with all the available low-resolution structural data describing the shape, the volume and the arrangement of the individual components. The low-resolution data integrated in this approach can be obtained from small-angle X-ray scattering (SAXS) experiments, medium or low-resolution cryo electron microscopy maps and they can be combined with biochemical and biophysical information. Studying such complex systems using an integrative approach able to combine experimental data with computational analysis appears to be a powerful approach to understand their behavior in the cellular environment. The work presented in this thesis describes the application of a novel IM approach for the structural characterization of the full-length KRAB-domain associated protein 1 (KAP1).

Structural and functional characterization of KAP1. In mammalian cells, KAP1 plays a fundamental role in regulating chromatin state and gene expression. KAP1 interacts with multiple chromatin-remodeling enzymes inducing the deposition of post-translational modifications and regulating the spreading of heterochromatin. Despite its emerging fundamental role, KAP1 is still poorly characterized from a structural and functional standpoint. In order to shed light on the overall structural organization of the protein, we developed a novel integrative modeling approach that combines low-resolution structural biology techniques, molecular modeling, biophysical techniques and functional assays. We demonstrate that KAP1 is an elongated antiparallel dimer with a native asymmetry within its C-terminal half. This intrinsic asymmetry in solution has key functional implications for the network of interactions of KAP1. We show indeed that KAP1 structural organization has unexpected consequences on the recruitment of the heterochromatin protein 1 (HP1), which occupies only one of the two putative HP1 binding domains (HP1BD) on the KAP1 dimer. In the new emerging structural organization, the proximity of the C-terminal and the N-terminal domains revealed a novel role of the RING domain as a second catalytic site for the autoSUMOylation of the protein. In this thesis, we present for the first time the overall structural organization of full-length KAP1, we characterize its interaction with full-length HP1 and we suggest a novel and previously unexplored functional role of the RING domain.

Discovering small molecule compounds targeting the KAP1 Bromodomain. Bromodomain (BRD) modules are a class of evolutionary conserved domains found in proteins involved in the recognition of the acetylated lysine residues of the histone tails. This post-translational modification regulates a broad range of biological functions such as DNA damage response, chromatin compaction, DNA-protein interactions and signal transduction. KAP1 presents a C-terminal tandem plant homeodomain (PHD) and Bromodomain (Br).

However it does not recognize the acetyl lysine of histone H3 and H4 but it instead plays a unique function acting as an E3 SUMO ligase and promoting both the auto-SUMOylation of the protein and the SUMOylation of other substrates. The auto-SUMOylation of KAP1 is necessary for the binding to the chromatin remodelling enzymes resulting in the deposition of post-translational modifications on the chromatin and the spreading of heterochromatin. BRD inhibitors are a class of drugs that includes anti-inflammatory and anti-cancer agents currently tested in several preclinical trials against a variety of cancer types including lung, breast, prostate, pancreatic, colon, hepatocellular cancers, medulloblastoma, glioblastoma and hematological malignancies. KAP1 represents a potential attractive therapeutic target given its involvement in tumorigenesis, control of behavioral stress and Parkinson disease. However, there are currently no inhibitors able to regulate its transcriptional activity. Consequently, we decided to screen 13'000 small molecules against the structure of KAP1 Bromodomain using a structure based virtual screening approach. At first, we retrieved 272 compounds presenting estimated low binding energy for the KAP1 Bromodomain and chemical groups common to the classical BRD inhibitors. We expressed and purified the PHD-Br domain of KAP1 and we tested these compounds together with 8 known BRDs inhibitors by differential scanning fluorimetry. We were able to identify small selective binders for KAP1 C-terminal domain that could represent the first tool available to modulate KAP1 function both *in vitro* and *in vivo*.

Keywords: Post-translational modifications, transcription factors, KAP1, structural biology, SAXS, integrative modeling, normal mode analysis, SUMOylation, HP1, drug discovery, Bromodomains, anti-cancer drugs.

Riassunto

Se da un lato le tecniche canoniche di biologia strutturale faticano ancora a fornire strutture ad alta risoluzione di complessi biologici caratterizzati da grande flessibilità, altri approcci come il modellismo integrativo (IM) stanno emergendo in quanto soluzioni efficaci per la risoluzione della struttura di diversi complessi macromolecolari. Gli approcci di IM hanno come scopo quello di fornire una descrizione di questi sistemi dinamici partendo da strutture ad alta risoluzione delle singole subunità ed integrandole con metodi computazionali e con tutti i dati a bassa risoluzione che forniscono una descrizione della forma, del volume e della disposizione delle singole componenti in questi sistemi. I dati a bassa risoluzione integrati in questo tipo di metodo possono provenire da diverse fonti come diffrazione di raggi X a piccolo angolo (SAXS), mappe di crio-microscopia elettronica a media e bassa risoluzione utilizzate in combinazione a dati biofisici e biochimici. Lo studio di questi sistemi complessi tramite un approccio integrato in grado di combinare dati sperimentali ed analisi computazionali rappresenta una risorsa importante per comprenderne il comportamento nell'ambiente cellulare. Il lavoro presentato in questa tesi descrive l'applicazione di un nuovo approccio di modellistica integrativa per la caratterizzazione strutturale di KAP1.

Caratterizzazione strutturale e funzionale di KAP1. Nelle cellule di mammifero, KAP1 svolge un ruolo fondamentale per la regolazione dello stato della cromatina e per l'espressione genica. KAP1 interagisce con diversi enzimi di rimodellamento della cromatina inducendo la deposizione di modificazioni post-traduzionali e regolando la formazione di un ambiente eterocromatico. Nonostante il suo ruolo fondamentale, KAP1 è ancora scarsamente caratterizzato dal punto di vista strutturale e funzionale. Al fine di far luce sull'organizzazione strutturale della proteina, abbiamo sviluppato un nuovo approccio di modellistica integrativa che combina tecniche di biologia strutturale a bassa risoluzione, modellistica molecolare, tecniche biofisiche e saggi funzionali. Abbiamo dimostrato che KAP1 è un dimero antiparallelo allungato con un'asimmetria nativa nella sua porzione C-terminale. Questa asimmetria intrinseca della proteina in soluzione ha implicazioni funzionali fondamentali per la rete di interazioni di KAP1. Abbiamo dimostrato infatti che l'organizzazione strutturale di KAP1 ha ripercussioni inaspettate sul reclutamento della proteina 1 dell'eterocromatina (HP1), che occupa solo uno dei due domini di legame (HP1BD) su KAP1. Secondo la nuova organizzazione strutturale, la vicinanza dei domini C-terminale ed N-terminale ha rivelato un nuovo ruolo del dominio RING come secondo sito catalitico per l'autoSUMOilazione della proteina. In questa tesi, viene presentata per la prima volta l'organizzazione strutturale globale di KAP1, la sua interazione con HP1 e viene suggerito un ruolo funzionale nuovo e inesplorato per il dominio RING.

Scoperta di ligandi di piccole dimensioni per il bromo dominio di KAP1. I bromo domini (BRD) sono una classe di domini conservati nel corso dell'evoluzione presenti nelle proteine coinvolte nel riconoscimento delle lisine acetilate delle code istoniche. Questa modifica post-traduzionale regola una vasta gamma di funzioni biologiche come la risposta al danno del DNA, la compattazione della cromatina, le interazioni

DNA-proteina e la trasduzione del segnale. KAP1 presenta un dominio tandem C-terminale composto da un omeodominio (PHD) e un bromo dominio (Br). Tuttavia il bromo dominio di KAP1 non riconosce le lisine acetilate degli istoni H3 e H4 ma ha invece sviluppato una nuova funzione svolgendo l'attività di E3 SUMO ligasi e promuovendo sia l'auto-SUMOilazione di se stesso sia la SUMOilazione di altri substrati. L'auto-SUMOilazione di KAP1 è necessaria per il legame con gli enzimi di rimodellamento della cromatina con conseguente deposizione di modificazioni post-traduzionali sulla cromatina e per la formazione di un ambiente eterocromatico. Gli inibitori dei BRDs sono una classe di farmaci che include agenti antinfiammatori e antitumorali attualmente testati in numerosi studi preclinici contro una varietà di tipi di cancro tra cui tumori polmonari, mammari, prostatici, pancreatici, del colon, epatocellulari, medulloblastoma, glioblastoma e neoplasie ematologiche. KAP1 rappresenta un potenziale bersaglio terapeutico interessante dato il suo coinvolgimento nella tumorigenesi, nel controllo dello stress comportamentale e nel morbo di Parkinson. Tuttavia, attualmente non esistono inibitori in grado di regolare la sua attività trascrizionale. Di conseguenza, abbiamo deciso di schermare 13'000 piccole molecole contro la struttura di KAP1 Bromodomain usando un approccio di screening virtuale basato sulla struttura. Inizialmente, abbiamo recuperato 272 composti che presentavano una bassa energia legante stimata per il bromodomo KAP1 e gruppi chimici comuni agli inibitori della BRD classica. Abbiamo espresso e purificato il dominio PHD-Br di KAP1 e abbiamo testato questi composti insieme a 8 inibitori noti di BRD mediante fluorimetria a scansione differenziale. Siamo stati in grado di identificare piccoli leganti selettivi per il dominio C-terminale KAP1 che potrebbero rappresentare il primo strumento disponibile per modulare la funzione KAP1 sia in vitro che in vivo.

Parole chiave: modifiche post-traduzionali, fattori di trascrizione, KAP1, biologia strutturale, SAXS, modellistica integrativa, analisi di normal mode, SUMOylazione, HP1, scoperta di nuovi farmaci, bromo domini.

Chapter 1 Introduction

1.1 Gene regulation

Although every cell within an organism contains almost the same genetic material (mature immune cells excepted) and the same set of genes, only a small portion of them is expressed in each cell type at a specific moment. The molecular mechanism implemented by the cell to regulate the expression of specific genes, known as gene regulation, occurs at two interconnected levels: the first level includes the transcription factors (TFs) and the recruitment of the transcriptional machinery and the second level includes modifications of the nucleosome and chromatin structure such as the deposition of specific post-translational modifications (PTMs) that allow the transcriptional apparatus including the RNA polymerase II to bind the promoter of specific genes^{1,2}.

TFs and the recruitment of the baseline transcription apparatus. The regulation of gene expression is an extraordinary complicated process based on the interaction between proteins known as *trans*-regulatory elements and DNA sequences called *cis*-regulatory elements. The *cis*-regulatory elements include (i) the core promoter, (ii) the proximal promoter and (iii) the distal promoter. The core promoter is the element common to all the genes and represents the minimal portion of the promoter in which the transcriptional machinery including the RNA polymerase II interacts with the DNA to initiate the transcription. The core promoter includes two main elements: the initiator element (Inr) that represents the site in which the transcription begins also known as transcription start site (TSS), preceded by a region rich in adenine and thymine located 25-30 base pairs (bp) before the Inr element. Moreover, it can include binding sites for general TFs such as the TATA box bound by the TATA-binding protein (TBP, cofactor of the RNA polymerase II) and by the transcription factor II D (TFIID) and the B recognition element (BRE) bound by the transcription factor II B (TFIIB).³ The proximal promoters are located 250 bp before the start site and they include the GC box (consensus sequence: GGGCGG) a common binding site for zinc finger proteins (ZFPs) and the CAAT box (consensus sequence: GGCCAATCT).⁴

The distal promoters are regulatory DNA elements that can be many kilobases distant from the core promoters. They can either increase (enhancers) or decrease (silencers) the expression of specific genes. The portion of the DNA involved in the binding of activator or repressor proteins bends such that the distal promoters together with their binding elements can contact the proximal promoter and the transcriptional machinery⁵.

The factors needed for transcription and known as *trans*-regulatory elements can be classified in (i) general transcription factors or baseline factors that recognize the core promoter and recruit the RNA polymerase II to the target genes, (ii) activators or repressors that bind the distal promoters and regulate the efficiency with

which the baseline apparatus interact with the core promoter regulating gene expression (iii) cofactors that do not bind directly the DNA but provide a platform for the interaction between the baseline apparatus and activators or repressors and (iv) regulatory proteins that induce chromatin modifications such as chromatin remodeling enzymes⁶ (**Fig. 1A**).

PTMs and gene expression. Gene expression is accompanied by the recruitment of ATP-dependent chromatin remodeling complexes such as SWI/SNF complex necessary for the mobilization of nucleosomes that facilitate the access of the promoter regions to the transcription apparatus⁷ (**Fig. 1B**). In addition TFs and the basal transcription apparatus are involved in the recruitment of chromatin remodeling enzymes that are involved in the deposition of PTMs on the chromatin such as DNA methylation, acetylation, ubiquitination, SUMOylation contributing to transcriptional control⁸ (**Fig. 1C**). The eukaryotic genome is organized into chromatin fibers whose fundamental unit is the nucleosome composed of a core of 8 histone proteins (two copies of H2A, H2B, H3, H4) and ~146 bp of DNA wrapped around this functional core. Histones expose unstructured N-terminal tails protruding from the protein core that represent the site of the majority of the histone PTMs^{9,10,11,12}. Histone PTMs especially DNA methylation affect the chromatin structural organization by (i) adding or removing charges which leads to an increased histone mobility reducing local protein-DNA contacts in the histone core^{13,14,15}, (ii) modulating inter-nucleosome contacts and preventing the formation of high-order chromatin structures¹⁶ and (iii) binding of effector proteins that recognize and interpret specific PTMs that induce the recruitment of chromatin remodeling enzyme altering the chromatin state by the deposition of chemical modifications^{17,18,19}. The integration between effector proteins and chromatin modifications regulates several biological pathways such as transcription²⁰, replication²¹, gene silencing^{20,22}, gene expression during development²³, apoptosis²⁴, tumorigenesis²⁵, DNA damage response²⁶.

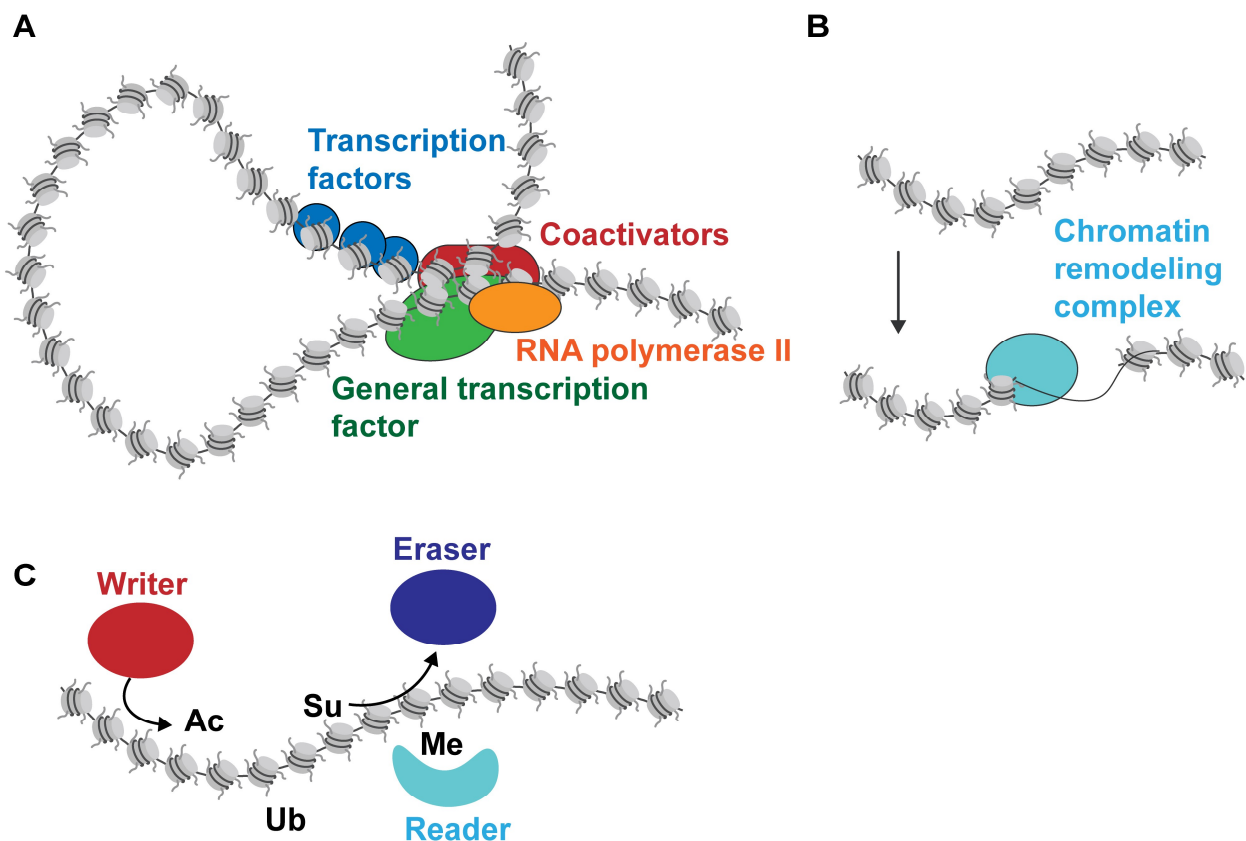


Figure 1| Schematic representation of proteins involved in gene transcription. **A.** Recruitment at the core promoter of the RNA polymerase II and general transcription factors complex, transcription factors bound to distal promoters (enhancers and silencers) and co-activators. The portion of the DNA involved in the binding of activator or repressor proteins bends such that the distal promoters together with their binding elements can contact the proximal promoter and the transcriptional machinery **B.** ATP-dependent chromatin remodeling enzymes can mobilize the nucleosome facilitating the access of the transcriptional apparatus to the DNA sequence **C.** Gene transcription is regulated by proteins modifying the histone tails with post-translational modifications. Some proteins add the PTMs (writers), some remove them (erasers) and some others read and interpret them (readers). These modifications include acetylation (Ac), methylation (Me), ubiquitination (Ub), sumoylation (Su). A typical PTM pattern found in actively transcribed genes includes H3 lysine 27 acetylation (H3K27ac), H3 lysine 4 trimethylation (H3K4me3), H3 lysine 79 dimethylation (H3K79me2) and H3 lysine 36 trimethylation (H3K36me3).

1.2 KRAB zinc finger transcription factors

KRAB (Krüppel associated box) zinc finger transcription factors (KRAB-ZFPs) constitute the largest family of TFs with over 700 members present only in tetrapod vertebrates. Their function was for a long time not well defined until recent data have shown that some members are actively involved in the silencing of transposable elements (TE_s). TE_s account for 50% of the human genome and they can perturb the genomic

environment in many ways. They can disrupt gene sequences or they can alter gene expression binding the promoter regions or enhancers elements. Cases of hemophilia A and B are associated with the insertions of TE_s into gene sequences encoding for coagulation factors. Moreover, TE_s drive the expression of oncogenes in human lymphoma and they are associated with multiple sclerosis and autoimmune diseases²⁷.

KRAB-ZFPs present an N-terminal KRAB domain that spans approximately 50-75 amino acids and it is predicted to fold into two amphipathic α -helices, organized into two boxes, the KRAB-A box and the KRAB-B box²⁸. The two boxes are always encoded by individual exons separated by introns of variable sizes allowing the generation of different splicing products. Most of the members of the KRAB ZFPs family present both the A box and the B box (known as KRAB-AB), although others present only the A box (KRAB-A) or an alternative form of the B box, named b-box (KRAB-Ab). The A box is longer and more conserved than the B box and it is involved in the binding of the KRAB-binding cofactor KAP1 while the B box enhances this binding through an unknown mechanism. In the KRAB-AB, both the A and the B boxes are necessary for the binding, while in the KRAB-A or KRAB-Ab only the A box is involved in this interaction²⁹. Moreover some members of the family present an additional N-terminal SCAN domain, also known as LeR domain (leucine rich domain) that seems to be involved in the homo or hetero dimerization of the family members³⁰. **(Fig. 2A)** The C-terminal region containing the C₂H₂ zinc finger repeats is encoded only by a single exon. The C₂H₂ zinc finger repeat is a classical protein structural motif that consists of an α -helix connected to an antiparallel β -sheet in which two zinc ions are coordinated by two cysteine and two histidine residues. The zinc finger repeats bind the DNA in a sequence specific manner, wrapping around the outside of the DNA double helix and interacting with the bases located in the major DNA groove. Many ZFPs interact with the proximal promoter GC box³¹.

The number of zinc finger repeats in the KRAB ZFPs family is variable ranging from 4 to 40, with an average number of 12. Each finger is separated by a highly conserved 7 residues linker sequence and can bind 3 to 4 DNA nucleotides via amino acids at position -1, 3 and 6 of the C₂H₂ α -helix³² (numbered with respect to the start of the α -helix). These amino acids form hydrogen bonds, water-mediated hydrogen bonds or van der Waals interactions with the DNA bases of the target sequence. **(Fig. 2B)** KRAB-ZFPs recruit the co-repressor KAP1 to specific genetic loci, via the interaction between the KAP1 N-terminal RBCC domain and the KRAB repression module³³. In turn, KAP1 recruits chromatin-modifying enzymes such as the methyltransferase SET domain bifurcated 1 (SETDB1), the nucleosome remodeling and histone deacetylation complex (NuRD), the histone deacetylase (HDAC), the nuclear co-repressor (N-CoR), the heterochromatin protein 1 (HP1), DNMT (DNA methyltransferase) inducing the deposition of PTMs (methylation and de-acetylation of the histone tails) and consequently the formation of an heterochromatin environment leading to gene silencing.^{34,35,36} **(Fig. 3)**

Many KRAB-ZFPs act as transcriptional repressors through KAP1-induced heterochromatin deposition. However not all the KRAB-ZFPs bind KAP1. On the contrary some of them, in particular those carrying the SCAN domain, interact with transcriptional activators. For example ZNF317 interacts with CARM1 (Co-activator Associated Arginine Methyltransferase 1) and KRAB-ZFPs ZFP28, ZFP273 and ZFP677 interact with the activator TP53BP1 (tumor-suppressor p53 binding protein 1).³⁷

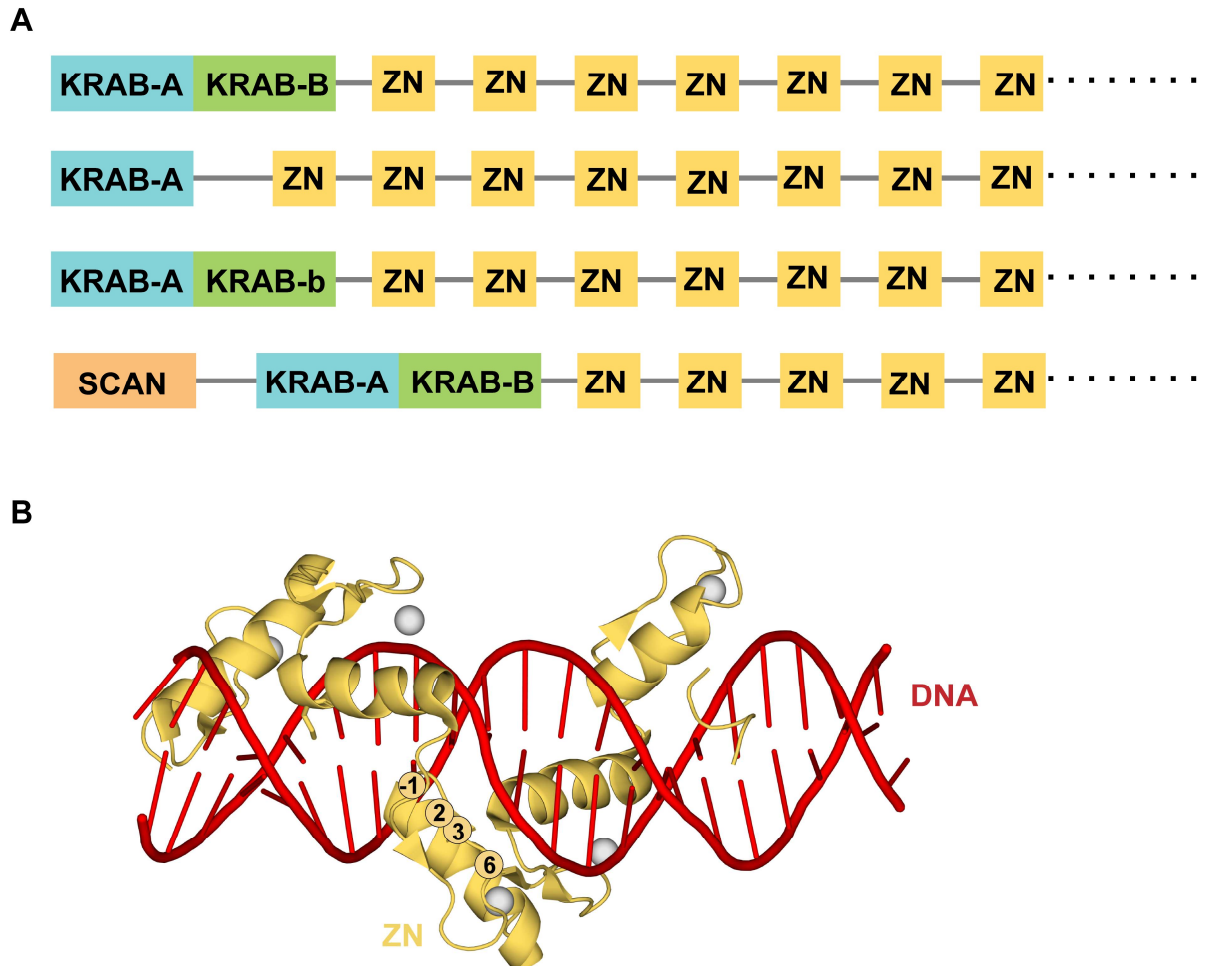


Figure 2| KRAB ZFPs families. **A.** The mammalian KRAB ZFPs can be divided into four subfamilies on the basis of the primary sequence of the amino-terminal KRAB repressor domain: those with a combination of the A and B boxes (KRAB-AB), those that contain an A-box alone (KRAB-A), those with an A box combined with a divergent B-box, called b-box (KRAB-Ab) and those that contain an addition SCAN domain. **B.** Structure of Aart zinc finger in complex with DNA (PDB: 2I13)³⁸. The zinc finger positions involved in the binding (-1, 2, 3 and 6) of the DNA major groove are highlighted with yellow spheres.

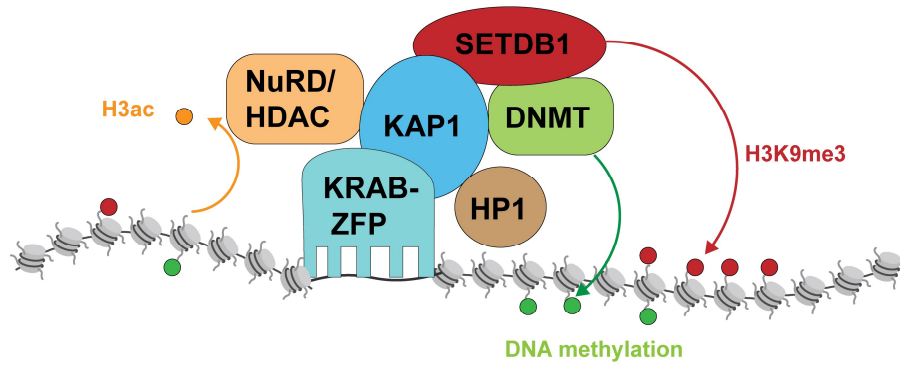


Figure 3| KRAB-ZFPs KAP1 repressor complex. KRAB ZFPs (in light blue) recruit KAP1 (in dark blue) to specific DNA loci. In turn, KAP1 acts as a scaffold for the binding of several chromatin remodeling enzymes such as the methyltransferase (SETDB1, in red) that induce the methylation of the lysine 9 of the histone H3 (H3K9me3), the DNA methyltransferase (DNMT, in green), the heterochromatin protein 1 (HP1, in brown) that binds the methylated histone tails with the N-terminal chromo domain and the nucleosome remodeling deacetylase complex (NuRD/HDAC, in orange) that removes the acetylation from histone H3 (H3ac).

1.3 KAP1: a master regulator of the genome

KAP1 (Krüppel-associated protein 1) also known as TIF1 β (Transcription Intermediary Factor 1 β) or TRIM28 (Tripartite Motif containing protein 28) was identified almost 20 years ago as the co-repressor of the KRAB-ZFPs.³⁹ KAP1 is classified among the superfamily of the tripartite motif-containing (TRIM) proteins that includes more than 60 members in humans and mice⁴⁰. The TRIM family is defined by the presence of a highly conserved N-terminal tripartite motif known as RBCC combined with a very diverse C-terminal domain including more than 10 different motifs. The composition of the C-terminal region in relation to the RBCC motif resulted in the classification of the TRIM family members into 9 sub families (C-I to C-IX)⁴¹ (**Fig. 4**).

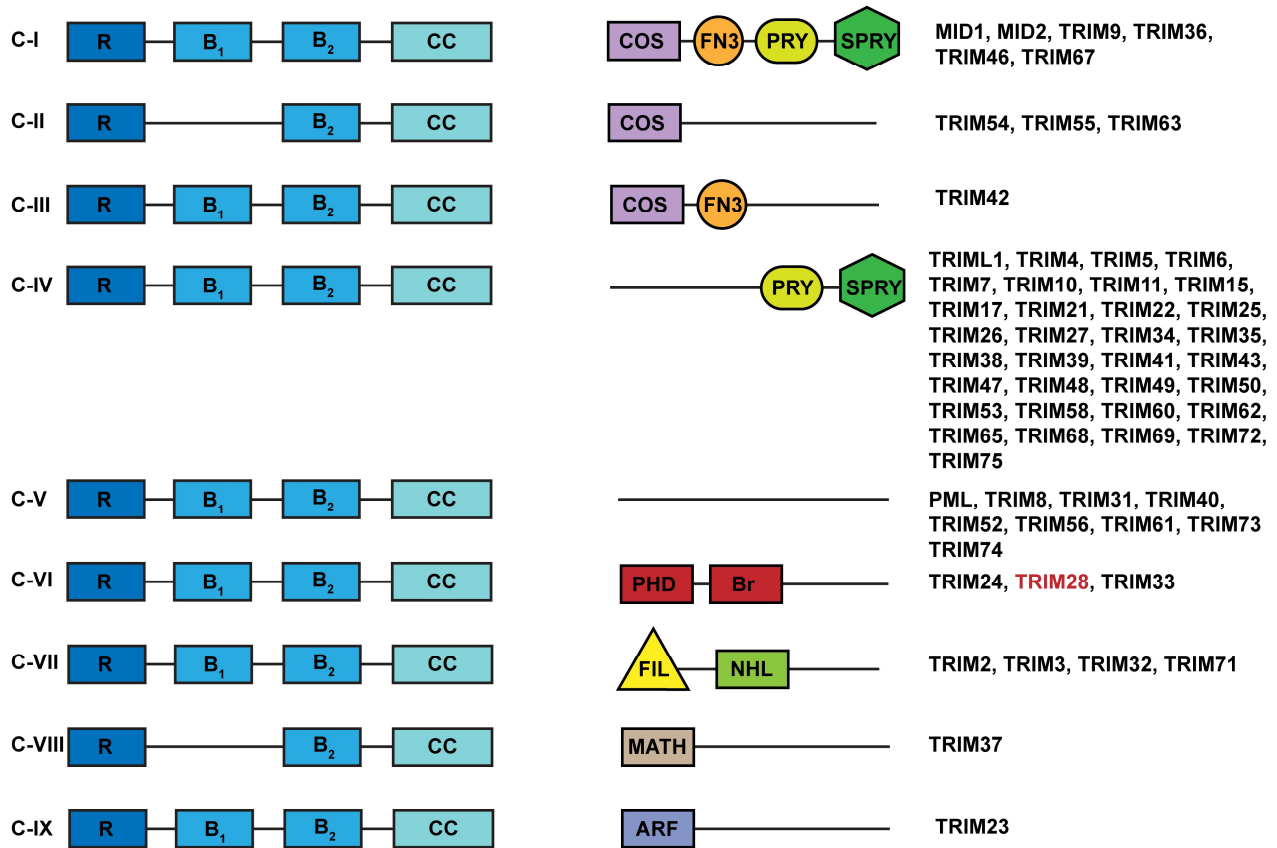


Figure 4| TRIM family members classification. The TRIM proteins share the same N-terminal RBCC domain and they are classified into 9 subfamilies based on the composition of their C-terminal domain. KAP1 or TRIM 28 belongs to the C-VI subfamily together with TRIM24 and TRIM33 characterized by the presence of a tandem plant homeodomain (PHD) and bromodomain (Br).

As a co-repressor, KAP1 is essential for the gene repression activity of the KRAB-ZFPs. The zinc finger repeats interact with specific DNA sequences while the N-terminal KRAB domain recruits KAP1 to specific DNA loci. Here, TRIM28 act as a scaffold for the recruitment of chromatin remodeling enzymes leading to the deposition of H3K9me3 mark on the nearby nucleosomes together with the deacetylation of the histone tails. Moreover it has been proved that the chromatin structural modifications induced by the recruitment of KAP1 lead to transcriptional repression of the RNA polymerase I, II, and III promoters^{28,42,43}.

Several studies associate KAP1 with tumor development and progression suggesting anti-KAP1 drugs as a cancer therapy⁴⁴. Specifically, high TRIM28 gene expression levels have been detected in prometastatic cervical cancer and in gastric cancers associated with poor prognosis. Specifically, TRIM28 significantly increases the phosphorylation of mTOR and its downstream molecule S6K1 critical for cervical cancer cell growth. These studies revealed that KAP1 plays a crucial role in the progression of cervical cancer representing a possible target for future therapies⁴⁵. KAP1 is proved to be involved in cell proliferation and differentiation, in the maintenance of genomic integrity and in the regulation of the embryonic stem cells

pluripotency.^{46,47} As a master regulator, KAP1 has been linked to fundamental cellular processes such as gene silencing, transcription regulation, DNA damage response, establishment of DNA methylation during development^{33,48,49}. Moreover, the association of KAP1 with MDM2-p53 complex abolishes the apoptosis independently from its transcriptional repression activity and promoting neoplastic transformation⁴⁴. Several studies reported a very different pattern of genomic recruitment for KAP1, independent of KRAB-ZFPs. KAP1 binds promoters of actively transcribed genes via its central region (residues 380-618, excluding the HP1 binding site). However, whether and how KAP1 regulates the expression of endogenous cellular genes is still under debate. There is no correlation between the few genes that are responsive to KAP1 knockdown and those with promoters with strong KAP1 binding sites⁵⁰.

Other studies explored the central role of KAP1 in the DNA damage response (DDR). In response to genotoxic stresses, the kinase ATM (Ataxia Telangiectasia Mutated) phosphorylates KAP1 at Ser824, repressing KAP1 SUMOylation and leading to the de-repression of a subset of genes involved in promoting cell cycle and apoptosis^{51,48}. Ser824-phosphorylated KAP1 co-localizes with several DNA repair factors and it is responsible for chromatin relaxation, a necessary step for DNA double-strand break repair³⁸. Moreover the accumulation of phosphorylated KAP1 in complex with the tumor suppressor p-53-binding protein (P53BP1) at the heterochromatic DNA break points helps to maintain activity of the ATM kinase at this sites where more time for DNA reparation is required⁵². Another important phosphorylation site is Ser473. The phosphorylation of this residue occurs in the late G1 phase mediated by the protein kinase C delta and affects the association of KAP1 with HP1⁵³.

Further studies suggested that KAP1 is involved in the release of RNA polymerase II (Pol II) paused at promoter-proximal regions. Promoter-proximal pausing of Pol II represents a genome-wide cellular mechanism to regulate gene expression. Typically, Pol II enzymes pause at around +20 to +100 nucleotides to the transcriptional start site until cellular signals activate the elongation. A number of proteins have been identified as regulators of Pol II pausing, such as P-TEFb (positive transcription elongation factor b). P-TEFb is composed of Cdk9 (cyclin-dependent kinase 9) and a regulatory cyclin subunit. Cdk9 phosphorylates the Pol II C-terminal domain, relieving the promoter-proximal pausing at many genes and promoting gene activation. P-TEFb is tightly regulated and a large fraction of P-TEFb is found reversibly associated into the snRNP (7SK small nuclear ribonucleoprotein) complex, which is composed of the 7SK small nuclear RNA, P-TEFb, kinase inhibitor Hexim1/2 and two RNA-binding proteins, Larp7 (La-related protein 7) and MePCE (methylphosphate-capping enzyme). A fraction of P-TEFb is recruited to promoter-proximal regions in a catalytically inactive state bound to the 7SK snRNP complex, until the kinase is needed for gene activation. Regarding the role of KAP1 in Pol II pause release, there are two very divergent proposed models. According to the first model, unphosphorylated KAP1 stabilizes Pol II pausing at promoter proximal regions. Upon transcriptional activation, KAP1 becomes phosphorylated at Ser824 by the ATM and DNA-PK kinases, leading to the release of paused Pol II. On the other hand, the second model suggests that KAP1 mediates the

placement of the 7SK snRNP (and therefore P-TEFb) on promoter-proximal regions of over 70% of Pol II paused genes, facilitating the release of paused Pol II. The recruitment could be mediated by the physical interaction between KAP1 and the Larp7 subunit^{54,55,56}.

Structural organization. The N-terminal domain of KAP1 known as RBCC consists of a RING (Really interesting new gene) finger domain, two B-box domains (B-box 1 and B-box 2) and a long coiled-coil domain (CCD) likely used to mediate protein-protein interactions⁴⁰. The RING domain contains a regular arrangement of cysteine and histidine residues that coordinate two zinc ions tetrahedrally in a unique “cross-brace” fold and it acts as an E3 SUMO (Small Ubiquitin Modifier) and E3 Ubiquitin ligase. The B-box domain shares the RING domain fold and may bind one or two zinc ions^{57,58}. The coiled-coil domain of KAP1 is estimated to be very long (~200 Å) and together with the B₂ is likely used to mediate protein-protein interactions specifically with the KRAB-ZFPs⁴². The structure of the RBCC domain has been recently solved by Stoll et al. (PDB: 6QAJ⁵⁹) and Lim et al. (PDB: 6H3A⁶⁰) revealing an antiparallel homodimer conformation (**Fig. 5A,B**). The RBCC domain is followed by the TIF1 signature sequence (TSS) domain, rich in phenylalanines and tryptophanes and known to regulate the transcriptional repression activity of TIF1γ⁶¹. A long loop of 200 amino acids, without any predicted structure, connects the N-terminal domain to a C-terminal tandem plant homeodomain (PHD) and bromodomain (Br)⁶². Only KAP1, TIF1α and TIF1δ include a central heterochromatin protein 1 binding domain (HP1BD). The HP1 chromo shadow domain binds KAP1 as a dimer and the chromo domain recognizes the trimethylated lysine of histone H3 (H3K9me3). The KAP1-HP1 interaction is crucial for chromatin state regulation, heterochromatin formation and gene silencing^{63,36} (**Fig. 6**).

PTMs associated with different KAP1 activities. KAP1 is a member of the TRIM C-VI subfamily, together with TRIM24 and TRIM33, characterized by the presence of a tandem plant homeodomain (PHD) and bromodomain (Br) typically involved in the recognition of acetyl lysine of histone H3 and H4^{65,66}. However, the C-terminal tandem PHD-Br module of KAP1 shows a unique function acting as an E3 SUMO ligase and promoting both the auto-SUMOylation of the protein and the SUMOylation of other substrates^{67,68}. The NMR structure of the KAP1 PHD-Br domain elucidated how the two domains cooperate as one E3 SUMO ligase unit. The auto-SUMOylation of the C-terminal PHD-Br domain is necessary for the binding of KAP1 to the chromatin remodeling enzymes (SETDB1, NuRD, HDAC, N-CoR) inducing the deposition of PTMs and consequently the spreading of heterochromatin^{34,35,36}. Six lysines have been identified as SUMOylation sites (lysine 554, 575, 676, 750, 779, 804) (**Fig. 6**) and distinct SUMOylation combinations affect differently the interaction with the histone modifiers SETDB1 and CHD3 (Mi-2 α , subunit of the NuRD complex together with HDAC). The SUMOylation level of KAP1 is regulated at two levels by the sentrin specific peptidase 1 (SENPI) involved in the deSUMOylation of KAP1, and by the phosphorylation status of the Ser824. Phosphorylation of Ser824 is crucial for the DNA damage response (DDR)⁶⁹. KAP1 phosphorylated at Ser824 is responsible for the activation of DNA damage checkpoints and for chromatin relaxation required for the accessibility of the DNA damage foci to the enzymes involved in the repair process such as BRCA1 (breast cancer type 1 susceptibility protein) and TP53BP1⁷⁰ (**Fig.7**). Three kinases are involved in this process: ATM (ataxia-telangiectasia mutated kinase), ATR (ataxia-telangiectasia Rad3-related kinase), DNA-PKcs (DNA-dependent protein kinase catalytic subunit). After the repair of the DNA damage, the phosphorylation at Ser824 is removed by the phosphatase PP4 (protein phosphatase 4) and by the phosphatase PP1 (protein phosphatase 1) that interacts with the PP1 binding motif located in the coiled coil domain of the RBCC domain^{71,49,53}. Another interesting phosphorylation site is represented by Ser473, which is important for efficient DNA repair and cell survival upon DNA damage. Contrary to KAP1 phosphorylated at Ser824 that accumulate at the damage sites forming DNA repair foci, KAP1 phosphorylated at Ser473 remains diffuse in the nucleus and regulates cell cycle progression. Chang et al. demonstrated that KAP1 is phosphorylated at Ser473 during the S phase of the cell cycle inducing the binding of HP1 γ ⁵³. The KAP1-HP1 complex promotes the progression through the S phase re-repressing the cyclin A2. Moreover, different studies demonstrated that KAP1 can be acetylated and the level of acetylation is directly modulated by HDAC10 (histone deacetylase 10). KAP1 de-acetylation promotes the release of the transcription for MITF (microphthalmia-associated transcription factor) and TRP1 and 2 (tyrosinase-related protein 1 and 2) promoting the cellular melanogenesis program⁷².

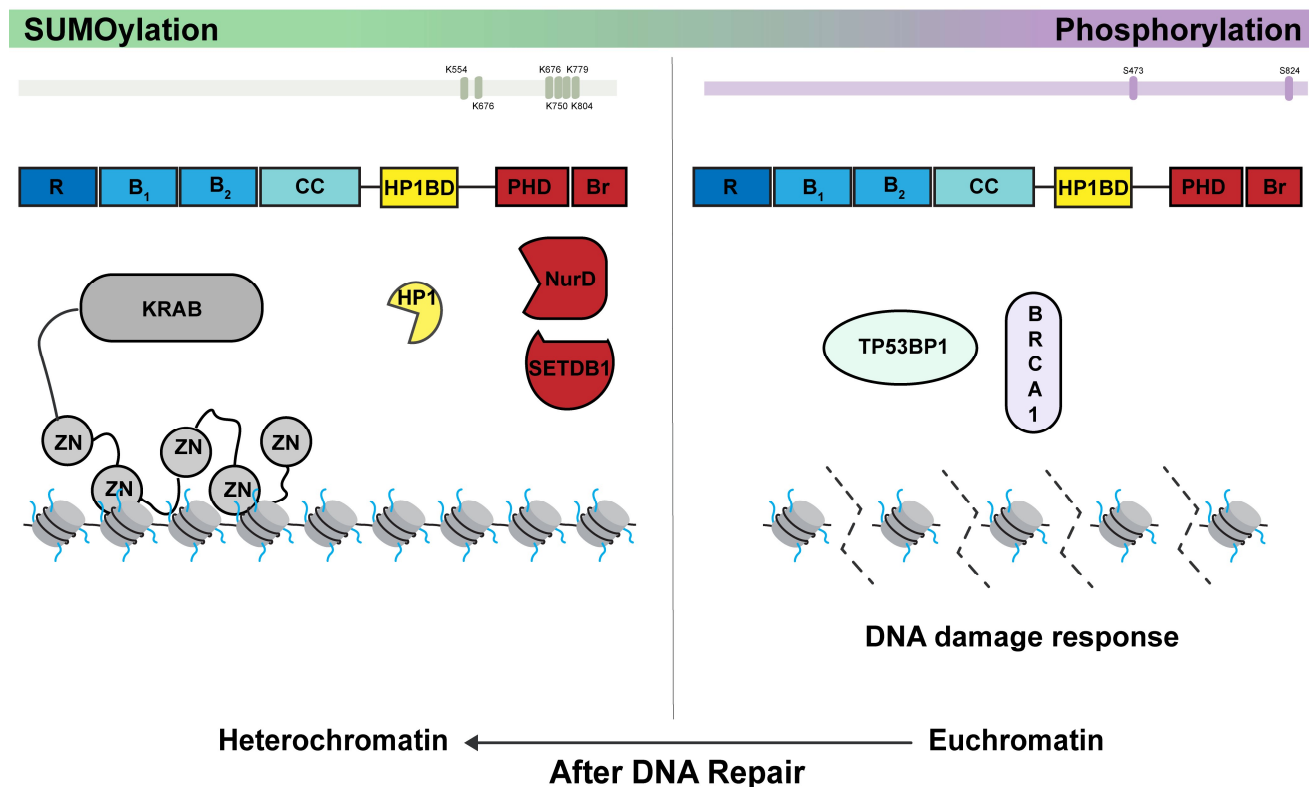


Figure 7| SUMOylation and phosphorylation regulate different KAP1 functions. KRAB ZFPs (in grey) recruit KAP1 to specific DNA loci. Upon SUMOylation of the C-terminal PHD-Br domain (in red), KAP1 recruits the chromatin remodeling enzymes such as the nucleosome remodeling deacetylase complex (NuRD, in red), the methyltransferase SETDB1 (in red) and the heterochromatin protein 1 (HP1, in yellow) that interact with the HP1BD (in yellow) located in the loop region between the N-terminal and the C-terminal domains. The RING domain, the B-box 1 (B₁), the B-box 2 (B₂) and the coiled-coil domain (CC) are shown in different shades of blue. Doing so, KAP1 induces the formation of heterochromatin leading to gene silencing (left part of the panel). Upon DNA damage (right part of the panel), KAP1 is phosphorylated at Ser824 and Ser473 and interacts with different binding partners such as the breast cancer type 1 susceptibility protein (BRCA1, in purple) and the tumor-suppressor p53 binding protein 1 (TP53BP1, in light green) relaxing the chromatin and increasing the accessibility to the chromatin repairing enzymes. After DNA repair, KAP1 is dephosphorylated and can be once again SUMOylated to regulate the gene expression.

The interaction with the heterochromatin protein 1. The N-terminal domain and the C-terminal PHD-Br domain are connected by a long loop of ~200 amino acids without any predicted structure, where the heterochromatin protein 1 (HP1) is recruited by KAP1 on a specific binding domain (HP1BD, **Fig. 6**)^{36,63}. HP1 is an evolutionarily conserved chromosomal protein that is essential for forming higher-order chromatin structures and, in humans, is present as three isoforms α , β , and γ . HP1 plays a widespread role in gene silencing, both through formation of heterochromatin and at euchromatic promoters. It is also required for the normal expression of some heterochromatic and euchromatic genes⁷³. HP1 presents an N-terminal chromo domain (chromatin organization modifier) that is connected to a slightly larger, homologous domain termed

the chromo shadow domain that is unique to the HP1 family¹⁹. While the chromo domain is monomeric, the chromo shadow domain forms a tightly associated symmetrical dimer that interacts with the binding motif PXVXL (where X is any amino acid) mediating the interaction with other proteins (**Fig. 8A**). The hinge region linking the two domains is unstructured in solution, enabling the chromo and the chromo shadow domains to move independently of each other in the intact protein⁷⁴. The methylation of histone H3 at Lys-9 (H3-metK9) is specifically recognized and bound by the HP1 chromo domain, resulting in the spreading of heterochromatin and HP1-mediated gene repression. A high local H3K9me3 density results in a prolonged HP1 residence time due to rapid rebinding and dissociation. Especially the dimeric state of the protein prolongs its chromatin residence time, and it results in accelerated chromatin association. Importantly, this behavior has been observed in the presence of the penta-peptide motif-containing interaction partners, which stabilize the HP1 dimeric state⁷⁵. The interaction with HP1 is crucial for KAP1 to regulate the chromatin state, carry out its gene silencing function and respond to DNA damage⁵³. The cryo-EM structure of HP1 bound to a H3K9me3-containing di-nucleosome shows how the HP1 dimer bridges the two nucleosomes without interacting with the linker DNA. The CSD is located in an accessible position where binding partners can attach such as KAP1 that contains a HP1BD (residues 483-497, **Fig. 6**), where a HP1 dimer can in principle bind⁷⁶ (**Fig. 8B**).

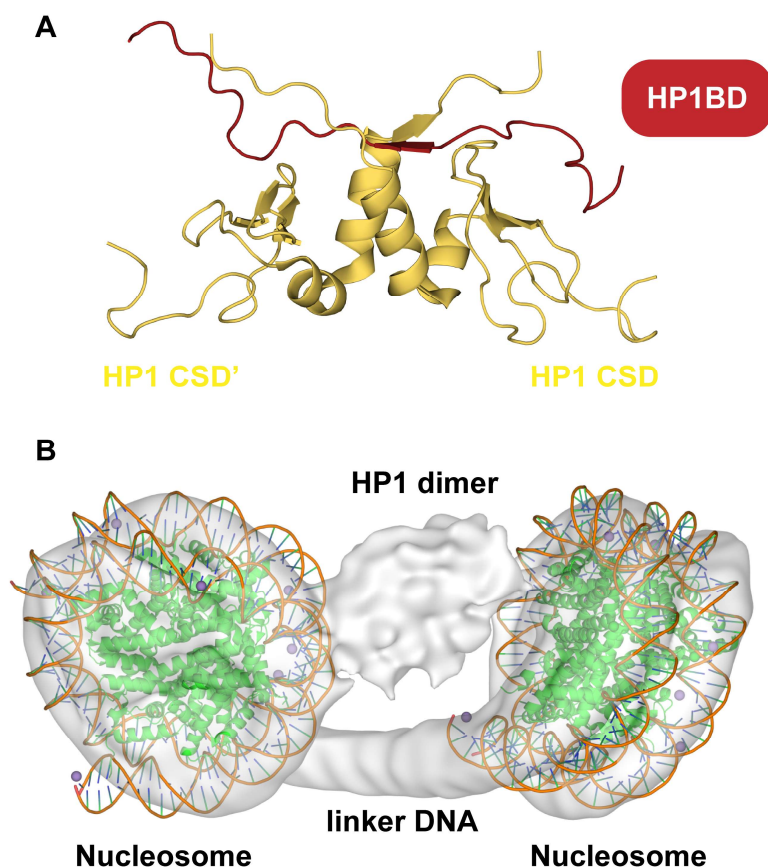


Figure 8| HP1 structures. **A.** Cartoon representation of the dimer of the HP1 CSD bound to a peptide containing the PXVXL HP1 binding motif (PDB 1S4Z⁷⁵) **B.** Cryo-EM structure of the HP1 α - di-nucleosome complex (EMDB: 6738⁷⁶) superimposed with the crystal structure of nucleosome core particle (PDB: 3LZ0⁷⁷) .

1.4 Integrative modelling to solve flexible large complexes

KAP1 is expected to be a very dynamic and flexible macromolecule presenting multiple conformational states. Thus we need to complement the classical structural biology techniques with an integrative modelling approach to combine all the available experimental and computational data with the ultimate goal of shed light on the overall structural organization of the full-length protein.

Around 35% of the human proteome does not fold into ordered and stable three-dimensional structures instead it is mainly composed of proteins with flexible regions or intrinsically disordered proteins (IDPs). The regions of flexibility are essential for the protein functions and they are involved in a vast range of physiological pathways including signal transduction, replication and transcription^{78,79}. IDPs play a crucial role in several human diseases such as neurodegenerative disorders (e.g Alzheimer's disease) and cancer^{80,81}. Despite the importance of characterizing this class of proteins, standard structural biology techniques still struggle to provide a high-resolution structures of systems of this kind, where multiple domains are connected by

disordered linkers producing natively flexible complexes capable of plastic, dynamic interactions modulated by specific environmental conditions. Moreover, the characterization of large molecular complexes is a challenging issue, determined by the chemical and conformational heterogeneity of these multi-component systems, which is exacerbated by their large molecular weight, enhanced flexibility and co-existence of multiple functional states. Studying such dynamic systems using an integrative modeling (IM) approaches able to combine experimental data and molecular modeling appears to be a powerful resource to understand their behavior in the cellular environment^{82,83}.

Experimental techniques able to provide low-resolution molecular information can support the understanding of spatial connectivity and reciprocal arrangement of individual protein subunits part of larger assemblies. The low-resolution information can be retrieved from chemical crosslinking coupled to ion mobility mass spectrometry (XL-MS)⁸⁴ experiments revealing protein contacts, mutagenesis experiments or coevolution analyses⁸⁵. Furthermore, small-angle X-ray scattering (SAXS) data, as well as low-resolution (i.e. > 5-6 Å) cryo-electron microscopy (cryo-EM) maps can be used to understand the spatial relationships of the single components, as well as their conformational states. Recent technological advances in image-processing technology and high-resolution image-capturing hardware are driving the next revolution in structural biology, enabling cryo-EM to determine the structure of several large complexes in the recent years^{86,87,87}. For example, in 2015 Campbell et al. published the cryo EM reconstruction of the 20S proteasome complex (700 kDa) of *Thermoplasma acidophilum* solved at 2.8 Å⁸⁹. Other examples are represented by the cryo-EM structures of the β-galactosidase complex (465 kDa) at 2.2 Å published in 2015 by Bartesaghi and collaborators⁹⁰ and the apoferritin (480 kDa) at 2.5 Å published in 2019 by Li and collaborators⁹¹.

However, when dealing with complexes lacking symmetry, presenting dynamic properties and/or multiple states, synergistically combining high and low-resolution experimental data with computational modeling in an IM approach has become the golden strategy to efficiently enhance the structural resolution of macromolecular assemblies. A successful example is represented by the solution of the yeast nuclear pore complex (NPC), a 50 MDa assembly involved in the selective transport of proteins and ribonucleoparticles between the nucleus and the cytoplasm. This result was achieved thanks to a novel integrative modeling platform (IMP)⁹² that allowed the combination of different experimental data such as SAXS envelopes, cryo-EM density maps, cross-linking and mutational data. These diverse data were converted into spatial restraints used to produce a scoring function specific for the NPC. Afterwards, a structural ensemble was generated to optimally fit this scoring function⁹². Other examples are the combination of solution X-ray scattering and computational modeling to explore multiple conformations of the proliferating cell nuclear antigen (PCNA) bound to ubiquitin⁹³ and the combination of negative stain EM images with integrative computational approach to solve the complex PCNA, Rad9-Hus1-Rad1 (9-1-1) and flap endonuclease 1 (FEN1) specialized in DNA repair⁹⁴.

1.5 Virtual screening to accelerate the finding of high affinity binders

KAP1 presents a C-terminal Br domain representing a potential target for drug development. BRD inhibitors were discovered in early '90s as potent anti-inflammatory and anti-cancer agent. Nowadays, many of these molecules are tested in preclinical and clinical trials against a broad range of human diseases. However, no binding compounds for KAP1 Br domain are known to date, thus we can apply molecular modelling, virtual screening together with experimental validation to overcome this problem.

Virtual screening (VS) is a computational method used to identify new molecules presenting a biological activity against a selected target. VS can be divided into two categories: ligand based virtual screening (LBVS) and structure based virtual screening (SBVS). LBVS uses the structures of known active compounds to identify new molecules displaying the same set of chemical properties (pharmacophore). On the other hand, SBVS utilizes the structure of a biological target to dock candidate compounds inside the identified binding pocket. SBVS is based in three pillars: (i) the target identification (ii) the chemical compound library preparation and (iii) the protein-ligand docking^{95,96}.

Target identification. The receptor 3D structure is normally solved by X-ray crystallography, NMR or retrieved from the Protein Data Bank (PDB). However, it can also be modeled after a homologous templates by homology modeling. SBVS begins with the identification of the active site meaning the portion of the receptor where the ligands are suppose to bind; it is generally a hydrophobic cavity on the protein surface characterized by the presence of hydrogen bond donors and acceptors. It can be visually identified looking at the crystal structure of the protein bound to a known inhibitor or can be predicted based on homologous protein structures.

Library preparation. A chemical compound library is a collection of chemical entities available for SBVS. Before docking the compounds are generally protonated at physiological pH and the tridimensional structure is generated. Currently, many public and commercial chemical libraries are available such us PubChem⁹⁷, ChEMBL⁹⁸, ZINC⁹⁹, ChemSpider¹⁰⁰, ChemBridge (<http://chembridge.com>) and Maybridge (<http://www.maybridge.com>).

Docking. Protein-Ligand Docking¹⁰¹ is the computational method that mimics the binding of a ligand to a protein. It predicts the “poses” of a ligand in the binding pocket, meaning the geometry (position, orientation, torsional angles) of a ligand within the binding site and it provides a value for the binding affinity representing the strength of the protein-ligand binding (affinity in kcal/mol). Docking algorithms are based on two key components: (i) the *searching algorithm*, to generate the poses of a ligand within the binding site and to explore its conformational space and to optimize the ligand geometry and (ii) the *scoring function*, to identify the most likely poses of a ligand in the binding pocket and to estimate for each compound the strength of the

binding (the compounds are then ranked based on this affinity value). From docking, each ligand will obtain a docking score as an estimation of the binding affinity for the target. By ranking all the compounds based on the docking scores, a small set of ligands with better score will be selected for further experimental evaluation. The hits are confirmed in a secondary screening process including functional assays and tests in cellular environment. The choice of the assay will depend on the biology of the target protein. Generally cell-based assays are used for membrane proteins, ion channels and nuclear receptors while biochemical assays are more indicated to study enzymes. In the case of the GPCR (G protein-coupled receptor), the affinity of the binding is measured using radio or fluorescently labelled ligands monitoring (i) the guanine nucleotide exchange, (ii) changes in one or more secondary metabolites such as calcium, inositol phosphate or (iii) activation of a downstream reporter gene¹⁰² (**Fig. 9**).

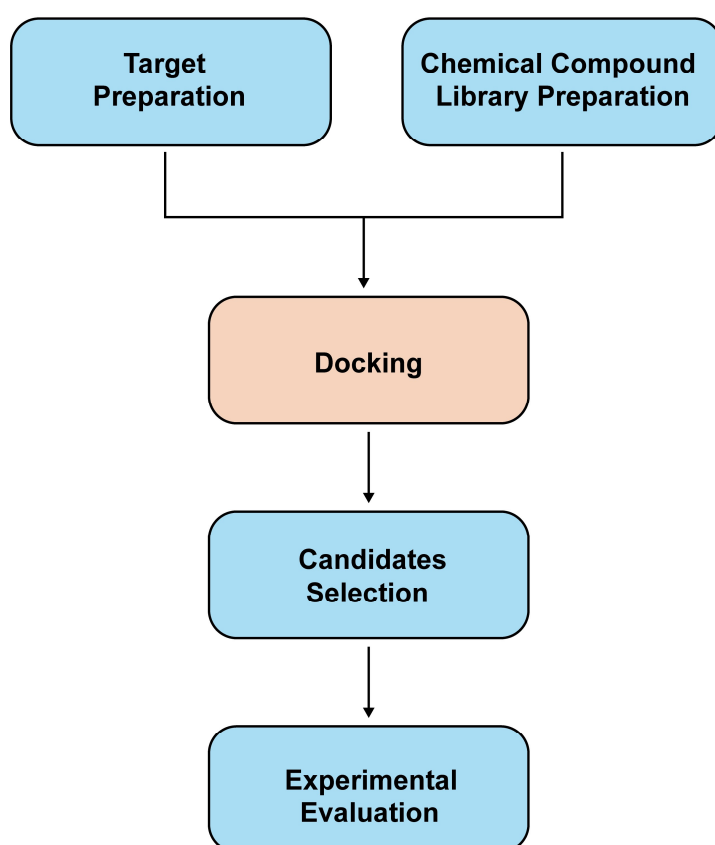


Figure 9| Workflow for structure-based virtual screening (SBVS). The structure of the target is retrieved and prepared for docking adding polar hydrogen and checking the integrity of the side chains. In the same way, the ligands are prepared with the addition of polar hydrogen and the 2D structures usually present in the chemical libraries are converted in 3D structures. The protein-ligand docking is performed and the compounds are scored and classified based on their binding energies. Subsequently, a small set of ligands presenting the better scores is selected for further experimental validation.

1.6 Thesis objectives

1.6.1 Aim 1: KAP1 structural and functional characterization

KAP1 was identified almost twenty years ago as a transcriptional co-repressor factor because of its interaction with the KRAB-ZFPs⁴². Several studies have demonstrated that KAP1 is involved in the control and in the regulation of many biological pathways and cell activities such as gene expression in mammalian cells¹⁰³, recruitment of different transcription factors and regulation of the chromatin state¹⁰⁴. KAP1 acts both as a platform for macromolecular interactions and as an E3 SUMO ligase⁶⁸. Despite its emerging fundamental role, little is known about the structure of the full-length protein and about its functional relationships with binding partners. The first aim of this thesis is to combine experimental techniques and computational methods to reveal the three-dimensional organization of full-length KAP1. To achieve this goal, we decided to combine high-resolution structural biology techniques (X-ray crystallography, Cryo EM), biochemical experiments (multi-angle laser light scattering, analytical ultracentrifugation, circular dichroism), low-resolution techniques (small angle X-ray scattering data), molecular modelling, single-molecule experiments and a novel integrative modelling protocol to treat elongated and highly flexible systems. In doing so, we also characterized the interaction of full-length KAP1 with HP1 both in vitro and in the chromatin context to have additional information about its functional activities.

1.6.2 Aim 2: Discovery of small-molecule inhibitors for KAP1 Bromodomain

Bromodomain (BRD) modules are a class of evolutionary conserved domains found in proteins involved in the recognition of the acetylated lysine residues of the H3 and H4 histone tails^{65,105}. Also KAP1 as TRIM24 and TRIM33 (**Fig. 4**) is characterized by the presence of a tandem plant homeodomain (PHD) and Bromodomain (Br). However, the C-terminal Bromodomain of KAP1 does not recognize the post-translational modifications of the histone tails but it shows a unique function acting as an E3 SUMO ligase⁶⁸. BRDs inhibitors were discovered in the early '90s as potent anti-inflammatory and anti-cancer agents. Since then, many new molecules have been developed and tested in clinical trials such as JQ1¹⁰⁶ and I-BET 762¹⁰⁷ for the treatment of acute myelogenous leukemia (AML), multiple myeloma (MM) and acute lymphoblastic leukemia (ALL) and other hematological malignancies. Regarding KAP1, no selective ligands able to bind the C-terminal Bromodomain regulating the protein transcriptional activity are currently available. Given the emerging fundamental importance of KAP1 in the regulation of gene expression and given the structural similarity with the other BRDs, we decided to start a structural-based drug discovery project with the ultimate goal of discovering novel selective compounds for KAP1 Bromodomain.

This thesis is organized as follows. First the experimental and computational techniques used in this work will be introduced and explained in **Chapter 2**. Next, I will present the biophysical characterization of KAP1 together with the SAXS data collection and flexible SAXS data fitting using a novel integrative modeling approach presenting for the first time the overall structural organization of the full-length protein (**Chapter 3**). Afterwards, I will discuss the drug discovery study to identify novel compounds targeting the KAP1 C-terminal domain and I will present the results discussing the structure and the binding mode of the most promising molecules (**Chapter 4**). Finally, conclusions and future perspectives of these studies will be presented in **Chapter 5**. The thesis ends with a series of **Appendices** showing the crystals obtained for the different constructs, the cryo EM images and the *ab initio* SAXS models.

Chapter 2 Methods

This chapter is organized as follows. First I will present the conditions to express and to purify the different constructs. Afterwards, I will discuss the methods for the biophysical characterization of KAP1 (MALS, AUC and CD). I will introduce the high-resolution techniques used in this work (crystallization trials and the cryo-EM studies). Then, I will discuss the low-resolution SAXS data collection together with all the analyzed parameters and the information retrieved from each of them. I will present the modeling of the single domains of full-length KAP1 and the flexible SAXS data fitting using a novel normal mode analysis approach (NOLB-NMA). Afterwards, I will present the experiments to characterize the function of the RING domain (SUMOylation assay) and the interaction with HP1 (single-molecule assay). I will conclude the chapter with the presentation of the virtual screening approaches (structure-based and ligand-based virtual screening) to find new small ligands for KAP1 Br domain. For each technique I will present a theoretical part and a practical application.

2.1 Expression and purification of recombinant proteins

All the constructs reported in this thesis belong to human proteins cloned inside the pETDuet1 vector (**2.1.1**). We tested the expression of each construct in four *E. coli* strains (**2.2**) and after finding the expression condition that maximized the level of soluble proteins we performed a large-scale purification in which we purified the proteins by affinity chromatography (HisTrap) and by two steps of gel filtration chromatography (HiLoad 16/600 Superdex 200 pg and Superose 6 Increase 10/300 GL) (**2.3**). Afterwards, we characterized these constructs applying biophysical techniques such as MALS (Multi-angle light scattering, **2.4**), CD (Circular Dichroism, **2.5**) and AUC (AnalYTical ultracentrifugation, **2.6**).

2.1.1 Cloning

The KAP1 and the RBCC constructs (**Table 1**) were cloned inside the first MCS (Multi Cloning Site I) of the pETDuet-1 vector (Novagen), between SACI and HINDIII restriction sites. The pETDuet-1 vector was designed for the co-expression of two target genes. It carries two MCSs (MCSI with His₆ tag and MCSII without His₆ tag). The vector presents the pBR322-derived ColE1 replicon, a LacI gene and an ampicillin resistance gene. The proteins were cloned into the MCSI preceded by a His₆ tag and a TEV (Tobacco Etch Virus) protease cleavage site. The RBCC domain was cloned in the laboratory and the presence of this construct inside the pETDuet-1vector (pETDuet1-RBCC) was confirmed by sequencing. The other DNA coding sequences were ordered from Genscript and optimized for *Escherichia coli* (*E. coli*) expression. The HP1 α FL construct was cloned inside the pET15b vector preceded by a His₆ tag and a Thrombin protease

cleavage site. The *E. coli* cells (BL21 (DE3)) expressing HP1 α FL were provided as glycerol stock by Professor Beat Fierz (Laboratory of Biophysical Chemistry of Macromolecules, LCBM, EPFL).

KAP1 constructs	Residues	Molecular Weight (kDa)
Δ KAP1	23-812	87
KAP1 FL	1-835	91
Δ Ring	128-835	79.5
PHD-Br	624-812	23.6
RBCC constructs		
RBCC	23-418	46
RBCC-T4Lysozyme	23-418+T4L	65
BBCC ₁	128-418	36.2
BBCC ₂	152-418	33.5
BCC ₃	201-418	28
BCC ₄	190-418	29.3
BBCC ₅	128-382	32
BBCC ₆	152-382	29.5
BCC ₇	201-367	22.4
BCC ₈	190-352	21.8
HP1		
HP1 α FL	1-191	22

Table 1| List of constructs used in this thesis. For each construct we report the molecular weight (kDa) and the residues cloned in the expression vector.

2.2 Expression, solubility and mini-purification tests

The expression test was performed for each construct using four *E. coli* stains (BL21 (DE3), BL21 (DE3) pLys, Rosetta (DE3), Rosetta (DE3) pLys (Promega), three growth temperatures (20°C over night, 30°C 3h, 37°C 4h) and three IPTG (IsoPropyl β -D-1-ThioGalactopyranoside) concentrations (0.1 mM, 0.5 mM, 1 mM). Each *E. coli* strain was transformed with the pETDuet1 vector carrying the different constructs, using a heat shock protocol. For each strain, 250 mL of fresh LB (Luria-Bertani) media were inoculated with 3 mL of an over night cell culture supplied with 100 μ g/mL of ampicillin (100 μ L from 100 mg/mL stock solution) and 17 μ g/mL of chloramphenicol (100 μ L from 17 mg/mL stock solution). The cells were grown till an O.D (Optical Density, 600 nm) of 0.6-0.7. 1 mL of NI sample (Non Induced sample, before the addition of IPTG) was collected and centrifuged at 4°C, 4'000 rpm, for 5 minutes. The 250 mL culture was then divided into 9 flasks of 20 mL each and the three IPTG concentrations (0.1 mM, 0.5 mM, 1 mM) were added for each temperature (20°C, 30°C, 37°C) for each *E. coli* strain. 1 mL of I sample (Induced sample, after addition of IPTG) was collected and centrifuged at 4°C, 4'000 rpm, for 5 minutes. The remaining culture (19 mL) was

centrifuged under the same condition and the pellet was frozen at -20°C. The following day, the 19 mL culture pellet was resuspended into 1 mL of Lysis Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol) with the addition of 1 tablet of cOmplete™ Protease Inhibitor Cocktail (Roche). For each strain and condition, we collected samples for total homogenate (T) and, after sonication, for pellet (P) and supernatant (SN). We applied the SN to a NiNTA resin, previously washed with 10X volume of water and 10X volume of Loading Buffer (corresponding to the Lysis Buffer). After 1h of incubation shaking at 4°C, we washed the resin with 5% of Elution Buffer (Loading Buffer + 500 mM Imidazole) and collected the flow-through (FT). We eluted the protein with 100% Elution Buffer and we collected the eluted (E) fractions.

2.3 Large scale purification

The constructs were all purified by affinity chromatography and one or two steps of gel filtration chromatography. The plasmids were transformed in Rosetta (DE3) (pETDuet-RBCC), BL21 (DE3) (pETDuet-RBCC-T4Lysozyme) and Rosetta (DE3) pLys cells (pETDuet-KAP1 constructs). Cells were grown to an optical density of 0.7-0.8 in LB media. Protein expression was induced by the addition of 1 mM IPTG and subsequent growth over night at 20°C (RBCC and KAP1 constructs) or 30°C for 4h (RBCC-T4Lysozyme). Three liters pellet were resuspended into 120 mL of Lysis Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol) with 2 mM TCEP (Tris (2-carboxyethyl) phosphine), 5 mM Imidazole and 1 tablet of cOmplete™ Protease Inhibitor Cocktail and then lysed using high pressure homogenizer (Avestin Emulsiflex C3). The resulting suspensions were centrifuged at 13'000 rpm for 35 minutes at 4°C after the addition of 0.1% P.E.I (Poly Ethylene Imine) for 5 minutes RT to precipitate nucleic acids. The supernatant was filtered through a 45 µm PVDF (Polyvinylidene difluoride) membrane filter and applied to an HisTrap HP column (GE Healthcare) previously equilibrated with Loading Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol). The proteins were eluted with a gradient over 40 column volumes of Elution Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 500 mM Imidazole). Subsequently, pure fractions were buffer exchanged into Final Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM TCEP) using a HiPrep 26/10 Desalting column (GE Healthcare) and stored over night at 4°C in Final Buffer. The His₆ tag located at the N-terminus of the protein was cut, incubating the sample over night at 4°C with a ratio 1:20 (mg/mg) of TEV protease. The sample was then concentrated into 6 mL using Amicon filters with a 3K cutoff (for the RBCC construct) or 30K cutoff (for the KAP1 constructs) and injected into a Superdex 200 10/300 GL column (GE Healthcare). The run was performed in the same buffer conditions. To guarantee the maximum purity of our samples, the KAP1 constructs were purified with an additional step of size exclusion chromatography. The samples were concentrated as previously described and applied to a Superose 6 Increase 10/300 GL column (GE Healthcare). The run was performed in the Final Buffer. Afterwards the proteins were flash frozen in liquid nitrogen and stored at - 20°C.

HP1 α FL was expressed in BL21 (DE3) cells grown at 37°C, 200 rpm, to an optical density of 0.7-0.8 in LB media. Protein expression was induced by the addition of 0.25 mM IPTG and subsequent growth at 20°C over night. The pellet was resuspended in 120 mL of Lysis Buffer (25 mM HEPES pH 7.5, 50 mM NaCl) and 1 tablet of cOmplete™ Protease Inhibitor Cocktail. The cells were lysed as previously described, centrifuged at 13'000 rpm, for 20 minutes at 4°C, filtered through a 45 μ m PVDF membrane filter and loaded into a 5 mL HisTrap HP column (GE Healthcare). The protein was eluted with a gradient over 40 column volumes of Elution Buffer (25 mM HEPES pH 7.5, 50 mM NaCl, 500 mM Imidazole). Afterwards the pure fractions were loaded into a HiTrap Q FF anion exchange chromatography column (GE Healthcare) previously equilibrated in Loading Buffer (25 mM HEPES pH 7.5, 50 mM NaCl). The salt concentration was increased from 50 mM to 1M NaCl over 20 column volumes. The protein was eluted with 160 mM NaCl and diluted with 25 mM HEPES pH 7.5 until a final salt concentration of 50 mM was reached. The protein was then incubated over night at 4°C with Thrombin (1U/mg), to cut the N-terminal His₆ tag. The protein was then purified by size exclusion chromatography (Superdex 200 10/300 GL column) equilibrated with 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP. Finally the pure fractions were concentrated using 3K cutoff Amicon filters. 40% (v/v) glycerol was added to the pure fractions and the protein was flash frozen in liquid nitrogen and stored at - 20°C.

The PHD-Br domain construct was expressed in BL21 (DE3) cells grown at 37°C to an optical density of 0.7. The expression was induced with 0.25 mM IPTG at 20°C over night. The pellet was resuspended in a minimal amount of Lysis Buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 15 mM Imidazole) supplemented with 10% (v/v) glycerol, 1 tablet of cOmplete™ Protease Inhibitor Cocktail, 5 μ L of Benzonase and 10 μ M Zinc Chloride. The cells were lysed as previously described, centrifuged at 13'000 rpm, for 30 minutes at 4°C, filtered through a 0.45 μ m filter and loaded into a 5 mL HisTrap HP column (GE Healthcare). The protein was eluted in three steps with 20%, 60% and 100% of Elution Buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 500 mM Imidazole). The elution fractions were pooled and the His₆ tag located at the N-terminus of the protein was cut, incubating the sample for 2h at room temperature and then over night at 4°C with a ratio 1:20 of TEV protease, during dialysis in the Final Buffer (10 mM HEPES pH 7.5 and 300 mM NaCl). The protein was then purified by size exclusion chromatography (Superdex 75 16/60 GL column, GE Healthcare). The run was performed in the same buffer conditions. Finally the pure fractions were stored at - 80°.

2.4 Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)

2.4.1 Theoretical basis

Multi-angle light scattering (MALS) is a technique for determining the exact molecular weight (M_w) of proteins in solution by detecting the way they scatter the light (**Fig. 10**).

$$R_\theta/KC = M_w (1 - (16\pi^2/3\lambda^2) \langle R_{rms} \rangle^2 \sin^2 (\theta/2)) \quad (1)$$

In **eq. 1** R_θ is the Rayleigh ratio, K is an optical constant, C is the sample concentration, M_w is the molar mass, λ is the wavelength, R_{rms} is the root mean square radius, θ is the scattering angle.

The R_θ value measured by the light scattering detector (MALS detector) is proportional to the ratio between the scattered intensity at an angle θ ($I_s(\theta)$) and the incident intensity (I_i) as in **eq. 2**.

$$R_\theta = I_s(\theta) / I_i \quad (2)$$

The ideal way to obtain the M_w of the sample would be to use the scattered intensity at angle θ equals to 0. In **eq 1** when $\theta = 0$ the term $(16\pi^2/3\lambda^2) \langle R_{rms} \rangle^2 \sin^2 (\theta/2)$ becomes 0 and the R_θ is directly proportional to the M_w of the sample. However the scattered intensity at angle θ equals to 0 is not experimentally measurable. Instead the scattered intensity can be measured at low angles, typically between 3° and 10°. At these angles the scattered intensity can be approximated to be equal to the one measured at 0°. It is possible to extrapolate the value of the scattered intensity at 0° from higher angle data (scattered intensity measured at many angles where $\theta > 0^\circ$). This extrapolation is frequently done in the Debye plot in which the R_θ/KC is plotted as a function of $\sin^2 (\theta/2)$. In this plot the intercept with the ordinate corresponds to the molar mass (M) of the protein and the slope at $\sin^2 (\theta/2)=0$ is a function of the molar mass (M) and the root mean square radius (R_{rms})¹⁰⁸ (**Fig. 11**).

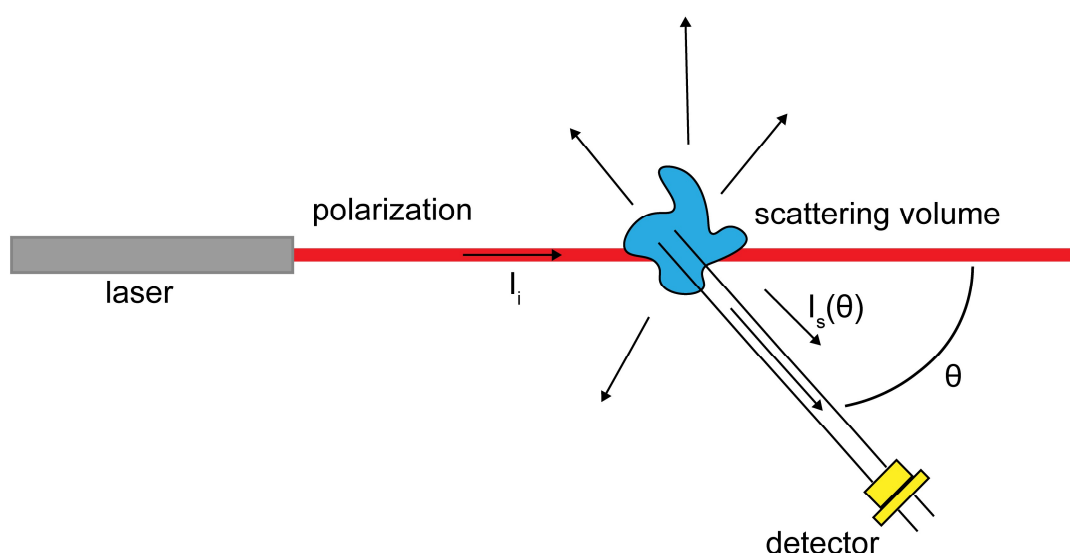


Figure 10| Detection of the scattered intensity from a biological sample. Schematic representation of the alignment between collimated beam of monochromatic light (laser) plane-polarized perpendicularly to the plane of measurement and the detector position to provide accurate measurements of the scattered intensity. The scattering volume is represented in blue, the incident light is reported as I_i and the scattered light at angle θ is reported as $I_s(\theta)$. Figure adapted from “Characterization of proteins and protein conjugates by Multi-Angle Light Scattering coupled to SEC”, Wyatt Technology Europe GmbH.

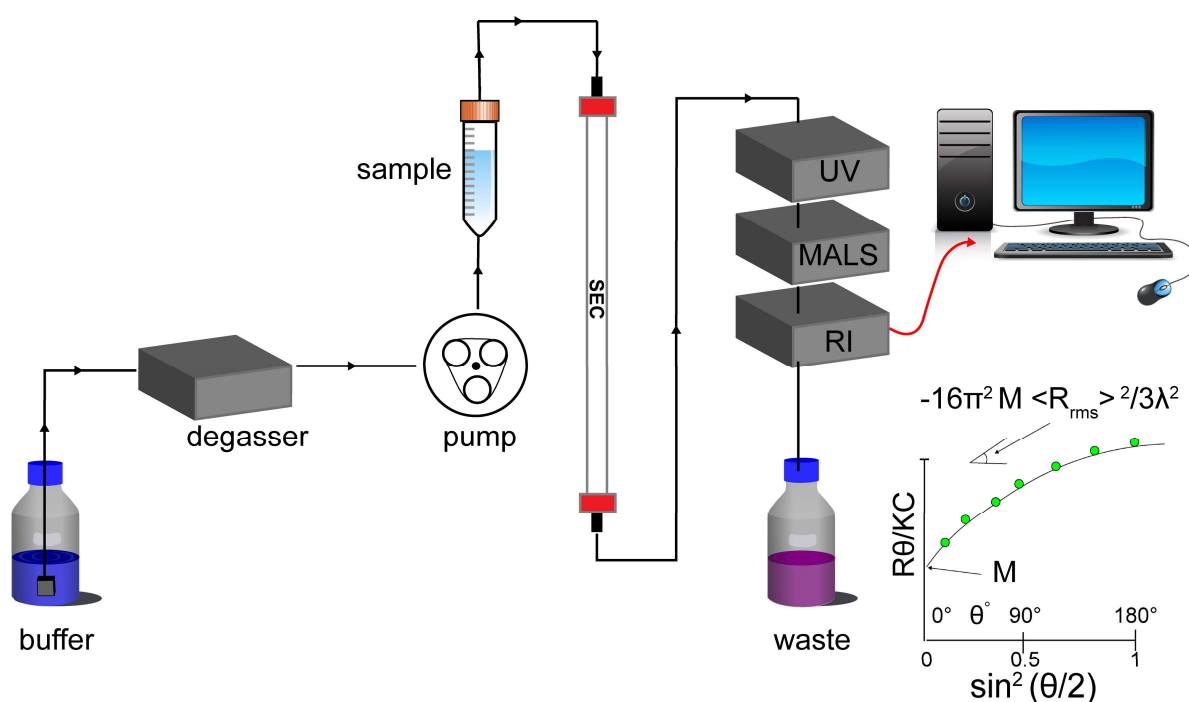


Figure 11| Schematic representation of a triple detector SEC-MALS setting. The sample passes through the size exclusion chromatography column (SEC) that separates the different species based on their molecular weight. The UV absorbance detector and the RI detector monitor the changes in protein concentration during the run while the light scattering detector (MALS detector) monitors the changes in scattering intensity at multiple scattering angles. The data acquisition is control by a software package (ASTRA). The value of the scattered intensity at angle $\theta=0^\circ$ is then

extrapolated in the Debye plot in which the R_g/KC is plotted as a function of $\sin^2(\theta/2)$. Figure adapted from “Selection of Biophysical methods for characterization of membrane proteins”¹⁰⁹.

2.4.2 Practical application

The molecular weights of the RBCC domain, KAP1 FL and Δ KAP1 constructs were determined by size exclusion chromatography coupled to a multi angle light scattering detector (SEC-MALS) (3.2, Fig. 27). The mass measurements were performed on a Dionex UltiMate3000 HPLC system equipped with a 3 angles miniDAWN TREOS static light scattering detectors (Wyatt Technology). The sample volumes of 100 μ L at a concentration of 40 μ M, were applied to a Superose 6 10/300 GL column (GE Healthcare) previously equilibrated with 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mM TCEP at a flow rate of 0.5 mL/min. Standard proteins (Ovalbumin and Aldolase) were run to confirm the accuracy of the molecular weight calculation and to compare elution volumes. Given the sensitivity of the system to the presence of external particles, the equilibration buffer was filtered with 0.1 μ m filters. To determine the sample concentration, a UV detector was used in the case of single protein measurements. Furthermore we studied the complex KAP1 FL-HP1 α FL using a refractive index (RI) detector to calculate the molecular weight of the samples. KAP1 FL at a concentration of 80 μ M was incubated with 360 μ M of HP1 α FL and 72 μ M of H3K9me3 (1-14) peptide¹ on ice for 30 minutes and then centrifuged at 4°C for 10 minutes at 13'000 rpm. A sample volume of 20 μ L was applied to a Superose 6 10/300 GL column equilibrated with 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mM TCEP at a flow rate of 0.5 mL/min. The data were analyzed using the ASTRA 6.1 software package (Wyatt technology), using the refractive index of the buffer as a baseline and the refractive index increment for protein $dn/dc = 0.185$ mL/g.

See in Chapter 3.2, Fig. 27 A| SEC-MALS analyses of the RBCC, KAP1 FL and Δ KAP1 constructs.

¹ Three methylated Histone 3 peptide bound to the Chromo domain of HP1 α FL

2.5 Circular Dichroism (CD) spectra

2.5.1 Theoretical basis

Circular Dichroism (CD) is a spectroscopic technique largely used to evaluate the conformation and the stability of proteins. CD uses a source of polarized light to probe secondary structure of proteins. This is possible since the peptide bonds are asymmetric and molecules without a plane of symmetry show the phenomenon of circular dichroism. CD spectra are normally collected between 180 nm and 250 nm due to the fact that the amides of the peptide bonds dominate CD spectra of proteins below 250 nm showing two main electronic transitions. One of these transitions is responsible for the negative peak at 222 nm characteristic of a α -helix spectrum and for the negative band at 216-218 nm typical of a β -sheet spectrum. The second electronic transition defines the positive band at 190 nm and the negative one at 208 nm found in the α -helix spectrum and the positive peak at 198 nm in the β -sheet spectrum¹¹⁰ (Fig. 12).

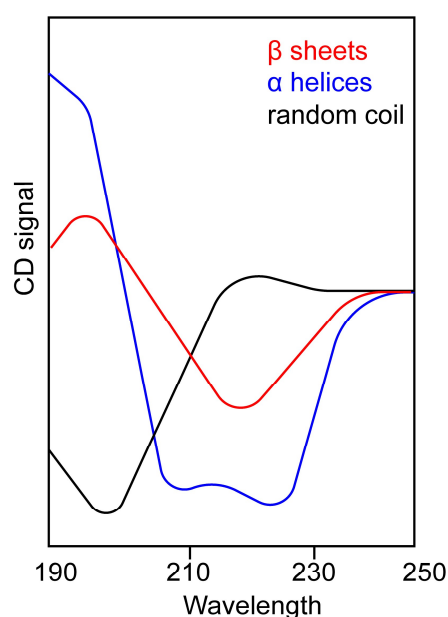


Figure 12| Circular Dichroism spectra of representative secondary structures. In blue a typical α -helix spectrum with 2 minima at 222 nm and 208 nm and a maximum around 190 nm. In red, a β -sheet spectrum with a minimum at 216-218 nm and a maximum around 198 nm. In black a random coil spectra with a negative band of great magnitude around 200 nm. Figure adapted from “The use of circular dichroism spectroscopy to study protein folding, form and function”¹¹¹

2.5.2 Practical application

Continuous scanning CD spectra were recorded from 200 to 250 nm at 4°C in a 300 μ L quartz cell (0.1 cm path length) using a Chirascan CD Spectrometer at a sample concentration ranged from 0.2 mg/mL to 0.8 mg/mL in Phosphate Buffer (1X PBS). Scans were measured from 200 to 250 nm, under continuous scanning mode with a 1 nm data pitch, a scan speed of 50 nm/min and a response time of 1 s. Variable temperature CD spectra were recorded from 200 to 250 nm at 2°C interval with a ramp rate of 1°C/min and tolerance of 0.2°C. Spectra were recorded from 20°C to 95°C in triplicates.

See in Chapter 3.2, **Fig. 28**| CD analysis of the RBCC, KAP1 FL and Δ KAP1 constructs, **Fig. 29**| RBCC variable temperature CD spectra, **Fig. 30**| KAP1 FL variable temperature CD spectrum and in Chapter 3.7, **Fig. 48D**| Δ RING variable temperature CD spectrum.

2.6 Analytical ultracentrifugation (AUC)

2.6.1 Theoretical basis

Analytical ultracentrifugation (AUC) monitors the sedimentation of macromolecules and allows their thermodynamic and hydrodynamic characterization without any interaction with a matrix or a surface (as in SEC). The analytical centrifuge is a centrifuge equipped with an optical system, a rotor and specialized centrifuge cells that together allow the measurement of the sample distribution within the cell while the rotor is spinning. The application of a centrifugal force causes the depletion of the protein sample at the meniscus of the cell and the formation of a concentration boundary that moves toward the bottom of the centrifuge cell as a function of time (**Fig. 13**).

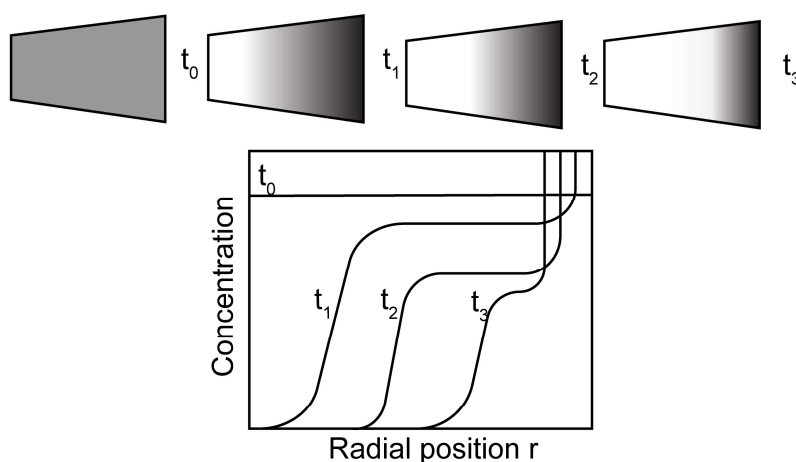


Figure 13| Schematic diagram of the sample cell used in AUC experiment. The scans diagram represents the sample concentration (measured at 280 nm) at each radial position as a function of time (t_0 to t_3). Figure adapted from

The velocity at which the bands move correspond to the sedimentation coefficient (s) of the macromolecular sample ¹¹².

$$s = v/\omega^2 r = M (1-\upsilon\rho) / N_A f \quad (3)$$

In **eq. 3** v is the velocity (change of position v as a function of time $t=d_r/d_t$), ω is the angular velocity, r is the distance from the center of rotation, M is the molar mass, υ is the partial specific volume (for proteins the accepted value is 0.73 mL*g⁻¹), ρ is the density of the buffer at 20°C, N_A is the Avogadro's number, f is the frictional coefficient. This equation defines the sedimentation coefficient (s) in Svedberg units (10⁻¹³ sec=1S) as the ratio of the velocity (v) to the field strength ($\omega^2 r$).

The diffusion coefficient (D) corresponds to the spreading of the concentration boundary and can be determined from the shape of the sedimentation band.

$$D = kT/f = RT / N_A f \quad (4)$$

In **eq. 4** T is the absolute temperature and R is the gas constant, N_A is the Avogadro's number, f is the frictional coefficient.

Together s and D allow the calculation of the exact molecular weight of the sample according to the *Svedberg equation*:

$$s/D = (M(1-\upsilon\rho)/N_A f) / (RT/N_A f) = M(1-\upsilon\rho)/RT \quad (5)$$

In **eq. 5** M is the molar mass, υ is the partial specific volume, ρ is the density of the buffer at 20°C, T is the absolute temperature and R is the gas constant.

2.6.2 Practical application

Sedimentation velocity data of the RBCC domain (0.5 mg/mL) were obtained using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter) with an absorbance optical system. Epon double-sector centerpieces with sapphire windows containing 200 μ L of protein solution and 300 μ L of sample buffer were centrifuged at 55'000 rpm and 20°C for 8 h and data acquired with a radial increment of 0.003 cm with no delay between scans. The sedimentation of the protein was monitored at 292 nm and 280 nm. The

experimental data were fitted with a continuous distribution model using Sedfit¹¹³. Buffer viscosity (0.01002 cp), density (1.04 g/cm³) and protein partial specific volumes (0.72 g/cm) were estimated using the program Sednterp¹¹⁴.

See in Chapter 3.2, Fig. 27B| RBCC sedimentation velocity experiment.

2.7 High resolution techniques

KAP1 plays a fundamental role in regulating gene expression in mammalian cells¹⁰⁴, by recruiting different transcription factors and altering chromatin state¹⁰³. Despite its emerging fundamental role, little is known about the structure of the full-length protein. At the time this work was performed, the X-ray structure of the RBCC domain was not available and the only structural information related to the protein was the NMR structure of the C-terminal domain solved in 2008⁶². Moreover almost twenty years ago the stoichiometry of the KAP1-KRAB domain complex was inferred from inconclusive SEC experiments⁴². We decided to shed light on the structural organization of the protein performing crystallization trials for the RBCC domain, for a variant called RBCC-T4-Lysozyme and for KAP1 FL. Moreover we attempted to obtain a high resolution Cryo-EM map of KAP1 FL to perform 3D reconstruction.

2.7.1 Crystallization trials

The MB Class Suite (MBC-I) screen (Quiagen) was used to crystallize the RBCC domain of KAP1. Crystals were obtained in (i) 500 mM NaCl, 50 mM Tris-HCl pH 8.5, 10% (w/v) PEG (Poly Ethylene Glycol) 4'000 and (ii) 100 mM Lithium Sulfate, 100 mM Na Citrate pH 5.6, 12% (w/v) PEG 6'000. Another crystallization condition was found in the Clear Strategy I (CS-I) screen (Molecular Dimensions). Crystals grown in 800 mM Na Formate and 25% (w/v) PEG 2'000 MME. The RBCC was crystallized at a concentration of 10 mg/mL in 20 mM HEPES pH 7.5, 300 mM NaCl, 3% (v/v) glycerol, 2.5 mM TCEP. We crystallize both the His₆ tag and the untagged protein. We used the vapor diffusion method, either sitting drop or hanging drop. Crystallization drops were set in 96 well plates using mosquito®crystals for protein nanoliters pipetting (TTP Labtech) with 150 nL of protein, 150 nL of buffer, 40 µL of reservoir. Crystals were grown at 18°C or 4°C. We improved the crystallization conditions varying (i) the protein concentration (ranging from 5 to 10 mg/mL) (ii) the buffer composition (from 300 to 150 mM NaCl, from 5% to 0% (v/v) glycerol, different buffer pH (iii) the ratio protein-reservoir (1:1, 1:2, 2:1) (iv) the cryo-protectant solutions (glycerol from 20% to 25%, Peg 400, MPD) (v) additives from the Additive Screen (Hampton Research).

Positive crystallization conditions for the RBCC-T4 Lysozyme construct were found in in the Clear Strategy I screen (CS-I). Crystals grown in (i) 800 mM Na Formate, 100 mM Tris-HCl pH 8, 8% (w/v) PEG 20'000, 8%

(w/v) PEG 550 MME and in (ii) 800 mM Na Formate, 100 mM Na Acetate pH 5.6, 8% (w/v) PEG 20'000, 8% (w/v) PEG 550 MME. Other crystals formed in the MB Class I screen (MBC-I) in 500 mM KCl, 50 mM MOPS pH 7, 12% (w/v) PEG 4'000, 20% (v/v) glycerol. The protein crystallized at 10 mg/mL in 20 mM HEPES pH 7.5, 300 mM NaCl, 3% (v/v) glycerol, 2.5 mM TCEP.

Finally we set crystallization drops for KAP1 FL using CS-I, MBC-I, PACT premier™ (Molecular Dimensions) and JCSG-plus™ screen (Molecular Dimensions). The protein was crystallized at 10 mg/mL and crystals grown in 0.2 M Na citrate tribasic dehydrate, 20% (w/v) 3'350 (PACT premier™).

2.7.2 Electron Microscopy experiments

2.7.2.1 Negative staining electron microscopy

KAP1 FL was diluted to 0.05 mg/mL in 20 mM HEPES pH 7.5, 300 mM NaCl and 2 mM TCEP and crosslinked with 0.1% glutaraldehyde for 2h at 23°C. The reaction was stopped by addition of 100 mM Tris-HCl pH 7.5. The sample was diluted 20 times and adsorbed to a glow-discharged carbon-coated copper grid (EMS, Hatfield, PA, USA), washed with deionized water and stained with a solution of 2% uranyl acetate. The grids were observed using an F20 electron microscope (Thermo Fisher, Hillsboro, USA) operated at 200 kV. Digital images were collected using a direct detector camera Falcon III (Thermo Fisher, Hillsboro, USA) with 4098 X 4098 pixels. The magnification of work was 29'000X (px=0.35 nm), using a defocus range of -1.5 µm to -2.5 µm. After manual picking of 400 particles, Relion¹¹⁵ was used to sort them into 2D class averages.

2.7.2.2 Cryo Electron Microscopy

KAP1 FL was buffer exchanged in 20 mM HEPES pH 7.5, 300 mM NaCl and 2 mM TCEP. The sample with an initial concentration of 6 mg/mL was diluted 5, 30, 100 and 200 times. 4 µL of diluted sample were adsorbed to a C flat™ Holey Carbon Grids (300 mesh). The sample was applied to the grid on the lab bench. Afterwards, the tweezers were mounted on the Vitroblot apparatus (ThermoFisher, Hillsboro, USA) to perform the blotting for 2 sec and plunge freezing in a bath of liquid ethane maintained at liquid nitrogen temperature (0 sec waiting time). The grids were observed on a Tecnai F20 Cryo microscope (Thermo Fisher, Hillsboro, USA) operated at 200 kV. Digital images were collected using both a direct detector camera Falcon III (Thermo Fisher, Hillsboro, USA) and Ceta 16M camera (ThermoFisher, Hillsboro, USA). The magnification of work was 50'000, 62'000 and 80'000 X.

2.8 Small-Angle X-Ray Scattering data collection and analysis

Given the intrinsic flexibility of KAP1 FL, characterized by a long loop connecting the N-terminal and the C-terminal domains, high resolution techniques were proved to be unsuitable to solve the three dimensional structure of the protein. For this reason, we developed a novel integrative modelling approach combining solution scattering data and flexible SAXS data fitting to reveal the molecular architecture of full-length KAP1.

2.8.1 Theoretical basis

2.8.1.1 SAXS parameters: s / $I(s)$ / $I(0)$ / Guinier / R_g / $P(r)$ / Kratky

Small-angle X-ray scattering (SAXS) is a biophysical method to structurally characterize proteins in solution. SAXS provides low-resolution information (12-20 Angstrom) on the shape and conformation of single macromolecules (few kDa) or proteins ensemble (GDa complexes). The technique can be used to study (i) ordered proteins (ii) multi-domain proteins with regions of flexibility (iii) intrinsically disordered proteins (IDPs) (iv) protein-protein complexes (v) protein-nucleic acids complexes¹¹⁶.

The sample in solution is generally placed in a quartz capillary and illuminated by a collimated X-ray beam (Incident intensity ($I(0)$) or k_0) and the scattering intensity ($I(s)$ or k_1) of the sample solution is collected by an X-ray detector (**Fig. 14**).

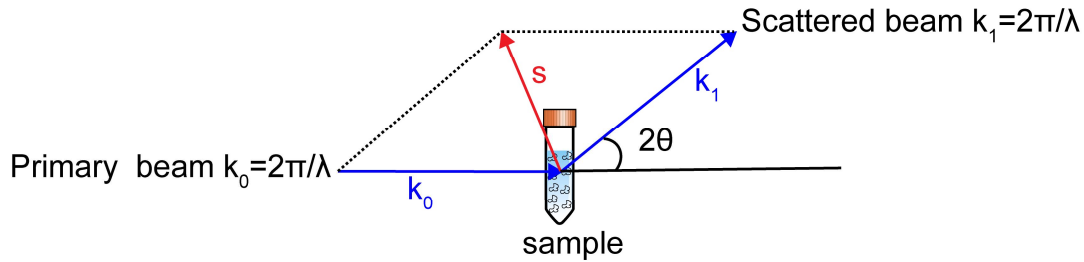


Figure 14| Schematic representation of a SAXS experiment. k_0 is the incident beam intensity, k_1 is the scattered intensity and s (or q) is the momentum transfer or the length of the scattering vector. Figure adapted from “A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins”¹¹⁶

$$|k_0| = |k_1| = 2\pi/\lambda \quad (6)$$

$$s = k_1 - k_0 \quad (7)$$

$$|s| = 4\pi\sin(\theta)/\lambda \quad (8)$$

In eq. 6 and in eq. 8 θ is the scattering angle, λ is the wavelength and s is the momentum transfer or the length of the scattering vector.

Afterwards, the scattering intensity is plotted as a function of the momentum transfer s (or q) (**Fig. 15**).

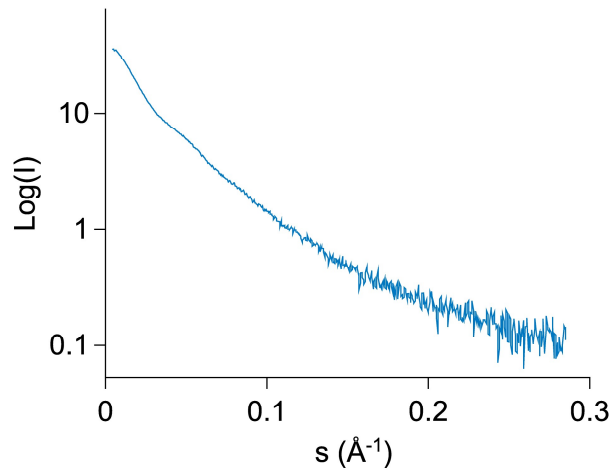


Figure 15| SAXS scattering curve. The scattering intensity ($I(s)$ or kI) is plotted as a function of the momentum transfer (s).

At the same time the scattering pattern of the pure solvent is collected and subtracted, leaving only the signal from the particle of interest. The final scattering signal is derived from the subtraction of the electron density of the solvent from the sample solution.

$$\Delta\rho(r) = \rho(r) - \rho(s) \quad (9)$$

In **eq. 9** $\rho(r)$ is the electron density of the sample and $\rho(s)$ is the electron density of the solvent.

The scattering intensity ($I(s)$) is a function of the momentum transfer.

$$I(s) = 4\pi \int_0^{D_{\max}} P(r) \frac{\sin(sr)}{sr} dr \quad (10)$$

In **eq. 10** D_{\max} is the maximum length of the macromolecular sample and $P(r)$ is the pair distance distribution function. The above equation relates the scattered intensity I to the $P(r)$ distribution as a function of the scattering angle, s . The observed scattering intensity I at a specific scattering angle $s = 4\pi\sin(\theta)/\lambda$, is the integration of the product of the $P(r)$ function with the $(\sin(sr)/sr)$ function, over all the possible distances, r .

In the region near $s=0$ (small angles) the *Guinier* approximation states that:

$$I(s) \approx I(0)\exp(-1/3 R_g^2 s^2) \quad (11)$$

According to **eq. 11** the scattering intensity plotted as $\ln I(s)$ vs s^2 is a linear function. From the slope of this function we can retrieve the radius of gyration (R_g). This parameter represents the quadratic mean of distances to the center of the mass and provides information about the overall size of the macromolecule. Moreover, the intercept of this function instead gives the $I(0)$, the scattering intensity at angle θ equals to 0. This value is proportional to the molecular weight (M_w) and to the concentration of the protein. $I(0)$ represents the square of the number of electrons in the scatterer unaffected by particle shape. The $I(0)$ value plotted vs C_p (protein concentration) is a linear function.

The Guinier approximation is valid for very small angle up to $sR_g < 1.3$. A lack in the linearity of the Guinier fit indicates sample aggregation or repulsive inter-particle interactions (**Fig. 16**). Scattering from aggregated sample influences the whole data set and no further analysis should be performed. The sample aggregation can be prevented by varying the buffer condition or by centrifugation and filtration of the sample.

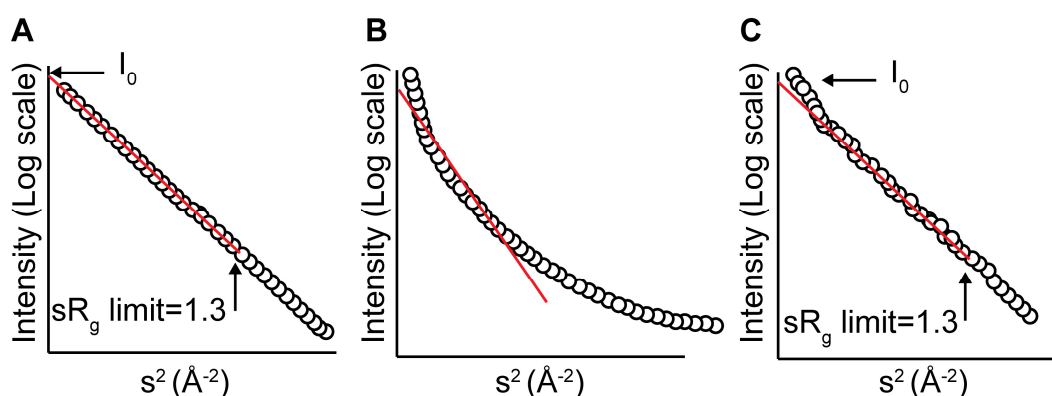


Figure 16| Guinier plots of different samples. A. non-aggregated sample B. aggregated sample C. partially aggregated sample. The plots show the $I(0)$ value as the intercept with the ordinate and the sR_g limit for Guinier calculation. Lack of linearity in the Guinier plot as in B. indicates sample aggregation. Figure adapted from “X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution”¹¹⁷.

Larger s values contain information on the shape of the protein. In real-space the sample can be described by the $P(r)$ (Pair Distance Distribution) function, which is directly calculated through an IFT (Indirect Fourier Transformation) of the scattering curve. The scattering intensity and the pair distance distribution function are related to each other by the **eq. 12**.

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 I(s) \sin(sr)}{sr} ds \quad (12)$$

The $P(r)$ function describes the distances (r) of all the pairs of electrons in the macromolecular sample and it is a useful tool to detect conformational changes within a macromolecule. As previously described for the scattering intensity (I_s) in **eq.10**, the $P(r)$ function is the integration of the product of the scattering intensity $I(s)$ with the $(\sin(sr)/sr)$ function, over all the possible s values. The function equals 0 at $P(0)$ terminates

smoothly at the maximum distance D_{\max} of the protein and it is 0 when $r > D_{\max}$. The function is measured in real-space so the x-axis is in Å. Since the function describes all the possible distances within a macromolecule, small changes in the relative position of residues or domains resulting in a change of mass distribution around the center of gravity (or radius of gyration, R_g) produce a clear and detectable change in the $P(r)$ function (**Fig. 17**).

The radius of gyration can be obtained from the $P(r)$ function by integrating the function with r^2 over all the possible values of r as described in **eq. 13**.

$$R_g^2 = \frac{\int_0^{D_{\max}} r^2 P(r) dr}{2 \int_0^{D_{\max}} r P(r) dr} \quad (13)$$

The R_g calculated from the Guinier approximation only considers a small portion of data while the R_g calculated from the $P(r)$ function (often referred ad Real-Space R_g) uses all the experimental data and it is determined in real-space. Large differences between the R_g values calculated from the $P(r)$ function and the Guinier approximation could indicate small amount of sample aggregation.

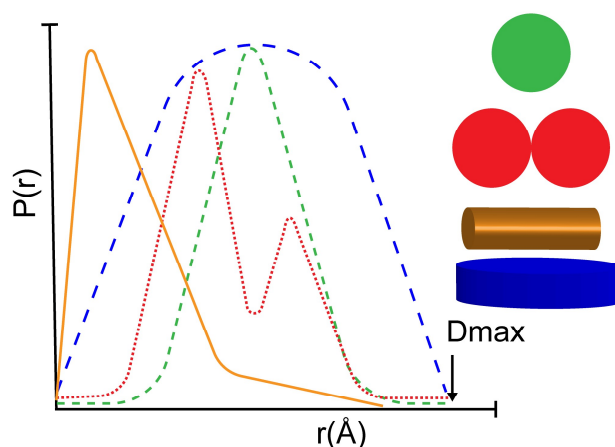


Figure 17| Pair distance distribution functions representative of different sample shapes. The $P(r)$ functions of proteins presenting different shapes are quite diverse however they are all smooth positive functions that converge to 0 at D_{\max} . In green the $P(r)$ function of a globular protein, in red of dumbbell-shape protein, in orange of a cylindrical protein, in blue of a flat elongated protein. Figure adapted from “Small Angle X-ray Scattering of group V Polyoxometalates” ¹¹⁸

SAXS is useful to identify and characterize proteins containing disordered regions. Specifically the Kratky plot in which $s^2 I(s)$ is plotted as a function of s , is used to identify unfolded states and distinguish them from globular proteins. Globular macromolecules have a bell-shaped curve that reach 0 in the larger s range, unfolded protein lack this peak and present a plateau at large s value, partially unfolded or flexible proteins present the peak of folded molecules but the Kratky function only partially reach the 0 value (**Fig. 18**).

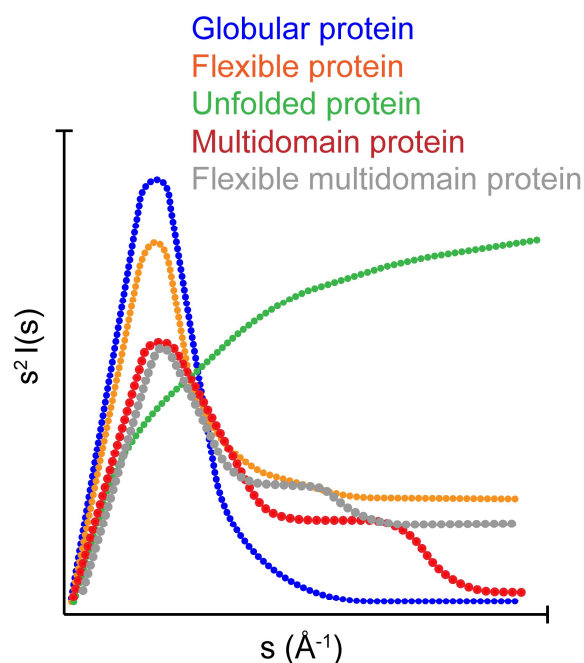


Figure 18| Schematic representation of typical Kratky plots. The curvature of the plot depends on the degree of flexibility of the protein. In blue the Kratky plot for a globular protein, in orange for a flexible protein and in green for an unfolded protein. The Kratky plots for multidomain flexible and globular proteins are shown in light grey and in red. Figure adapted from “X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution”¹¹⁷.

The summary of the preliminary analysis reported in this chapter is shown in **Fig. 19**.

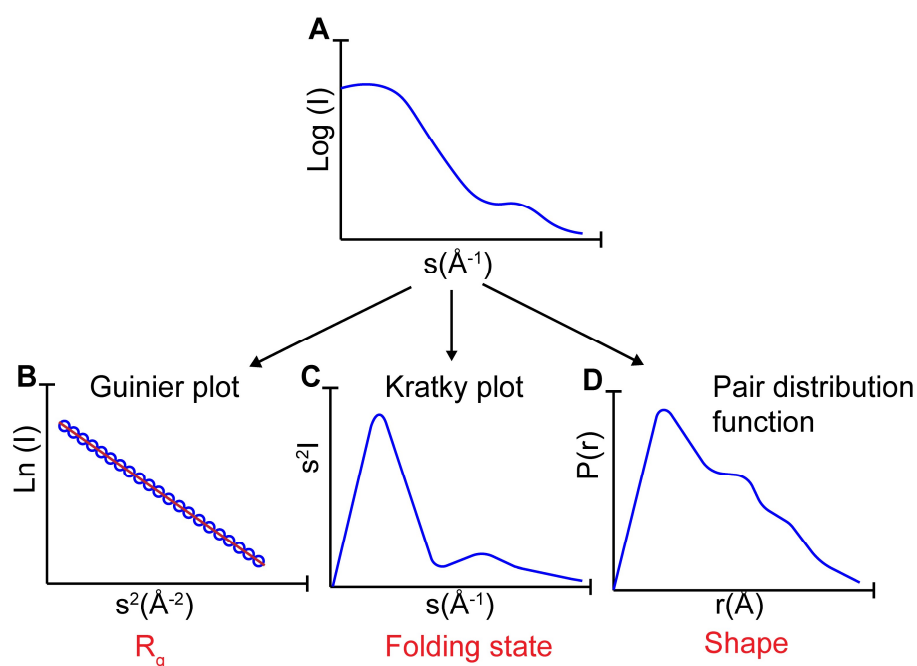


Figure 19| Summary of the different plots analyzed in this thesis obtained from SAXS data. A. Scattering curve B. Guinier plot C. Kratky plot D. P(r) distribution function. For each plot is shown the general representation of a possible result, label x and y-axis and in red the final parameter retrieved from each analysis.

2.8.1.2 *Ab initio* shape determination: DAMMIN/F and Gasbor

The scattering curve is proportional to one of a single particle averaged all over the possible orientations. For this reason it is possible to determine from the scattering profile the shape of a protein at low resolution. We can apply the *ab initio* shape determination approach to generate shapes that will fit the scattering data without using predefined input structures but generating models made of dummy beads. This approach applies Monte Carlo algorithm, genetic algorithm or simulated annealing to search for models in agreement with the experimental data¹¹⁹.

*DAMMIN/F*¹²⁰ is a modeling algorithm that attempts to generate the shape of a scattering particle from the SAXS profile by applying the simulated annealing approach (**Fig. 20**). Issues with buffer subtraction are reduced automatically by removing from the analysis the data in the high s -range ($s > 0.2 \text{ \AA}^{-1}$). This approach does not reduce the resolution of the final model since DAMMIN uses a volume approximation (Porod-scaling) that emphasizes the data in the low s -range. The program searches for models with the most compact configuration made of dummy beads by minimizing the discrepancy (χ^2) between the experimental scattering intensity ($I_{\text{exp}}(s)$) and the calculated scattering curve $I(s, X)$ for each model (X). Moreover the program applies a penalty function to ensure compactness and connectivity, $P(X)$ in **eq. 14**. Doing so the models containing loose and disconnected beads are removed.

$$f(X) = \chi^2 [I_{\text{exp}}(s), I(s, X)] + \alpha P(X) \quad (14)$$

where $\alpha > 0$ is a weight function

Since the program uses a simulated annealing schedule (stochastic method), 10 to 15 runs must be performed. The generated models are then superimposed, averaged and a probability density map is computed. The most probable model is then selected as the most compact volume close to the expected excluded volume of the scattering particle.

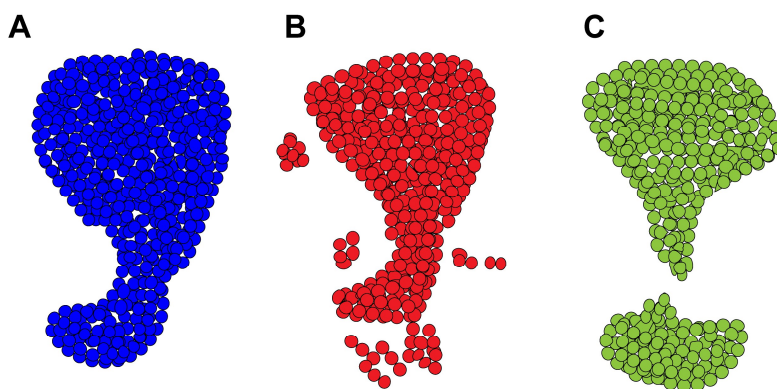


Figure 20| Example of Dammin models. A. compact model B. model presenting loose beads C. model with two disconnected parts. The program searches for the most compact volume excluding models that present loose or even disconnected dummy atoms. Figure adapted from “SAS model determination, *ab initio* modeling”, EMBL.

*Gasbor*¹²¹ is another program used in *ab initio* shape determination. The most important difference when compared to DAMMIN/F is that the protein structure is not represented as dummy beads not corresponding to real atoms but as an ensemble of dummy residues forming a chain compatible model. Each dummy residue corresponds to the average density of each protein residue and its center aims to approximate the position of the C α atoms in the real structure. For this reason the number of dummy residues used in the reconstruction should be equal to that in the protein. Moreover while in DAMMIN/F it is recommended to exclude from the analysis the data in the high s-range value region with $s > 0.2 \text{ \AA}^{-1}$, in Gasbor this is not required since the program is able to fit the data up to a resolution of 5 Å corresponding to a momentum transfer:

$$|s| = 4\pi\sin(\theta)/\lambda = 1.2 \text{ \AA}^{-1} \quad (15)$$

2.8.1.3 Why SEC SAXS (Size Exclusion Chromatography Small-Angle X-ray Scattering)

Problematic samples can present (i) aggregation, frequently happening at the re-concentration of the sample (ii) dissociation during the transit through the quartz capillary part of the SAXS setting, typical of complexes characterized by a weak K_d (dissociation constant) (iii) mixture of different species or presence of multimers. In these cases the method of choice to investigate self-associating systems and aggregated heterogeneous samples is SEC-SAXS (size exclusion chromatography in line with small-angle X-ray scattering). The sample passes through the size exclusion chromatography column before being illuminated by the collimated X-ray beam. This will allow the separation of aggregates (they elute in void volume of the column) from the non-aggregated sample. In this way the scattering signal produced by the aggregate portion will not influence the whole data set, covering the signal coming from the non-aggregated sample.

2.8.2 Practical application

To exclude sample aggregation, the proteins were analyzed by size-exclusion chromatography in line with small-angle X-Ray scattering (SEC-SAXS). The data were collected at the European Synchrotron Radiation Facility (ESRF), beamline 29 (BM29) at a wavelength of 0.99 Å with a sample to detector distance of 2.867 m and a PILATUS 1M detector, covering a momentum transfer of $0.0025 < s < 0.6 \text{ \AA}^{-1}$ [$s = 4\pi\sin(\theta)/\lambda$]. Measurements were made at 18°C. The samples were applied to a Superose 6 10/300 GL column (GE Healthcare) at a concentration of 9-12 mg/mL and run at a flow rate of 0.75 mL/min in 20 mM HEPES pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM TCEP. During the elution, 2160 scattering measurements were taken with 1s time-frames. The in-house software *BsxCuBE* (Biosaxs Customized Beamline Environment) connected to a data processing pipeline (EDNA)¹²² was used to control the real time data display (two dimensional and one dimensional) and to provide the first automatic data processing up to a preliminary *ab initio* model. SAXS data were analyzed using the ATSAS package version 2.8.3¹²³ and ScÅtter¹²⁴. For each

sample, using Chromixs¹²⁵, an elution profile was generated with the integrated intensities plotted versus the recorded frame number. Using Chromixs, 30 buffer frames were averaged and used to (i) subtract the buffer average from each frame of the sample peak selected and (ii) calculate the corresponding Radius of Gyration (R_g). The subtracted peak region was selected in Chromixs and averaged to generate the final scattering curve used for subsequent analysis. The scattering curves were initially viewed in *PRIMUS*¹²⁶ where the R_g was obtained from the slope of the Guinier plot within the region defined by $q_{\min} < q < q_{\max}$ where $q_{\max} < 1.3/R_g$ and q_{\min} is the lowest angle data point included by the program. When one dimension of a scattering particle is greater than the other two (e.g. a rod particle), “Rod” Guinier analysis in *PRIMUS* can be used to calculate the radius of gyration of its cross-section, R_{gc} . The $P(r)$ function, the distribution of the intra-atomic distances (r) in the particle, was generated using the indirect transform program *GNOM*¹²⁷. The maximum distance (D_{\max}) was selected by letting the $P(r)$ curve decay smoothly to zero. As our molecules are rod-like and flexible, the R_g was also estimated from the $P(r)$ function, such that, unlike the Guinier R_g estimation, the $P(r)$ R_g calculation takes the whole scattering curve into account. *Ab initio* models were produced using GASBOR imposing a P1 symmetry with prolate (elongated protein) anisotropy until a χ^2 of 1 was reached. Moreover 20 additional *ab initio* models were generated using DAMMIF, superimposed with SUPCOMB, averaged with DAMAVER and refined with DAMMIN. We compared the results obtained for the RBCC and KAP1 FL with and without imposing 2-fold symmetry.

DATPOROD, DATMOW and DATVC within the ATSAS package were used to estimate the Porod Volume (V_p) and the concentration-independent estimation of the MW for the proteins. The final figures were generated using VMD¹²⁸, PyMOL (Schrödinger, LLC) and Chimera¹²⁹.

See Chapter 3.4, **Fig. 36**| SAXS data collection and Guinier analysis, **Fig. 37**| Kratky plots of KAP1 constructs, **Fig. 38**| RBCC and KAP1 FL $P(r)$ functions and *ab initio* models.

2.9 Molecular modeling and flexible SAXS data fitting

2.9.1 Theoretical basis

In the case of a monodisperse sample the theoretical scattering curve computed from a single high-resolution structure fits and consequently explains the experimental scattering curve. However this is not valid for proteins that contain flexible loops or IDP (Intrinsically Disordered Proteins) that cannot be represented by a single rigid model. Flexible systems are by definition polydisperse and the measured scattering intensity $I(s)$ is an averaged intensity coming from the different protein conformations in the sample.

$$I(s) = \sum_{k=1}^K v_k I_k(s) \quad (16)$$

In **eq. 16** v_κ and I_κ are respectively the volume and the scattering intensity of the κ -component present in the sample solution.

SAXS is a useful technique to analyze flexible proteins and to study equilibrium systems like oligomeric protein solutions. Several approaches have been developed to correctly interpret and describe scattering data coming from polydisperse samples. These methods are based on a common strategy divided in two main steps: (i) generation of a large pool of initial random models and (ii) selection of a subset of models that together describe the experimental scattering profile.

The next sections provide an introduction to two of the most popular approaches: NMA (Normal Mode Aalysis) and EOM (Ensemble Optimization Method).

2.9.1.1 Normal mode analysis for flexible SAXS data fitting

As previously described, SAXS data allow for *ab initio* shape determination without using predefined atomistic input structures. However if an atomistic structure of the macromolecule is available or if homologous structures can be used to generate high-resolution models, the experimental SAXS data can be combined with NMA (Normal Mode Aalysis) to sample the conformational space explored by the macromolecule in solution.

NMA is a well-established coarse-grained technique used to study protein conformational changes. In fact the crystallographic structures often represent a static conformational snapshot of the protein that can differ from the one in solution resulting in a discrepancy between the theoretical scattering curve calculated from the crystal structure and the experimental SAXS data¹³⁰.

In this thesis, the NMA methodology is used to (i) explore the conformational space of high-resolution models obtained by homology modeling, (ii) fit them to SAXS data minimizing the discrepancy between the experimental and the theoretical scattering curve and (iii) find an ensemble of conformations that describe the SAXS data obtained for a flexible system.

The normal modes associated with a molecule are obtained from the Hessian (H) matrix of the potential energy function $V(r)$ calculated for each set of C α and diagonalized to obtain the eigenvectors each of them associated with an eigenvalue, given the directionality and the strength of the movement. The potential energy function V is described as the sum of the harmonic springs of equal strength κ that link the C α atoms within a distance R_c of 10 Å¹³¹.

$$V = \sum_{\substack{r_{ij}^0 < R_c \\ i < j}} \kappa (r_{ij} - r_{ij}^0)^2 \quad (17)$$

In **eq. 17** r_{ij}^0 is the Euclidean distance between the atoms i and j , κ is the strength of the harmonic springs equal to $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

The normal modes are classified according to their frequencies in (i) low-frequency normal modes (large amplitudes) responsible for global rearrangements of the protein structure and (ii) high-frequency normal modes corresponding to small and local changes. These movements are considered not realistic since we use a force-field (elastic network model) which is only good at low-frequency range. Moreover at high-frequencies all the local details such as correct electrostatics start to matter.

The initial models are displaced in the Cartesian space according to one or a combination of multiple normal modes and new conformers are generated. At this stage the new conformers are completely unrestrained meaning that the stereochemistry as well as some secondary structures could get distorted.

To minimize the amount of structural distortions in the generated models, the structures of the unrestrained conformers can be partitioned into “pseudo-domains” defined by the user or defined automatically as continuous segment that move in a coordinated manner according to NMA. Once a “pseudo-domain” is defined, this region is taken from the original structure and superimposed as a rigid-body on the corresponding region of the unrestrained model generating in this way a new pool of restrained conformers that conserve the local topology of the initial models. This superimposition allows to explore large conformational movements conserving at the same time the intra molecular interactions, stereochemistry and secondary structures of the original models. The described approach is followed by the software SREFLEX¹³².

The new conformers are scored against the experimental SAXS data by computing a theoretical scattering profile for each of them.

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N_p} \left(\frac{I_e(s_i) - c I(s_i)}{\sigma(s_i)} \right)^2 \quad (18)$$

In **eq. 18** I_e and I are the experimental and theoretical scattering intensities, N_p is the number of experimental points, σ is the experimental error and c is a scaling factor. The conformers with the lowest χ^2 value are selected and the procedure continues iteratively.

This work uses a novel procedure based on the nonlinear Cartesian NMA method called NOLB¹³³. This new method approximates the system as n rigid blocks characterized by 6 DOF (6 degrees of freedom) (translation and rotation along the three axis). The Hessian matrix (3X3 rotation-translation matrix) is computed for each block. The total displacement is obtained by summing up the displacement of the single rigid-blocks obtaining the RTB-projected matrix (rotations translations of blocks). The normal modes are computed from the diagonalization of this matrix that gives a set of eigenvectors composed of an *instantaneous linear velocity* and an *instantaneous angular velocity*. This method generates a twist motion applied to the system that has

been proved to conserve the total kinetic energy. Only the low frequencies or large amplitudes normal modes are selected and applied specifically to flexible regions or loops. Once the normal models are calculated and applied to the system generating new translated models, a novel SAXS profile calculator called Pepsi-SAXS¹³⁴ creates the theoretical scattering curves and calculates the χ^2 value corresponding to the discrepancy between the experimental and the theoretical scattering curves. Only the mode that leads to the lowest χ^2 value is selected and the resulting translated model is taken and used again to compute the Hessian matrix. The iteration is repeated 100 times. At the end of each step a steepest-descent minimization algorithm is used to keep the bonds and the angles in agreement with the original structure (**Fig. 21**). In this method the normal modes are considered separately and not as a combination because the behavior of the χ^2 value along a combination of normal modes is not linear but oscillating, leading to a inefficient analysis of the gradient along a combination of modes.

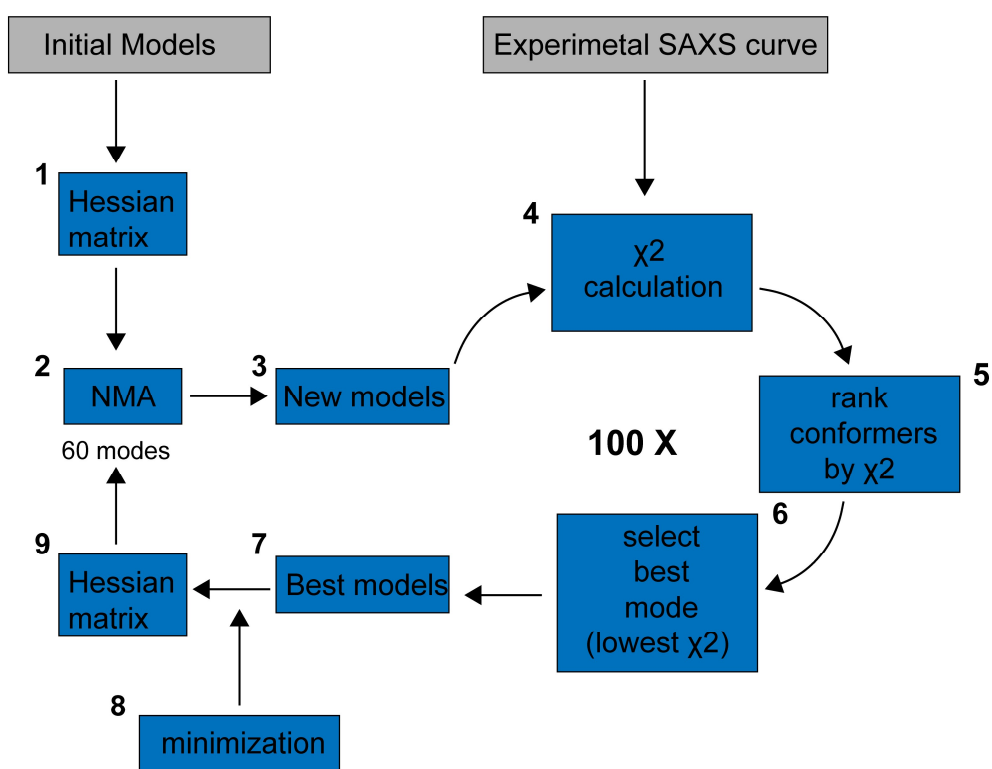


Figure 21| Flowchart of the non-linear Cartesian normal mode method¹³³ for SAXS data fitting. In this chart: **1.** The Hessian matrix is computed for the initial models **2.** The normal modes are computed from the diagonalization of this matrix **3.** New initial displaced models are generated by the application of NMA to the original structures **4.** The χ^2 value is computed for each new model **5.** The conformers are ranked based on this value **6.** The normal mode that leads to the lowest χ^2 value is selected together with the best conformer **7.** The new models with the lowest χ^2 value are selected **8.** A minimization step is applied to maintained the initial topology in agreement with the original structure (angles and bonds) and **9.** A new Hessian matrix is computed for the best conformers and the process is repeated 100 times.

2.9.1.2 Ensemble optimization method

In the EOM approach the experimental scattering profile of a flexible protein is fit with the averaged theoretical scattering profile calculated from an ensemble of different conformations. At the beginning a pool of n initial models (typically 10'000) is generated. For multi-domain proteins, the high-resolution structures of the domains (when available) can be used as rigid-body or constrains in the initial model generation. For each domain it is possible to specify the crystallographic symmetry (i.e. P1, P2 ..), the oligomerization interface between domains using distance constrains and the inter-domain contacts. For intrinsically disordered protein no structural information is required and a random C α chain will be generated for each model based only on the protein sequence. The theoretical scattering curve for each conformer is pre-computed using CRY SOL¹³⁵. Afterwards, a genetic algorithm is used to select a sub-ensemble of 10-50 conformers from the initial pool and the scattering data for each ensemble is computed by summing the individual scattering profile of each conformer. Finally the sub-ensemble with the best agreement (presenting the lowest χ^2 value) to the experimental data is selected. An ensemble of 10 to 50 conformers is too small to describe the complexity of a flexible system, however it is sufficient to define general properties of the system such as R_g and D_{max} (shape and dimensions).

2.9.2 Practical application

The preliminary KAP1 model was created using existing, homologous structures of the RING, B-box and coiled-coil α -helical domains, and the solved NMR structure of the PHD-Br domain (PDB: 2RO1)⁶². Specifically, the SWISS Model server¹³⁶ was used to construct the RBCC domain using the following templates: the RING domain was based on the dimer of Rad18 (PDB: 2Y43)¹³⁷, the B-box1 was based on the B-box domain of MuRF1 (PDB: 3DDT)¹³⁸, the B-box 2 was based on the B-box domain of TRIM54 (PDB: 3Q1D), and the coiled-coil domain was based on the coiled-coil of TRIM69 (PDB: 4NQJ)¹³⁹. To model the TSS domain in the central part of KAP1 we used the Robetta server¹⁴⁰, while MODELLER v9.14¹⁴¹ was used for the HP1-binding domain, assigning as a template the structure of EMSY protein in complex with HP1 (PDB: 2FMM¹⁴²) and the structure of the small subunit of the mammalian mitoribosome (PDB: 5AJ3¹⁴³). Finally all the missing loops were modeled by Rosetta loop modeling application v3.5¹⁴⁴. In the process of generating starting models for fitting to SAXS data, each PHD-Br domain was randomly rotated and translated at a maximum distance of 140 Å (distance D in **Fig. 43A**) from the RBCC domains using VMD¹²⁸. Afterwards, the linker region between each PHD-Br domain and its respective RBCC domain was modelled as a random coil using MODELLER¹⁴⁵. 1000 models were generated and flexibly fitted to the SAXS data through a χ^2 –minimizing optimization procedure based on the nonlinear Cartesian NMA method called NOLB¹³³ and a novel SAXS profile calculator called Pepsi-SAXS¹³⁴. More precisely, for each initial model

we performed 100 optimization iterations. Each iteration comprised the computation of 60 slowest normal modes (using the NOLB tool), nonlinear structure deformation along these modes, and choosing the deformation with the least χ^2 value to the experimental scattering profile (using Pepsi-SAXS). A steepest-descent minimization algorithm was used at the end of each iteration to keep the local topology (bonds and angles) in agreement with the initial structure. The flexible fitting method is available as a standalone executable called Pepsi-SAXS-NMA at (<https://team.inria.fr/nano-d/software/pepsi-saxs/>). The choice of 60 normal modes was based on the observation of high flexibility of the linker region. This number of modes allowed us to explore plenty of plausible conformations of the linker. The chosen slowest modes did not change the structure of the rigid domains, but only changed their relative orientation. ~60% of the initial models converged to low values of χ^2 within the wall-time assigned for fitting (24h). The resulting structures produced a statistically relevant ensemble and were clustered using a new method tailored for clusters exhibiting Gaussian distributions, typical of structural ensembles. As for collective variables, the distance between the center of mass of the PHD-Br domain and that of the closest RB₁B₂ module was computed for each protomer. The models were clustered according to these two distances and the cluster center was extracted.

See Chapter 3.6, **Fig. 40**| Overall architecture of the RBCC domain, **Fig. 42**| Flexible SAXS data fitting of KAP1 FL models, **Fig. 43**| Native asymmetry of KAP1, **Fig. 44**| Asymmetric domain arrangement in KAP1 FL, **Fig. 45**| Analysis of the KAP1 FL models in the first density level, **Fig. 47**| Integrative modeling analysis of Δ KAP1.

2.10 KAP1 and the binding partners

After exploring the architecture of KAP1 alone we decided to characterize its interaction with HP1 α , using for the first time full-length purified proteins. The interaction was first characterized by MALLS and then the stoichiometry of the complex was confirmed by ITC (Isothermal titration calorimetry). Afterwards, in collaboration with Beat Fierz Laboratory we decided to explore the interaction between the two proteins in the context of chromatin fibers to understand whether the unexpected stoichiometry was maintained and whether KAP1 was playing a role in the interaction between HP1 and the histone tails.

2.10.1 Isothermal titration calorimetry (ITC)

ITC experiments were performed using a MicroCal PEAQ ITC from Malvern. KAP1 FL and HP1 α FL were buffer exchanged into 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mM TCEP and concentrated respectively to 15 and 170 μ M. HP1 α FL containing solution was injected into the KAP1 FL solution (2 μ L of HP1 per injection at an interval of 180 sec, a total of 19 injections into the cell volume of 300 μ L, with stirring speed of

800 rpm, at 25°C). HP1 α FL was also injected into buffer to determine the unspecific heat of dilution. However, subtracting this experimental heat of dilution was not sufficient to get a good fit, so the last injections were used to better estimate the heat of dilution and subtract a straight line from our data. The experimental data were fitted to a theoretical titration curve (“one set of sites” model) using software supplied by Microcal, with n (number of binding sites per monomer), ΔH (binding enthalpy, kcal/mol) and K_d (dissociation constant, M), as adjustable parameters. Thermodynamic parameters were calculated from the Gibbs free energy equation, $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$, where K_a is the association constant, and ΔG , ΔH and ΔS are the changes in free energy, enthalpy, and entropy of binding, respectively. T is the absolute temperature, and $R = 1.98 \text{ cal mol}^{-1} \text{ K}^{-1}$.

See Chapter 3.8, Fig. 51A| Titration of HP1 α FL into KAP1 FL.

2.10.2 Single molecule experiments

2.10.2.1 H3K9me3 synthesis

H3K9me3 was synthesized as previously described¹⁴⁶. In short, the peptide H3(1–14) K9me3-NHNH₂ (carrying a C-terminal hydrazide) was synthesized by solid phase peptide synthesis (SPPS). The truncated protein H3(Δ 1–14)A15C was recombinantly expressed as N-terminal fusion to small ubiquitin like modifier (SUMO), the N-terminal SUMO was cleaved by SUMO protease and the protein purified by RP-HPLC. For ligation, in a typical reaction 3 μ mol H3(1–14)K9me3-NHNH₂ was dissolved in ligation buffer (200 mM phosphate pH 3, 6 M GdmCl) at -10°C. NaNO₂ was added dropwise to a final concentration of 15 mM and incubated at -20°C for 20 min. H3(Δ 1–14)A15C was dissolved in mercaptophenyl acetic acid (MPAA) ligation buffer (200 mM phosphate pH 8, 6 M GdmCl, 300 mM MPAA) and added to the peptide, followed by adjustment of the pH to 7.5. The ligation was left to proceed until completion (as observed by RP-HPLC). The product (H3K9me3A15C) was purified by semi-preparative RP-HPLC. For desulfurization, H3K9me3A15C was dissolved in tris(2-carboxyethyl)phosphine (TCEP) desulfurization buffer (200 mM phosphate pH 6.5, 6 M GdmCl, 250 mM TCEP). Glutathione (40 mM) and a radical initiator, VA-044 (20 mM), were added, followed by a readjustment of the pH to 6.5. The reaction mixture was incubated at 42°C until completion (verified by RP-HPLC and ESI-MS). The final product, H3K9me3, was purified by semi-preparative RP-HPLC, lyophilized and kept at -20°C for further use. A final characterization of the modified histones was done by analytical RP-HPLC and ESI-MS.

2.10.2.2 Chromatin preparation

As previously described¹⁴⁶, chromatin arrays were reconstituted at a concentration of around 1 μ M per mononucleosome, at a scale of 100 pmol. Array DNA (12x601 with 30 bp linker DNA) was mixed with 1 equivalent (per nucleosome positioning sequence) of recombinant human histone octamer, containing H3K9me3, in reconstitution buffer (10 mM Tris pH 7.5, 10 mM KCl, 0.1 mM EDTA) containing 2 M NaCl. 0.5 molar equivalents of MMTV DNA was added to prevent oversaturation. In the case of reconstituted chromatin containing linker histone, 0.5, 1 or 1.5 equivalents H1.1 were also added to the DNA/octamer mixture. The reactions were gradually dialyzed from high salt buffer (10 mM Tris pH 7.5, 1.4 M KCl, 0.1 mM EDTA) to reconstitution buffer over 12 h with a two-channel peristaltic pump. After the dialysis the reconstituted chromatin arrays were analyzed by non-denaturing 5 % polyacrylamide gel electrophoresis (PAGE) in 0.5 x Tris-Borate-EDTA (TBE) running buffer or on a 0.6 % agarose gel following *ScaI* restriction digest.

2.10.2.3 HP1 labeling

HP1 α was labeled as described in ref.¹⁴⁶. A short tripeptide (Thz-G₂-C₃-CONH₂, Thz: thiazolidine) was synthesized by SPPS. 1 mg Atto532-iodoacetamide (5 eq.) was used to label the tripeptide in 200 mM phosphate pH 7.3, 5 M GdmCl. The reaction was followed by analytical-HPLC and MS, after completion quenched by addition of TCEP, purified by semi-preparative RP-HPLC, and lyophilized. The thiazolidine was opened by treatment with 2 M methoxylamine at pH 5. The labeled tripeptide was finally purified by semipreparative RP-HPLC, lyophilized and kept at -20°C until further use. HP1 α was expressed as a fusion to the *Npu*^N split-intein at its C-terminus, followed by a hexahistidine tag. Expression was induced in *E. coli* BL21 DE3 cells with 0.25 mM IPTG overnight at 18° C. After cell lysis in lysis buffer (25 mM phosphate pH 7.8, 50 mM NaCl, 5 mM imidazole and 1x protease inhibitor/50 mL), HP1 α was purified over Ni-affinity resin and eluted with elution buffer (25 mM phosphate pH 7.8, 50 mM NaCl, 400 mM imidazole). Eluted fractions were pooled and further purified by anion exchange chromatography, using a 1 mL HiTrap Q FF column and a gradient from low (25 mM phosphate pH 7.8, 50 mM NaCl) to high salt buffer (25 mM phosphate pH 7.8, 1 M NaCl). A total of 500 μ l of the expressed HP1-Npu^N fusion constructs at a concentration of 50–100 μ M were applied to a column of 125 μ l SulfoLink resin slurry, containing an immobilized Npu^C peptide¹⁴⁷. After 5 min incubation, the column was drained, followed by washes with wash buffer (100 mM phosphate pH 7.2, 1 mM EDTA, 1 mM TCEP) containing high (500 mM NaCl), intermediate (300 mM NaCl) and low salt (150 mM NaCl). 1 mM of tripeptide in labeling buffer (100 mM phosphate pH 7.8, 50 mM MPAA, 200 mM MESNA, 150 mM NaCl, 10 mM TCEP, 1 mM EDTA) was added to the column and incubated for 16 h at room temperature. The column was drained and the eluate was collected. The column was further washed using elution buffer (100 mM phosphate pH 7.2, 200 mM MESNA, 150 mM NaCl, 10 mM TCEP, 1 mM EDTA). The eluted protein was finally purified by size exclusion chromatography

with a superdex S200 10/300 GL in gel filtration buffer (50 mM Tris pH 7.2, 150 mM NaCl, 1 mM DTT). Fractions containing purified labeled HP1 were pooled and concentrated, glycerol was added to 30 % (v/v), frozen aliquots were stored at -80°C. A final characterization of the purified and labeled HP1 proteins was done by analytical RP-HPLC and ESI-MS.

2.10.2.4 Single-molecule assays

Single-molecule measurements were performed as previously reported¹⁴⁶. Glass coverslips (40 × 24 mm) and microscopy slides (76 × 26 mm) containing four drilled holes on each side were cleaned by sonication in 10 % alconox, acetone and ethanol with washing in milliQ H₂O between each step. The slides/coverslips were incubated for 1 h in a mixture of concentrated sulfuric acid to 30 % hydrogen peroxide (3:1). The coverslips and slides were thoroughly washed with milliQ H₂O, sonicated in acetone for 15 min and then submerged in acetone containing 2 % (3-aminopropyl)triethoxysilane (APTES) for silanization. The slides and coverslips were dried with a nitrogen flow and strips of double-sided tape were sandwiched between a coverslip and a slide to create four channels. The glass coverslips were passivated with a solution of 100 mg/mL mPEG(5000)-succinimidyl carbonate containing 1 % biotin-mPEG-succinimidyl carbonate for 3 h. The channels were subsequently washed with water and T50 buffer (10 mM Tris, 50 mM KCl). For chromatin immobilization, 0.2 mg/mL neutravidin in T50 injected for 5 min, followed by extensive washes with T50 buffer. Then, 500 pM reconstituted chromatin arrays in T50 buffer were injected into the neutravidin treated flow chamber for 5 min, followed by T50 washes and imaging buffer (50 mM HEPES, 130 mM KCl, 10 % (v/v) glycerol, 2 mM 6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox), 0.005 % Tween-20, 3.2 % glucose, glucose oxidase/catalase enzymatic oxygen removal system). Chromatin coverage was observed with a TIRF microscope (Nikon Ti-E) by fluorescent emission in the far-red channel upon excitation by a 640 nm laser (Coherent Obis). Dynamic experiments were initiated by influx of 3 nM Atto532-labeled HP1, as well as the indicated KAP1 concentrations in imaging buffer. All smTIRF experiments were performed at room temperature (22°C). HP1 dynamics were observed with an EMCCD camera (Andor iXon) in the yellow/orange channel using a 530 nm laser line for excitation at 20 W/cm². 10k frames were acquired at 20 Hz over a 25 × 50 μm observation area at a resolution of 160 nm/pixel. Every 200 frames, an image of the chromatin positions in the far-red channel was recorded for drift correction. For each chromatin fiber, an individual trace was extracted using a custom-made semi-automated Matlab (Mathworks) script, as described in ref.¹⁴⁶. After an initial baseline correction and a drift correction, a peak-finding algorithm was employed to detect individual chromatin array positions. Fluorescence intensity traces for each chromatin position were obtained by integrating over a circle of 2-pixel radius. Individual HP1 fluorescence peaks were included based on point-spread-function (PSF) and distance cut-offs. To ensure that only single-molecules were analysed, peaks exhibiting step-wise bleaching kinetics were excluded from the analysis. Kinetics were extracted from fluorescence traces using a semi-automated thresholding algorithm. Cumulative histograms were constructed

from dark and bright intervals and fitted to mono or bi-exponential functions. For intensity analysis, normalized intensity histograms were constructed over several hundred kinetic traces.

The single molecule experiments have been performed in Beat Fierz Laboratory of Biophysical Chemistry of Macromolecules, EPFL.

See Chapter 3.8, Fig. 53| KAP1 FL stabilizes HP1-chromatin interaction.

2.11 In vitro SUMOylation assay

In vitro functional studies have been conducted to relate the novel KAP1 FL structure emerged from SAXS studies to the protein activity.

The in vitro sumoylation assay was conducted using a commercial kit from Abcam (ab139470) using 1 μ M target protein. The mixtures (E1, E2 ubc9, SUMO1, ATP-Mg, target proteins) were set up in 20 μ L of 10 X SUMO reaction buffer and incubated at 37°C for 2h. The reactions were collected every 20 minutes to study the progression of the SUMOylation process. The reaction mixtures were subsequently separated using sodium dodecyl sulfate–polyacrylamide gel electrophoresis SDS-PAGE (4-12%) and were subjected to western blot analysis using anti SUMO1 antibody. Gels were scanned with a Fusion FX 7 (Witec) and analysed using ImageJ¹⁴⁸. Experiments were performed in triplicate.

See Chapter 3.7, Fig. 49| Role of the RING domain in the auto-SUMOylation process of KAP1.

2.12 Virtual screening

Currently, no selective inhibitors able to regulate the transcriptional activity of KAP1 are known. Given the emerging fundamental importance of KAP1 in the regulation of gene expression and given the structural similarity with the other BRDs, we decided to performed structural-based drug discovery on KAP1 Bromodomain with the ultimate goal of discovering a novel selective compounds.

2.12.1 Theoretical basis

Virtual screening is a computational method used to discover new ligands that can later be optimized to give new drug candidates. The main goal of the virtual screening is to reduce the number of chemical compounds to synthesize, screen and test against biological targets. A typical pharmaceutical compound collection contains ~1.5 millions chemical entities and overall the known compounds are estimated to be ~65 millions (UniChem¹⁴⁹, PubChem⁹⁷). HTS (High Throughput Screening) methods allow the screening of hundreds of

thousands of compounds (up to 1 million) against a given target using assays for biological activity. The HTS approach has been adopted by many pharmaceutical companies with the aim to test as many compounds as possible increasing the probability of finding hits among them. However the amount of compounds that can be tested by HTS methods represent just a small portion of the overall available chemical entities. Moreover large screening are expensive and not all the targets are suitable for HTS. To overcome these difficulties, virtual screening has been introduced in the common practice as a tool to analyze large compound collections in order to select the most promising compounds and prioritize them for synthesis or assay.

Virtual screening is divided into (i) Structure based virtual screening (Docking) and (ii) Ligand based virtual screening.

2.12.1.1 Structure based virtual screening

Structural based drug discovery (SBDD¹⁵⁰) begins with the identification of the binding site or active site on the structure of the target protein solved by X-ray crystallography, NMR or modeled after an homologous templates (homology modeling). The active site is the part of the protein where the ligands bind; it is generally a hydrophobic cavity on the protein surface characterized by the presence of hydrogen bond donors and acceptors. Can be visually identified looking at the crystal structure of the protein bound to a known inhibitor or can be predicted based on homologous protein structures.

Protein-Ligand Docking¹⁰¹ is the computational method that mimics the binding of a ligand to a protein. It predicts the “poses” of a ligand in the binding pocket, meaning the geometry (position, orientation, torsional angles) of a ligand within the binding site and it provides a value for the binding affinity representing the strength of the protein-ligand binding (affinity in kcal/mol). Docking algorithms are based on two key components: (i) the *searching algorithm*, to generate the poses of a ligand within the binding site and to explore its conformational space and to optimize the ligand geometry and (ii) the *scoring function*, to identify the most likely poses of a ligand in the binding pocket and to estimate for each compound the strength of the binding (the compounds and then ranked based on this affinity value).

The searching algorithm produces for each ligand different poses and each time a new pose is produced with the help of the scoring function it is compared to the previous one and accepted or rejected according to the score. Then new ligand conformations are generated and the iteration process continues. Most of the software use a stochastic searching algorithm and for this reason an iteration process is recommended.

Based on the *searching methods*, the docking algorithms can be divided into three categories: (i) Algorithms that search the conformational space of the ligand during docking, (ii) Algorithms that search the conformational space of the ligand before docking and (iii) Incremental docking methods.

(i) *Searching the conformational space during docking:*

The optimization of the ligand geometry is explored within the binding pocket. Due to the complexity of the combined approach (searching of the conformational space while docking), stochastic algorithm, such as simulated annealing, must be implemented to sample the conformational flexibility of the ligands. In the simulated annealing approach a ligand starts at a random conformation, it undergoes random changes in a constant-temperature annealing cycle. The energy of the new conformation is compared to the previous one and accepted or rejected. The cycle ends when the maximum number of accepted moves is reached. At this point the temperature is reduced and a new cycle starts until the maximum number of accepted cycle is reached. Docking programs included in this category such as AutoDock¹⁵¹, DOCK¹⁵², GOLD¹⁵³, and DARWIN¹⁵⁴ explore many degrees of freedom: the translation and rotation of one molecule respect to the other (6 DOF), the degrees of freedom of the ligand and the protein (i.e. torsional angles) and the degrees of freedom given by the solvent that can play an important role in the protein-ligand geometry (even if it is usually neglected). These docking methods give very accurate solutions even for flexible and large ligands.

(ii) *Searching the conformational space before docking:*

The optimization of the ligand geometry is performed before docking the molecule in the active site of the target protein and all the low-energy conformations found are rigidly placed inside the binding pocket. The docking algorithms included in this second category only explore the degrees of freedom of the ligand alone. Docking programs included in this category are Slide¹⁵⁵ and Fred¹⁵⁶.

(iii) *Incremental docking methods:*

The ligand is divided into fragments based on rotational bonds, the largest fragment is placed inside the binding pocket and afterwards the other fragments are added in various orientations and scored until the whole ligand is reconstructed. Docking programs included in this category are DOCK¹⁵², FlexX¹⁵⁷, Hammerhead¹⁵⁸ and HOOK¹⁵⁹.

The *scoring function* has two main objectives: discriminate between the different poses of a ligand and estimate the binding affinity of each ligand in order to classify them. Several factors contribute to form the receptor-ligand binding complex. Among them, the most important are hydrophobic, electrostatic and van der Waal's interactions, hydrogen bonding and solvation effects. The scoring functions implemented in the docking programs usually simplify the evaluation of the binding complexes in order to reach a compromise between accuracy and computational effort, not taking fully into account all the parameters previously mentioned.

An important parameter to consider during docking is the flexibility of the ligands. Almost all the docking programs allow the flexibility of the compounds altering them with several rotatable bonds to predict the binding interaction with a protein target usually maintained fixed in its crystal structure conformation. Moreover some programs generate, upon binding of the ligands, discrete sets of receptor conformations used to perform flexible-ligand rigid-target docking.

The docking program used in this thesis, AutoDock Vina¹⁵¹, is currently the most cited docking methods with a total number of 6'100 citation including 3'200 in the past two years. AutoDock Vina¹⁵¹ has displayed great accuracy in predicting the bounds conformation of ligands in the active site as well as great precision in estimating their binding energies. These achievements are due to the implementation of a carefully calibrated global scoring function composed of linear added energies terms including dispersion/repulsion coefficients, hydrogen bonding terms, electrostatic interactions, deformation terms (i.e. torsional angles), hydrophobicity. Weight constants are then assign to each term to balance their contributions. The scoring function has been calibrated with 1'300 complexes with an estimated standard error of 2.5 kcal/mol. AutoDock Vina¹⁵¹ explores the conformational space of the ligands in the binding pocket during the docking process applying a stochastic searching method, specifically the simulated annealing algorithm previously described.

2.12.1.2 Ligand based virtual screening

Ligand based virtual screening (LBVS¹⁶⁰) represents the method of choice to identify active compounds when no high-resolution structure of the target protein is available. The method is based on the assumption that ligands similar to an active compound are more likely to show activity. Active compounds can be selected following two approaches (i) *shape-based similarity screening* and (ii) *pharmacophore-based approach*.

The *shape-based similarity* methods search for novel active compounds chemically similar (i.e. same scaffold) to known active ligands used as a query structure. However, active ligands with different shape compared to the query structure are likely to be missed during the virtual screening and more importantly the screening performances for some ligands have been shown to be equivalent to the random identification of active compounds.

According to the IUPAC definition a pharmacophore is defined as “an ensemble of steric and electronic features necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”. In drug discovery a pharmacophore is a set of features found in a series of active compounds including hydrogen-bond donors and acceptors, positively and negatively charged groups and hydrophobic regions. The *pharmacophore-based methods* creates patterns of molecular properties

(pharmacophores) found in active ligands and use them to screen chemical libraries searching for new active molecules. Each time a specific pattern is found in a new compound a similarity value is calculated.

2.12.2 Practical application

Structural based virtual screening was performed to discover new compounds of KAP1 PHD-Br domain (2RO1⁶², residues 624-812). The protein structure was solved by nuclear magnetic resonance (NMR) in 2008 (2RO1⁶²). Two protein conformations were used in the screening: The KAP1 structure in which the ZA loop was maintained in the original position found in the NMR structure called “open” conformation and another structure in which the ZA loop was modeled after the common conformation found in the other BRDs (6HDN¹⁶¹, 4WHU (*to be published*), 6HDQ¹⁶¹, 6MO9¹⁶², 5H1T¹⁶³, 5VOM¹⁶⁴, 6HDO¹⁶¹) (*close conformation*) called “close” conformation. Specifically the TRIM24 Bromodomain (5H1T¹⁶³) was used to perform homology modeling using SWISS Model server¹³⁶. The screening can be divided into four steps including (i) *Protein preparation*, (ii) *Ligand preparation*, (iii) *Docking* and (iv) *Analysis*.

- (i) The target was prepared for docking using AutoDockTolls¹⁶⁵ (ADT4.2), a GUI interface for setting and running docking simulations. The Windows version of AutoDockTools¹⁶⁵ is included in the MGLTools package (MGLTools-1.5.6). AutoDockTools¹⁶⁵ was used to add the polar hydrogen to the protein structure (file.pdb) and save it in file.pdbqt format. The binding site was identified comparing the structure of the KAP1 PHD-Br domain with the other X-ray structures of BRDs in complex with inhibitors (6HDN¹⁶¹, 4WHU (*to be published*), 6HDQ¹⁶¹, 6MO9¹⁶², 5H1T¹⁶³, 5VOM¹⁶⁴, 6HDO¹⁶¹). Moreover AutoDockTools¹⁶⁵ was used to select the searching space, meaning the region where the docking is performed. This operation known as *grid generation* aims to define a docking box that maximizes the accuracy of the binding pose prediction. The center of the grid was defined in x, y, z coordinates (-8.073, -11.952, -8.409) and a grid box 20x22x22 was created with a grid point spacing of 1.0 Å.
- (ii) In this study, 13'433 compounds were screened against the structure of KAP1 PHD-Bromo domain. These compounds are part of a bigger chemical library, The Chemically Diverse Collection (CDC) available at the Biomolecular Screening Facility (BSF), EPFL. The 13'433 compounds cover the 85% of the chemical diversity of the entire library and for this reason they have been selected for the first round of the screening. If active compounds are identified the scaffold will be used as a probe to screen again the whole library using a ligand based approach. Open Babel ¹⁶⁶(openbabel-2.4.1), a chemoinformatics tool used for the conversion of different chemical files formats, was used to convert the 2D structures (file.sdf) into file.mol2 format and to create the corresponding 3D conformations using the command gen3d. Afterwards, MGLTools

package was used to convert the 3D file.mol2 into file.pdbqt format adding at the same time the polar hydrogen.

- (iii) The docking was performed using AutoDock Vina¹⁵¹. A configuration file was prepared containing the structure of the receptor (file.pdbqt) the center and dimension of the grid box (x, y, z) and the number of modes (9) corresponding to the number of poses to be generated. At the end of the process a text file containing the binding energy for each of the 9 poses for each ligand was created.
- (iv) The compounds were classified based on the value of their binding energy. Specifically the best pose for each compound showing the lowest binding energy was retrieved. Afterwards the first 136 compounds (with binding energies between 10 and 8 kcal/mol), corresponding to ~1% of the total number of ligands used in the screening, were visually analyzed with PyMOL (PyMOL 2.3). The preferred binding pocket and the orientation of the ligands in the binding site were compared to the crystallized BRDs in complex with inhibitors. The first 136 compounds found for the “open conformation” and the “close conformation” were also compared to find common structures. The selected compounds were retrieved and they were analyzed with Scaffold Hunter¹⁶⁷, a Java-based open source tool for clustering and classification of chemical compounds based on their scaffold. The most representative scaffolds were retrieved and compared with the structures of the active BRDs inhibitors to find chemical similarities.

See Chapter 4.4, **Fig. 67**| ZA loop of KAP1 compared to the other BRDs, **Fig. 71**| Grid box preparation, **Fig. 72**| Binding mode of KAP1 inhibitors, **Fig. 73**| and **Fig. 74**| Scaffold Hunter analysis.

2.13 Differential Scanning Fluorimetry

Finally, 272 positive compounds, 136 compounds for each PHD-Br conformation, were selected together with 10 compounds with high binding energy (-4 kcal/mol) and 8 BRDs inhibitors (BIC1¹⁶⁸, NP1¹⁶⁹, Ischemin¹⁷⁰, JQ1¹⁰⁶, I-BET762¹⁰⁷, I-BET151¹⁷¹, PF-1¹⁷², RVX-208¹⁷³). The compounds were ordered from the Biomolecular Screening Facility. The ECO-550 dispenser (LABCYTE) for nanoliter pipetting was used to prepare 384 well PCR plates, in triplicates, with 25 nL of compound in each well at an initial concentration of 10 mM in 100% DMSO. The concentration of the compounds in the final assay volume of 25 μ L was 10 μ M and 0.1% DMSO. The compounds were test by Differential Scanning Fluorimetry (DSF).

Differential scanning fluorimetry (DSF)¹⁷⁴ is a scanning method used to detect changes in the melting temperature of a protein upon binding of inhibitors. The temperature at which the protein unfolds, the melting

temperatures or T_m , it is measured by an increase in fluorescence emission of a dye with affinity for the hydrophobic regions exposed when the protein unfolds. The difference in T_m in the presence or absence of an inhibitor is related to the binding affinity of the compound that can stabilize the protein structure leading to a higher melting temperature value (**Fig. 22**). DSF can be performed in a conventional real-time PCR machine using 96 or 384 well plates (FrameStar®, 4Tititude). DSF do not show direct binding but stabilization mainly due to binding events. The binding will be confirmed with direct methods such as ITC or Biocore.

In this study, 21 μ L of KAP1 PHD-Br domain at a concentration of 0.2 mg/mL in 10 mM HEPES pH 7.5 and 300 mM NaCl together with 4 μ L of Sypro Orange Dye (Sigma-Aldrich) diluted in water to a final working concentration of 5X, were added to a 384 well PCR plate containing 25 nL of the selected compounds in each well. The ideal protein concentration for the thermal shift assay was previously determined measuring the T_m of the protein without the addition of ligands testing three protein concentrations (2 mg/mL, 0.2 mg/mL and 0.02 mg/mL) with the addition of 1% or 5% DMSO in the same buffer condition as the one used for the final assay (10 mM HEPES pH 7.5, 300 mM NaCl).

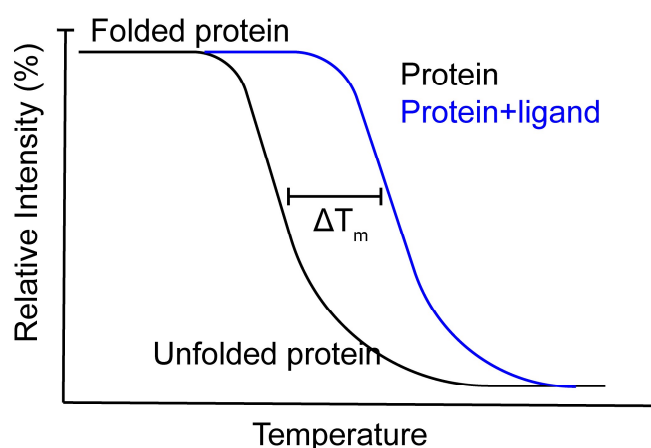


Figure 22| Schematic representation of a thermal shift assay. The melting temperature curve of a protein without the presence of an active ligand is shown in black. The T_m curve of a protein after binding an active compound that stabilizes the target structure leading to a higher melting point is shown in blue.

See Chapter 4.5, **Fig. 75|** PHD-Br domain purification, **Table 3|** List of compounds that bind KAP1, **Fig. 76|** Boltzmann fit of the melt curve data, **Fig. 77|** F2031-0348 binding poses.

Chapter 3 Biophysical and structural characterization of KAP1

Disclaimer: This chapter is adapted from the following article, with permission of all the co-authors:

Giulia Fonti, Maria J. Marcaida, Louise C. Bryan, Sylvain Traeger, Alexandra S. Kalantzi, Pierre-Yves J.L. Helleboid, Davide Demurtas, Mark D. Tully, Sergei Grudinin, Didier Trono, Beat Fierz, Matteo Dal Peraro.

“KAP1 is an antiparallel dimer with a functional asymmetry”

Life Science Alliance, 2019.

My contribution: I performed and designed all the experiments apart smTIRF. I analyzed and interpreted the data and prepared the original draft of the paper.

The single-molecule experiments have been performed by Bryan Louise in Beat Fierz laboratory.

Sylvain Traeger from Dal Peraro laboratory developed the clustering algorithm and performed the clustering analysis on KAP1. Details can be found in the following paper: “CLONe: Clustering Based on Local Density Neighborhoods for Application to Integrative Dynamic Modeling” Traeger Sylvain, Tamo Giorgio, Aydin Deniz, Fonti Giulia, Dal Peraro Matteo, in preparation.

This chapter is organized as follows. First I will discuss the biophysical characterization of KAP1 together with the SAXS data collection and flexible SAXS data fitting using a novel integrative modeling approach. I will present the structure of full-length KAP1 and I will discuss the implications of the unexpected organization on the protein function and on the interaction with binding partners, specifically the heterochromatin protein 1.

KAP1 (KRAB (Krüppel-associated box)-domain associated protein 1)) also known as TIF1 β (Transcription Intermediary Factor 1 β) or TRIM28 (Tripartite Motif containing protein 28) is a central regulator that controls the fate of the genetic material by recruiting transcription factors and altering the chromatin environment^{71,104}. KAP1 is thus essential for early development, and has been linked to fundamental cellular processes such as cell differentiation^{46,175}, gene silencing^{39,176,177,43}, transcription regulation^{178,179,180,54} and DNA damage response^{181,70,49,51,48}. Despite its emerging fundamental importance, KAP1 and its interaction with binding partners are still poorly characterized from a structural standpoint. Crystal structures of the coiled-coil domain with or without additional B-box or C-terminal domains have been solved for some TRIM family members such as TRIM25^{182,183}, TRIM5^{184,185}, TRIM20¹⁸⁶ and TRIM69¹³⁹) showing that the RBCC is an elongated domain with an antiparallel homodimer conformation. However, there is insufficient structural information

about full-length proteins of the TRIM family. We applied biophysical techniques, small angle X-ray scattering data, molecular modelling and single-molecule experiments to present for the first time the structure of KAP1 full-length and to elucidate its interaction with full-length HP1.

3.1 Expression and purification of recombinant proteins

3.1.1 The RBCC domain

Cloning. The RBCC construct contains an N-terminal His₆ tag followed by the TEV (Tobacco Etch Virus) protease cleavage site. The construct starts in position 23 and it ends in position 418, at the end of the predicted TSS domain (M_w 46 kDa). The RBCC construct includes the RING domain, the B-box 1 and the B-box 2 domains and a long coiled-coil domain (**Fig. 23A**). Previous experiments shown that this construct encodes for a soluble protein⁴². The RBCC domain was successfully cloned inside the first MCS (Multi Cloning Site) of the pETDuet-1 vector, between SACI and HINDIII restriction sites. The PCR-amplified sequence was ligated into the pETDuet-1 vector. Transformation of HB101 cells (RecA deficient and streptomycin resistant), was performed for plasmids replication. Positive clones were selected and their plasmids were extracted with mini-preps (66 ng/μL). The presence of the RBCC construct inside the pETDuet-1 vector (pETDuet1-RBCC) was confirmed by sequencing (Microsynth) (**Fig. 23B**).

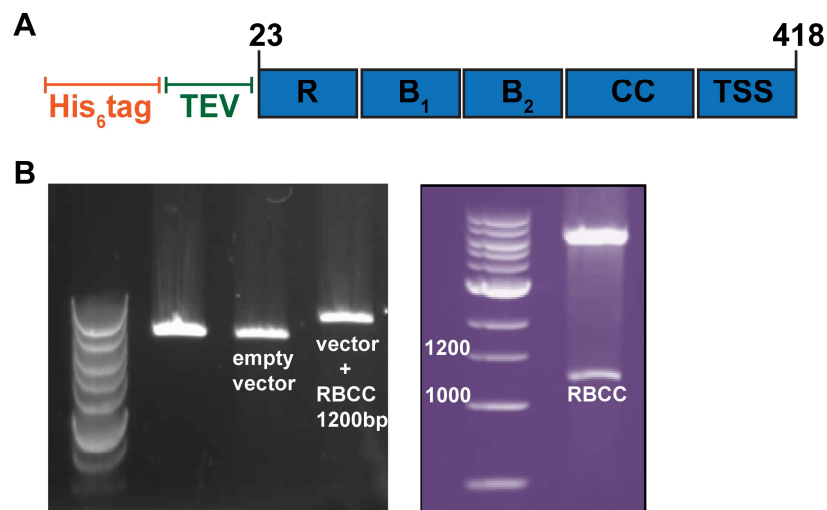


Figure 23| Cloning of RBCC construct. **A.** From the N-terminus: poly-histidine tag (His₆ tag) in orange, TEV protease cleavage site in green, RING domain (R), B-box 1 (B₁) and B-box 2 (B₂), coiled-coil domain (CC), TSS domain in blue. The construct starts in position G23 and it ends in position T418 **B.** pETDuet1-RBCC vector digested with SacI and HIND III restriction enzymes: 1% (v/w) agarose gel stained with Gel Red, 1kb DNA ladder. The digestion shows the presence inside the pETDuet-1 vector of the RBCC construct of 1200 bp.

Expression and large-scale purification. The expression test for the RBCC domain was performed using four *E. coli* stains (BL21 (DE3), BL21 (DE3) pLys, Rosetta (DE3), Rosetta (DE3) pLys), three growth temperatures (20°C over night, 30°C 3h, 37°C 4h) and three IPTG (IsoPropyl β -D-1-ThioGalactopyranoside) concentrations (0.1 mM, 0.5 mM, 1 mM). Each *E. coli* strain was transformed with the pETDuet1-RBCC vector, using a heat shock protocol. The RBCC domain was nicely expressed in all the cell lines at every induction condition. Moreover we found that the maximum level of soluble and purified protein was obtained from Rosetta (DE3) strain grown at 20°C over night after the induction of protein expression with 1 mM IPTG (**Fig. 24**).

Afterwards, 6 L of Rosetta (DE3) cells expressing the RBCC domain were purified using affinity chromatography and gel filtration chromatography. The supernatant was applied to a 5 ml HisTrap HP column (GE Healthcare). The protein was buffer exchanged using a HiPrep 26/10 Desalting column (GE Healthcare). The His₆ tag located at the N-terminus of the protein was cut incubating the sample over night at 4°C with a ratio 1:20 of TEV protease. The sample was then concentrated and injected into a Superdex 200 10/300 GL column (GE Healthcare) (**Appendix 1A**).

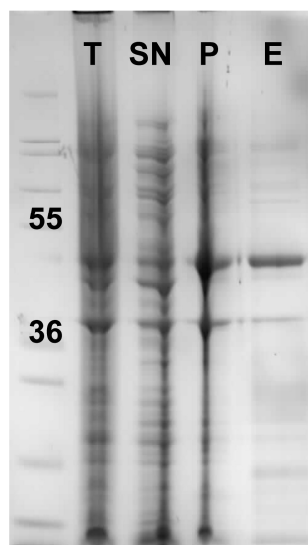


Figure 24| RBCC mini-purification. The expression of the protein was induced in Rosetta (DE3) strain with the addition of 1 mM IPTG. The cells were grown over night at 20°C. The total extract (T), supernatant (SN) and pellet (P) samples were collected. The SN was applied to the NiNTA resin and the eluted fractions were collected and pooled together. Afterwards, the protein was purified by affinity chromatography.

3.1.2 KAP1 FL constructs (Δ KAP1, KAP1 FL)

Cloning. KAP1 FL has never been purified before. To maximize the probability of obtaining a soluble protein, we first designed a construct in which the N-terminal domain was the same as the previously purified

RBCC domain (starting in position G23) and the C-terminal domain was the same as the PHD-Bromo domain construct solved by NMR (2RO1)⁶² (ending in position P812). The construct called Δ KAP1 was preceded by an N-terminal His₆ tag and followed by the TEV protease cleavage site. The protein was cloned inside the first MCS of pETDuet-1 vector (M_w 87 kDa) (**Fig. 25A**).

Subsequently we designed a second construct, named KAP1 FL, in which the whole protein (M1-P835) was cloned inside the first MCS of pETDuet-1 vector (M_w 91 kDa) preceded by the N-terminal His₆ tag and the TEV cleavage site (**Fig. 25B**). The DNA coding sequences were ordered from GenScript.

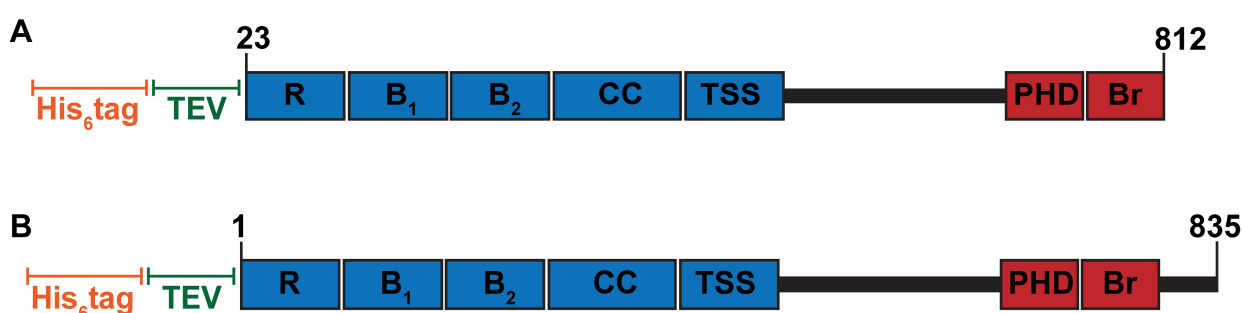


Figure 25| KAP1 FL constructs. **A.** Δ KAP1 starts in position G23 and ends in position P812 **B.** KAP1 FL starts in position M1 and ends in position P835. For both constructs from the N-terminus: poly-histidine tag (His₆ tag) in orange, TEV protease cleavage site in green, RING domain (R), B-box 1 (B₁) and B-box 2 (B₂), coiled-coil domain (CC), TSS domain in blue, PHD-Bromo domain (PHD-Br) in red.

Expression and large-scale purification. For Δ KAP1 and KAP1 FL the expression, solubility test and mini-purification tests were performed under the same buffer condition and following the same protocol as described previously in this thesis (see Method Chapter 2.2). Δ KAP1 was expressed only in Rosetta (DE3) pLys, and the conditions that maximized the production of a soluble protein were 20°C, growth over night with the addition of 1 mM IPTG (**Fig. 26A**). KAP1 FL was nicely expressed in all the cell lines at every induction condition. Moreover we found that the maximum level of soluble and purified protein was obtained from Rosetta (DE3) strain grown at 20°C over night after the induction of protein expression with the addition of 1 mM IPTG (**Fig. 26B**). Afterwards, 6 L of each construct were purified using affinity chromatography and gel filtration chromatography as described in Method Chapter 2.3 (**Appendix 1B,C**).

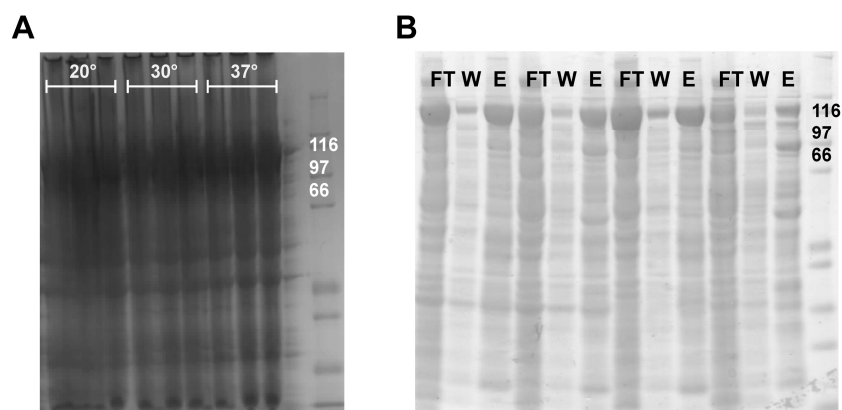


Figure 26| KAP1 FL constructs mini-purification test. **A.** Δ KAP1 was expressed only in Rosetta (DE3) pLys strain. The other three strains did not show any protein expression (gels not reported). The protein was equally expressed at 20°C, 30°C and 37°C with the addition for each temperature of 0.1 mM, 0.5 mM or 1 mM IPTG **B.** KAP1 FL was expressed in all the cell lines at 20°C over night with the addition of 1 mM IPTG. Here is reported the gel showing the mini-purification of the protein from Rosetta (DE3) pLys (lanes 1-3), from BL21(DE3) (lanes 4-6), from BL21 (DE3) pLys (lanes 7-9) and from Rosetta (DE3) (lanes 10-12). FT: flow-through, W: wash, E: elution.

3.2 Biophysical characterization

3.2.1 KAP1 is a homodimer in solution

The TRIM family members have been shown to form homodimers in solution through their coiled-coil domain.^{182,183,184,186,187} Biochemical and bioinformatics analyses suggest that this organization could be a common feature for all the TRIM family members¹⁸³. Despite these evidences, KAP1 has been previously shown to fold preferentially as a homotrimer alone in solution and in complex with the KRAB domain of the KRAB ZFPs.⁴² Furthermore, heterodimers with TRIM24 and further associations as hexamers have been shown.¹⁸⁸ To shed light on the oligomerization state of KAP1, size exclusion chromatography coupled to multi angle light scattering (SEC-MALS) experiments were performed. Three different constructs (**Fig. 23A**, **Fig. 25A,B**) have been used: (i) the RBCC domain (RBCC, 23-418), (ii) a KAP1 construct nearly covering the whole sequence (Δ KAP1, 23-812) and (iii) the complete KAP1 FL protein (KAP1 FL, 1-835). SEC-MALS experiments were performed across a concentration range of 10 to 40 μ M. No concentration dependence in either elution volume or mass estimation was observed in these conditions. The resulting molecular weights (M_w s) were determined to be 88 kDa for the RBCC domain, 183 kDa for Δ KAP1 and 190 kDa for KAP1 FL, in agreement with dimeric species of expected M_w s of 92 kDa (RBCC), 175 kDa (Δ KAP1) and 182 kDa (KAP1 FL) (**Fig. 27A**). Moreover, to confirm this result, a sedimentation velocity analytical ultracentrifugation (SV-AUC) experiment was performed for the RBCC construct. The data analysis was

performed using the software Sedfit.¹¹³ The data were fitted with a continuous concentration vs molecular mass, $c(M)$ model. The rmsd (root-mean-square deviation) for the fit was 0.0546, and the sedimentation coefficient was 2.3 S with an estimated M_w of 89 kDa (**Fig. 27B**).

Under the concentration range tested, both SEC-MALS and SV-AUC consistently indicated a dimeric conformation for KAP1.

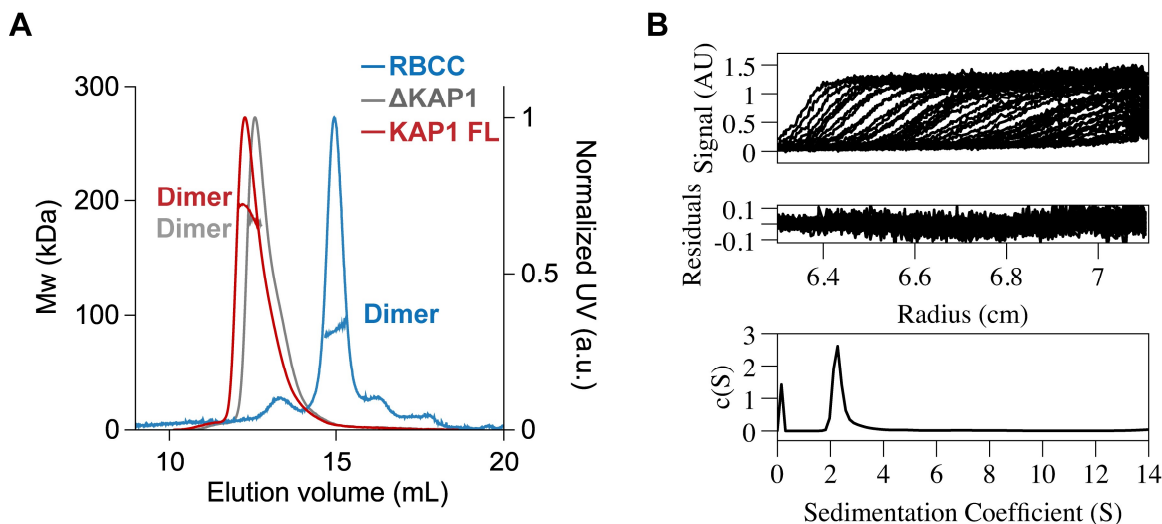


Figure 27| Biophysical characterization of KAP1 FL constructs. **A.** SEC-MALS analyses of the KAP1 constructs. The traces show the normalized elution profile measured at 280 nm (right axis) and the calculated molecular weight of the selected peaks in kDa (left axis): 88 kDa for the RBCC domain in blue, 183 kDa for Δ KAP1 in light grey and 190 kDa for KAP1 FL in red **B.** Analytical ultracentrifugation analysis of the RBCC domain. The first panel shows the raw sedimentation velocity data, shown as fringe displacements, which is proportional to the protein concentration, plotted against the distance from the center of rotation. The second panel shows the residuals after fitting to a continuous concentration vs molecular mass, $c(M)$, model. The third panel shows the distribution of the s values obtained from the sedimentation velocity data.

3.2.2 KAP1 is well folded and show an α -helical structure

Circular dichroism was used to obtain some preliminary information regarding the secondary structure of the purified constructs and to assess whether they were correctly folded under the current purification conditions. The shape of the CD spectra obtained for the RBCC, Δ KAP1 and KAP1 FL constructs were associated with a α -helix structure, showing two minima at 208 nm and at 222 nm (**Fig. 28A,B,C**).

In order to check the folding state of the RBCC construct and to calculate its melting temperature (T_m), the wavelength was fixed at 220 nm and the temperature was raised from 20°C to 90°C, with an increment of 2°C/min and recording data at every 1°C. At the end of the data collection, the temperature was restored at 20°C and the sample was supplied with 1mM DTT (Dithiothreitol) and 50 μ M ZnSO₄. Since the RING domain as well as the two B-box domains are zinc finger domains in which the zinc ions are coordinated by histidine and cysteine residues another CD spectrum was recorded to check whether the protein structure was changing upon addition of reducing agents or zinc ions (**Fig. 29A,B**). In the case of KAP1 FL variable temperature CD spectra were collected as previously described in the Method Chapter 2.5.5 (**Fig. 30**).

For both constructs, when the temperature was raised the secondary structures started to denature providing a strong indication that both proteins were folded. Moreover, we were able to calculate a melting temperature of 45°C for the RBCC domain and 47°C for KAP1 FL.

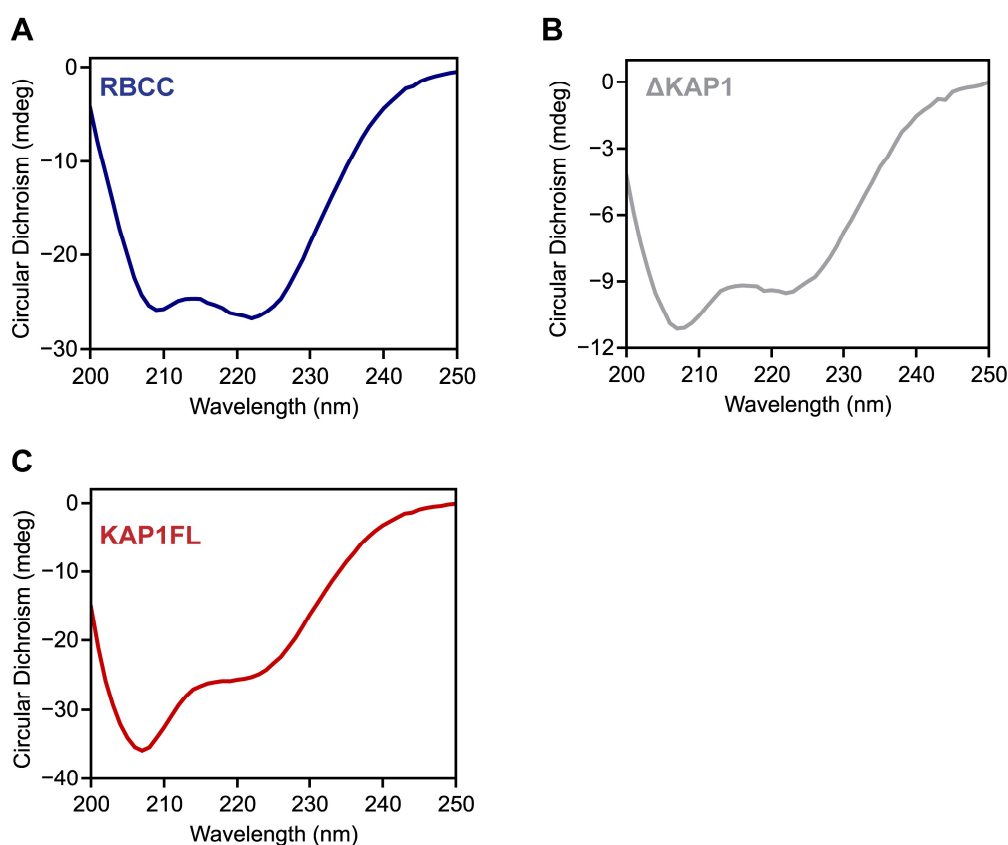


Figure 28| Circular Dichroism (CD) analysis of the KAP1 constructs. CD spectra of **A.** RBCC domain **B.** Δ KAP1 **C.** KAP1 FL. CD data were collected using a Chirascan CD Spectrometer and a 0.1 cm quartz cell. Sample concentration ranged from 0.2 mg/mL to 0.8 mg/mL in Phosphate Buffer (1x PBS). Scans were measured from 200 to 250 nm, under continuous scanning mode with a 1 nm data pitch, a scan speed of 50 nm/min and a response time of 1 s. Each spectrum corresponds to the average of three spectra buffer subtracted. The spectra confirmed that all the constructs were well folded.

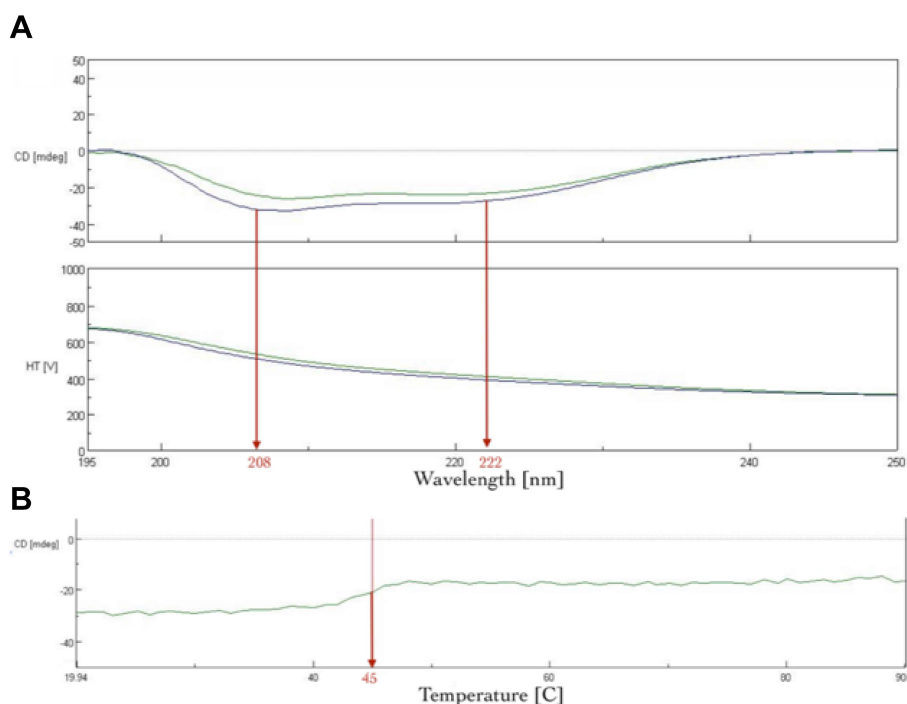


Figure 29| RBCC variable temperature CD spectra: A. The top graph shows in black the CD spectrum obtained from the mean of three measurements performed on the purified RBCC construct in PBS buffer at a concentration of 0.2 mg/mL, and in green the CD spectrum of the same sample supplied with 1 mM DTT and 50 μM ZnSO₄. Both curves are baseline subtracted. The bottom graph shows the variation of the current measured in Volts (V) **B.** The graph shows the results of the variable temperature CD run. In order to detect the temperature at which the secondary structures started to denature, the wavelength was fixed at 220 nm while the temperature was increasing. A conformational change happened at 45°C.

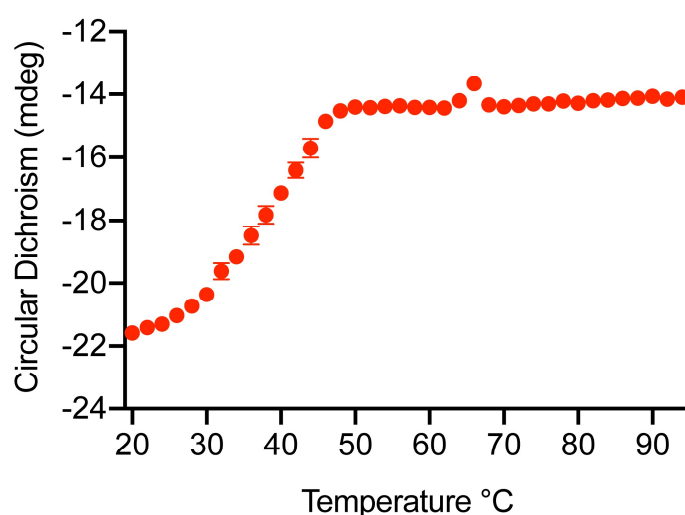


Figure 30| KAP1 FL variable temperature CD spectra: Variable temperature CD spectra were recorded from 200 to 250 nm within 20°C and 95°C in triplicates. The graph shows the result of the three repetitions at 220 nm. The conformational change happened at 43°C.

3.3 Crystallization trials

3.3.1 The RBCC domain formed crystals but they diffracted at low resolution

First we crystallized the RBCC domain (**Fig. 23**). We found two positive crystallization conditions in the MB Class Suite (MBC-I) screen (Quiagen). The first condition was made of 500 mM NaCl, 50 mM Tris-HCl pH 8.5, 10% (v/v) PEG 4'000; the second was made of 100 mM Litium Sulfate, 100 mM Na Citrate pH 5.6, 12% (v/v) PEG 6'000. Moreover we found another positive crystallization condition in the Clear Strategy Screen I screen (Molecular Dimensions), made of 800 mM Na Formate and 25% PEG 2'000 MME. The protein crystallized at a concentration of 10 mg/mL in 20 mM HEPES pH 7.5, 300 mM NaCl, 3% (v/v) glycerol, 2.5 mM TCEP. We crystallize both the His₆ tag and the non-tag protein. We used the vapor diffusion method, either sitting drop or hanging drop. We obtained crystals with both methodologies but the crystals with bigger volume and a length of 100 μ m were obtained by sitting drop method. However the crystals diffracted at a low resolution (9-10 Å) (**Appendix 2**).

Afterwards, we tried to improve the crystallization conditions. We changed (i) the protein concentration (range from 5 to 10 mg/mL), (ii) the buffer composition (300-150 mM NaCl, 5%-0% (v/v) glycerol, Buffer pH), (iii) the ration protein: reservoir (1:1, 1:2, 2:1), (iv) the cryo-protectors (glycerol 20-25%, PEG (Polyethylene glycol), MPD (2-Methyl-2,4-pentanediol)). Moreover we tried to improve the results adding additives from the Hampton Additive Screen. We were able to obtain crystals with a length of 800-900 μ m, however they diffracted at low resolution. Consequently, we decided to attempt the crystallization of constructs with different length and fused with proteins such as lysozyme to assist the crystallization process.

3.3.2 RBCC-T4 Lysozyme

We were not able to solve the structure of the RBCC domain since the crystals diffracted at low resolution. Our first hypothesis was that the coiled-coil domains were interacting between each other in the crystal lattice perturbing the formation of a well order repeated structure. We went back to literature and we found that in 2014 was solved the antiparallel crystal structure of the Rhesus Trim5alpha B-box 1 coiled-coil domain in complex with T4-Lysozyme (4TN3)¹⁸⁴. The T4-Lysozyme was located in the N-terminal part of the construct, separated from the coiled-coil domain by an unsolved loop of 42 residues. The lysozyme interacts with the coiled-coil domain, stabilizing the crystal lattice giving crystals that diffracted at 3.4 Å.

We decided to reproduce this result and we designed our own construct based on the similarity between the KAP1 RBCC domain and this TRIM family member. We observed that we had 40 residues between the end

of the coiled-coil domain (M378) and the end of the construct (T418). Consequently we decide to attach the T4-Lysozyme sequence immediately at the end of the TSS domain (M378), obtaining the new construct RBCC-T4-Lysozyme (pETDuet1-RBCC-T4L, M_w : 65 kDa) (**Fig. 31A,B**).

Expression and purification. The RBCC-T4L was cloned inside the MCS-I of pETDuet1 vector (pETDuet1-RBCC-T4L). The expression test, the solubility test and the mini purification test were performed following the same protocol previously described in the Method Chapter 2.2. We observed that the protein was expressed only in one strain, BL21 (DE3) and the conditions that maximized the production of a soluble protein were 30°C, growth for 4h with the addition of 1mM IPTG (**Fig. 31C**).

The RBCC-T4L formed crystals but they diffracted at low resolution. We were able to crystallize the RBCC-T4L construct. The crystals showed the same morphology as the previous RBCC crystals. We found two positive crystallization conditions in the Clear Strategy I screen (CS-I) (Molecular Dimensions). The first condition was made of 800 mM Na Formate, 100 mM Tris-HCl pH 8, 8% (v/v) PEG 20'000, 8% (v/v) PEG 550 MME; the second was made of 800 mM Na Formate, 100 mM Na Acetate pH 5.6, 8% (v/v) PEG 20'000, 8% (v/v) PEG 550 MME. Moreover we found another positive crystallization condition in the MB Class I (MBCI) screen (Quiagen), made of 500 mM KCl, 50 mM MOPS pH 7, 12% (v/v) PEG 4'000, 20% glycerol. The protein crystallized at a concentration of 10 mg/mL in 20 mM HEPES pH 7.5, 300 mM NaCl, 3% glycerol, 2.5 mM TCEP. Also in this case crystals diffracted at low resolution (10-20Å) (**Appendix 2**).

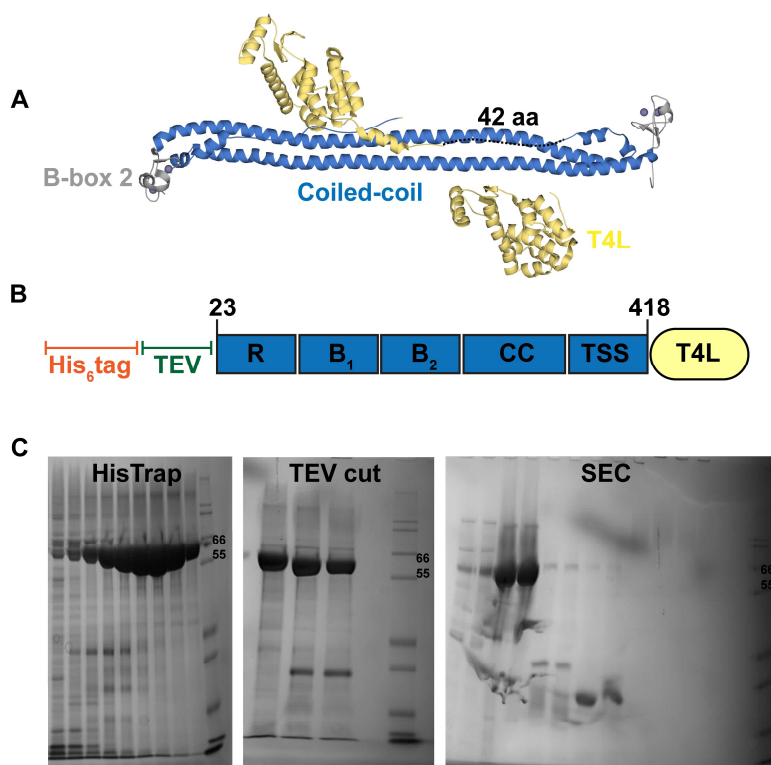


Figure 31| RBCC-T4L sequence and purification: **A.** Rhesus Trim5alpha structure: The B-box 2 (light grey) and the coiled-coil domain (blue) were crystallized in complex with the T4-Lysozyme (4TN3¹⁸⁴). The coiled-coil and the

lysozyme are separated by an unsolved loop of 42 residues **B**. The cartoon model represents the new RBCC-T4L construct with the C-terminal T4-Lysozyme separated from the end of the coiled-coil domain by a length of 40 residues (TSS domain) **C**. SDS-PAGE (10% polyacrylamide gels) showing the pure fractions after affinity chromatography purification (HisTag, gel on the left), cut of the N-terminal His₆ tag with TEV protease (gel in the middle) and gel filtration chromatography (S200 16/600, gel on the right). The RBCC-T4L construct was successfully purified.

3.3.3 New constructs for RBCC

We performed limited proteolysis experiments on the RBCC domain to identify new protein fragments for crystallization trials. This approach was based on the assumption that the enzymes cut in areas displaying heightened structural flexibility. Therefore, we considered the resulting protein fragments to be structurally robust making them ideal candidates for the new crystallization trials.

We used three different proteases: (i) Elastase (Promega) (ii) Trypsin (Thermo Scientific) and (iii) Chymotrypsin (Roche). The initial stock solutions at 1 mg/mL were diluted to 1:10, 1:100 and 1:1000 in 20 mM HEPES pH 7.5 and 150 mM NaCl. Afterwards, 4 μ L of the final solutions were added to 15 μ L of protein at 0.6 mg/mL in the same buffer and incubated on ice for 30 minutes. The reactions were stopped by the addition of 1 μ L of PMSF (Phenylmethylsulfonyl fluoride) at 100 mM. The samples were analyzed by SDS-PAGE (**Fig. 32**).

The fragments appearing with different proteases were subsequently analyzed by mass spectrometry to retrieve their molecular weights. Afterwards we screened the RBCC sequence to find protein fragments corresponding to each molar mass (**Fig. 33**). Following this procedure we were able to identify the constructs listed in **Table 1**.

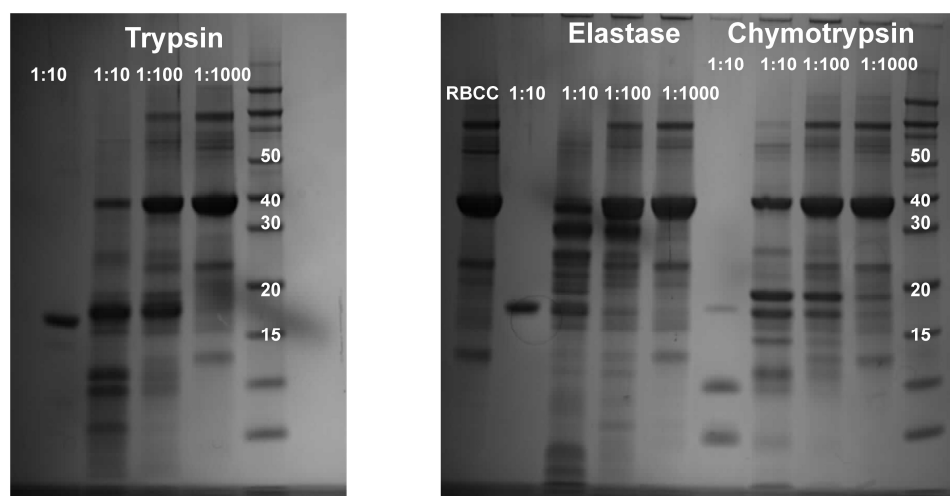


Figure 32| Limited proteolysis experiment. SDS-PAGE analysis of the fragments obtained from the digestion of the RBCC with Trypsin (gel on the left), Chymotrypsin and Elastase (gel on the right). The gels show clear bands between

30 kDa and 15 kDa. The expected molecular weights for constructs containing two B-box domains, one B- box domain or only the coiled-coil domain were 34 kDa, 26 kDa and 21 kDa.

A

GEGSAGGEKRSTAPSAAASASASAAASSPAGGGAELELLEH**CGVCR**
ERLRPEREPRLLPCLHSACSACLGPAAPAAANSSGDGGAAGDGTVV
DCPVCKQQCFSDKIVENYFMRDSGSKAATDAQ**DANQCCTSCEDNAP**
ATSYCVECSEPLCETCVEAHQRVKYTKDHTVRSTGPAKSRDGE**ERTVY**
CNVHKHEPLVLFCESCDTLT**CRDCQLNAHKDHQYQFLEDAVRNQKRL**
LASLVKRLGDKHATLQKSTKEVRSSIRQVSDVQKRVQVDVKMAILQIM
KELNKRGRVLVNDAQKVTEGQ**QERLERQHWMTMTKIQKHQEHILRFAS**
WALESDNNTALL**LSKKLIYFQLHRAL**KMIVDPVEPHGEMKFQWDLNAW
 TKSAEAFGKIVAERP GTNST

Elastase 1:1000

Peak 3.57 21'726.0010 Da

B

Residues : 197-382

PAKSRDGE**ERTV**CNVHKHEPLVLFCE**SCDTLT**CRDCQLNAHKDHQYQFL
EDAVRNQKRLLASLVKRLGDKHATLQKSTKEVRSSIRQVSDVQKRVQV
DVKMAILQIMKELNKRGRVLVNDAQKVTEGQ**QERLERQHWMTMTKIQKH**
QEHILRFASWALESDNNTALL**LSKKLIYFQLHRAL**KMIVDP

C

Residues : 114-305

VVDCPVCKQQCFSDKIVENYFMRDSGSKAATDAQ**DANQCCTSCEDNAP**
ATSYCVECSEPLCETCVEAHQRVKYTKDHTVRSTGPAKSRDGE**ERTVYC**
NVHKHEPLVLFCESCDTLT**CRDCQLNAHKDHQYQFLEDAVRNQKLLA**
SLVKRLGDKHATLQKSTKEVRSSIRQVSDVQKRVQVDVKMAILQIMKE

Figure 33| New RBCC fragments. **A.** Complete RBCC sequence with the RING domain highlighted in red, the B-box 1 in yellow, the B-box 2 in green, the coiled-coil domain in blue. The loops connecting the different domains are shown in light grey. In this example after digestion with Elastase we obtained, among the others, one fragment with a molecular weight of ~ 21.7 kDa. The RBCC was then screened to find the sequence corresponding to this molecular weight and two possible solutions were found. **B.** The first solution is a fragment that starts 7 residues above the B-box 2 domain (green) and ends 6 residues after the end of the coiled-coil domain (blue). **C.** The second solution is a fragment that includes part of the RING domain (red) and it ends in the middle of the coiled-coil domain (blue). This second solution was discarded since the integrity of two domains was compromised. This approach was followed for all the other identified fragments.

Half of the final constructs were designed with the same C-terminus as the previous soluble RBCC domain (T418) but with a new N-terminus identified by the limited proteolysis experiment (D128, C152, H190). One of them (BCC 201-418) was designed to carry the same N-terminus as the KAP1 RING domain structure solved in 2008 (2YVR) (unpublished paper). The other half of the constructs, were designed with the new N-terminus (D128, C152, H190, R201) and the new C-terminus identified in the limited proteolysis experiment (P382, L367, A352). (**Table 1**)

Expression and purification. The DNA coding sequences were ordered from GenScript and they were cloned inside the pETDuet-1 vector within the SacI and HindIII restriction sites in the first MCS. The four different *E.coli* strains were transformed with the recombinant vectors and the expression and mini purification tests were performed as previously described in the Method Chapter 2.2.

It was not possible to detect any expression of these recombinant proteins in the four *E. coli* strains. For this reason we could not perform any crystallization trial on these constructs given the impossibility to obtain soluble proteins.

3.3.4 KAP1 small binders selection for crystallization trials

The last attempt to crystallize KAP1 FL will be done by complexing the protein with small binders selected specifically to interact with KAP1 and to ideally stabilize its flexible regions, specifically the long loop connecting the N-terminal and C-terminal domains. In collaboration with Andreas Plückthun Laboratory, University of Zurich, we are selecting DARPins (Designed Ankyrin Repeat Proteins), engineered antibody mimetic proteins proven to be valuable tools in the crystallization of many difficult proteins^{189,190,191}. The selection process involves the chemical or enzymatic immobilization of the target protein on a solid plastic surface (referred as immobilized target) and subsequent exposure of the target to a phage DARPins library through the phage display method. In the case of KAP1 we decided to use an N-terminal Biotinylated tag (Avi-Tag) to directly bind the Streptavidin located on the solid plastic support guaranteeing the immobilization of the target protein. The new construct, KAP1 Avitag, has been cloned inside the pETDuet-1 vector, in the first MCS between the SACI and HINDIII cloning sites, with an N-terminal His₆ tag followed by the AviTag sequence (GLNDIFEAQKIEWHE) and TEV cleavage site (**Fig. 34A**). The construct has been ordered from GenScript and purified by affinity chromatography (HisTrap) and gel filtration chromatography (Superdex 200 and Superose6). Details on the purification process can be found in the Method Chapter 2.3, pETDuet-KAP1 constructs) (**Fig. 34B**).

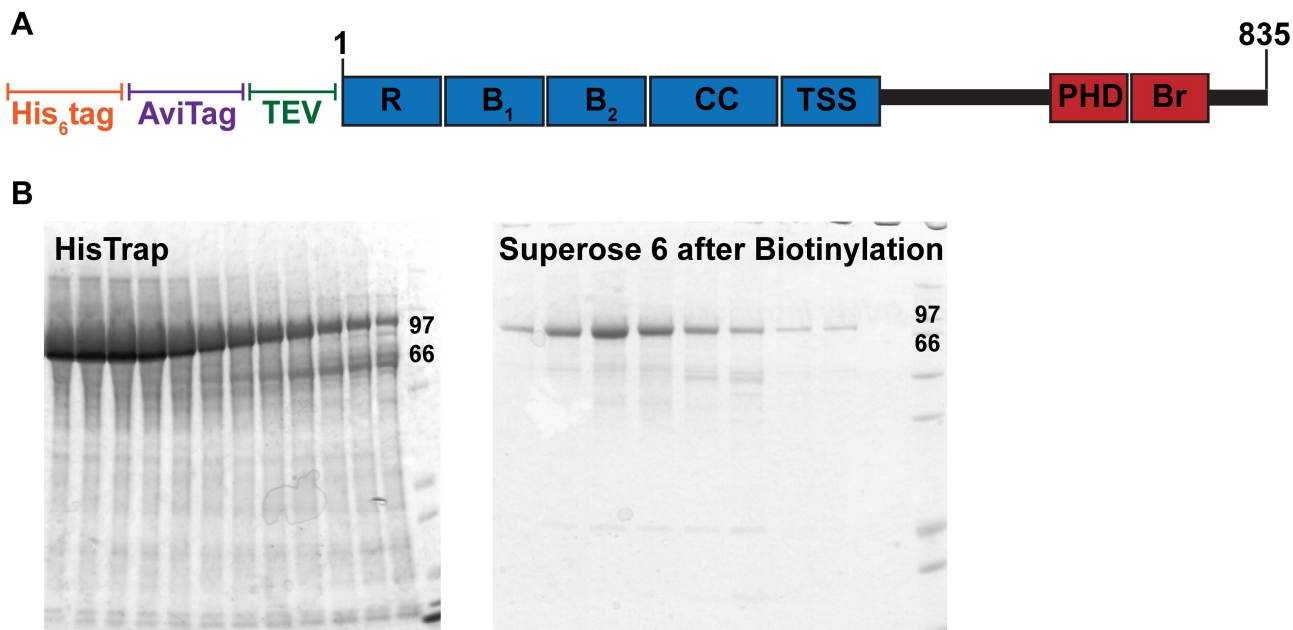


Figure 34| KAP1 AviTag. **A.** KAP1 AviTag cartoon model. From the N-terminus: poly-histidine tag (His₆ tag) in orange, AviTag in purple, TEV cleavage site in green, RING domain (R), B-box 1 (B₁) and B-box 2 (B₂), coiled-coil domain (CC), TSS domain in blue, PHD-Bromo domain (PHD-Br) in red. The construct starts in position M1 and ends in position P835. **B.** KAP1 AviTag purification. The protein was purified by affinity chromatography (HisTrap, right) and size exclusion chromatography (Superdex 200 16/600, and Superose 6 10/300 GL). Here we report the pure fractions purified by size exclusion chromatography (Superose 6 10/300 GL) after Biotinylation in vitro, gel on the left.

KAP1 AviTag was biotinylated in vitro in a total volume of 1 mL adding 756 μ L of protein at 11 μ M, 5 μ L of MgCl₂ (1M), 40 μ L of ATP (50 mM), 16 μ L of BirA (Biotin ligase) and 183 μ L of Biotin (10 mg/mL). After 3 hours of incubation on ice the biotinylation efficiency was tested. All the reagents except KAP1 were provided by Oliver Hantschel laboratory, EPFL. 200 μ L of Streptavidin magnetic beads were re-suspended in 200 μ L of TBS (Tris-buffered saline) buffer (free beads) and other 200 μ L were re-suspended in the same buffer with the addition of 2 μ L of 50 mM Biotin (blocked beads). 10 μ L of sample were added to the two different beads and incubated shaking for 15 min at 4°C. The supernatant was collected at the magnetic stand as the unbound fraction, the beads were washed with TBS buffer and the dry beads were re-suspended in 10 μ L of TBS buffer and immediately prepared for SDS-PAGE (**Fig. 35**).

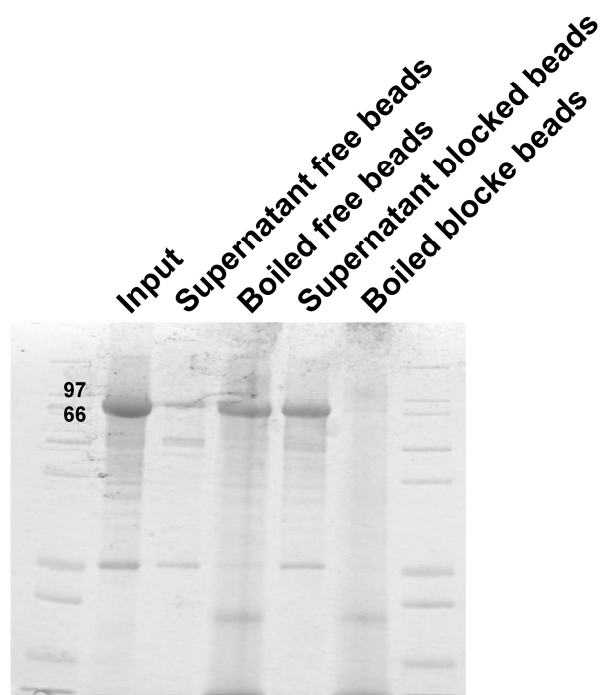


Figure 35| KAP1 AviTag in vitro biotinylation test. From the right: KAP1 Avitag 11 μ M (Input); supernatant of the free beads (KAP1 did not eluted as unbound fraction), boiled free beads with biotinylated KAP1 AviTag bound to Streptavidin, supernatant of the blocked beads (KAP1 eluted as unbound fraction), boiled blocked beads without KAP1 bound to Streptavidin since the beads have been previously blocked with 50 mM Biotin. The gel confirms the correct in vitro Biotinylation of KAP1 FL.

Given the presence of flexible regions in the structure, KAP1 full-length is probably not a suitable candidate for crystallization trials. For this reason we decided to study the protein with low-resolution techniques to obtain information on its shape and to confirm the dimeric species previously observed in AUC and MALLS experiments.

3.4 Small-angle X-ray scattering

3.4.1 Data collection and preliminary analysis

High-resolution techniques (X-ray crystallography and Cryo-EM) did not provide any relevant information about the overall structural organization of KAP1 FL. In order to gain more information on the protein architecture, we performed small angle X-ray scattering (SAXS). Size-exclusion chromatography was coupled in line with SAXS to avoid aggregation effects during the experiments (Method Chapter, **2.8**). SAXS data were collected across a concentration range of 9 to 15 mg/mL (100-200 μ M) for the RBCC domain, KAP1 FL and Δ KAP1 (**Fig. 36A**).

The Guinier analysis (Method Chapter, **2.8**) of the scattering curves showed good linearity indicating neither aggregation nor polydispersity effects and gave an estimated radius of gyration (R_g) of 83 Å for the RBCC domain, 90 Å for KAP1 FL and 89 Å for Δ KAP1 (**Table 4, Fig. 36B**). Moreover, the value for the cross sectional R_g (R_{gc}), was similar for KAP1 FL (35.8 Å) and Δ KAP1 (38.8 Å), while it was smaller for the construct containing only the N-terminal RBCC domain (20.2 Å) (**Table 4**) pointing to an elongated structure for KAP1 (**Fig. 36C**). This value can be calculated only for particles with an elongated cylindrical or rod-like shape.

Additionally, the SV-AUC experiment on the RBCC construct showed the protein to sediment as a single species with an $S_{20,w}^2$ of 2.3 and a frictional ratio (f/f_0 , where f_0 is the frictional coefficient of a smooth compact sphere) well above 1.3, indicating the elongated nature of the molecule and confirming the SAXS observation (**Fig. 27B**).

² The viscosity and the density of the buffer influence the sedimentation coefficient. In order to take into account the differences in density and viscosity between different solvents it is common practice to calculate the standard sedimentation coefficient, meaning the sedimentation value of the system in water at 20°C.

$$s_{20,w} = s_{exp} (\eta_{T,w}/\eta_{20,w}) (\eta_s/\eta_w) (1 - v\rho_{20,w}/1 - v\rho_{T,s})$$

$s_{20,w}$: sedimentation coefficient in water at 20°C, s_{exp} : sedimentation coefficient in buffer at experimental temperature T, $\eta_{T,w}$ and $\eta_{20,w}$: water viscosity at T or at 20°, η_s and η_w : solvent and water viscosity at common temperature, v : partial specific volume, $\rho_{20,w}$ and $\rho_{T,s}$: density of water at 20° and density of the solvent at T. The ratio between s_{exp} and $s_{20,w}$ gives the frictional ratio of the particle (f). If f/f_0 has a value of 1.2-1.3 the protein has a globular shape, but if the value is higher than 1.2-1.3 the protein is elongated as in the case of KAP1.

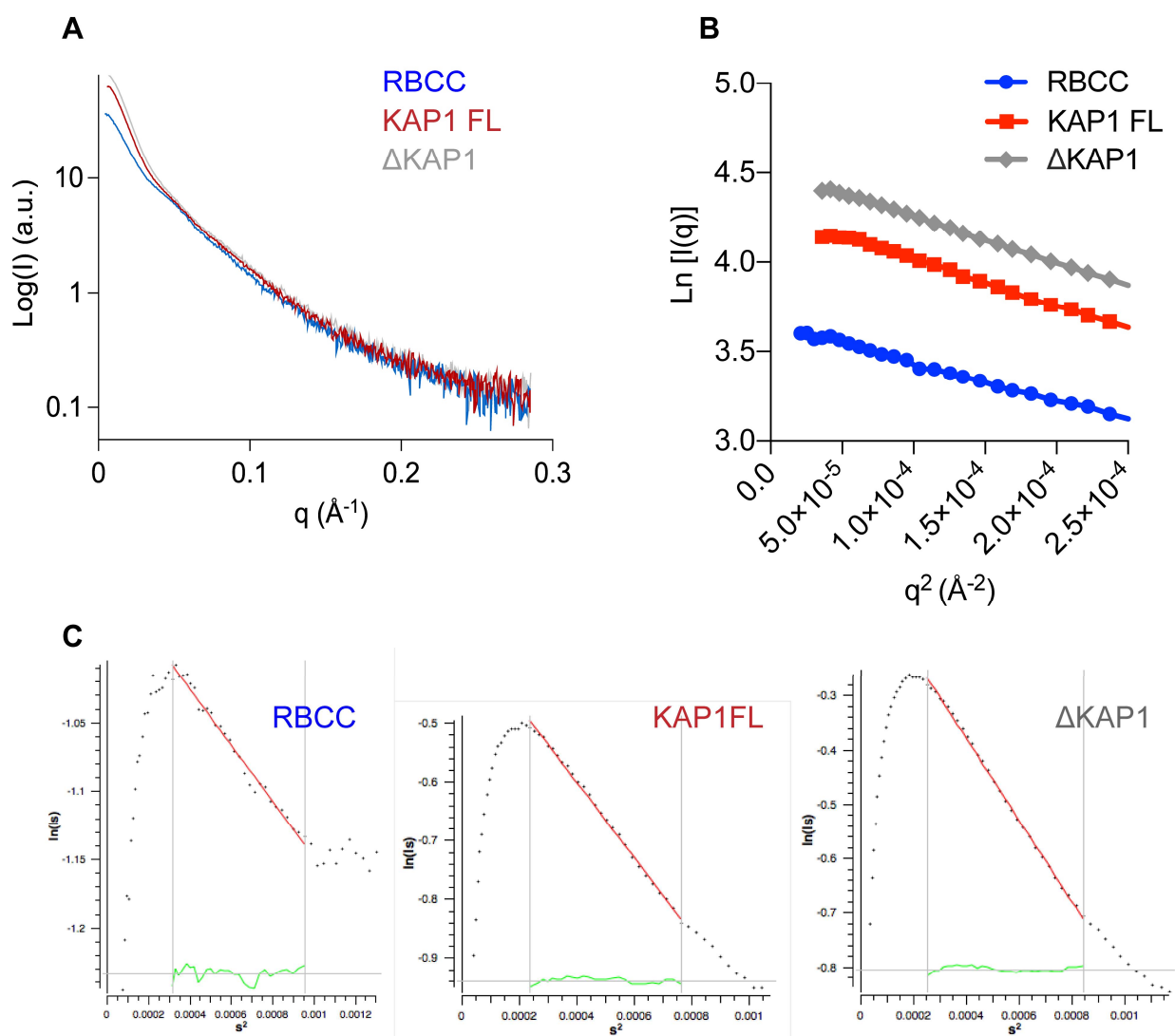


Figure 36| SAXS data collection and Guinier analysis. **A.** SAXS scattering curves for RBCC in blue, KAP1 FL in red and Δ KAP1 in light grey. **B.** Guinier plots displaying good linearity indicating non-aggregated samples. **C.** Cross-section R_g (R_{gc}) plots show that the KAP1 variants are elongated rod-like molecules. Values are shown in **Table 2** and were calculated using Primus¹²⁶. $s = q = 4\pi \sin(\theta)/\lambda$, where θ is the scattering angle and λ is the wavelength of the x-rays.

Furthermore, the analysis of the Kratky plot (Method Chapter 2.8) indicates a large degree of flexibility for KAP1 (**Fig. 37**). This observation agrees with the prediction of an unstructured 200 residue-long loop connecting the N-terminal RBCC domain and the C-terminal PHD-Br domain.

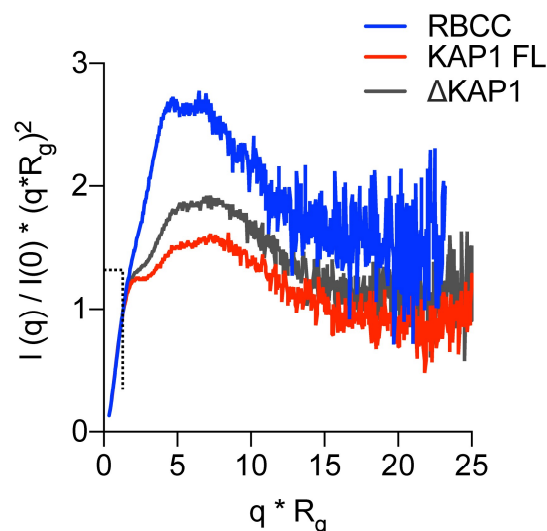


Figure 37| Kratky plot of the KAP1 constructs. Dimensionless Kratky Plots produced with ScÅtter¹²⁴ showing that all KAP1 constructs are flexible multi-domain proteins. In dimensionless Kratky Plots, globular proteins always have a maximum at $q \cdot R_g = 3^{1/2}$ and $(q \cdot R_g)^2 I(q)/I(0) = 1.104$, which is shown by the dotted lines.

MALS and preliminary SAXS analyses reveal that KAP1 is an elongated flexible dimer. Therefore, to gain insight into its mass distribution, we compared the pair distribution functions, $P(r)$ (Method Chapter, **2.8**), of the RBCC domain and KAP1 FL. Both show the signature of an elongated molecule with very large maximum dimension (D_{max}) values and a main peak at a shorter radius (**Fig. 38A**). The $P(r)$ function for the RBCC is clearly bimodal, containing two peaks, one centered at 40 Å and the other one at 160 Å with a D_{max} of 370 Å (**Fig. 38A** and **Table 4**) suggesting a dumbbell-shaped macromolecule with the RBCC domain arranged in an antiparallel fashion separating the two N-terminal RING/B-box 1/B-box 2 (RB₁B₂) modules by its length (160 Å). The $P(r)$ function of KAP1 FL is also characteristic of an elongated molecule but it shows a similar D_{max} (380 Å) as for the RBCC domain alone.

This observation, together with the increase in R_{gc} , leads to the conclusion that the extra domains present in the full-length protein are not fully extended but are in proximity to the N-terminal RBCC domain. This conclusion is also supported by the loss of the bimodality in the $P(r)$ function compatible with a more compact protein in which the extra C-terminal domains are close to the RBCC domains (**Fig. 38A**).

Ab initio bead models (**2.8**) calculated directly from the SAXS scattering curves (**Fig. 38B**) provided conformations of similar length (~320 Å), but different width, in agreement with the larger R_{gc} of KAP1 FL. These models have been initially calculated using GASBOR¹²⁷ without imposing 2-fold symmetry (**Fig. 38B**). Furthermore, we generated 20 additional bead models with DAMMIF¹¹⁹, superimposed them with SUPCOMB¹²⁰, averaged them with DAMAVER¹⁹² and refined them with DAMMIN¹⁹³ with the aim to compare the results with and without imposing 2-fold symmetry (**Appendix 4**).

For the RBCC domain, we observed that both the bead models generated with and without 2-fold symmetry similarly fit our SAXS data. This observation leads to the conclusion that at this low resolution, the RBCC domain can be considered an antiparallel dimer with a 2-fold symmetry axis in the middle of the coiled-coil domain such that the RB₁B₂ modules are located at opposite ends and ~160 Å apart. The RB₁B₂ modules can explore different conformations that can be not necessarily symmetric in solution, but they might be considered as symmetric at the low resolution accessible to SAXS.

For KAP1 FL, the bead models generated without imposing any symmetry (P1) present a well-conserved shape among them with an NSD³ value of 1.43±0.07 (**Appendix 4**). When applying 2-fold symmetry, the models show a significantly broader range of conformations and a higher NSD value of 2.98±0.64, indicating that an asymmetric bead model generation better fits our KAP1 FL SAXS data. This is a first indication of a possible asymmetric organization in KAP1 full-length protein.

³ The NSD (Normalized spatial discrepancy) value is a measure of quantitative similarity between sets of three-dimensional points (bead models) implemented in SUPCOMB¹⁹⁴. Considering two models (X and Y), the program represents each model as a set of points and calculates the distances between each point in model X and each point in model Y. The minimum distance value is then retrieved for each point in model X and the same procedure is repeated for model Y. The minimum distances are then added and normalized against the average distances between the neighboring points for the two models.

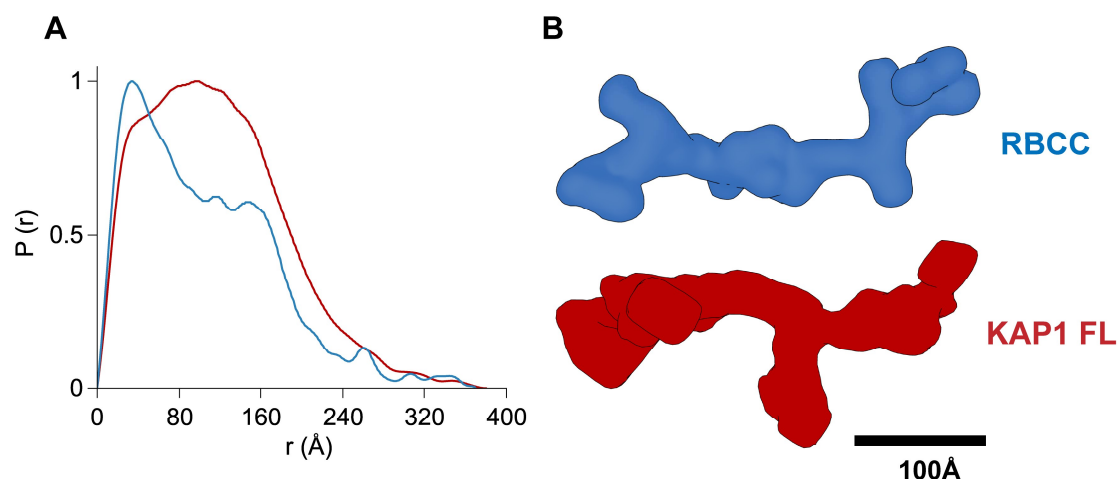


Figure 38| RBCC and KAP1 FL $P(r)$ function and *Ab initio* models. **A.** Pair distance distribution function of KAP1 FL in red and RBCC domain in blue. **B.** *Ab initio* bead models created using GASBOR⁸ shown in surface representation.

The RBCC domain being a dimer with an elongated dumbbell-shaped conformation is only compatible with an antiparallel organization in which the two tandem RB₁B₂ modules are separated by the Coiled coil domain.

3.5 Negative staining and cryo-electron microscopy

We performed negative staining EM experiments on KAP1 FL at a diluted concentration (0.1 mg/mL) and at the same concentration used in the crystallization trials (10 mg/mL). We were not able to observe a repeated shape due to protein flexibility and heterogeneity (**Appendix 3**).

Cryo EM studies have been conducted on the KAP1 FL protein without the addition of crosslinkers (Method Chapter, **2.7**). The protein at an initial concentration of 6 mg/mL was diluted several times (5X-200X) and then observed on a Tecnai F20 cryo microscope. The best images were acquired with the Falcon camera, at a magnification of 80'000X and dilution 200X (0.03 mg/mL). We were able to observe elongated protein structures. The sample looked homogeneous and not aggregated however it was conformationally heterogeneous. We were not able to observe repeated conformations and the 3D reconstruction could not be performed. Finally, to enhance the signal to noise ratio, we performed a cryo electron tomography experiment that did not show improvements in the acquired signal (**Appendix 3**).

Finally, we visualized negatively stained crosslinked KAP1 FL to independently measure its size and shape (**Fig. 39**). Rod-like particles with a multi-domain organization could be readily observed in the micrograph when sorted into 2D class averages. In agreement with the SAXS data, the particles were around ~300 Å (D_{\max} of the RBCC domain defining the total length of the protein).

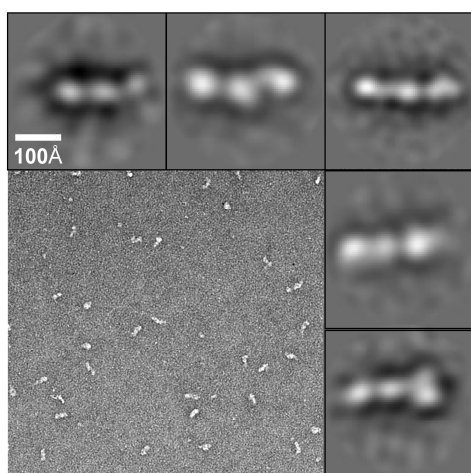


Figure 39| KAP1 FL TEM gallery. Representative TEM image of negatively stained KAP1 FL and gallery of 2D class averages (2.15, Appendix 3, negative staining A).

3.6 Molecular modeling and flexible SAXS data fitting

The RBCC domain is an antiparallel elongated dimer determined by the coiled-coil domain. We could build an initial atomic model of the RBCC domain using homologous TRIM family members structures as templates (PDBs: 4TN3¹⁸⁴, 5EIU¹⁹⁵, 3Q1D *to be published*, 4NQJ¹³⁹, 4CFG *to be published*). Despite the low sequence homology among the family, model building was facilitated by the conservation of the canonical left-handed coiled-coil motif and the zinc-coordinating His/Cys pairs of the B-box 2 domain among 54 human TRIM proteins¹⁸³.

In order to sample the flexibility of the RBCC domain in solution, an ensemble of 10 different RBCC models were created by varying the positions of the RING, B-box 1 and B-box 2 domains with respect to the Coiled-coil domain whose position is conserved in several crystal structures. These models were fitted to the SAXS data using a χ^2 -minimizing optimization method (Method Chapter, 2.9 NOLB¹³³) resulting in highly similar structures that explore a tight conformational space (**Fig. 40A,B**). In all the fitted models, the RBCC domain shows an elongated structure of ~ 310 Å in length, with the RB₁B₂ modules separated by ~ 160 Å, the dimension of the dimeric antiparallel coiled-coil domain (**Fig. 40C**).

Furthermore, the sedimentation and diffusion coefficients of a representative RBCC model were computed using HYDROPRO¹⁹⁶ resulting in an R_g value of 79.8 Å and an s value of 2.59 S, in agreement with the experimental values obtained from the SV-AUC experiment (**Fig. 27B**).

Interestingly, when compared to published SAXS data of other TRIM family members, the R_{gc} of the RBCC domain is ~ 20 Å, remarkably smaller compared to the R_{gc} of RBCC-TRIM25 (~ 31 Å) and RBCC-TRIM32 (~ 32 Å). In the first case, the difference in R_{gc} value can be due to the fact that the RB₁B₂ modules of TRIM25

are thought to fold back on the structure of the coiled coil domain, while for KAP1 RBCC domain the RB₁B₂ modules can be fitted only at the extremity of the coiled-coil domain. Moreover, the short loop connecting the coiled-coil domain with the RB₁B₂ does not allow this module to fold back on coiled-coil structure. In the second case, the higher R_{gc} value can be due to the fact that RBCC-TRIM32 behaves as a tetramer in solution while KAP1 is exclusively as a dimer in solution at the sampled concentrations (**Fig. 27A**).

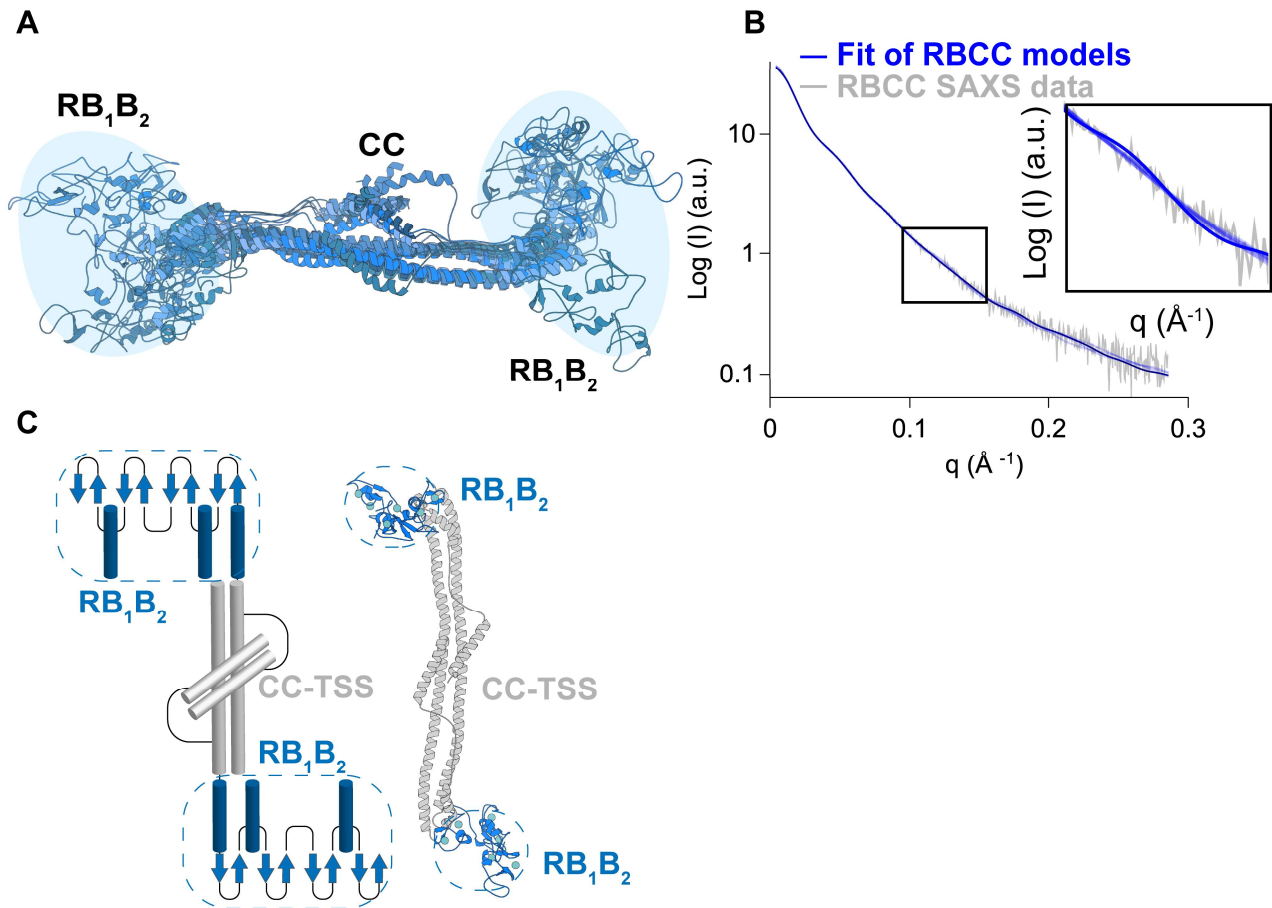


Figure 40| Overall architecture of the RBCC domain. **A.** Ensemble of 10 representative RBCC models that fit the SAXS data with χ^2 values between 0.8 and 1 **B.** Fit of the scattering profiles calculated for the 10 atomic models (in blue) to the SAXS scattering data (in grey) using Pepsi-SAXS¹³⁴ (fitting with $0.8 < \chi^2 < 1$) **C.** A representative atomic model of the antiparallel RBCC dimer (the two protomers are colored in blue and grey) capturing the main features of the ensemble of models able to fit the SAXS data. The end of the RBCC contains the TIF1 signature sequence (TSS) domain, which is rich in phenylalanine and tryptophan residues¹⁰⁴ and is modeled as a helix that packs against the coiled-coil domain forming overall a 4-helix bundle.

KAP1 FL is natively asymmetric in solution. The overall organization of KAP1 FL is unknown as well as the structural stabilization or interaction between the N-terminal RBCC domain and C-terminal PHD-Bromo domain. Therefore, to gain deeper insight into the domain organization of KAP1, very flexible according to the Kratky plot (Fig. 37), we developed a novel integrative modeling strategy that uses nonlinear Cartesian Normal Mode Analysis (NOLB¹³³ NMA) for SAXS data fitting (Fig. 41). At first 1'000 models of KAP1 FL were generated as described in the Method Chapter 2.9. These models conserved the structure of the RBCC domain previously modeled on the SAXS data and the structure of the C-terminal PHD-Br domain solved by NMR (2RO1⁶²). The linker region (~200 amino acids) between each PHD-Br domain and its respective RBCC domain was modeled as a random coil using MODELLER¹⁴⁵. The models were flexibly fitted to the SAXS data through NMA using 60 low frequency normal modes to explore the conformational space of the flexible regions and detect finally the position of the C-terminal domain respect to the N-terminal RBCC domain. The final models converged to an ensemble of structures well consistent with the SAXS data ($\chi^2 = 1.1 \pm 0.3$), reducing their R_g values from $\sim 94 \pm 7$ to $\sim 88 \pm 2$ Å, a value similar to the experimental R_g value for KAP1 FL (90 Å) (Table 4, Fig. 42).

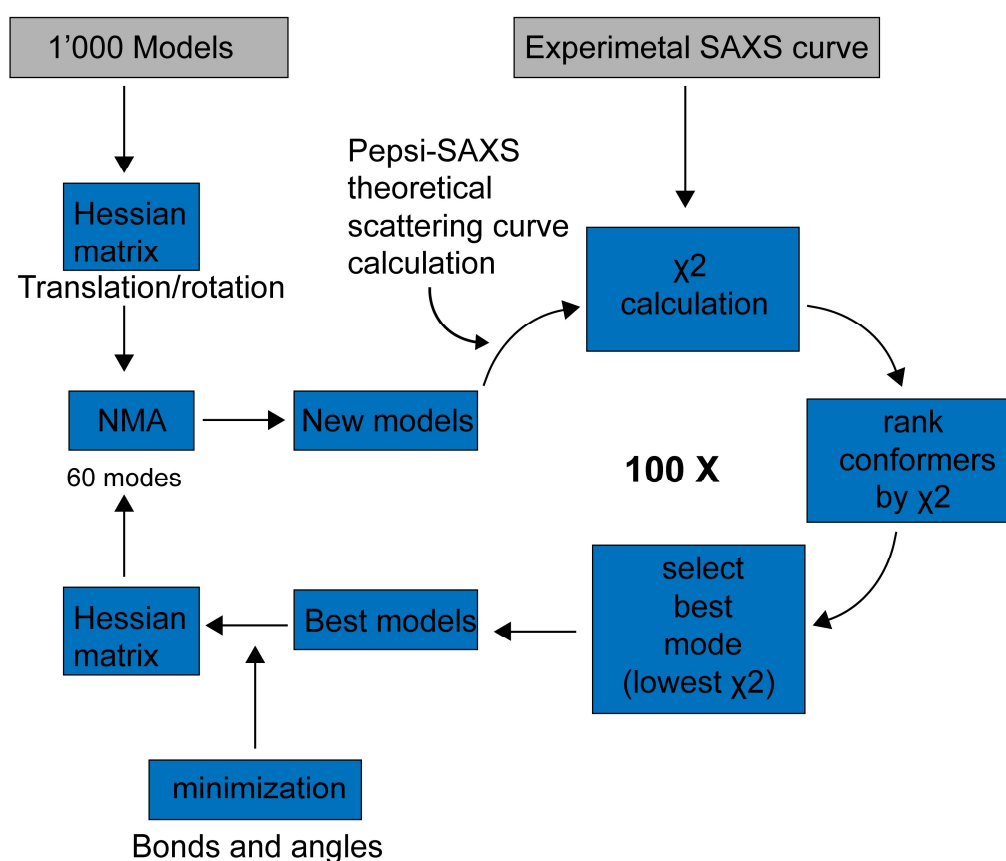


Figure 41| Flowchart of the NOLB¹³³ NMA for SAXS data fitting. The Hessian matrix is computed for the initial 1'000 models. From the diagonalization of this matrix 60 low frequency normal modes are computed and new initial displaced models are generated by the application of NMA to the original structures. Pepsi-SAXS¹³⁴ generates the theoretical scattering curve for each displaced model and calculates a χ^2 value comparing the experimental and theoretical SAXS curves. The conformers are ranked based on this value and the normal mode that leads to the lowest χ^2

value is selected together with the best conformer. A minimization step is applied to keep the initial topology (angles and bonds) and the new best conformer undergoes further optimization steps for a total of 100 iterations.

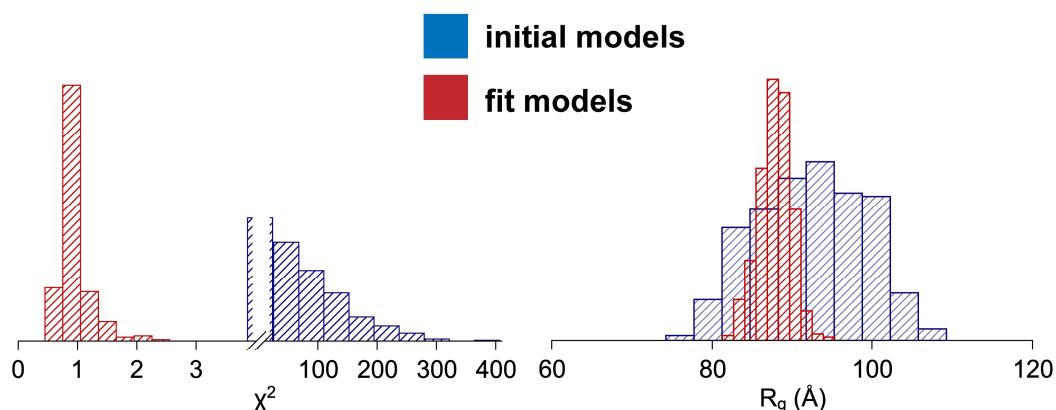


Figure 42| Flexible SAXS data fitting of KAP1 FL models. χ^2 (left) and R_g (Å) plots (right) of the initial (blue) vs. final (red) model structures, with final distributions centered around $\chi^2=1$ and $R_g=90$ Å.

In order to better characterize the global architecture of the KAP1 FL, the distances between the COM of the RB_1B_2 modules and the COM of the PHD-Br domains of the two protomers were calculated (dA_1 , dA_2 , dB_1 , dB_2) (**Fig. 43A**). The minimum dA and dB distances have been plotted. As expected, the initial models had a very heterogeneous ensemble of conformations, however after flexible fitting two distinct clusters could be identified (**Fig. 43B**). The two resulting clusters are symmetric and they, as well as their centers⁴ (white dots), can be considered equivalent. This cluster contains models that are characterized by a large displacement (~ 120 Å) of one PHD-Br domain from the RB_1B_2 module and a small displacement (~ 60 Å) of the other PHD-Br domain from the remaining RB_1B_2 . This is highlighted in the 1D plot (**Fig. 43C**) where the fit models are not randomly distributed according to the distance between the N- and C-terminal domains but are organized as two separated Gaussian distributions centered at ~ 60 Å and ~ 120 Å.

⁴ Cluster center: point of highest density of the cluster according to the definition of local density from the algorithm.

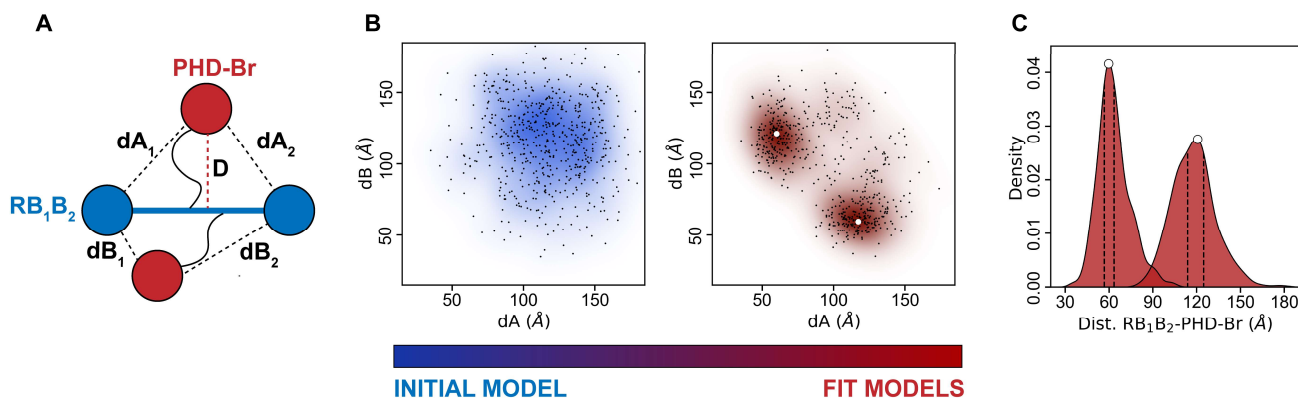


Figure 43| Native asymmetry of KAP1. **A.** Schematic representation of the KAP1 FL dimer where the distances (D) $dA_{1,2}$ and $dB_{1,2}$ are displayed. **B.** Plot of the initial vs. final models according to their dA vs. dB minimum distance values. The clusters centers are highlighted by white dots. **C.** 1D density distribution of the clustered models according to the minimum distances between the PHD-Br and the RB₁B₂ domains showing two clear Gaussian distributions. The dotted lines show the limits of the first density level containing 10% of the models and the two white dots represent the clusters centers.

To further explore the asymmetric organization of KAP1 FL, we superimposed the two symmetric clusters previously identified (**Fig. 44A**). We selected and analyzed the density level containing 300 models close to the cluster center or the peak of the Gaussian distribution and highlighted as a solid black line in **Fig. 44A**. The models belonging to this density level have been extracted and superimposed to recover common asymmetric features. In **Fig. 44B** is reported the ensemble of these 300 models in which the PHD-Br domain close to the RB₁B₂ module is highlighted in red and the PHD-Br domain far from the module RB₁B₂ is highlighted in blue. This representation shows a clear general organization of the C-terminal domains that is common for all the models with one PHD-Br domain far and one close to the RB₁B₂ domains. Finally the first density level containing 10% of the cluster models and highlighted as a dotted line in **Fig. 43C** was selected. In **Fig. 44C**, the atomic structure representing the cluster center is presented in red and some representative models included in the first density level are shown in grey shade. Again the C-terminal PHD-Br domains close and far to the RB₁B₂ module are highlighted respectively in red and in blue. In these models the position of the loops and the C-terminal domain itself is not the same and given the low resolution of the SAXS data we cannot discriminate among them.

Thus we can conclude that the asymmetry previously observed in the bead models without imposing 2-fold symmetry for KAP1 FL is most likely dictated by this different arrangement of the PHD-Br domain, as the RBCC domain can be considered symmetric at this level of resolution.

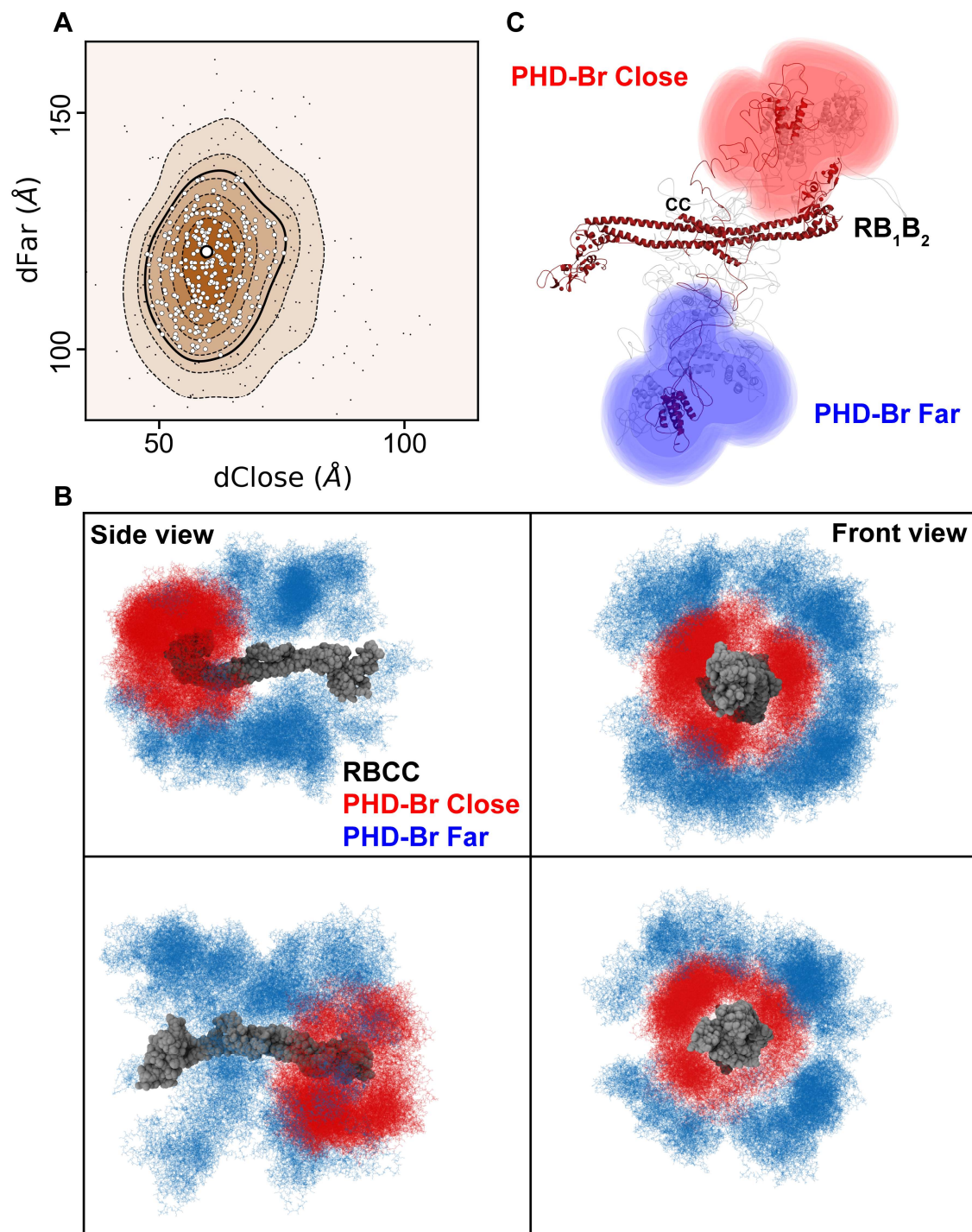


Figure 44| Asymmetric domain arrangement in KAP1 FL. **A.** Superposition of the two equivalent clusters of NMA fitted models (see **Fig. 43B**). The density level depicted by a solid line contains the 300 closest models to the cluster center, shown as white dots. The cluster center is shown as a bigger white dot. **B.** Superimposition of the 300 closest models to the cluster center. For each model, the PHD-Br domain nearest to an RB₁B₂ module is shown in red, while its far counterpart is shown in blue. The loops between the RBCC and PHD-Br domains have been omitted for visualization. **C.** Cartoon representation of a selection of models belonging to the first density level.

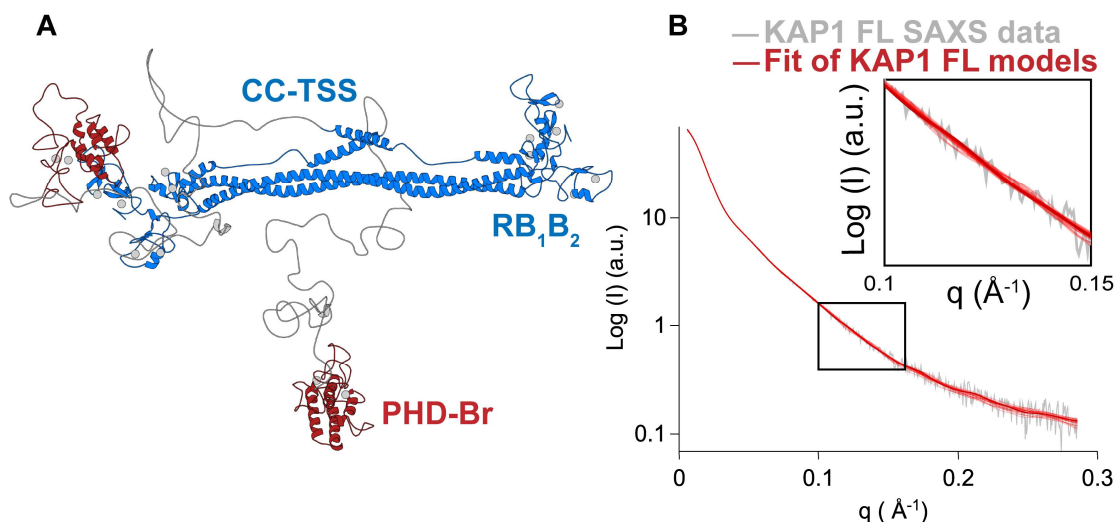


Figure 45| Analysis of the KAP1 FL models in the first density level. **A.** Cartoon representation of the KAP1 FL model (cluster center) representative of the ensemble that optimally fits the SAXS data. The ensemble of selected models belonging to the first density level is shown in **Fig. 44B** and **B.** Comparison between the SAXS scattering profile (in grey) and the calculated scattering profiles (in red) from the models in the first density level using Pepsi-SAXS¹³⁴ ($0.8 < \chi^2 < 2$).

The models contained in the first density level better describe the conformation of KAP1 FL in solution (**Fig. 43**, **Fig. 44** and **Fig 45**). Within the ensemble of models describing the conformation of KAP1 FL in solution, the two PHD-Br domains are never found simultaneously close to the RB₁B₂ modules, contrary to what was proposed for the TRIM25 PRYSPRY domain and the NHL repeats of TRIM32¹⁸⁷. If such conformation is imposed, this induced large distortions of the coiled coil domain and poor data fit ($\chi^2 > 80$). Similarly, flexible fitting never selected conformations where the two C-termini are fully extending away from the RBCC domains ($\chi^2 > 75$). Consistently, P(r) functions calculated *a posteriori* from models in close or open conformations (meaning the two PHD-Br domains close or far away at the same time from the RB₁B₂ modules) do not fit the experimental P(r) function. The fit to the experimental P(r) function can only be obtained for the structure of the cluster center model (**Fig. 46**).

Therefore, SAXS data strongly indicate that the organization of KAP1 FL is a natively asymmetric dimer, arranged with one C-terminal domain close to the RB₁B₂ module and the other one far away. The asymmetry seems to originate exclusively from the C-terminal domains, as the ensemble of the RBCC structures can be considered symmetric, at least at this level of resolution. *Ab initio* bead models generated for the RBCC using the SAXS scattering profiles support in fact this conclusion as we obtained quasi-symmetric structures also without imposing any symmetry (**Appendix 4**). However in the case of KAP1 FL, imposing 2-fold symmetry generates a set of bead models with a much divergent range of conformations than that generated without imposing symmetry restraints (**Appendix 4**).

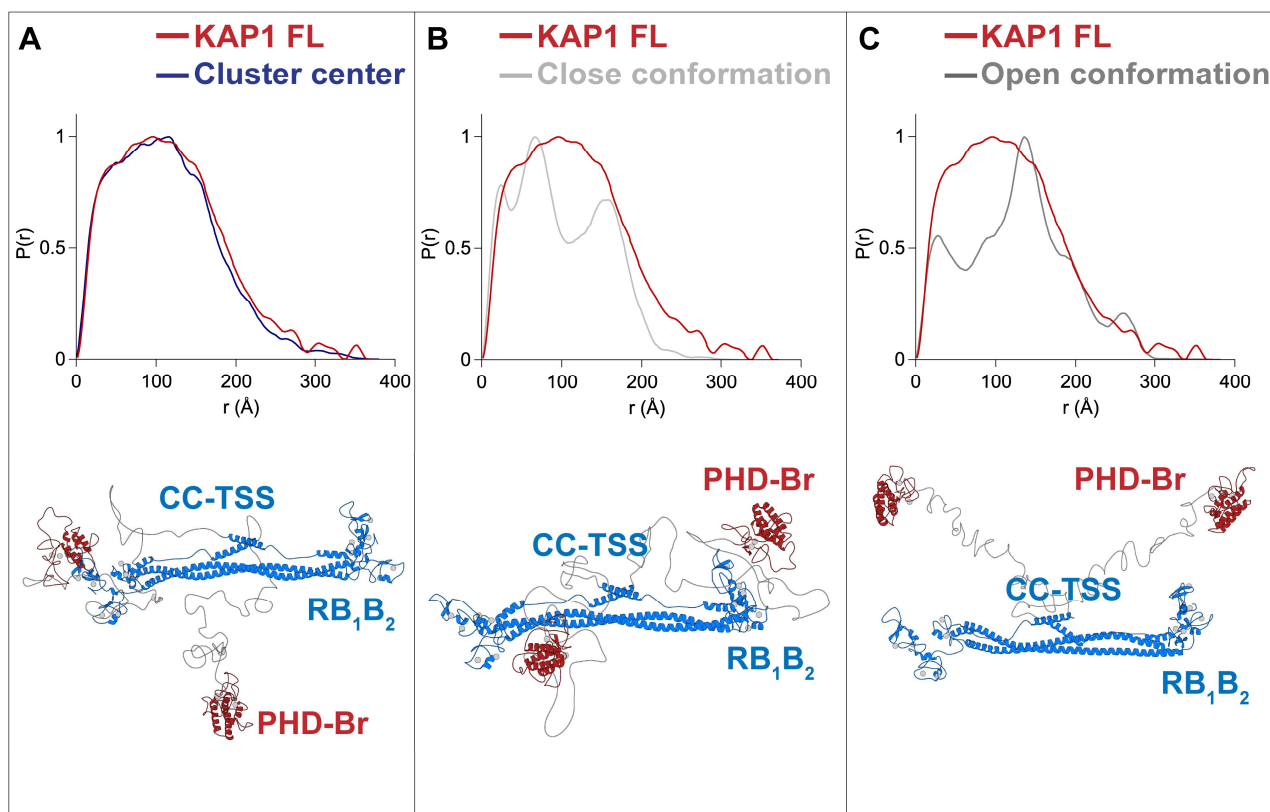


Figure 46| Comparison of $P(r)$ functions. Superposition of experimental (red) and calculated (blue/light grey/dark grey) $P(r)$ functions for models of KAP1 FL where the PHD-Br domains are in different positions, as follows: (left) as in the cluster center structure, (middle) both close to the RB₁B₂ module and (right) both away from the RB₁B₂ module. The $P(r)$ functions were calculated using ScÅtter¹²⁴.

ΔKAP1 is an asymmetric antiparallel dimer as KAP1 FL. The integrative modeling strategy described in Method Chapter 2.9 has been applied to the SAXS data collected for ΔKAP1 construct. This construct (23-812) misses 20 residues at the N-terminus and at the C-terminus when compared to the full-length protein (1-835). This independent data set gave the same result as the one obtained for KAP1 FL, confirming the protein asymmetric structural organization. Again, within the ensemble of models describing the conformation of ΔKAP1 in solution, the two PHD-Br domains are never found simultaneously close or far from the RB₁B₂ modules but always organized in an asymmetric configuration (**Fig. 47**).

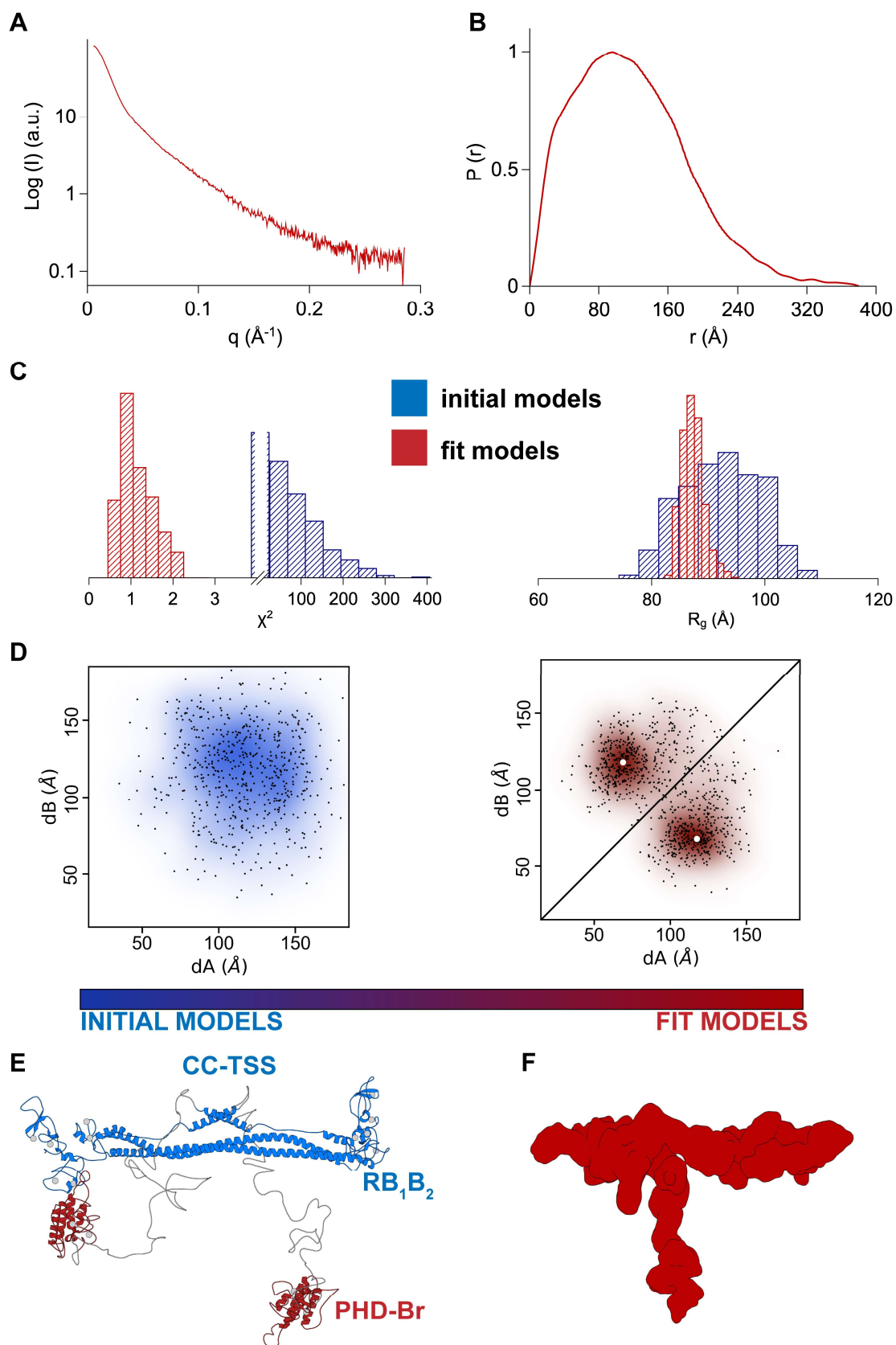


Figure 47| Integrative modelling analysis of Δ KAP1. **A.** 1-dimentional scattering curve **B.** Pair distance distribution function calculated from the SAXS scattering profile **C.** χ^2 plot (left) and R_g plot (right) of the initial models in blue and the final structures after NMA in red, with narrower distributions centered on $\chi^2=1$ and on the experimental $R_g=90$ \AA

(Table 2) **D.** Plot of the initial randomized models in blue and the final fit models in red according to their dA vs dB distance values. The centers of the clusters are highlighted with white dots **E.** Cartoon representation of the cluster center model of Δ KAP1 representative of the ensemble with the most probable domain arrangement **F.** GASBOR¹²⁰ *ab initio* bead model of Δ KAP1 shown in surface representation.

3.7 Functional studies: KAP1 architecture allows RING-dependent autoSUMOylation

After discovering the proximity of the N-terminal domain to one of the PHD-Br domains, we investigated if this unexpected conformation of the KAP1 C-terminal domain in solution could have direct implications for its function. The cysteine 651, located in the PH domain, has been shown to be the key residue for the auto-SUMOylation of the C-terminal Bromo domain⁶⁸. This PTM is fundamental because it allows KAP1 to interact with chromatin remodeling enzymes inducing heterochromatin formation.^{34,197}. However, previous studies reported that the intact PH domain was necessary for the auto-SUMOylation of many but not all the sites in KAP1⁶⁸, implying the existence of a second catalytic site. Similarly, the RING domain has been shown to strongly interact with the Ubc9 E2 SUMO ligase and be fundamental for the SUMOylation process of KAP1 substrates such as the Interferon Regulatory Factor 7 (IRF7)¹⁹⁸ and the neurodegenerative disease driving proteins tau and α -synuclein¹⁹⁹. These findings as well as the proximity of one of the C-terminal domain to the RBCC domain proposed by our SAXS based ensemble (**Fig. 45A**) hints to a role for the RING domain in the auto-SUMOylation of KAP1. In order to prove this hypothesis, we performed a SUMOylation assay *in vitro* using purified proteins, where KAP1 acts as an E3 SUMO ligase, auto-SUMOylating itself (Method Chapter, **2.11**). We compared the auto-SUMOylation of KAP1 FL, RBCC, KAP1 FL C651A mutant and a deletion mutant missing the N-terminal RING domain (Δ RING).

We designed the mutant C651A to prove the presence of residual autoSUMOylation activity of KAP1 after mutating the residue (C651) shown to be fundamental for this process. The aim was to replicate and confirm the results previously obtained by Ivanov and collaborators⁶⁸. This residual activity would have implied the existence of a second catalytic site outside the PH domain. The C651 was mutated on the sequence of KAP1 FL using the Site-Direct Mutagenesis Kit QuickChange (Agilent). Details on the primers and the PCR protocol are provided in **Appendix 5**. Moreover, we designed a second mutant missing the N-terminal RING domain, (Δ RING) to prove the involvement of the RING domain in the SUMOylation process. Following the hypothesis that the RING domain represents the second catalytic site, without this domain we would have observed an effect on the autoSUMOylation of KAP1 *in vitro*. The optimized DNA coding sequence was ordered from Genscript.

The new constructs were expressed and purified as previously described in the Method Chapter 2.3 (pETDuet-KAP1 constructs) (Appendix 1D). We checked the folding state of the proteins using circular dichroism and comparing their melting temperature (T_m) with the one previously measured for KAP1 FL. (Fig 48) to make sure that the differences in activity were not due to protein misfolding.

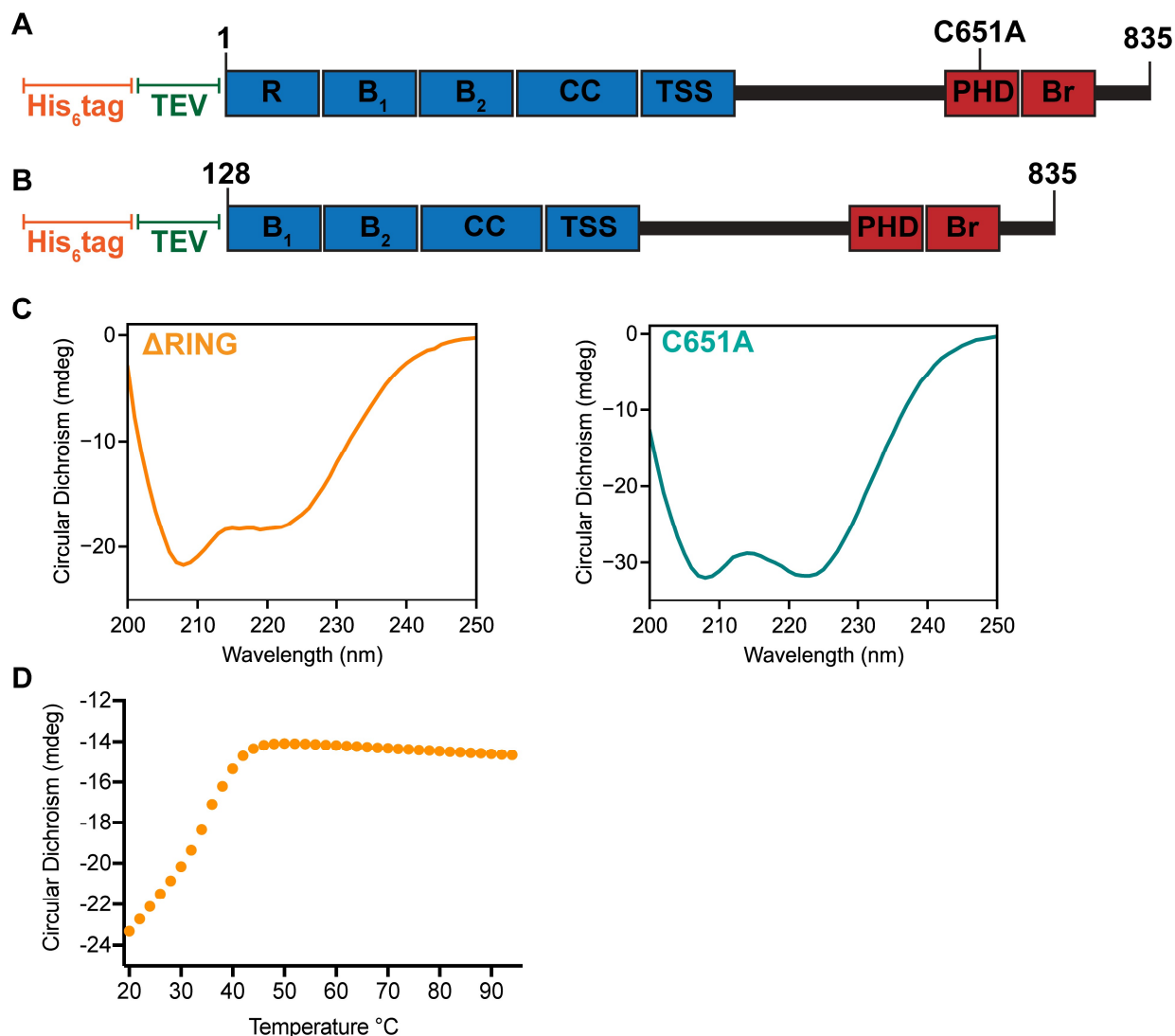


Figure 48| KAP1 constructs for the autoSuMOylation activity. **A.** KAP1 FL C651A mutant. From the N-terminus: poly-histidine tag (His₆ tag) in orange, TEV protease cleavage site in green, RING domain (R), B-box 1 (B₁) and B-box 2 (B₂), coiled-coil domain (CC), TSS domain in blue, PHD-Bromo domain (PHD-Br) in red. The construct starts in position M1 and ends in position P835 and presents the mutation C651A located in the PH domain **B.** ΔRING construct. From the N-terminus: poly-histidine tag (His₆ tag) in orange, TEV protease cleavage site in green, B-box 1 (B₁) and B-box 2 (B₂), coiled-coil domain (CC), TSS domain in blue, PHD-Bromo domain (PHD-Br) in red. The construct starts in position D128 and ends in position P835 **C.** CD data collected using a Chirascan CD Spectrometer and a 0.1 cm quartz cell for sample concentration ranging from 0.2 mg/mL to 0.8 mg/mL in Phosphate Buffer (1x PBS). The scans were measured from 200 to 250 nm, under continuous scanning mode with a 1 nm data pitch, a scan speed of 50 nm/min and a response time of 1 s. The figure shows the average of three spectra buffer subtracted, for the KAP1 FL C651A mutant in light blue and for ΔRING construct in orange. The proteins present a clear alpha helical CD spectra with two minima at

208 nm and at 222 nm **D**. Variable temperature CD spectra for KAP1 Δ RING construct were recorded from 200 to 250 nm within 20°C and 95°C in triplicates. The graph shows the result of the three repetitions at 220 nm. The conformational change happened at 40°C.

As expected, the RBCC domain was not SUMOylated and the KAP1 C651A mutant showed residual SUMOylation activity with respect to the wild-type protein as previously reported.⁶⁸ However, differences in the time-dependent SUMOylation activity of the other constructs (KAP1 FL and Δ RING) were observed (**Fig. 49**). Specifically, we found that the Δ RING construct presented a much lower auto-SUMOylation activity than KAP1 FL (about one third of KAP1 FL activity at 120 min **Fig. 49**). The autoSUMOylation assay was repeated three times, for each gel the bands were quantified using ImageJ¹⁴⁸ and the intensities were normalized with respect to the sum of all the considered intensities (i.e., 100% intensity is the sum of all the bands). The graph in **Fig. 49** shows the mean value of each data point for the three repetitions with the standard error of the mean. The internal control is represented by an SDS page gel, demonstrating how KAP1 FL and Δ RING have been equally loaded meaning that the difference in activity is not due to protein concentration. The uncut triplicates SUMOylation assay gels together with the internal control SDS page gel and the RBCC-KAP1 FLC651A mutant gel are reported in **Appendix 6**.

Taken together, these results show that not only the PHD but also the N-terminal RING domain is implicated in the auto-SUMOylation process of the C-terminal domain. The conformation discovered by the SAXS-based modeling (**Fig. 45**) is thus consistent with a potential communication between the two terminal domains, which can both cooperate in the E3 SUMO ligase activity of KAP1. The RING domain can thus act as a second active site, as it is able to SUMOylate other substrates^{198,199}. Interestingly, RING-mediated auto-SUMOylation has been observed also in another TRIM family member, the protein PML (Promyelocytic leukemia, TRIM19), which is involved in the formation of nuclear structures called PML-nuclear bodies implicated in a variety of cellular processes^{200,201}.

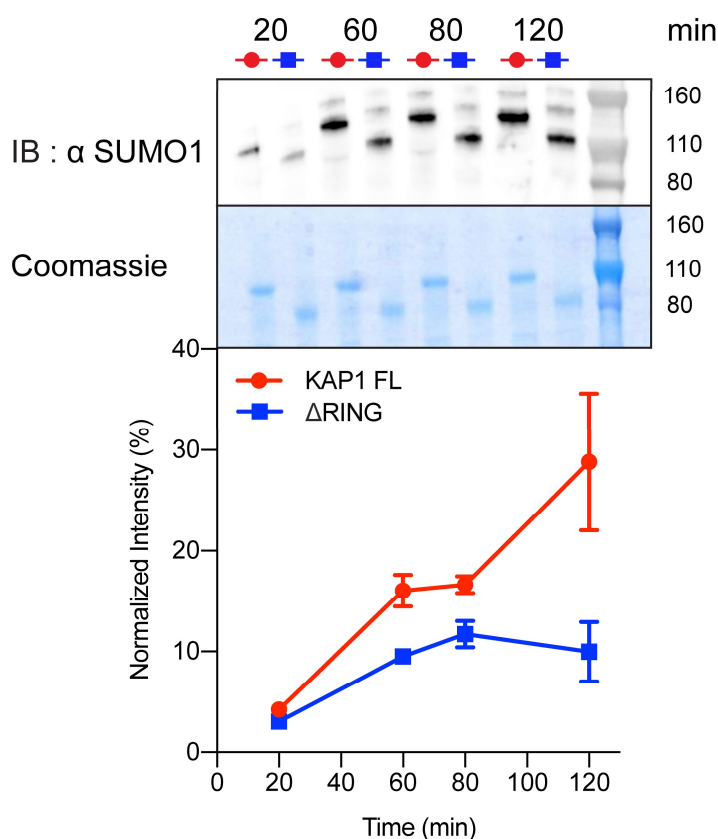


Figure 49| Role of the RING domain in the auto-SUMOylation of KAP1. *In vitro* SUMOylation of KAP1, where reactions containing E1, E2 (Ubc9), SUMO1 and KAP1 variants (KAP1 FL and ΔRING) were incubated for the time indicated (20 to 120 min). The samples were analyzed by western blot with anti-SUMO1 antibody (top) and Coomassie blue staining (center). The intensity of the bands was quantified using ImageJ¹⁴⁸ and normalized with respect to the sum of all intensities. The graph shows the mean value and the standard error of the mean for each time step for the three technical replicas (uncropped images shown in **Appendix 6**). Auto-SUMOylation for RBCC, KAP1C651A mutant, and controls are also reported in **Appendix 6**.

3.8 Binding partners: KAP1 asymmetry is functional for recruiting HP1α

To understand if KAP1 structural asymmetry has functional implications when interacting with other binding partners, we explored the direct interaction between KAP1 FL and full-length HP1α (HP1α FL). Previous studies have identified a fragment of 15 residues within the central unstructured region of KAP1 (in the loop between the RBCC domain and the PHD-Br domain) as responsible for HP1 binding (residues 483-497, HP1BD)^{202,203,36}. Moreover, mutations of the HP1BD abolish the HP1 binding and significantly reduce the repression activity of KAP1.³⁶ This interaction is mediated by the CSD (Chromo Shadow Domain) of HP1 that binds as a dimer the HP1BD of KAP1. All previous studies explored the binding between the two proteins using purified fragments, specifically the CSD of HP1 and the HP1BD peptide of KAP1. Instead we studied

for the first time the binding between KAP1 and HP1 using full-length proteins (KAP1 FL and HP1 α FL). HP1 α FL was expressed and purified as described in the Method Chapter 2.3 (Fig. 50).

Afterwards, using full-length HP1 α and KAP1 FL we determined their binding affinity by isothermal titration calorimetry (ITC, Fig. 51A), observing an equilibrium dissociation constant (K_d) of 176 nM for the KAP1-HP1 complex, value which is in line with previous measurements obtained using protein fragments and which confirms the strong affinity between the two proteins.

Furthermore, we determined the molecular weight of the complex by SEC-MALS, exploring a range of concentrations from 10 μ M to 127 μ M for KAP1 FL and from 50 μ M to 500 μ M for HP1 α FL, i.e. covering complex stoichiometries ranging from 1:1 to 1:10 in order to saturate the binding. The proteins elute as single symmetric peaks with retention volumes of 12 mL for KAP1 FL, 16 mL for HP1 α FL and 11.8 mL for the HP1 α -KAP1 complex (Fig. 51B,C), and no concentration dependence in either elution volume or mass estimation was observed. Thus, our measurements estimate a molecular weight of \sim 231 kDa for the complex, only compatible with a dimer of KAP1 FL (182 kDa) bound to one dimer of HP1 α FL (45 kDa), for a 2:2 stoichiometry of the KAP1-HP1 α complex. This unexpected stoichiometry is confirmed by the fit of the molar ratio in the ITC experiments to \sim 1 ($N=0.98$) (Fig. 51A). The expected stoichiometry 2:4, having two HP1 α dimers bound to the two HP1BDs of the KAP1 dimer (Mw of \sim 276 kDa) was never observed. HP1 α was always a dimer in the range of concentrations used in this study. We never observed higher oligomeric states. The same observation is valid for KAP1 FL as extensively discussed above.

Based on these findings, our hypothesis is that the asymmetric nature previously observed from SAXS data for KAP1 FL in solution affects the ability of the protein to expose the HP1BD and to recruit efficiently HP1 molecules. The ensemble of KAP1 FL structures fitting the SAXS data has in fact only one HP1BD fully exposed for HP1 recruitment, while the second is mostly compacted around the RBCC domain, thus less accessible for productive binding (Fig. 44B, 45A).

No high-resolution structures are available for HP1 α full-length or for KAP1-HP1 complex. Consequently, to model the interaction between the two proteins, we collected SEC-SAXS data for HP1 α FL construct (Fig. 52A). We analysed the $P(r)$ function, compatible with a globular protein, retrieving a D_{max} of 100 Å and an R_g of 23 Å (Fig. 52B). In this case we did not have a rod-like structure so was not possible to calculate an R_{gc} value. HP1 α FL model was built using the PDB files 1S4Z⁷⁵ for the CSD and 1Q3L for the CD (To be published). The linker between the two domains (43 residues) was first built using MODELLER¹⁴⁵. The final HP1 FL model was fit to the SAXS data using NOLB¹³³, as previously described for KAP1 FL. However, in this case the normal mode analysis was not able to find a good fit for this model, meaning that the SAXS curve could not be describe by only one protein conformation. For this reason we input the structures of the CD and the CSD in the EOM (Ensemble Optimization Method) software letting the program building the

flexible linker between the two domains and to creating an ensemble of dimeric conformations that describes the SAXS curve. The EOM ensembles contained only 2 structures: a close and an open conformations (**Fig. 52C**). The average of these two conformations described our data. Finally an *ab initio* model was built using GASBOR¹⁹³ (**Fig. 52D**). Moreover, we built the model for KAP1 FL and HP1 α FL complex, placing the CSD of HP1 to bind the accessible HP1BD in the extended linker (**Fig 52E**). The position of the CSD of HP1 on KAP1 binding domain was determined by superimposing the HP1BD with the peptide co-crystallized with the CSD (PDB: 1S4Z⁷⁵) and maintaining the reciprocal orientations.

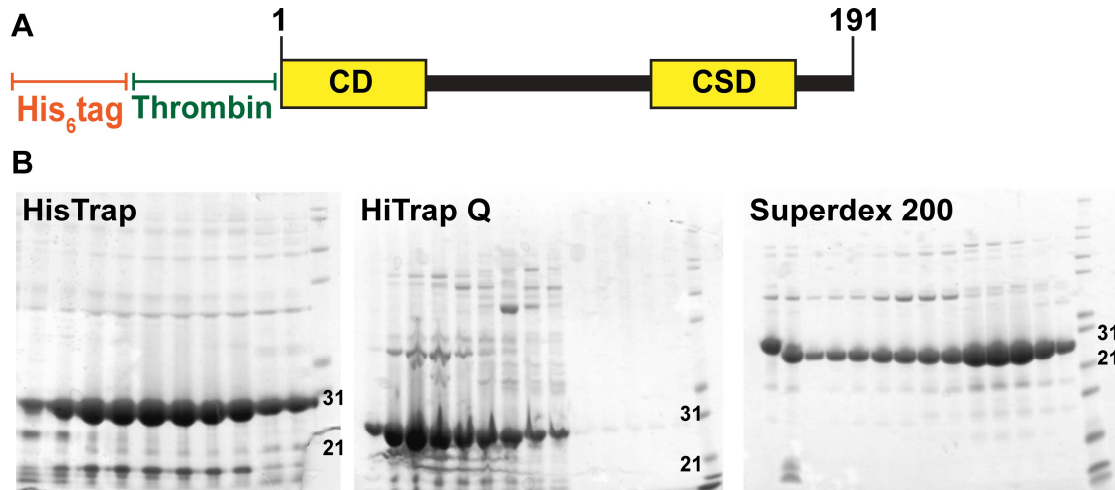


Figure 50| HP1 α FL construct **A.** HP1 α FL cartoon model. From the N-terminus: poly-histidine tag (His₆ tag) in orange, Thrombin cleavage site in green, Chromo domain (CD) and Chromo Shadow domain (CSD) in yellow. The loop of 43 residues between the two domains is shown as a black line. The construct starts in position M1 and ends in position S191 **B.** HP1 α FL purification. The protein was purified by affinity chromatography (HisTrap), ion exchange chromatography (HiTrap Q) and size exclusion chromatography (Superdex 200 16/600). The first two wells of the last gel present the protein before and after cutting the N-terminal His₆ tag with Thrombin.

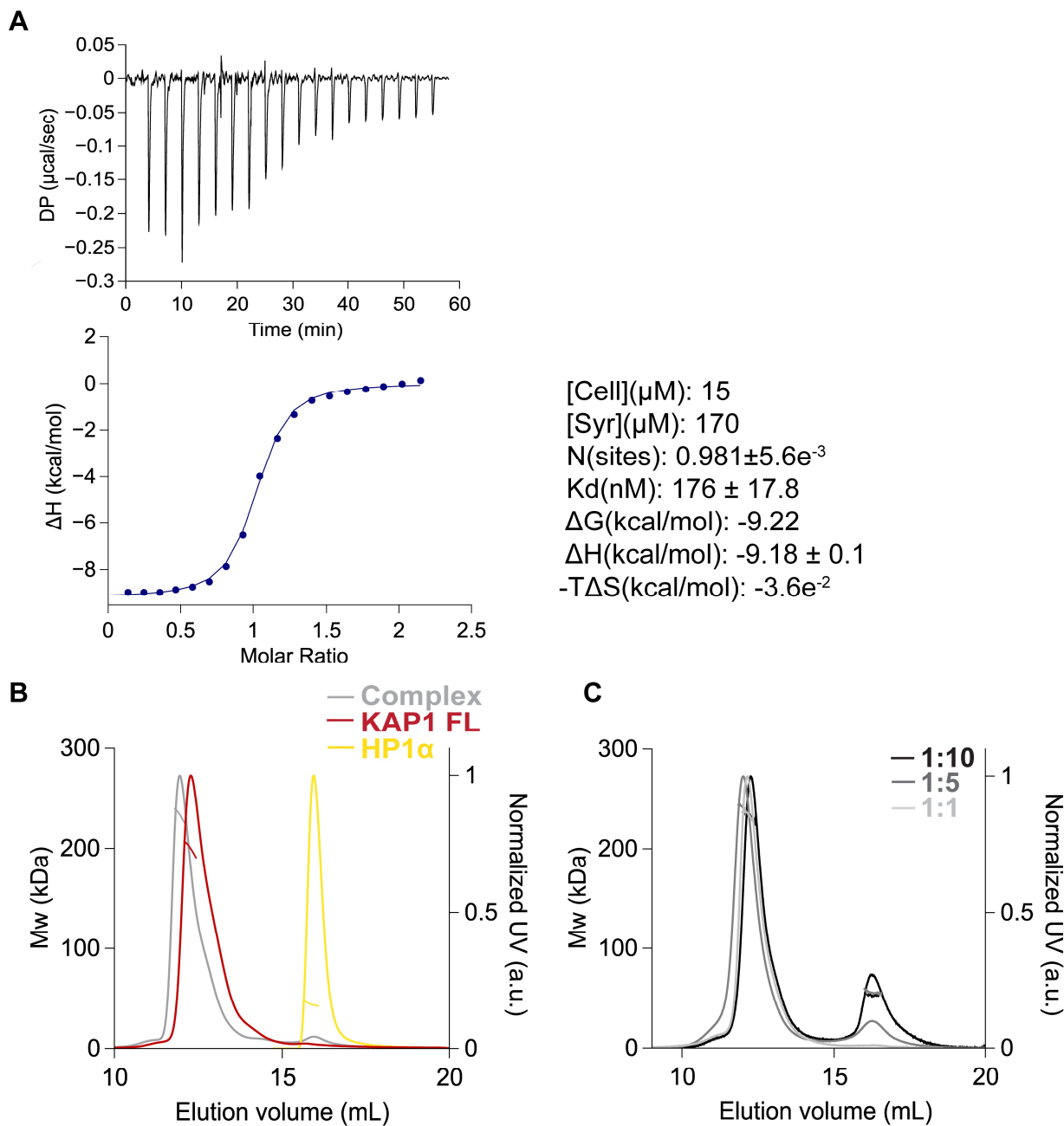


Figure 51| KAP1 dimer binds only one HP1 dimer. **A.** ITC experiment titrating HP1 α FL into KAP1 FL measures a tight interaction with a K_d of 176 nM and a molar ratio of 0.98. **B.** SEC-MALS analysis of KAP1 FL (red), HP1 α FL (yellow), and KAP1 FL- HP1 α FL complex (grey). Protein concentration has been measured with refractive index detector. **C.** SEC-MALS analysis of the complex at different concentration ratios. The estimated masses for the complexes eluting at 11.8 mL with KAP1: HP1 ratios of 1:1, 1:5 and 1:10, were 233 kDa, 239 kDa, and 232 kDa, respectively. The second peak eluting at 16 mL is the excess HP1 dimer with a mass of 45 kDa.

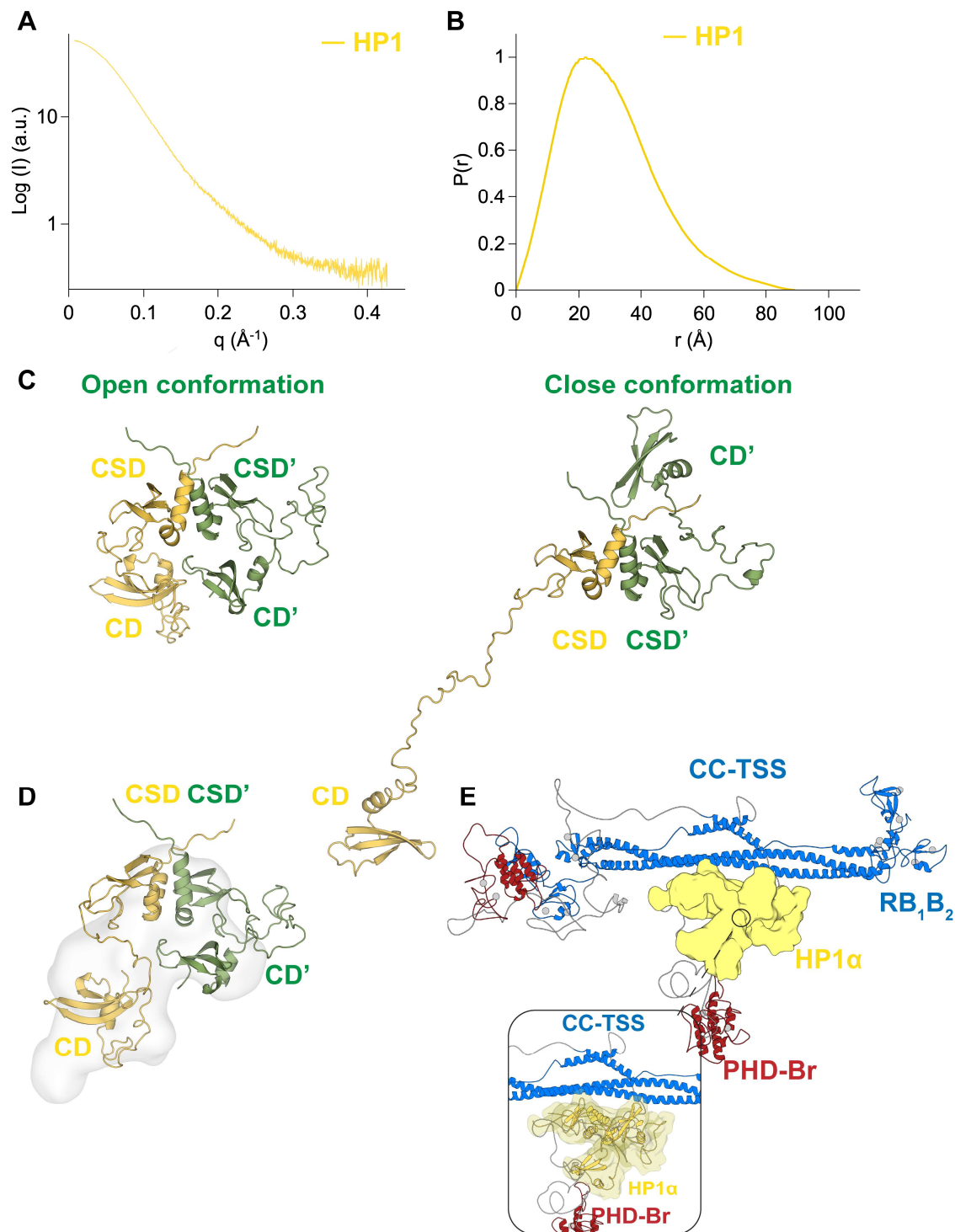


Figure 52| SAXS data analysis of KAP1-HP1 complex. **A.** SAXS scattering profile. **B.** Pair distance distribution function, $P(r)$. The R_g and D_{max} values were respectively 23 \AA and 100 \AA . **C.** HP1 α FL model built using PDB files 1S4Z⁷⁵ for the CSD and 1Q3L (*to be published*) for the CD. The two protomers are shown in yellow and green. EOM analysis revealed only 2 conformations: a close conformation (left) and open conformation (right). **D.** Closed HP1 α model superimposed to the *ab initio* shape created directly from the SAXS data using GASBOR¹²¹. **E.** Representative model of the interaction between a KAP1 FL dimer (RBCC domain in blue and PHD-Br domain in red) and HP1 α FL dimer (yellow). The CSD of HP1 has been placed as to bind the accessible HP1BD (see insert).

Further functional studies have been conducted by our collaborators (Beat Fierz, Laboratory of Biophysical Chemistry of Macromolecules) to determine how KAP1 alters the interaction dynamics of HP1 α with chromatin fibers, trimethylated at lysine 9 on histone H3 (H3K9me3) and to observe the behavior of the complex in the context of chromatin. They applied the single molecule total internal reflection fluorescence (smTIRF) imaging approach to observe directly the chromatin interaction kinetics of HP1 α in presence or absence of KAP1 FL^{204,205}. The H3K9me3 containing chromatin fibers (labeled with a fluorescent dye, Atto647N) were immobilized in a flow cell, and fluorescently labeled HP1 α (carrying an Atto 532 dye) was injected, in the presence or absence of KAP1 FL. Transient HP1 α binding interactions on individual chromatin fibers were then directly observed using smTIRF (**Fig. 53B,C**). Compared to the short dwell times of individual HP1 α molecules on chromatin (**Fig. 53B**), the presence of 100 nM KAP1 FL resulted in a higher frequency of long (> 4 s) binding events (**Fig. 53C**). Analyzing dwell-time histograms (**Fig. 53D**) revealed bi-exponential binding kinetics for HP1 α (characterized by a fast and slow residence time $\tau_{\text{off},1}$ and $\tau_{\text{off},2}$. (**Table 2**) It was previously shown that the slow exponential phase is attributed to bivalent chromatin binding by HP1 α dimers. KAP1 addition indeed stabilized HP1 α chromatin binding, exhibiting a larger percentage of bound molecules at times >1s (**Fig. 53D**). Conversely, binding events were less frequent in the presence of KAP1, as indicated in the slower binding kinetics (**Fig. 53E**). In the presence of KAP1, HP1 binds less to the chromatin but spends more time on in.

To more quantitatively elucidate the effect of KAP1 on HP1 α binding dynamics our collaborators tested the effect of KAP1 FL concentrations ranging from 50 nM to 400 nM. While the fast binding times $\tau_{\text{off},1}$ did not exhibit a systematic dependence on KAP1 FL concentration (**Fig. 53F**), $\tau_{\text{off},2}$ was increased by around three-fold at KAP1 > 100 nM (**Fig. 53G**). KAP1 stabilizes bivalent chromatin interactions of HP1 α dimers, consistent with the fact that a KAP1 dimer can only bind one HP1 dimer as shown by the SEC-MALS results.

Conversely, binding rate constants (**Fig. 53H**) were reduced by KAP1, consistent with slower diffusion dynamics of the HP1 α -KAP1 complex. Finally, an analysis of the fluorescence intensities of the single-molecule observations allowed the quantification of the oligomeric state of HP1 α . In the absence of KAP1, HP1 α was mainly monomeric (due to the low concentrations employed) and only 8% dimers were observed (**Fig. 53I**). Addition of 100 nM KAP1 resulted in an increase of dimers (up to 28%, **Fig. 53J**), consistent with our previous observations that KAP1 binds and stabilizes a single HP1 α dimer. KAP1 stabilizes HP1 dimers. In the presence of KAP1, a tetrameric form of HP1 was never observed consistently with the previous SEC-MALS results showing that only one HP1BD can be occupied by one HP1 dimer while the other binding site on KAP1 remains unoccupied.

[KAP1](nM)	$\tau_{off,1}$ (s)	% A ₁	$\tau_{off,2}$ (s)	% A ₂	k_{on} (M ⁻¹ s ⁻¹) x 10 ⁶	n
0*	0.25 ± 0.03	87 ± 7	2.26 ± 1.22	13 ± 7	3.64 ± 1.56	10
50	0.29 ± 0.05	89 ± 3	3.64 ± 0.47	11 ± 3	1.03 ± 0.37	4
100	0.28 ± 0.02	85 ± 3	5.79 ± 1.03	15 ± 3	1.42 ± 0.95	4
200	0.27 ± 0.04	87 ± 2	5.47 ± 0.93	13 ± 2	1.43 ± 0.45	3
400	0.23 ± 0.02	86 ± 11	5.03 ± 0.65	14 ± 11	0.77 ± 0.44	3

Table 2 smTIRF values. Fit results from smTIRF measurements of HP1 α in the presence of the indicated concentrations of KAP1. N denotes the number of independent experiments, each contributing > 100 kinetic traces.* values taken from ref²⁰⁵.

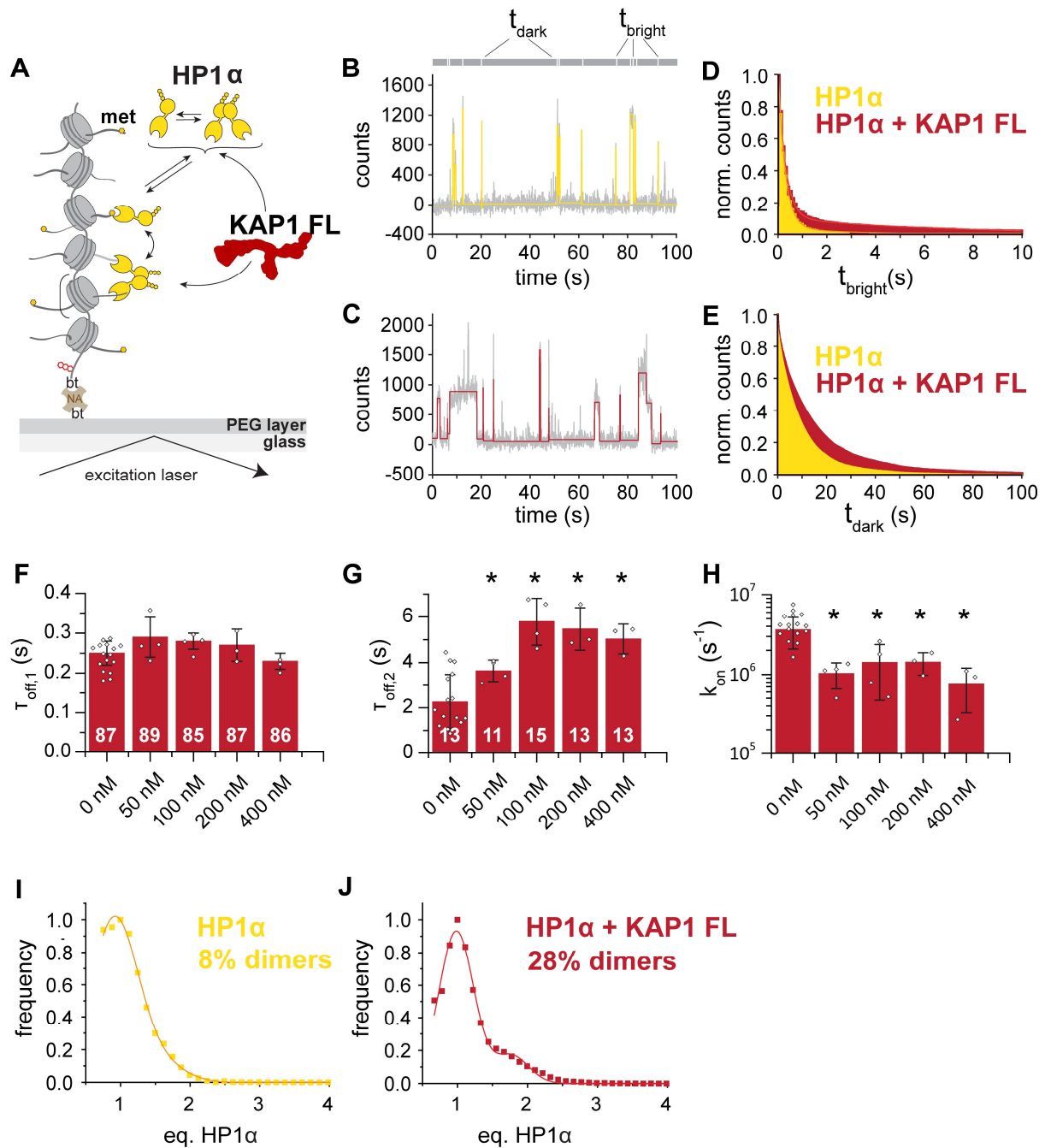


Figure 53| KAP1 stabilizes HP1 α –chromatin interactions. **A.** Schematic representation of the smTIRF imaging experimental setup, showing HP1 α interacting with chromatin fibers in the presence of KAP1. bt: biotin, NA: neutravidin **B.** Characteristic fluorescence time trace (grey) of HP1 α (3 nM) binding dynamics to a single chromatin fiber, in absence of KAP1. High fluorescence emission reveals the time an HP1 α molecule is bound (t_{bright}), whereas low fluorescence emission indicates the absence of any bound molecules over a given time (t_{dark}). The trace is fitted with a step function (yellow). Each intensity peak represents one binding event **C.** Fluorescence time trace of HP1 α – chromatin binding dynamics, in the presence of 100 nM KAP1. The trace is fitted with a step function (red) **D.** Dissociation kinetics: normalized, cumulative histograms of HP1 α dwell times in the absence of KAP1 (yellow) or presence of 100 nM KAP1 (red). Both histograms are fitted with a double-exponential function. For fit values, see **Table 2** **E.** Association kinetics: normalized cumulative histogram of times between binding events for HP1 α alone (yellow) or HP1 α in the presence of 100 nM KAP1 (red), fitted with mono-exponential function. For fit values, see **Table 2** **F.** Dwell times $t_{\text{off},1}$ and **G.** dwell

times $t_{\text{off},2}$, demonstrating that KAP1 stabilizes HP1 α -chromatin interactions. Numbers indicate % amplitude. Error bars: standard deviation (s.d.), n = 3 - 10 replicates, *: p<0.05 vs. 0 nM KAP1, student's t-test. For the fit values, see **Table 2**. **H.** KAP1 reduces on-rates k_{on} . Error bars: standard deviation (s.d.), n = 3 - 10 replicates, *: p<0.05 vs. 0 nM KAP1, student's t-test. For the fit values, see **Table 2 I**. A histogram of observed, normalized fluorescence peak intensities in kinetic traces reports on the oligomeric state of chromatin bound HP1 α (yellow). This demonstrates that HP1 α exists mostly as monomers at 3 nM concentration. Considering the observed labeling efficiency of 60%, the 8% of all observations correspond to dimers **J**. At 100 nM, KAP1 stabilizes HP1 α dimers (28%) (red).

SAXS data and models were deposited in the Small Angle Scattering Biological Data Bank SASBDB with accession codes SASDEV6, SASDER7 and SASDEW6 (**Appendix 9, Table 4-SAXS data**).

3.9 Conclusions

KAP1 is an elongated and flexible protein. At first, to shed light on KAP1 structural organization we applied high-resolution structural biology techniques such as X-ray crystallography and cryo-electron microscopy. We were able to crystallize the RBCC domain however the crystals diffracted at low resolution (**Appendix 2**). The first approach was to stabilize the coiled-coil domain as we thought these domains were interacting between each other in the crystal lattice perturbing the formation of a well order repeated structure. In 2016 was published the antiparallel crystal structure of the Rhesus Trim5 α B-box coiled-coil domain in complex with the T4 lysozyme (PDB: 5F7T¹⁸⁵). In this structure the T4 lysozyme interacted with the central part of the coiled-coil domain. Consequently we designed a new construct in which the RBCC domain was fused with the T4 lysozyme (**Fig. 31A,B,C** and **Appendix 2**). The protein was successfully purified and crystallized. However the crystals diffracted at low resolution and we concluded that the coiled-coil domain was not perturbing significantly the organization of the crystal lattice. The second approach was to performed limited proteolysis experiments on the RBCC domain to identify new protein fragments for crystallization trials. We were not able to detect any expression of these recombinant proteins in the four *E.coli* strains. For this reason we could not perform any crystallization trial given the impossibility to obtain soluble proteins (**Fig. 32, Fig. 33** and **Appendix 2**). For the first time we observed the impossibility to express and purify a construct lacking both the N-terminal RING domain and the C-terminal PHD-Br domain module, which may be linked to a role of the RING domain in protein stability. Subsequently we crystallized the full-length protein obtaining crystals with the same shape and length as the one observed for the RBCC domain and they diffracted at low resolution (**Appendix 2**). Given the unsuccessful crystallization trials, we decided to perform negative staining and Cryo EM studies on KAP1 FL. We were able to observe elongated structures (**Appendix 3**). The sample was homogeneous and not aggregated however it was conformationally too heterogeneous to perform any particle reconstruction. At the end of these studies we were able to conclude that KAP1 FL was an elongated dimer with an intrinsic flexibility that perturbed the formation of a stable crystal lattice and prevented the Cryo EM reconstruction.

The RBCC domain is a symmetric dimer. Given the intrinsic flexibility of KAP1, the classical structural biology techniques did not provide any high-resolution information on the overall protein organization. For this reason, we decided to follow an alternative IM approach combining low-resolution structural data with computational analysis and biophysical and functional assays. Doing so, we collected SEC-SAXS data for the RBCC domain, KAP1 FL and Δ KAP1 applying a novel integrative modelling approach that combines solution scattering diffraction data and flexible data fitting and we found an unexpected asymmetry in the KAP1 molecule. First we observed that, like the other TRIM family members, KAP1 contains an N-terminal RBCC domain that folds into an antiparallel dimer with a long coiled-coil domain that separates the two RB₁B₂ modules located at the opposite ends of it. This result was obtained from the analysis of the P(r) function of the RBCC domain, clearly bimodal and only compatible with an antiparallel organization in which the two tandem RB₁B₂ modules are separated by the length of the coiled-coil domain (**Fig. 38A**). The ensemble of solutions we obtained that optimally fit the SAXS data are highly consistent and similar as the main differences are given by slight modifications of the reciprocal orientation of the RING, B-box1 and B-box2 domains. The asymmetry of RBCC is minor in solution since the RB₁B₂ domains can likely explore slightly different conformations but at the level of resolution accessible by SAXS these deviations can be seen as symmetric states (**Fig. 40A,B,C**). We concluded that the RBCC domain could be considered a symmetric dimer.

KAP1 full-length is an asymmetric antiparallel dimer. In solution, the C-terminal halves, with the long flexible linker and the PHD-Br tandem domain, seem not to be randomly distributed around the coiled-coil domain, but they behave in an asymmetric fashion. Flexible fitting to the SAXS scattering curve of around 600 models resulted in a conformation where one C-terminal PHD-Br domain is close to one RB₁B₂ module, while the other is far away (**Fig. 43**, **Fig. 44** and **Fig. 45**). This analysis can be compared to the one performed for TRIM25²⁰⁶ both in the method and in the results obtained. The authors created an ensemble of structures where the linkers are disordered and the C-terminal PRYSPRY domains are either fixed in the crystallographic position or allowed to move freely, both in a symmetric and asymmetric fashion. They used an ensemble optimization method (EOM²⁰⁷) and they compared the calculated scattering profiles from the ensembles or the X-ray structure to the experimental SAXS data using Crysol¹³⁵. For TRIM25, SAXS indicates that in solution the C-terminal domains are freely moving constrained by the flexible linker (73 aa), allowing for a transient interaction with the coiled-coil domain that was observed by crystallography and that is important for substrate ubiquitination²⁰⁶. In fact, the ensemble of solution that fit the SAXS data contains both structures in which the C-terminal domains are in close proximity to the N-terminal domain and structures in which the PRYSPRY domains are completely detached from the N-terminus. On the contrary, for KAP1 a conformation in which the two C-terminal domains are found at the same time close or far away from the N-terminal domain was never observed and the protein maintained always an asymmetric organization.

In conclusion, our method guides the randomly generated initial models into conformations that fit the SAXS data applying nonlinear Cartesian Normal Mode Analysis (NOLB¹³³ NMA) for SAXS data fitting. We obtained a set of 600 similar structures that converged into a representative cluster center that is consistent with the SAXS data (**Fig. 45A**). In this ensemble of structures the PHD-Br domains are never found at the same time close or far from the RBCC domain, but they always behave in an asymmetric fashion with one PHD-Br domain far and one close to the RB₁B₂ module (**Fig. 44A,B,C**). In conclusion, our SAXS analysis of KAP1 FL (and the Δ KAP1 construct) showed evidence for an asymmetry in the conformational states of the two C-termini not previously reported for TRIM25^{206,208} or other TRIM family members. This novel protocol can be more generally applied to treat elongated and highly flexible systems and fit different data dealing with low-resolution volumetric maps reporting on protein complex architecture.

The asymmetry drives KAP1 SUMOylation function. The auto-SUMOylation assays have shown that the RING domain does play a role in the SUMOylation of the C-terminus. As shown by us and others⁶⁸, SUMOylation happens at lysine residues at the C-terminal end and some SUMOylation sites remain SUMOylated after the PHD-Br domain is compromised with catalytic incompetent mutations (e.g., C651A) (**Appendix 6**). The remaining SUMOylation could be due to (i) the existence of another E3 ligase active site or (ii) the PHD being still active despite C651A mutation. Unfortunately, KAP1 bearing RING truncations together with PHD-Br inactive mutations proved to be insoluble, not allowing us to distinguish between these two situations. We provide evidence that the RING domain can act as an additional catalytic domain SUMOylating those sites or can act synergistically with the PHD domain for the same purpose. RING domains in other TRIM family members dimerize to preserve E3 ligase activity (e.g., TRIM32²⁰⁹, TRIM5 α ²¹⁰ or TRIM23²¹⁰), but the RING domain of KAP1 appears to be monomeric¹⁸⁰. This finding supports our model in suggesting a possible synergy between the RING and the PHD domain, which are in proximity based on our SAXS-based models. Therefore, we conclude that KAP1 has two separate SUMO E3 ligase active sites^{212–214}, which might work independently or synergistically depending on the substrate available (**Fig. 49**).

The asymmetry is functional for the interaction with binding partners. We characterized the interaction between KAP1 and HP1 α , using for the first time full-length purified proteins. HP1 α is a dimer with two globular domains linked by a flexible linker that could be found in solution in the closed or in the open state as shown by our SAXS data and others²¹⁵ (**Fig. 52**). Previous data using the HP1 Chromo Shadow domain (CSD) and the KAP1 HP1 binding domain (HP1BD) showed that the CSD forms dimer that bind only one KAP1 peptide²¹⁶. According to our MALS data, the interaction between the two proteins is maintained when using full-length proteins but only one of the two KAP1 FL binding sites is available for the interaction. The observed stoichiometry 1:1 (1 KAP1 FL dimer bound to 1 HP1 α FL dimer) suggests that the asymmetric structure of KAP1 modulates the interaction with binding partners (**Fig. 51B,C**). As this binding asymmetry could be triggered upon HP1 binding, it can also be reasonable to think that the asymmetry of the C-terminal

domains found in solution is responsible to modulate HP1 binding accessibility. Moreover it is not possible to place two HP1 α FL proteins in both HP1BDs of the asymmetric KAP1 FL dimer without causing clashes while building the molecular model of the interacting proteins (**Fig. 52E**).

Within chromatin fibers, KAP1 dimers recruit HP1 as dimers and not as tetramers, to the trimethylated nucleosomes such that two Chromo domains (CD) per complex can reach for the histone tails. This observation confirmed again the accessibility of only one HP1BD bound by a dimer of the HP1 Chromo Shadow domain (**Fig. 53I,J**). This interaction is more durable than the one in the absence of KAP1, allowing us to speculate that in cells, the binding of KAP1 stabilizes HP1 dimerization in the chromatin facilitating heterochromatin formation and spreading (**Fig. 53B,C**). Moreover, in the presence of KAP1, HP1 binds less the chromatin indicating a regulatory effect of KAP1 on the HP1-chromatin interaction (**Fig. 53H**). Similarly, this functional asymmetry in recruiting partners with controlled stoichiometry can be present for other KAP1 interacting partners engaged by the C-termini, such as SETDB1, or that interact with the coiled-coil domain, such as the MAGE (melanoma antigen genes) protein or KRAB-ZFPs.^{43,217,34}. Recently it has been shown that KAP1 FL binds only one KRAB domain at the center of the coiled-coil domain⁵⁸. Finally we have speculated on the fact that this asymmetry can be linked to multiple KAP1 functions, such that one RING/PHD-Br unit of the dimer can be responsible for the SUMOylation activity, while the other PHD-Br domain is SUMOylated and able to interact with chromatin remodelling enzymes.

To summarize, we can conclude that:

- The **RBCC domain** is an elongated antiparallel symmetric dimer; The dimerization domain is given by the coiled-coil domain and the RB₁B₂ modules are located at its extremities separated by a length of 160 Å; The ensemble of solutions that optimally fit the SAXS data are highly consistent and similar as the main differences are given by slight modifications of the reciprocal orientation of the Ring, B-box 1 and B-box 2 domains; In solution, the RB₁B₂ modules can likely explore slightly different conformations but at the level of resolution accessible by SAXS these deviations can be seen as symmetric states.
- **KAP1 FL** is natively asymmetric in solution; A large ensemble of models that satisfies the SAXS data can be identified for any given χ^2 cutoff, for example with optimized χ^2 lower than 2 there are ~500 models that describe our SAXS data; The position of the loops and the C-terminal domain itself is not the same in these models and given the low resolution of the SAXS data we cannot discriminate among them; However all these models maintain a common configuration of the C-termini with one PHD-Br domain close and one far from the RB₁B₂ module.

Chapter 4 Discovering small molecule compounds targeting the KAP1 Bromodomain

Bromodomain (BRD) modules are a class of evolutionary conserved domains found in proteins involved in the recognition of the acetylated lysine residues of the histone tails. This PTM is controlled by two classes of proteins: HAT_s (histone acetyltransferases) and HDAC_s (histone deacetylases). Lysine acetylation regulates a broad range of biological functions such as (i) DNA damage response²¹⁸(ii) chromatin compaction¹⁶ (iii) DNA-protein interactions²¹⁹ and (iv) signal transduction²²⁰. The upregulation of the histone acetylation level is involved in cell proliferation, oncogenes expression and cancer development²²¹. The down regulation of HDAC promotes cell growth and the deregulation of tumor suppressor genes²²². HDAC_s inhibitors are used in clinics for cancer treatments however they are largely non-selective molecules and their mechanism of action is poorly understood^{223,224}. Hence, the need to develop novel selective compounds for proteins involved in the regulation of histone acetylation arose. BRD_s inhibitors have emerged as a new treatment strategy for cancer²²⁵, inflammation²²⁶ and viral infections²²⁷.

4.1 Bromodomains: structure and functions

Over 60 different BRDs are present in the human proteome. They are divided into eight subfamilies based on their structures²²⁸ (**Fig. 54**). BRDs are involved in several biological pathways including transcriptional regulation²²⁹, control of nuclear-factor- κ B (NF- κ B) involved in inflammatory processes²³⁰ and replication of viral genomes through regulation of viral proteins transcription²³¹. They all share the same structural organization consisting of four α -helices with a left-handed packing (α Z, α A, α B, α C) connected by two loops known as ZA loop between the α Z and α A helices and BC loop between the α B and α C helices (PDBs: 4WHU (*to be published*), 6HDN¹⁶¹, 6HDQ¹⁶¹, 6MO9²³², 5H1T¹⁶³, 5VOM¹⁶⁴, 2RO1⁶²) (**Fig. 55A,B**). High-resolution crystal structures have shown that the BRDs recognize the acetylated lysine in a central hydrophobic pocket located between the ZA loop and the BC loop (PDBs: 1E6I²³³, 2RNX²³⁴). The acetylated lysine interacts directly with a conserved asparagine residue (Asn140) located in this hydrophobic cavity^{235,233} (**Fig. 55B**). Moreover, the BET BRDs subfamily are able to recognize simultaneously two acetylated lysine residues. One acetylated lysine interacts directly with the conserved asparagine while the second one is stabilized by a network of hydrogen bonds formed by conserved water molecules present in the binding pocket (PDB: 3UVW²³⁶) (**Fig. 56A,B,C**).

KAP1 is a member of the TRIM C-VI subfamily, together with TRIM24 and TRIM33 and it is characterized by the presence of a tandem plant homeodomain (PHD) and Bromodomain (Br) typically involved in the

recognition of various histones modifications^{65,105}. The C-terminal tandem PHD-Br domain of KAP1 does not recognize the acetyl lysine of histone H3 and H4. On the contrary it shows a unique function acting as an E3 SUMO ligase and promoting both the auto-SUMOylation of the protein⁶⁸ and the SUMOylation of other substrates^{67,237}. The auto-SUMOylation of the C-terminal PHD-Br domain is necessary for the binding of KAP1 to the chromatin remodeling enzymes, for the PTMs deposition on the chromatin and for the heterochromatin spreading^{34,197,202,238,35}.

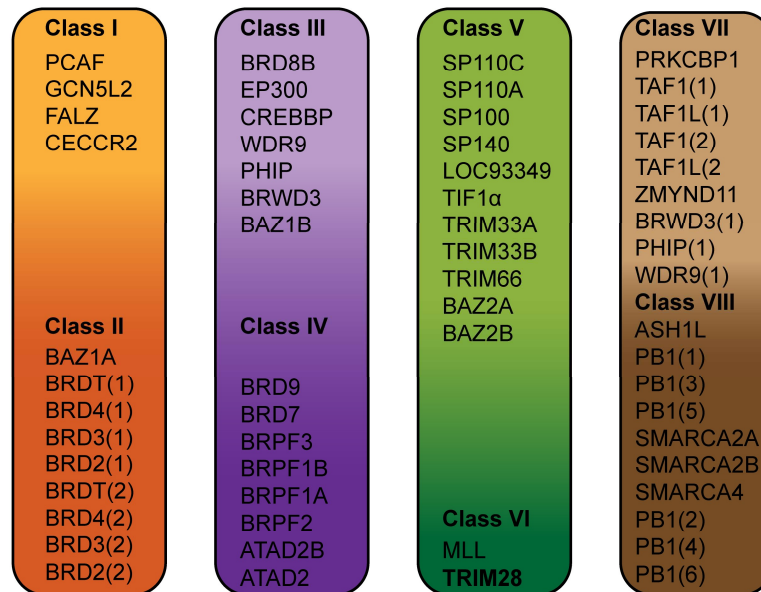


Figure 54| Bromodomains classification. The BRDs are classified into 8 subfamilies. These include among the others the general control of amino acid synthesis protein 5 like 2 (GCN5L2), Bromodomain-containing proteins (BRD), P300/CBP associated factor (PCAF), helicases (SWI/SNF-related matrix-associated actin-dependent regulators of chromatin subfamily A (SMARCA_s), histone methyltransferases (lineage leukemia protein, MLL), transcription factors (tripartite motif containing proteins (TRIM_s), ATP-dependent chromatin remodeling complexes (Bromodomain adjacent to zinc finger domain protein 1B (BAZ1B)), TBP-associated factors (TAF_s), transcriptional mediators (TAF1), nuclear scaffold protein (polybromo 1, PBRM1) and BET BRDs. KAP1 Bromodomain (TRIM 28) is part of the Class VI together with MLL.

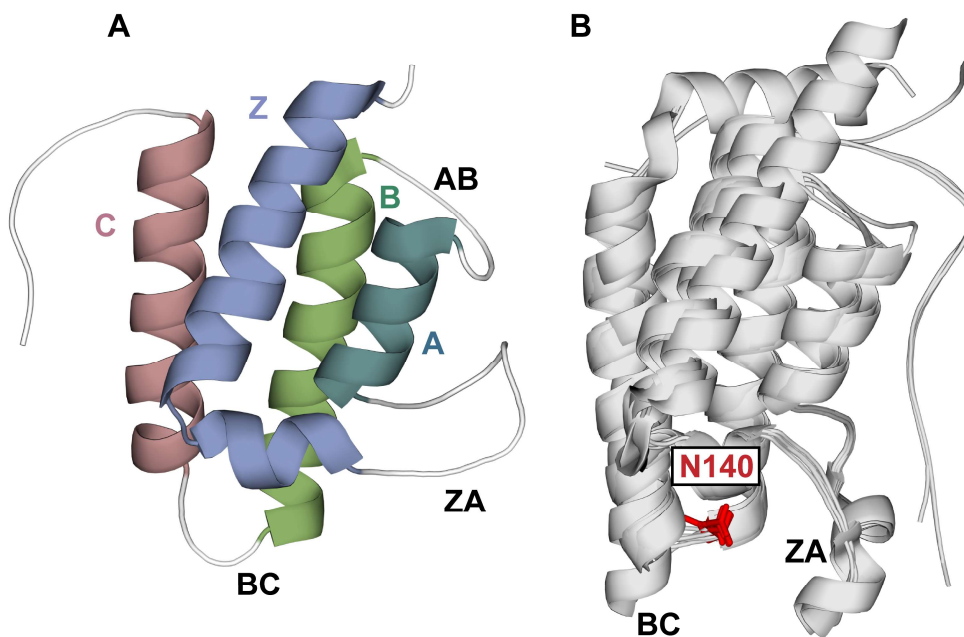


Figure 55| Bromodomain structure. **A.** NMR structure of KAP1 Bromodomain (2RO1⁶²). The four left-handed α -helices (α Z, α A, α B, α C) are highlighted with different colors. The loops between the α Z and α A (ZA loop), α B and α C (BC loop), α A and α B (AB loop) are shown in light grey **B.** Superimposition of 6 BRDs (PDBs: 6HDN¹⁶¹, 6HDQ¹⁶¹, 6MO9¹⁶², 4WHW *to be published*, 5H1T¹⁶³, 5VOM¹⁶⁴) showing the conserved four helical structures. The conserved asparagine residue (N140) is highlighted in red.

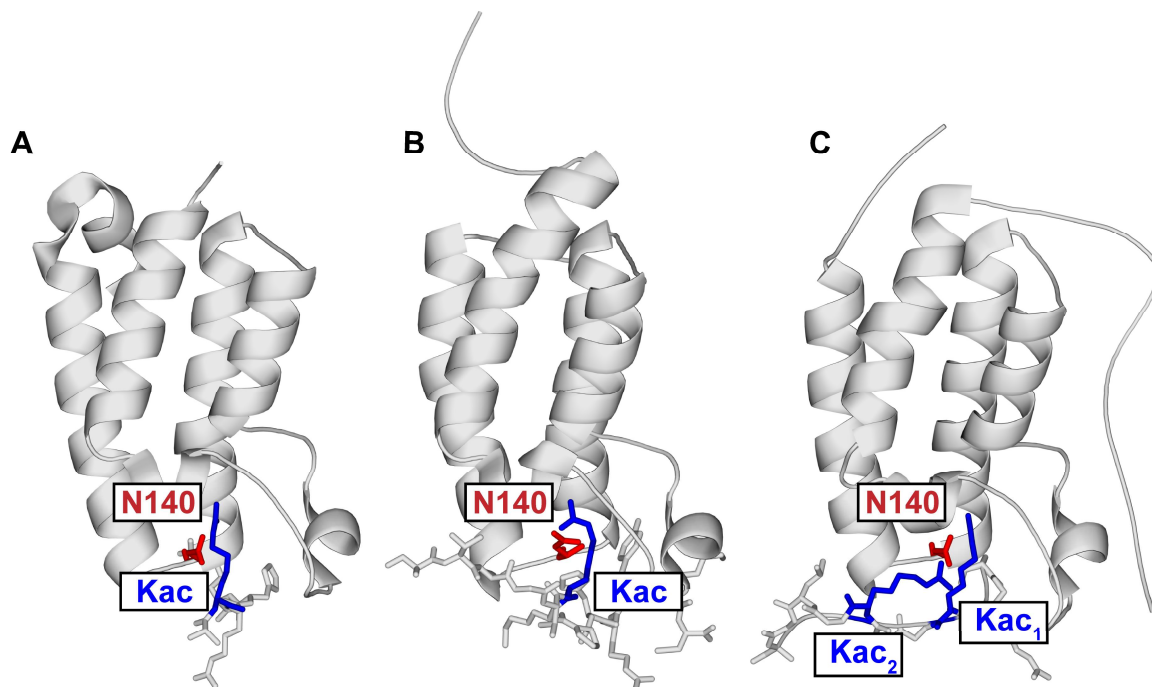


Figure 56| Bromodomains interaction with acetylated lysine. **A.** Crystal structure of GCN5 in complex with acetylated H4 peptide (PDB 1E6I²³³) **B.** Crystal structure of PCAF in complex with acetylated H3 peptide (PDB: 2RNX²³⁴) **C.** Crystal structure of BRD4 in complex with di-acetylated (Kac₁ and Kac₂) H4 peptide (PDB: 3UVW²³⁶). The conserved asparagine residue (N140) is highlighted in red, the acetylated lysine residues in blue (Kac).

4.2 Bromodomains inhibitors

The BRDs inhibitors can be classified based on how they engage the binding cavity (acetylated lysine mimetic and non acetylated lysine mimetic) or based on their target (BET BRDs (BRDT, BRD2, BRD3, BRD4) inhibitors and non-BET BRDs inhibitors).

Acetylated lysine mimetic compounds. This class contains molecules that engage the hydrophobic cavity between the ZA loop and BC loop and form a hydrogen bond with the conserved asparagine residue (N140). Usually, molecules belonging to this class bind deep in the binding pocket mimicking the binding of the acetylated lysine residues without displacing the water molecules located at the bottom of the cavity²³⁹ (**Fig. 57A,B**). Compounds belonging to this class include: triazolothienodiazepine and benzodiazepine, isoxazole-azepine, dimethyl isoxazole, quinoline, quinazoline and quinazolone templates.

Non-acetylated lysine mimetic compounds. The inhibitors found in this class engage the hydrophobic pocket without forming a hydrogen bond with the conserved asparagine residue but sterically preventing the binding of the acetylated peptide²⁴⁰ (**Fig. 57C,D**). Compounds belonging to this class include: azobenzene, benzoimidazole templates.

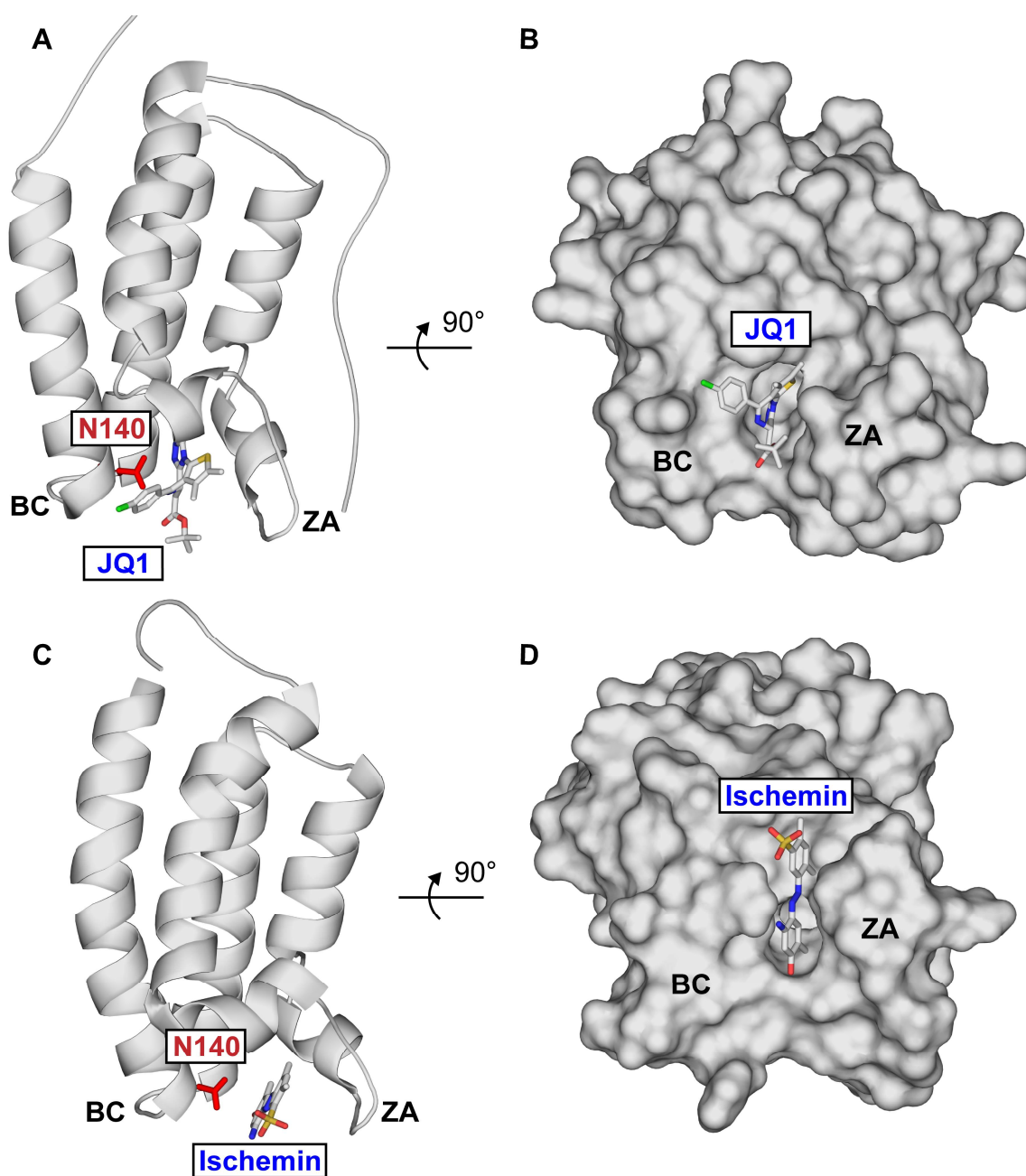


Figure 57| Bromodomains inhibitors. A. and B. show the cartoon and surface representation of a acetylated lysine mimetic compound (JQ1 in complex with BRD4, PDB: 3MXF²³⁹). The inhibitor binds deep in the hydrophobic pocket between the ZA and BC loop forming a hydrogen bond with the conserved asparagine (N140) shown in red. C. and D. show the cartoon and surface representation of a non-acetylated lysine mimetic compound (Ischemin in complex with CREBBP (CREB-binding protein) PDB: 2L84²⁴⁰). The inhibitor stays on the surface of the hydrophobic pocket between the ZA and BC loop without forming the hydrogen bond with the conserved asparagine (N140) shown in red.⁵

⁵ Nitrogen atoms are shown in blue, Oxygen atoms in red, Carbon atoms in grey, Sulfur atoms in yellow.

4.2.1 Triazolothienodiazepine and Benzodiazepine templates

Triazolothienodiazepine compounds present a triazolo ring connected to a diazepine ring and a thieno ring (**Fig. 58A**). They bind deeply in the hydrophobic pocket between the ZA loop and BC loop through the triazolo ring that forms a stable hydrogen bond with the conserve asparagine (N140) (PDB: 3MXF²³⁹). At the same time the diazepine ring occupies the same position as the second acetylated lysine of the H4 peptide (PDB: 3UVW²³⁶) (**Fig. 56C**) while the thieno ring is packed against the ZA loop resulting in a complete occupancy of the binding site.

JQ1 belongs to this class of BET inhibitors and show affinity for over 46 different BET BRDs with affinities ranging from 49 nM to 190 nM²³⁹. JQ1 is actually tested in preclinical studies against a variety of cancer types (lung, breast, prostate, pancreatic, colon, hepatocellular cancers, medulloblastoma, glioblastoma and hematological malignancies^{106,241,242,243}) (**Fig. 58B**).

MS417 shares the same triazolothienodiazepine scaffold as JQ1 but instead of a methyl ester group it contains a butyl ester group. Previous studies demonstrated that the compound binds the BRD of BRD4 with an affinity of 36 nM. The crystal structure of BRD4 in complex with the inhibitor has been solved (PDB: 4F3I²⁴⁴). The compound has a strong anti-inflammatory activity suppressing the activation pathway of NF-κB (**Fig. 58C**).

I-BET762 regulates the level of APOA1 (Apolipoprotein A1) through the inhibition of BRD2, BRD3 and BRD4^{245,246,247}. It is actually tested in clinical trials on subjects with relapsed, refractory hematologic malignancies (Trial NCT01943851).

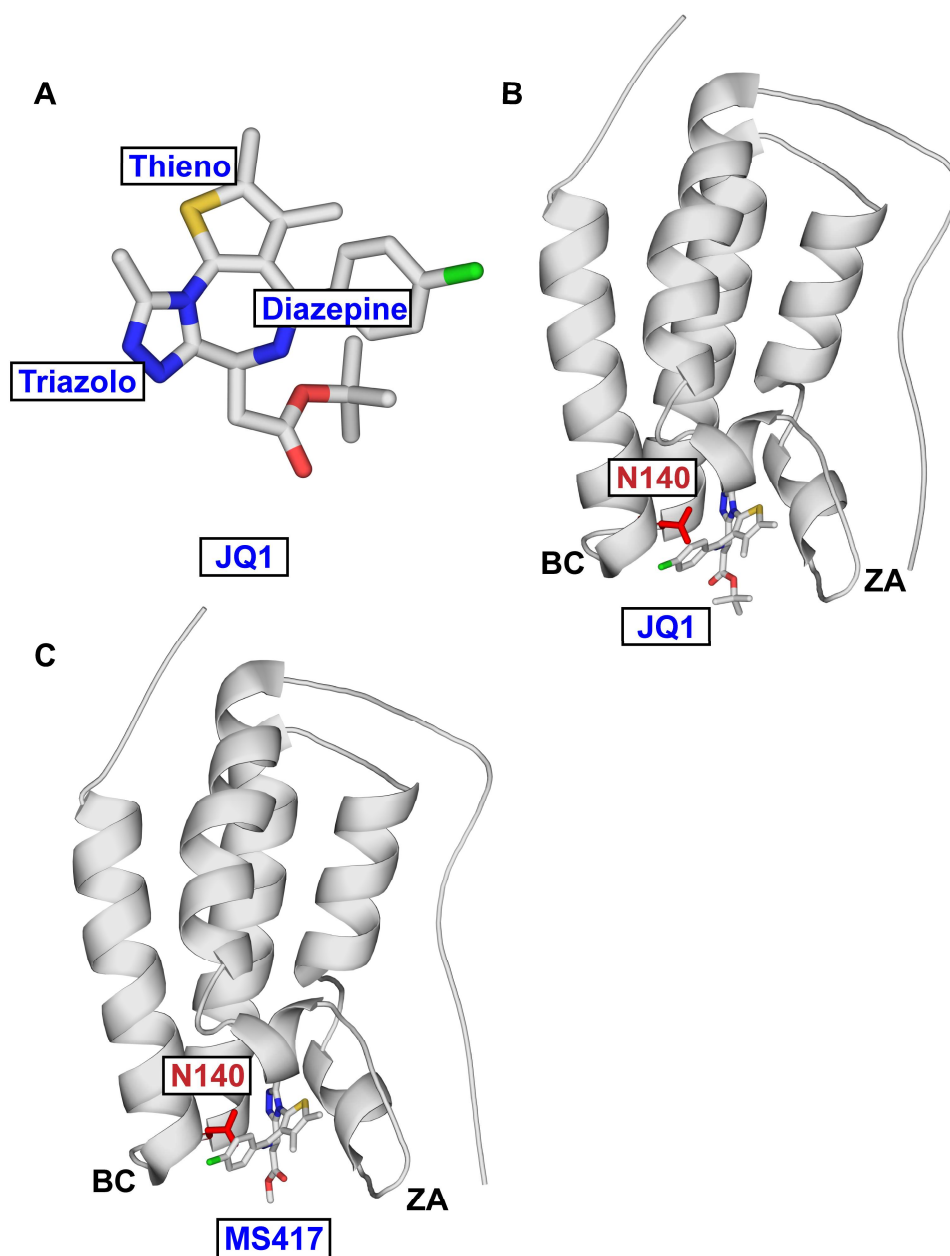


Figure 58| Triazolothienodiazepine templates. **A.** Structure of a typical triazolothienodiazepine compound (JQ1) showing the thieno ring, the diazepine ring and the triazolo ring **B.** JQ1 in complex with BRD4 (PDB: 3MXF²³⁹). **C.** MS417 in complex with BRD4 (PDB: 4F3I)²⁴⁴. The binding mode is the same as JQ1. The conserved asparagine (N140) is shown in red.

4.2.2 Isoxazole-azepine templates

The Isoxazole-azepine compounds present a combination of isoxazole and azepine rings (**Fig. 59A**). They are active against several BET BRDs engaging the binding pocket with the same acetylated lysine binding mode observed for JQ1 and I-BET762. The **compound 3** belongs to this second class. It was co-crystallized with

BRD4 (PDB: 4LRG²⁴⁸) (**Fig. 59B**) and in vivo studies proved its stability showing an IC₅₀ below of 500 nM when tested against several BET BRDs in assays that measured the IL-6 (Interleukin 6) secretion²⁴⁸.

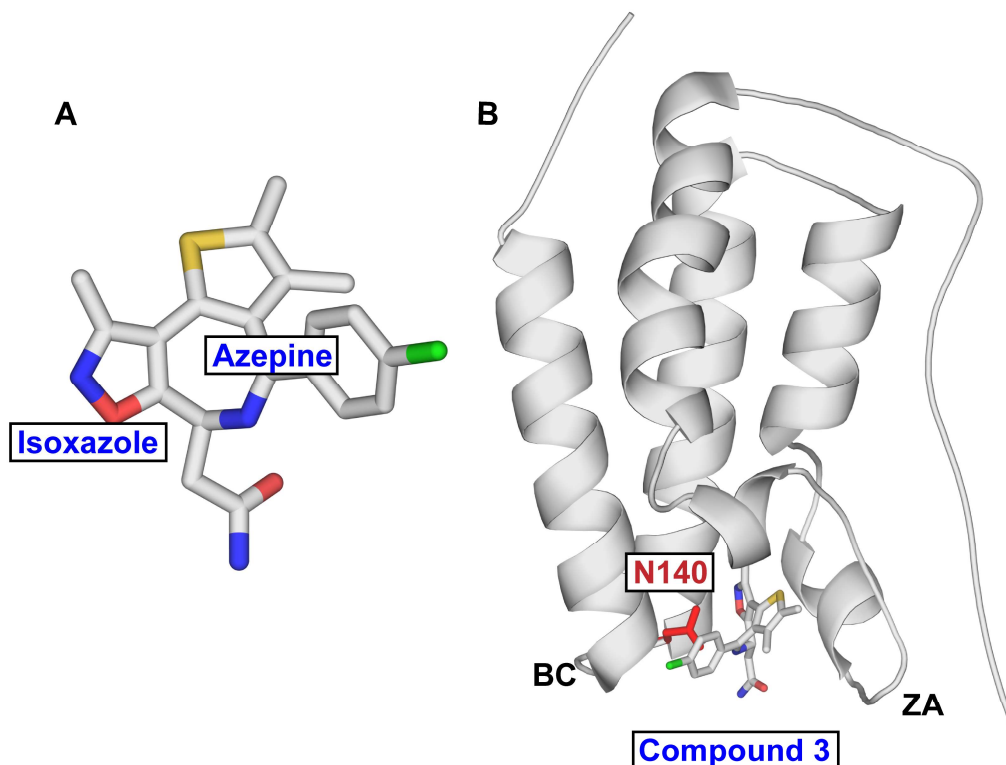


Figure 59| Isoxazole-azepine templates. **A.** Structure of compound 3 showing the isoxazole ring combined with the azepine ring **B.** Compound 3 in complex with BRD4 (PDB: 4LRG²⁴⁸). The conserved asparagine (N140) is shown in red.

4.2.3 Dimethyl Isoxazole templates

These compounds have been developed against BET BRDs using a structure-based approach in which the initial isoxazole scaffold was gradually optimized obtaining a new class of potent and specific inhibitors (**Fig. 60A**). The best representative compound in this class is **I-BET151**, a dimethylisoxazole small molecule that selectively binds the BET subfamily with a K_d below 100 nM (PDB: 3ZYU²⁴⁹) (**Fig. 60B**). This compound regulates the cell-cycle and apoptosis process in leukemia cells, blocking their growth^{249,250}. Moreover, I-BET151 has been shown to be active in assays that measured the IL-6 (Interleukin 6) and TNF (Tumor necrosis factor) secretions, increasing in mice the survival rate from 6% in the controls to 83% in the treated animals after 4 days of injections²⁵¹. I-BET151 is now tested in preclinical trial against mixed lineage leukemia, myeloma, glioblastoma, acute myeloid leukemia and melanoma^{252,253,172,254,255}. Other compounds belonging to this class are: **isoxazole 4d** (active against BRD2 and BRD4²⁵⁶, PDB: 3SVG (*to be published*)) (**Fig. 60C**), **compound 15** (a isoxazolylbenzimidazole molecule active against BRD4, PDB: 4Z1Q²⁵⁷) (**Fig. 60D**), **isoxazole 61** (also known as SGC-CBP30, active against CREBBP BRDs, PDB: 5BT3 (*to be*

published)) (**Fig. 60E**), **compound 36** and **compound 38** (isoxazoloquinoline and naphthyridine derivatives, showing anti-inflammatory effects in mouse models²⁵⁸).

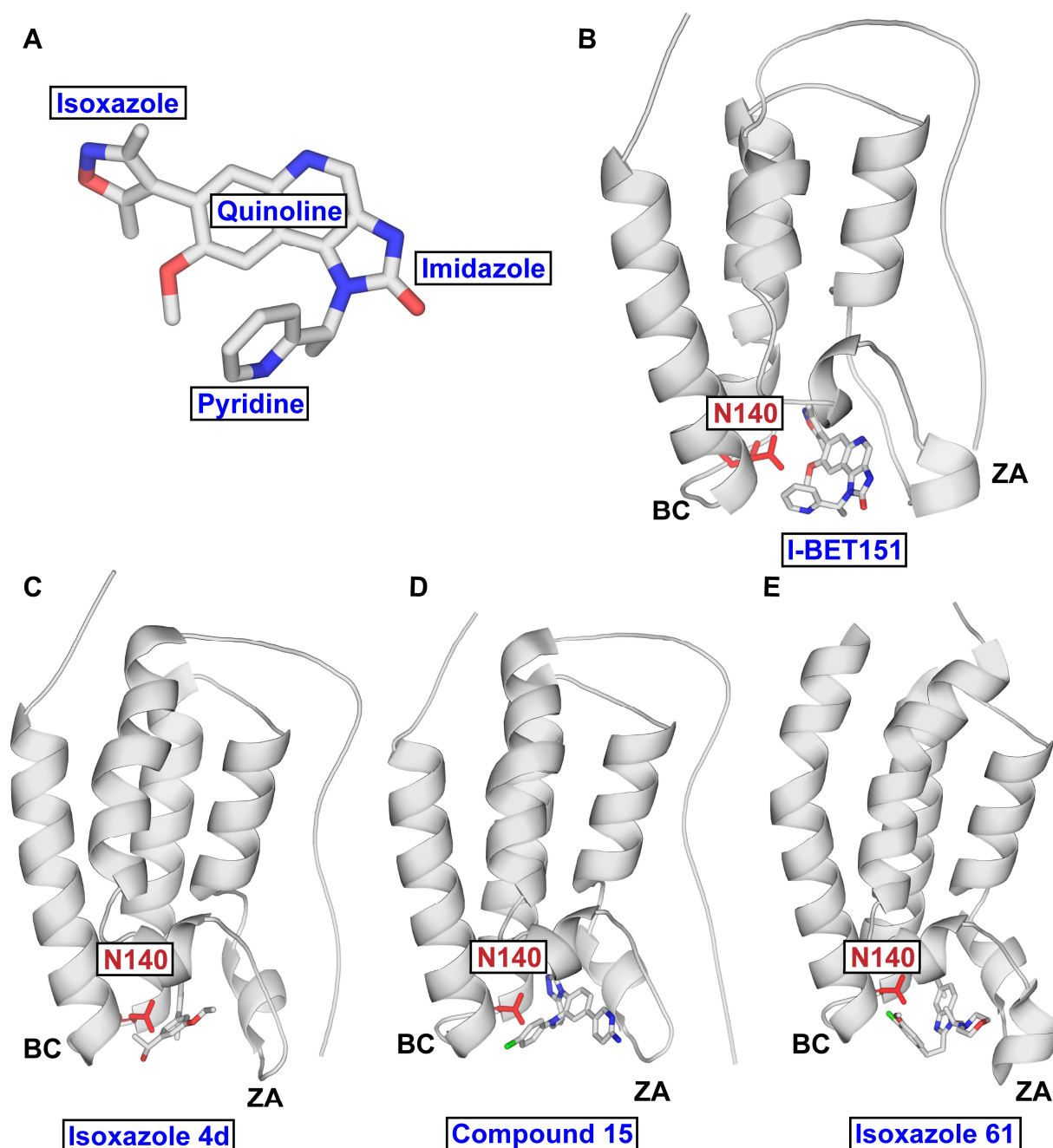


Figure 60| Isoxazole derivatives. **A.** Structure of I-BET151. The isoxazole scaffold has been optimized with the addition of a quinoline ring, pyridine ring and imidazole ring **B.** I-BET151 in complex with BRD4 (PDB: 3ZYU²⁴⁹) **C.** Isoxazole 4d in complex with BRD4 (PDB: 3SVG, *to be published*) **D.** Compound 15 in complex with BRD4 (PDB: 4Z1Q²⁵⁷) **E.** Isoxazole 61 in complex with CREEBP (PDB: 5BT3, *to be published*). The conserved asparagine (N140) is highlighted in red.

4.2.4 Quinoline, Quinazoline and Quinazolone templates

Compounds based on the *quinoline* scaffold and fused with an isoxazole ring (see I-BET151, **Fig. 61A**) have shown *in vitro* activity against BET BRDs (PDB: 4MEO)²⁵⁹. These compounds have been proved to reduce proliferation in cervical cancer cell lines infected with human papilloma virus (HPV)²⁶⁰ (**Fig. 61A,D**).

PFI-1 is a *dihydroquinazoline* compound with anti-proliferative effects in leukemia cell lines (PDB: 4E96)²⁶¹. PFI-1 down-regulates the expression of Aurora B kinase, induces cell cycle arrest in G1 phase and promotes cell apoptosis²⁶². PFI-1 is tested in preclinical trial against several forms of leukemia²⁶³ (**Fig. 61B,E**).

RVX-208 is a quinazolone molecule active against BRD2 (Kd: 4 μ M) and BRD3 (Kd: 0.19 μ M) (PDB: 4MR6²⁶⁴). The molecule inhibits BET BRDs inducing the up-regulation of APOA1 protein level²⁶⁴ (**Fig. 61C,F**).

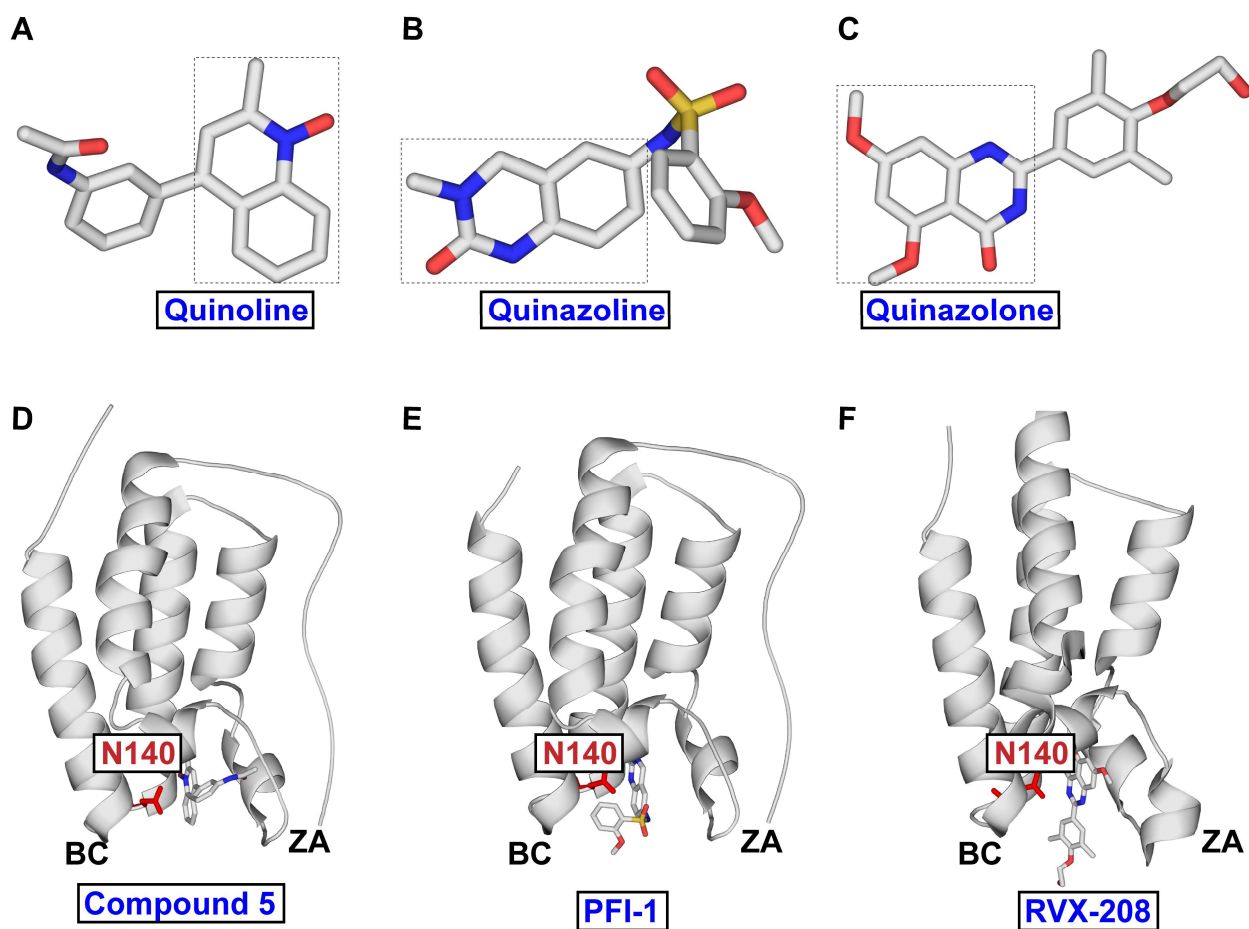


Figure 61| Quinoline, Quinazoline, Quinazolone derivatives. A. and D. Quinoline scaffold of compound 5 and complex with BRD4 (PDB: 4MEO)²⁵⁹ B. and E. Quinazoline scaffold of PFI-1 and complex with BRD4 (PDB: 4E96)²⁶¹ C. and F. Quinazolone scaffold of RVX-208 and complex with BRD2 (PDB: 4MR6²⁶⁴) The conserved asparagine (N140) is highlighted in red.

4.2.5 Other templates

Triazolophthalazine compounds show activity against BRD4 and non-BET BRDs (CREBBP and BRD9). Specifically, this new series of compound inhibit the association between CREBBP and chromatin in human osteosarcoma cell lines.²⁶⁵ (PDB: 4NQM, *to be published*) (**Fig. 62**).

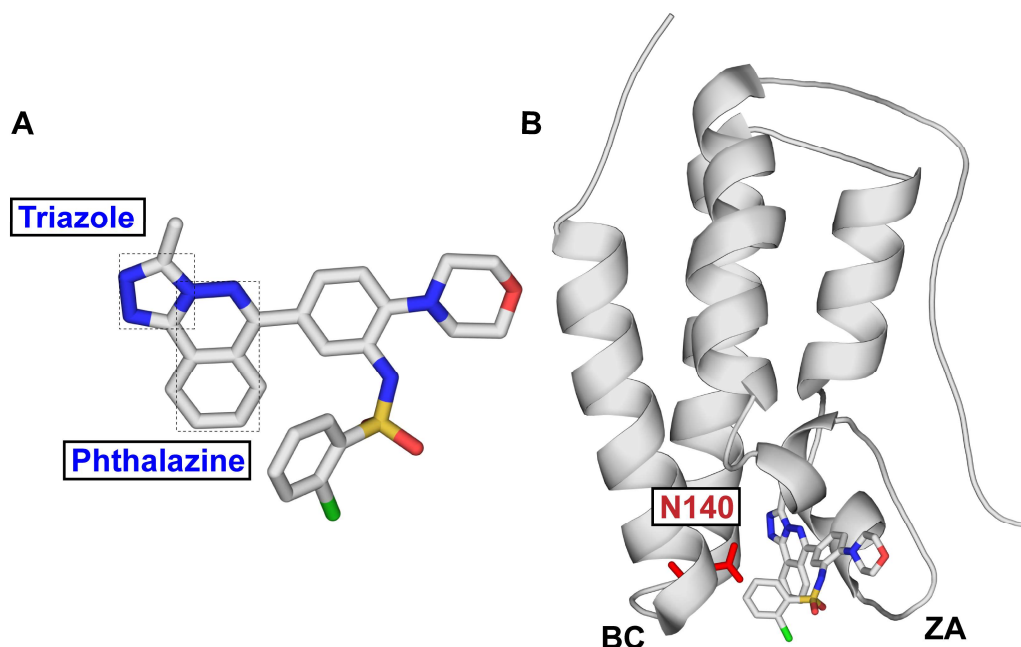


Figure 62| Triazolophthalazine templates. A. Triazolophthalazine scaffold B. Crystal structure of BRD4 in complex with Triazolophthalazine ligand (PDB: 4NQM, *to be published*). The conserved asparagine (N140) is highlighted in red.

Non-acetylated lysine mimetic compounds targeting BET and non-BET BRDs:

4.2.6 Azobenzene templates

Ischemin is active against the BRD of CREBBP binding the hydrophobic cavity in a non-acetylated lysine competitive way²⁴⁰ (PDB: 2L84²⁴⁰). The compound occupies the surface of the acetylated lysine pocket packing between the ZA loop and BC loop without entering deep inside the hydrophobic cavity and without forming a hydrogen bond with the conserved asparagine (**Fig. 57C,D**). Ischemin protects cardiomyocytes against apoptosis after myocardial damage²⁶⁶ (**Fig. 63B**).

MS436 blocks the pro-inflammatory production in IL-6, blocking the transcriptional activity of BRD4²⁶⁷. (PDB: 4NUD²⁶⁷) (**Fig. 63A,C**).

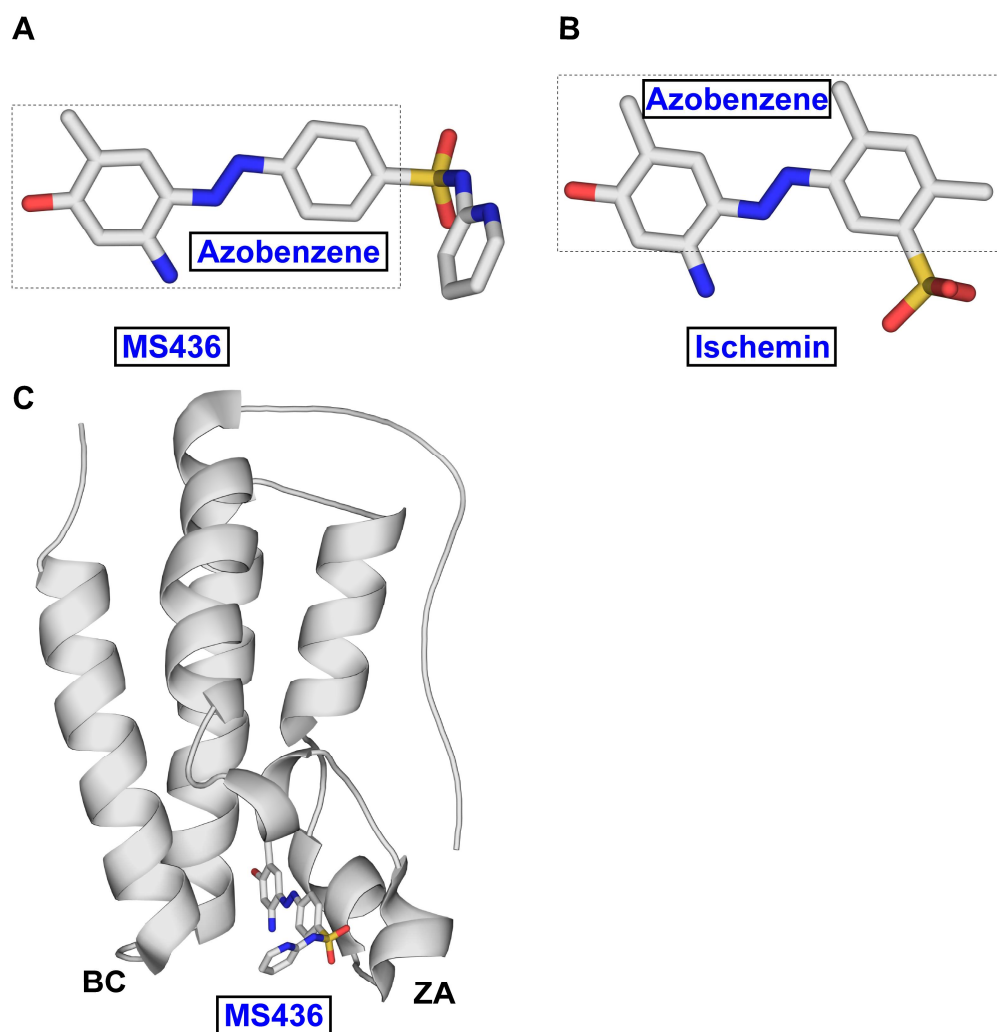


Figure 63| Azobenzene templates. **A.** and **C.** Azobenzene scaffold of MS436 and crystal structure of BRD4 in complex with MS436 (PDB: 4NUD²⁶⁷) **B.** Azobenzene scaffold of Ischemin (PDB: 2L84²⁴⁰).

4.2.7 Benzoimidazole templates

BIC1 prevents the binding of BRD2 to the di-acetylated (K_{ac5} and K_{ac8}) H4 peptide in a concentration dependent manner²⁶⁸. The compound binds BRD2 without forming hydrogen bonds but through a series of hydrophobic interactions (PDB: 3AQA²⁶⁹) (**Fig. 64**).

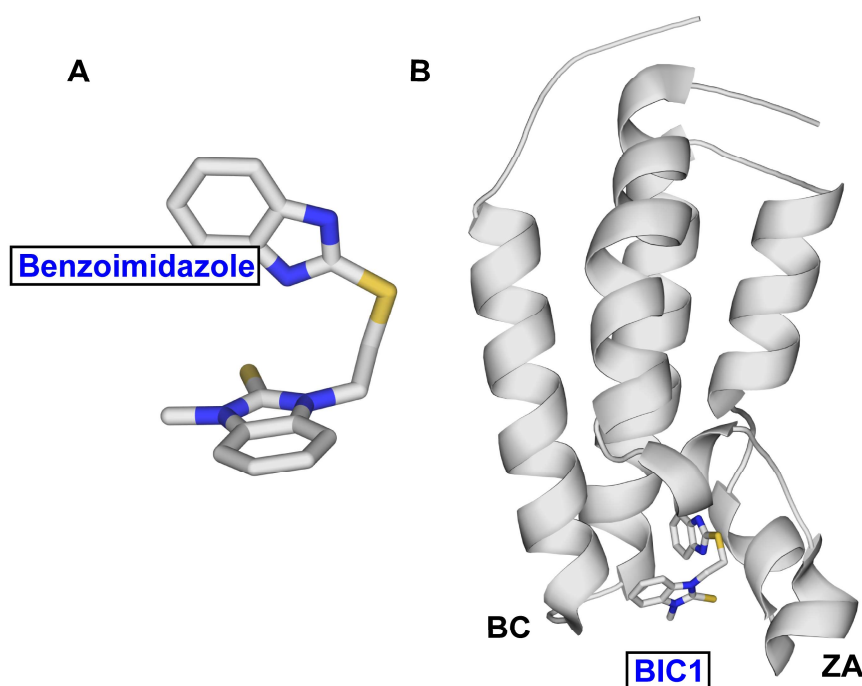


Figure 64| Benzoimidazole templates. A. and B. Benzoimidazole scaffold of MS436 and crystal structure of BRD2 in complex with BIC1 (PDB: 3AQA²⁶⁹).

4.2.8 Other templates

NP1 disrupts the interaction between the BRD of PCAF and HIV-1 Tat peptide (PDB: 1WUG²⁷⁰)

MS7972 prevents the interaction between CREBBP and the AcK-containing binding partner p53²⁷¹.

The efficiency of BET BRDs inhibitors in tumor models have justified the rapid transition of several molecules into preclinical and clinical trials and the development of a drug discovery program to find new inhibitors targeting non-BET BRD modules. To evaluate the benefit of targeting non-BET BRDs, a promiscuous BRDs inhibitor known as BSP (bromosporine inhibitor) has been developed in 2016 by Picaud and collaborators. BSP shows nanomolar affinity for 13 BRDs and low micromolar affinity for other 12 BRDs, similarly to the pan-BET inhibitor JQ1. The effect of BSP was evaluated in leukemic cell lines known to be sensitive to BET BRDs inhibition. BSP was proved to down-regulate growth promoting and anti-apoptotic genes such as the oncogene *c-Myc* and the treated cells revealed a pronounced BET signature in a genome-wide transcriptional analysis. The authors concluded that BET BRDs are the main regulators of primary transcription response in leukemic cells while non-BET BRD do not present a relevant role in gene expression. Non BET-BRDs seem to be required mainly for chromatin structural organization during development rather than for gene expression. Interestingly both JQ1 and BSP did not show any effect on the Br domain of KAP1 at 0.2 and 1 μ M. This is an indication that compounds specific for KAP1 Br domain will probably not be promiscuous with respect to the other BRDs²⁷².

4.3 KAP1 PHD-Bromodomain module

KAP1 is a member of the TRIM C-VI subfamily, together with TRIM24 and TRIM33 and it is characterized by the presence of a tandem plant homeodomain (PHD) and Bromodomain (Br) typically involved in the recognition of various histones modifications^{65,105}. However, the C-terminal tandem PHD-Br domain of KAP1 shows a unique function acting as an E3 SUMO ligase, promoting both the auto-SUMOylation of the protein⁶⁸ and the SUMOylation of other substrates^{67,237}. The NMR structure of the KAP1 PHD-Br domain elucidated how the two domains cooperate as one E3 SUMO ligase unit (PDB: 2RO1⁶²). The auto-SUMOylation of the C-terminal PHD-Br domain is necessary for the binding of KAP1 to the chromatin remodeling enzymes inducing the deposition of PTMs and consequently the creation of an heterochromatin environment.^{34,197,202,238,35}

Currently, no selective ligands for KAP1 Bromodomain able to regulate its transcriptional activity are known. Given the emerging fundamental importance of KAP1 in the regulation of gene expression and given the structural similarity with the other BRDs, we decided to performed structural-based drug discovery on KAP1 Bromodomain with the ultimate goal of discovering novel selective compounds.

4.3.1 KAP1 PHD-Br domain: function and structure

Tandem PHD-Br domain modules are found in methyltransferases (Mixed lineage leukemia protein 1, MLL1²⁷³), in transcription factors (William syndrome transcription factor, WSTF²⁷⁴), in chromatin remodeling complexes (WINAC), in nucleosome interacting proteins (p300/CBP⁶⁵), in proteins involved in viral infection and in the pathogenesis of acute promyelocytic leukemia (Nuclear body protein Sp140²⁷⁵) and in the TRIM family (TRIM28 or KAP1, TRIM 24 and TRIM 33). Mutations of the PHD finger domain are involved in tumorigenesis and genetic disorders²⁷⁶.

PHD fingers function as methylated lysine binding domains. They recognize the modified residues of histone H3 within a cage of four aromatic side chains located on the surface on one of the β -sheets (PDB: 2FSA²⁷⁷) (**Fig. 65A**). However this aromatic cage is absent in the corresponding position in KAP1 and instead the four aromatic amino acid cluster is rotated on the opposite side of the β -sheet and packed against the α Z helix of the adjacent Bromodomain (PDB: 2RO1⁶²) (**Fig. 65B**). This conformational change is related to the development of the novel protein function.

The KAP1 PHD finger domain functions as a SUMO E3 ligase mediating the auto-SUMOylation of KAP1 and regulating its transcriptional repression activity⁶⁸. The PHD finger domain interacts with the SUMO E2 conjugating enzyme (UBC9)⁶⁸ and controls the SUMOylation of four of the six SUMOylation sites located on

KAP1, specifically the Lys779 on the BC loop, the Lys 804 in the loop following the α C helix of the adjacent Bromodomain; and two minor sites, the Lys676 in the PB linker and the Lys750 in the Bromodomain AB loop (**Fig. 65C**). Upon SUMOylation the protein structure undergoes conformational changes that allow the interaction with chromatin remodeling enzymes such as SETDB1, NuRD complex and HDAC⁶⁸. In a SUMO-dependent manner KAP1 functions as a scaffold for the coordination of chromatin modifying and remodeling enzymes inducing the deposition of PTMs on the chromatin context and modulating transcriptional silencing.

BRDs recognize acetylated lysine H3/H4 peptides within a hydrophobic pocket located between the ZA loop and the BC loop. The acetylated lysine forms a hydrogen bond with a conserved asparagine located in the BC loop (**Fig. 56**). Two tyrosine residues located respectively in the ZA loop and in the BC loop are important for the recognition of the acetylated peptides (PDB: 5FE0)²⁷⁸. Although they all share the same structural organization (four-helical bundle, see Chapter 4.1), the ZA loop and BC loop are different in length and amino acids composition, except the three conserved residues. These differences define the ligand binding specificity of each Bromodomain.

The NMR structure of KAP1 Bromodomain (PDB: 2RO1⁶²) revealed how the three conserved residues are mutated respectively in glutamine, leucine and threonine leading to a complete different functionality. Moreover the positions of the ZA loop as well as the respective orientation of the BC loop are different in KAP1 compared with the BRD of PCAF (**Fig. 66**). These structural changes are consistent with the different function of the KAP1 tandem PHD-Br domain module that does not recognize anymore the acetylated lysine H3/H4 peptides.

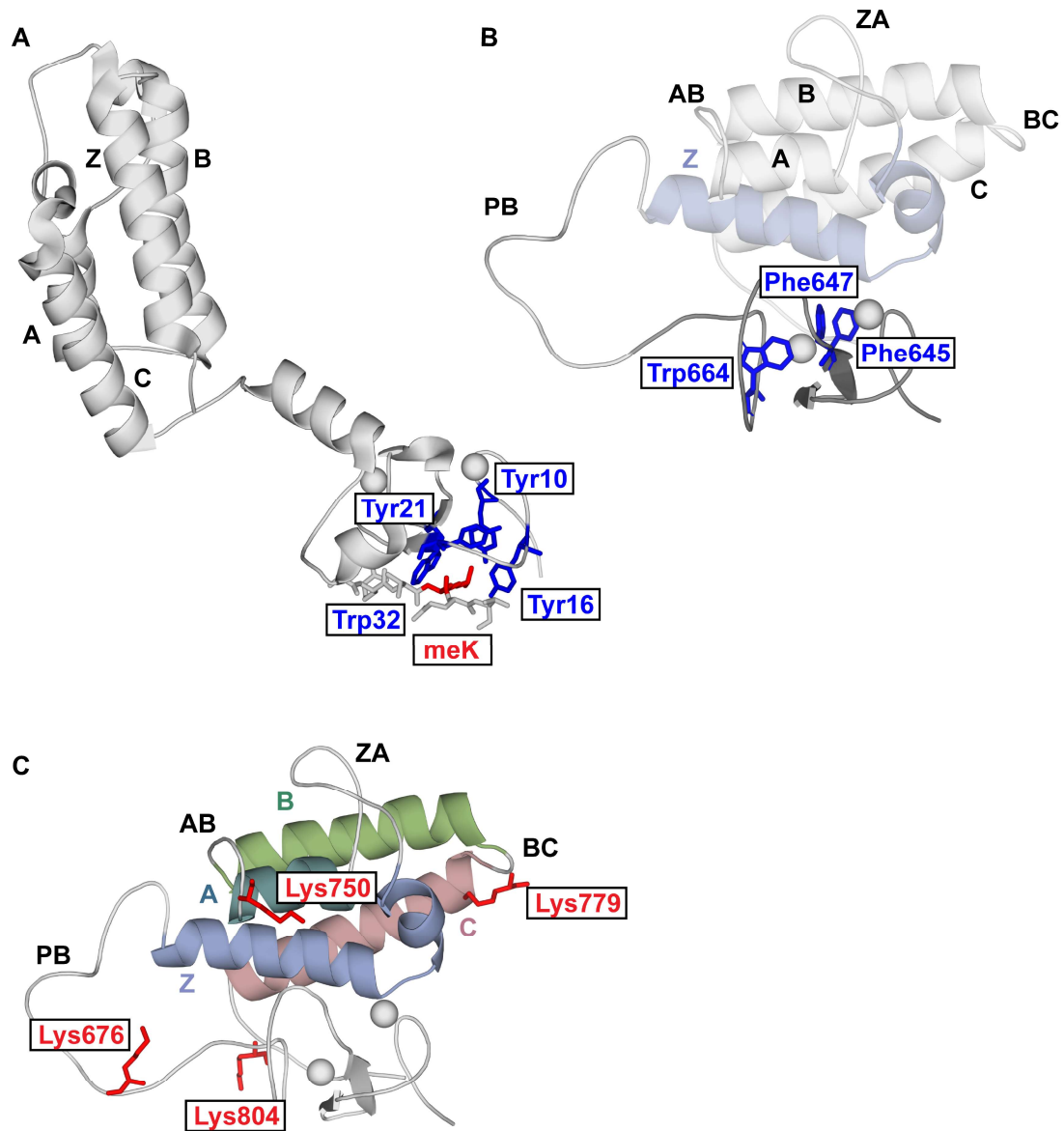


Figure 65| KAP1 and BPTG PHD finger domains comparison. **A.** PHD fingers domain of BPTF (Bromodomain PHD finger transcription factor). The methylated lysine of histone H3 (meK, in red) is recognize in a cage of four aromatic side chains (in blue) located on the surface on one of the β -sheets (PDB: 2FSA²⁷⁷) **B.** PHD-Br domain structure of KAP1 (PDB: 2RO1⁶²). The aromatic cage for the recognition of the methylated lysine is not present anymore and the aromatic residues (in blue) are packed against the α Z helix (in purple) **C.** PHD-Br domain structure of KAP1 (PDB: 2RO1⁶²). The SUMOylation sites (lysine 804, 779, 750 and 676) are shown in red.

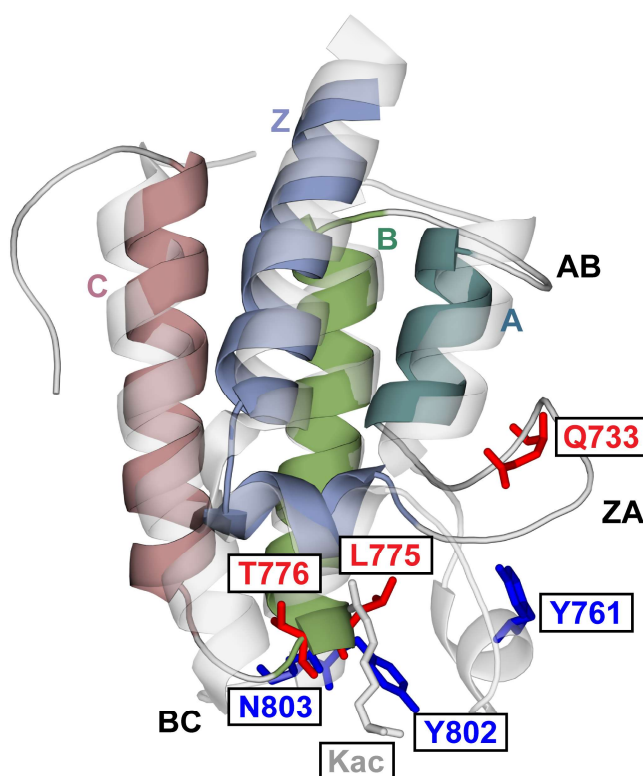


Figure 66| KAP1 and PCAF Bromodomains comparison. Superimposition of KAP1 Bromodomain (PDB: 2RO1⁶²) and PCAF BRD (PDB: 5FE0)²⁷⁸. The amino acids involved in the recognition of the acetylated lysine (K_{ac} in grey) are shown in blue on the PCAF structure. The amino acids substitutions are shown in red on the KAP1 structure.

4.4 Virtual screening: protein and ligands preparation

KAP1 Bromodomain conserves the same structural organization as the other family members (four left-handed α -helices). The main difference is given by the position of the ZA loop found in the 19 NMR structures (PDB: 2RO1⁶²) in an “open” conformation compared to the other BRDs in which the ZA loop forms together with the BC loop a close hydrophobic pocket where the acetylated lysine is recognized (**Fig. 67A**). Moreover KAP1 ZA loop is shorter compared to the other BRDs.

Protein preparation and preliminary analysis. The presence of this unique conformation can have different explanations: (i) the “open” conformation of the ZA loop together with the amino acids substitutions (**Fig. 66** and **Fig. 67A**) precludes the formation of the hydrophobic cavity with the BC loop preventing the recognition of the acetylated peptides (ii) given the flexibility of this protein region, the “open” conformation could be a transition state. Upon binding of an inhibitor the ZA loop could present the same “close” conformation as the other BRD structures leading to the formation of a binding pocket together with the BC loop. This second hypothesis is supported by the high degree of flexibility of KAP1 ZA loop shown in the 1H - ^{15}N heteronuclear NOE_s plot⁶².

Given the flexibility of the ZA loop, we decide to proceed with the screening of two protein conformations: the original NMR structure (PDB: 2RO1⁶²) in which the ZA loop is found in an “open” conformation and a new “close” conformation produce by homology modeling using Swiss-Model¹³⁶ based on the crystal structure of TRIM24 (PDB: 5H1T¹⁶³) (**Fig. 67B**).

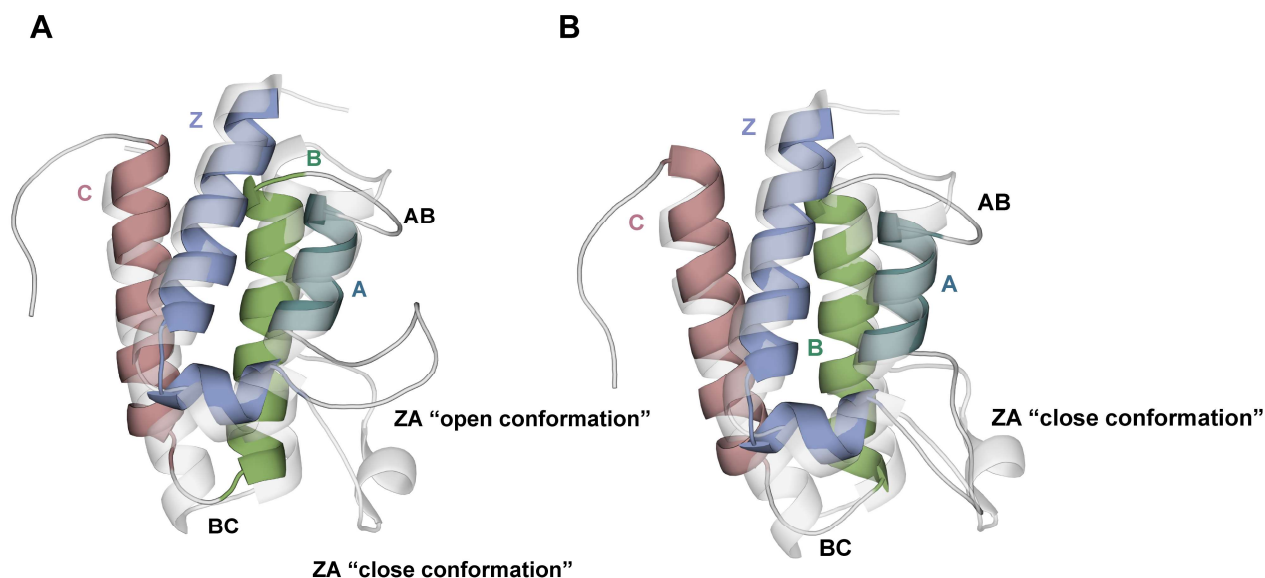


Figure 67| ZA loop of KAP1 compared with other Bromodomains. **A.** Superimposition of KAP1 Bromodomain (PDB: 2RO1⁶²) and TRIM24 BRD (PDB: 5H1T¹⁶³). The ZA loop of KAP1 is shown in the original “open” conformation while the ZA loop of TRIM24 is shown in the “close” conformation conserved in all the other BRDs. **B.** New ZA loop conformation of KAP1. Superimposition of KAP1 Bromodomain (PDB: 2RO1⁶²) and TRIM24 BRD (PDB: 5H1T¹⁶³). The ZA loop of KAP1 was modeled on the structure of TRIM24 (PDB: 5H1T¹⁶³) into the “close conformation” common to all the BRDs.

The hydrophobic nature of the cavity between the ZA loop and the BC loop is fundamental to accommodate the charged-neutralized acetylated lysine. To better clarify the differences between KAP1 Bromodomain and the other family members, we decided to perform a hydrophobicity analysis focusing on the portions possibly involved in the binding of inhibitors.

Interestingly, comparing the KAP1 structure with the ZA loop in a “close” configuration modeled on TRIM24 (PDB: 5H1T²³) with the Bromodomain BRD4 in complex with Bdi1 (PDB: 5XHY, *to be published*) we noticed the presence in KAP1 of the conserved hydrophobic cavity between the ZA loop and the BC loop. This could imply the propensity of this region to interact with acetylated lysine mimetic compounds or other ligands (**Fig. 68**).

The main difference is given by the nature of the αZ helix. In a typical Bromodomain this helix has an amphipathic nature packed on one side to the hydrophobic core of the helical bundle and on the other side exposed to the solvent. The KAP1 αZ helix consists of amino acids with large aromatic side chains and its hydrophobic C-terminal part is buried in the center of the tandem PHD-Br domain, packed on one side to the other three helices and on the other side to the PHD finger domain β sheet (residues Val630, Phe647, Leu653 and Leu668) (**Fig. 69**). The hydrophobic nature of the αZ helix keeps stable the tandem PHD-Br domain module⁶².

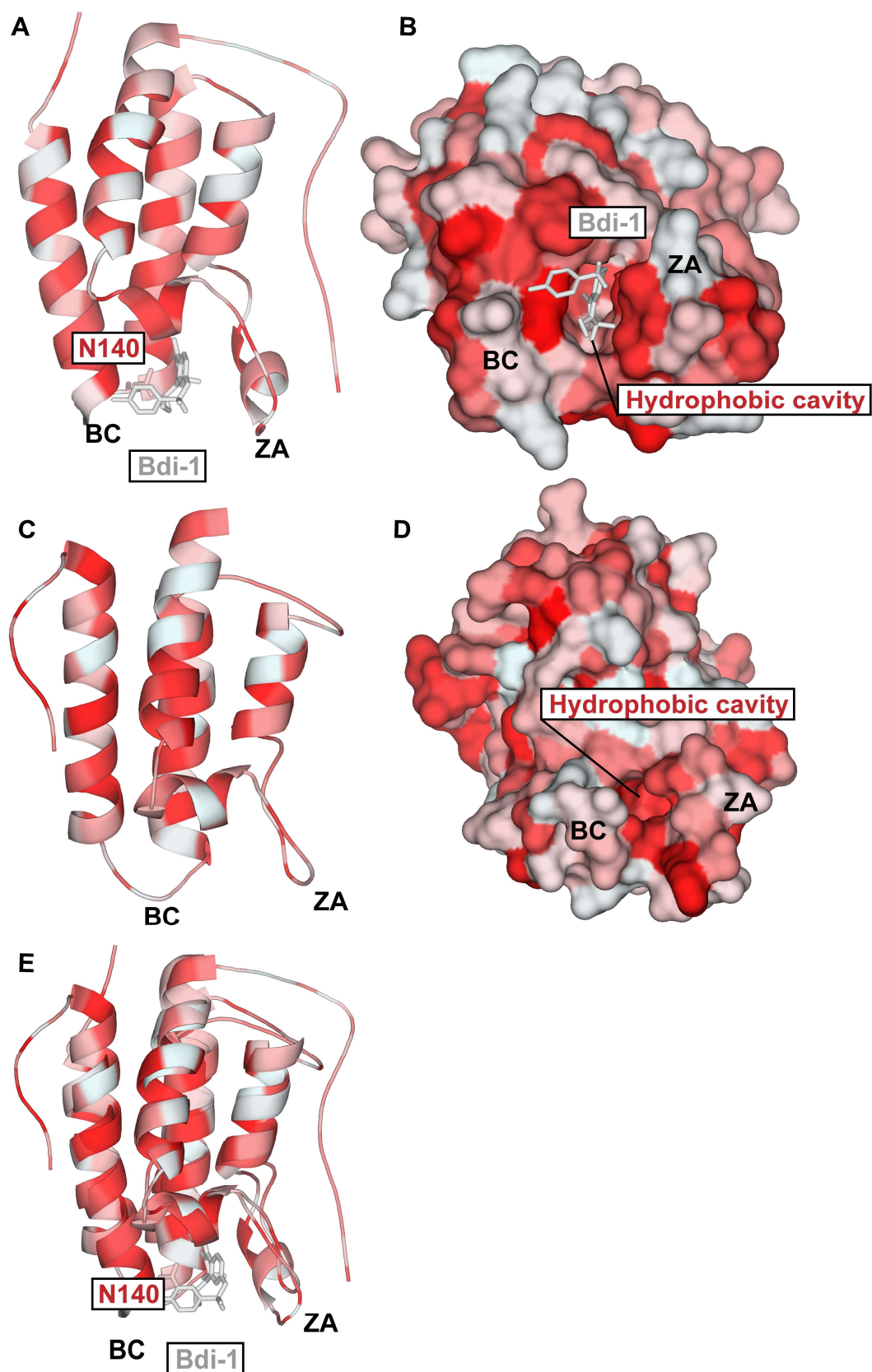


Figure 68| Hydrophobicity analysis. A. and B. Structure of BRD4 in complex with Bdi1 (PDB: 5XHY, *to be published*) shown as cartoon or surface representation C. and D. Structure of KAP1 with the ZA loop in a "close" configuration modeled on TRIM24 (PDB: 5H1T²³) E. Superimposition of KAP1 Bromodomain (PDB: 2RO1⁶²) and BRD4 Bromodomain (PDB: 5XHY, *to be published*). The hydrophobic cavities are highlighted in B. and D. We used a

PyMOL (PyMOL Molecular Graphic System, version 2.3, Schrödinger, LLC) script to color protein molecules based on the Eisenberg hydrophobicity scale in which the amino acids are assigned with the following values: Ala:0.620, Arg:-2.530, Asn:-0.780, Asp:-0.900, Cys: 0.290, Gln:-0.850, Glu:-0.740, Gly:0.480, His:-0.400, Ile:1.380, Leu:1.060, Lys:-1.500, Met:0.640, Phe: 1.190, Pro:0.120, Ser:-0.180, Thr:-0.050, Trp:0.810, Tyr:0.260, Val:1.080. Hydrophobic residues appear in bright red on the map while the polar ones appear in white.

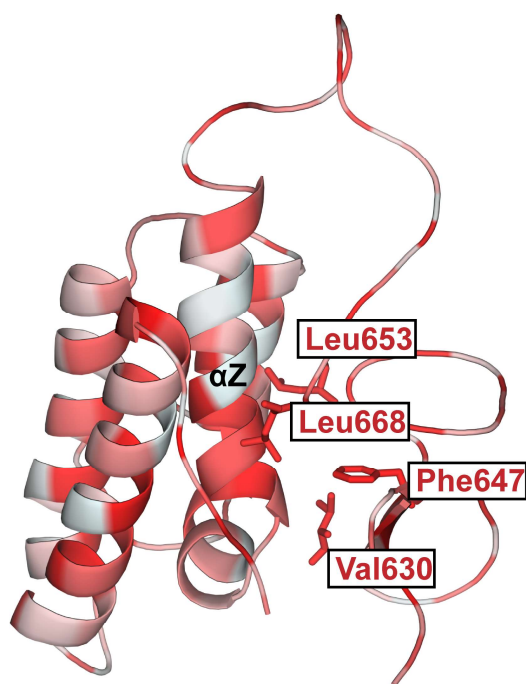


Figure 69| KAP1 α Z helix. The α Z helix has a hydrophobic nature in KAP1. It is packed on one side to the α B, α A and α C helices and on the other side to the PHD finger module. The residues of the PHD finger domain involved in the interaction with the α Z helix are highlight in red.

Finally the two protein configurations (open and close ZA loop positions) were prepared for docking using AutoDockTools¹⁶⁵ (ADT4.2), a GUI interface for setting and running docking simulations. Specifically, AutoDockTools¹⁶⁵ was used to add the polar hydrogen to the protein structure and save it in file.pdbqt.

Grid preparation. The putative binding site was identified comparing the structure of KAP1 Bromodomain with the other X-ray structures of BRDs in complex with inhibitors (PDBs: 5XHY, *to be published*, 2L84²⁴⁰, 1WUG²⁷⁰, 4NUD²⁶⁷, 4NQM *to be published*, 4E96²⁶¹, 4MR6²⁶⁴, 4MEO²⁵⁹, 5BT3 *to be published*, 3SVG *to be published*, 4F3I²⁴⁴, 4LRG²⁴⁸, 6HDN¹⁶¹, 4WHU *to be published*, 6HDQ¹⁶¹, 6MO9¹⁶², 5H1T¹⁶³, 5VOM¹⁶⁴, 6HDO¹⁶¹, 3MXF²³⁹). We concluded that a possible binding region was represented by the hydrophobic cavity between the ZA and the BC loop (**Fig. 70**).

AutoDockTools¹⁶⁵ was used to generate the grid box. The grid was centered in the middle of the four-helical bundle (-8.073, -11.952, -8.409) and the box dimensions (20x22x22 Å) were defined to maximize the

searching space between the ZA loop and BC loop where the binding pocket for BRDs inhibitors is located. These specific grid dimensions were chosen after several docking trials in which the grid was extended to include the loop regions and the PHD finger domain. In this extended grid condition the ligands not only interact with the hydrophobic pocket but also with the external portion of the αC and the αA helices in a non-specific manner (**Fig. 71**). Since we wanted to identify specific inhibitors for the hydrophobic pocket between the ZA loop and the BC loop we decided to keep the grid box restricted to this protein region.

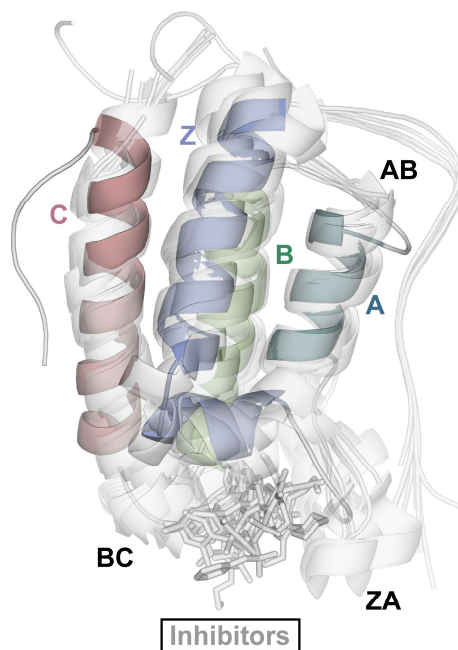


Figure 70| Bromodomains comparison. Superimposition of KAP1 Bromodomain and other BRD structures in complex with inhibitors. PDBs: 2RO1⁶² (KAP1), 5XHY, *to be published*, 2L84²⁴⁰, 1WUG²⁷⁰, 4NUD²⁶⁷, 4NQM *to be published*, 4E96²⁶¹, 4MR6²⁶⁴, 4MEO²⁵⁹, 5BT3 *to be published*, 3SVG *to be published*, 4F3I²⁴⁴, 4LRG²⁴⁸, 6HDN¹⁶¹, 4WHU *to be published*, 6HDQ¹⁶¹, 6MO9¹⁶², 5H1T¹⁶³, 5VOM¹⁶⁴, 6HDO¹⁶¹, 3MXF²³⁹). The putative binding pocket was identified in the region between the ZA loop and the BC loop.

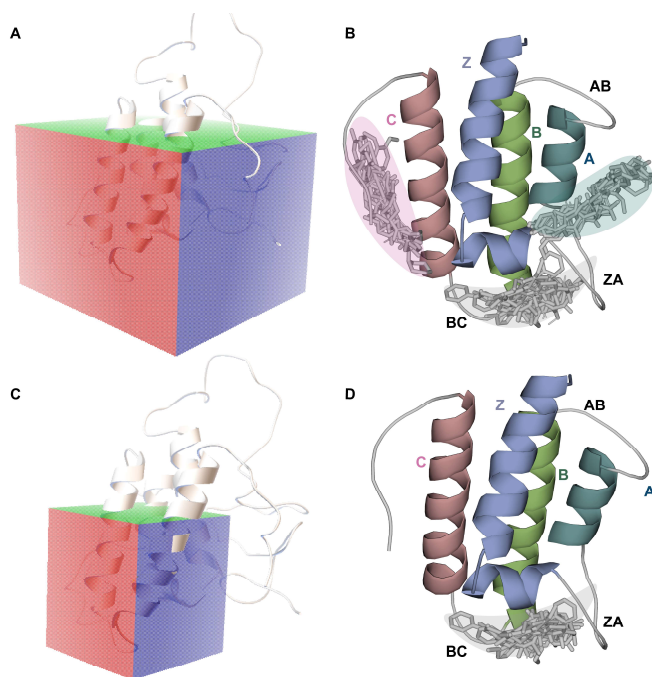


Figure 71| Grid box preparation. **A. and B.** The grid box is extended to the loop before the α C helix and to the PHD finger domain. The compounds bind the pocket between the ZA loop and the BC loop, the external portion of the α C and the α A helices **C. and D.** The grid box is centered in the middle of the four-helical bundle limited to the Bromodomain, in particular to the ZA loop and to the BC loop region. In this case the compounds interact only with the hydrophobic pocket between the ZA loop and the BC loop.

Ligand preparation. The 13'433 compounds screened against the structure of KAP1 PHD-Bromodomain module are part of the Chemically Diverse Collection (CDC) provided by the Biomolecular Screening Facility (BSF). They cover the 85% of the chemical diversity of the entire library. We used Open Babel¹⁶⁶ (openbabel-2.4.1) to convert the initial 2D structures of the compounds into 3D conformations. Afterwards MGLTools package was used to add the polar hydrogens. Finally the docking was performed with AutoDock Vina¹⁵¹ generating 9 poses for each ligand. The two Bromodomain conformations (open ZA loop and close ZA loop) were treated separately, performing two different screenings against the same library (CDC). The details on the ligands preparation and the docking process can be found in Method Chapter 2.12.

Docking results. The compounds were classified based on their binding energy. The best pose for each compound with the lowest binding energy was retrieved. The first 136 compounds with binding energies between -10 and -8 kcal/mol, corresponding to ~1% of the total number of ligands used in the screening, were selected and visually analyzed with PyMOL (PyMOL 2.3). The preferred binding pocket and the orientation of the ligands in the binding site were compared to the other crystal structures of BRDs in complex with inhibitors. The first 136 ligands found in the “open” conformation and in the “close” conformation” were compared to retrieve common structures. Among the 136 compounds, 18 ligands were found in both conformations (**Appendix 7**). The analysis of the binding poses revealed an interesting feature not observed

previously in the other BRDs. The compounds bind the hydrophobic pocket between the ZA loop and BC loop occupying the center of the four-helical bundle and they are packed on one side to the last portion of the α Z helix and on the other side to the α B helix. This binding mode seems to be unique for KAP1 (**Fig. 72**).

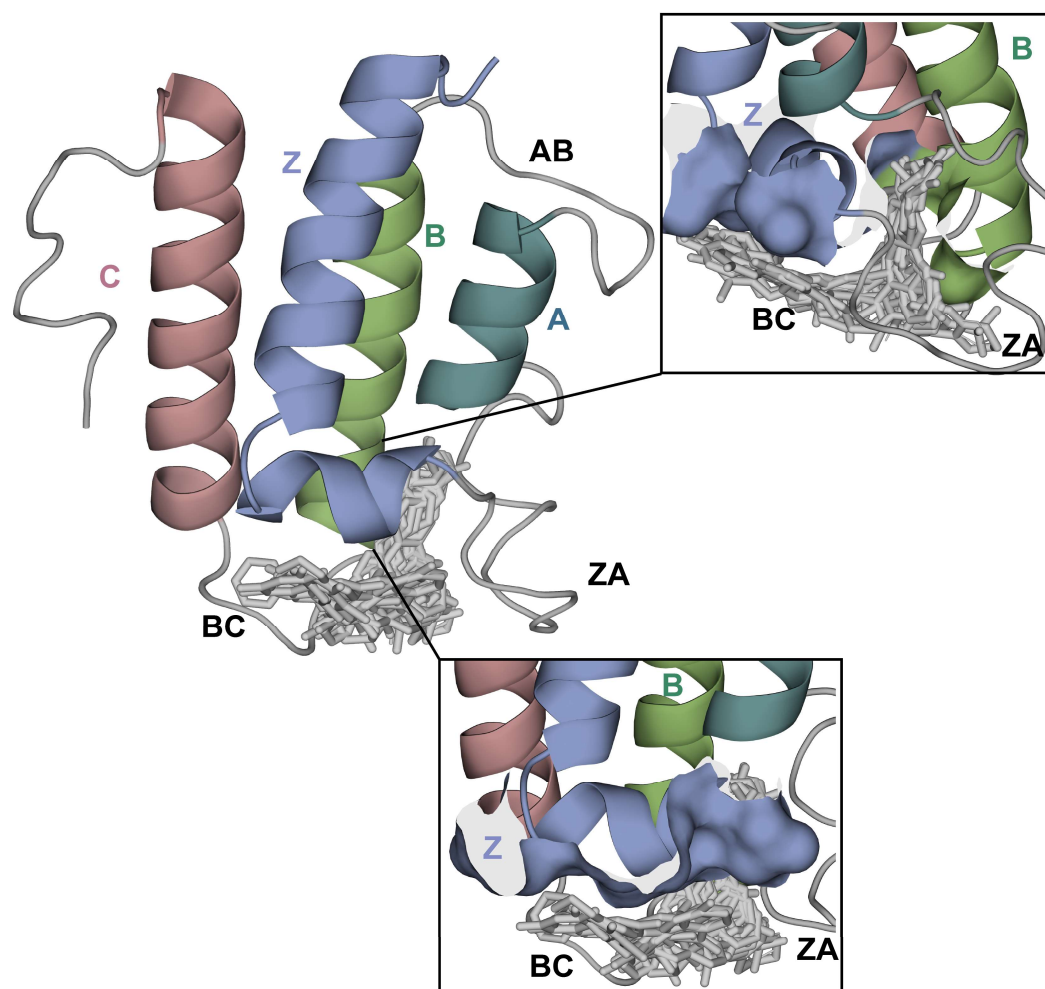


Figure 72| Binding mode of KAP1 inhibitors. KAP1 Bromodomain in complex with 18 inhibitors found both in the “open” and in the “close” ZA loop conformations. The inhibitors bind the pocket between the ZA and the BC loops and they are packed to the last portion of the α Z helix and to the α B helix. The two inserts show the surfaces respectively of the Z helix in purple (bottom insert) and the B helix in green (top insert).

Scaffold Hunter¹⁶⁷ was used to cluster and classify the 136 compounds based on their scaffold for both conformations (**Fig. 73** and **Fig. 74**). The most representative scaffolds were retrieved and compared with the structures of the BRD inhibitors to find chemical similarities.

The compounds binding the receptor in the “close” conformation were classified in 5 main clusters. The ligands belonging to each cluster were analyzed and common characteristics were retrieved. All these ligands

are characterized by the presence of 3-5 aromatic rings and they share chemical groups with the classical BRD inhibitors such as quinoline, benzene, benzamide, pyridine, carboxamide groups (**Fig. 73**).

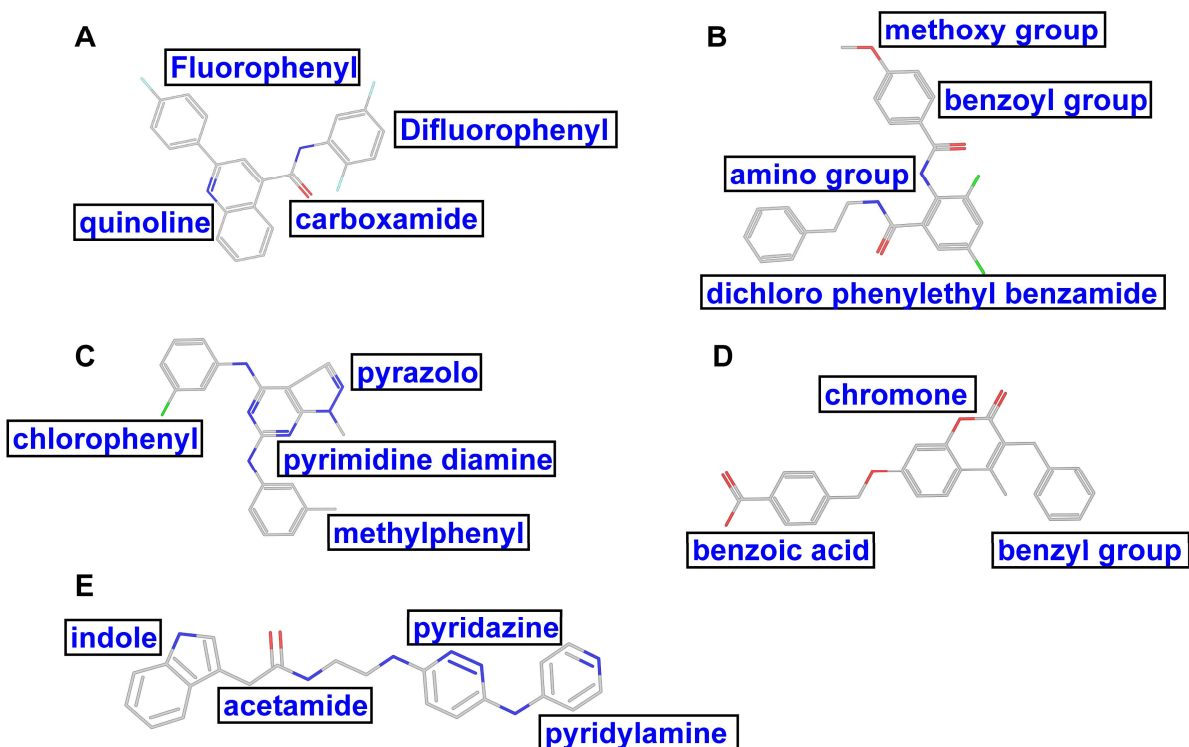
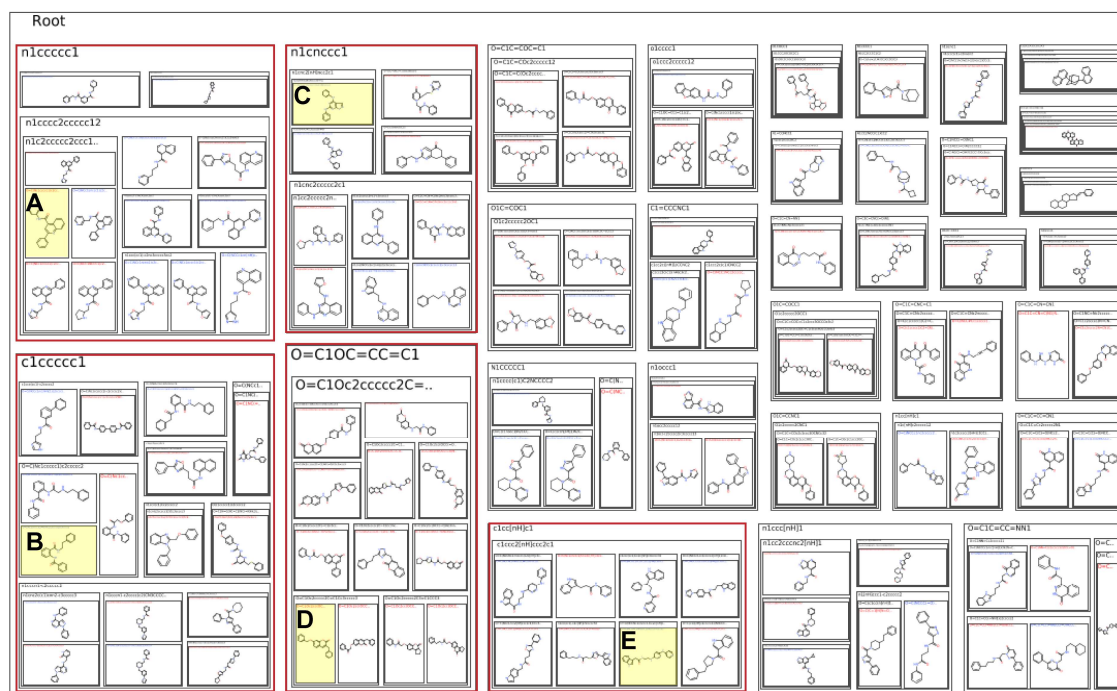


Figure 73| Analysis of KAP1 Bromodomain in the “close” conformation. The compounds are classified into clusters based on their scaffold. The five most populated clusters are highlighted in red and for each of them some representative compounds (in yellow) are shown in 2D representation. These compounds share some common chemical groups with the BRD inhibitors. For each representative ligand the chemical components are shown in blue. The compounds in **A**, and **D**, were found as high affinity ligands in both conformations. **A**. N-(2,5-Difluorophenyl)-2-(4-fluorophenyl)quinoline-4-carboxamide **B**. 3,5-dichloro-2-[(4-methoxybenzoyl)amino]-N-(2-phenylethyl)benzamide **C**.

N~4~(3-Chlorophenyl)-1-methyl-N~6~(3-methylphenyl)-1H-pyrazolo[3,4-d]pyrimidine-4,6-diamine **D**. 4-[[[(3-benzyl-4-methyl-2-oxo-2H-chromen-7-yl)oxy]methyl}benzoic acid **E**. 2-(1H-indol-3-yl)-N-[2-[[6-(4-pyridylamino)pyridazin-3-yl]amino]ethyl]acetamide

The ligands binding the receptor in the original ZA loop conformation were classified into 3 main clusters. The analysis revealed the presence of chemical groups previously found for the receptor in the “close” conformation and common to the BRDs inhibitors such as quinoline, benzene, pyridine, carboxamide groups (**Fig. 74**). These compounds can be considered promising hits for further experimental validation for the following reasons: (i) they are found in two independent screenings against KAP1 Br domain, (ii) they were not previously identified as binders of other BRDs (in fact KAP1 is not regulated by classical BRDs ligands) and (iii) they present chemical similarities with the classical BRDs inhibitors (KAP1 Br domain presents the same structural organization as the other BRDs modules).

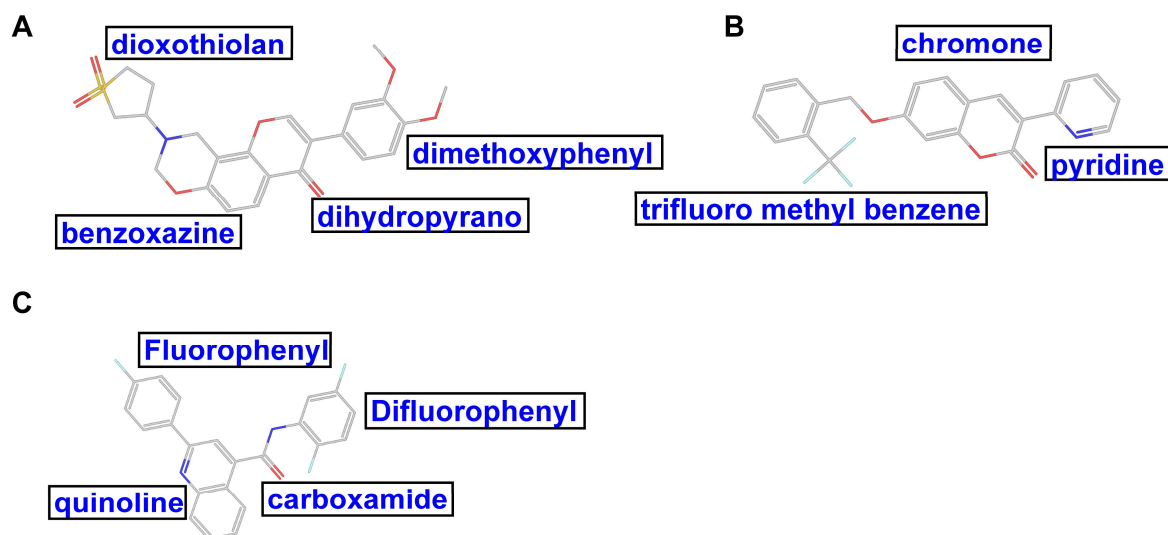
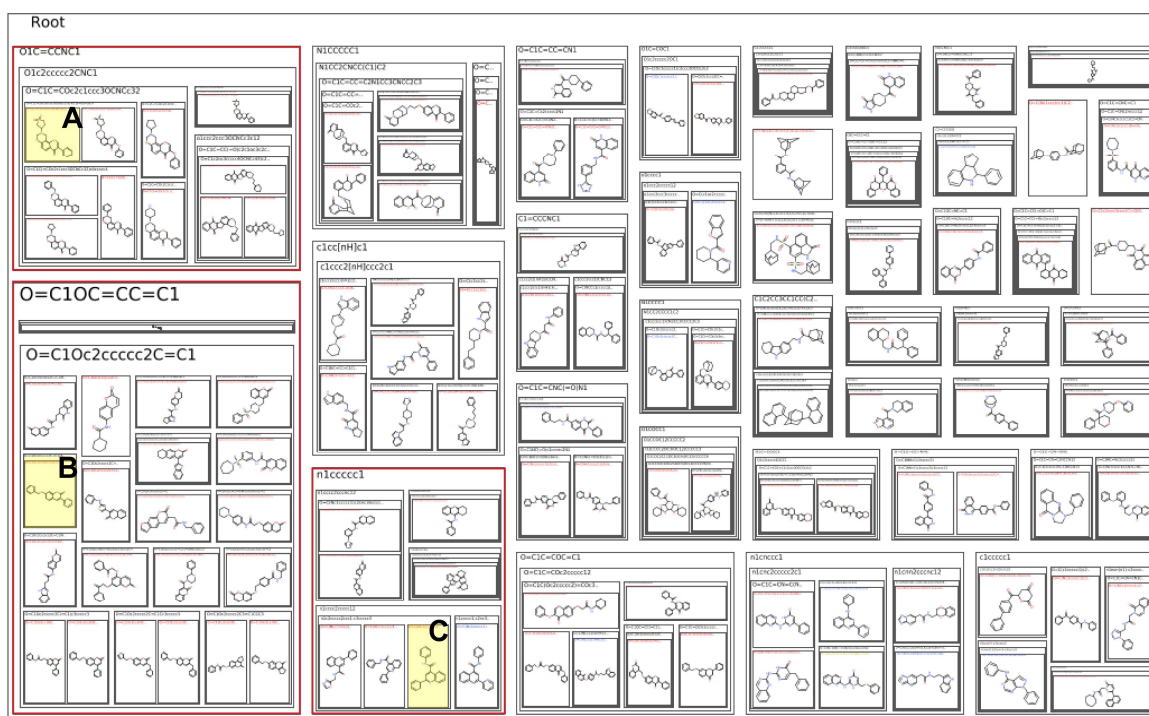


Figure 74| Scaffold Hunter analysis of KAP1 Bromodomain in the “open” conformation. The compounds are classified into clusters based on their scaffold. The three most populated clusters are highlighted in red and for each of them some representative compounds (in yellow) are shown in 2D representation. These compounds share some common chemical groups with the BRD inhibitors. For each representative ligand the chemical components are shown in blue. The compound in C. was found as a high affinity ligand in both conformations (**Fig. 73A**). **A.** 3-(3,4-Dimethoxyphenyl)-9-(1,1-dioxothiolan-3-yl)-8,10-dihydropyrano[2,3-f][1,3]benzoxazin-4-one **B.** 3-(2-Pyridinyl)-7-{[2-(trifluoromethyl)benzyl]oxy}-2H-chromen-2-one **C.** N-(2,5-Difluorophenyl)-2-(4-fluorophenyl)quinoline-4-carboxamide

These studies revealed that:

- The selected 272 compounds (136 for each KAP1 Bromodomain conformation) bind the hydrophobic pocket between the ZA loop and BC loop occupying the center of the four-helical bundle and they are packed on one side to the last portion of the α Z helix and on the other side to the α B helix.
- 18 common compounds were retrieved for both the “open” and the “close” conformation.
- The ligands included in these clusters were characterized by the presence of 3-5 aromatic rings and chemical groups common to the BRDs inhibitors such as quinoline, benzene, benzamide, pyridine, carboxamide.

4.5 Experimental validation

In an attempt to identify ligands able to modulate KAP1 transcriptional activity, the most promising compounds were selected for *in vitro* validation. Since no *in vitro* inhibition assay is available for KAP1, the binding site of the target protein is still unknown and no cofactors or ligands interacting with the putative binding pocket are known, we decided to test the compounds using an indirect method. We investigated the protein-ligand interaction by thermal shift assay (TSA) measuring the melting temperature of the target protein alone and in the presence of the different compounds. The protein-ligand interaction is indicated by a shift in the melting curve. TSA is a scanning method to determine changes in the T_m value of a protein upon binding of a selective inhibitor, however, the technique does not provide any K_d value that needs to be measured with a direct method such as ITC or Biacore. We choose this technique for a variety of technical advantages including the fact that the experiment is fast (30 min), inexpensive, precise and can be performed in a conventional qPCR machine with 384 PCR well plates that allow the screening of multiple compounds simultaneously.

Protein expression and purification. The DNA coding sequence of KAP1 PHD-Br domain was ordered from GenScript. We decided to purify the same construct previously characterized by NMR (2RO1⁶²) starting in position S624 and ending in position T812. The construct was cloned inside the pETDuet-1 vector between the SacI and the HINDIII restriction sites in the MCS1, preceded by a His₆ tag and TEV cleavage site. The PHD-Br domain module was expressed in BL21 (DE3) and purified by affinity chromatography and gel filtration chromatography (**Fig. 75**). All the details on the protein expression and purification are reported in the Method Chapter 2.3. We obtained soluble protein to be tested by Differential Scanning Fluorimetry. In the final assay the temperature was maintained at 25 °C for 2 minutes and then raised to 99 °C with a ramp of 0.05°C/sec in a continuous ramp mode. Fluorescence data were collected for each temperature.

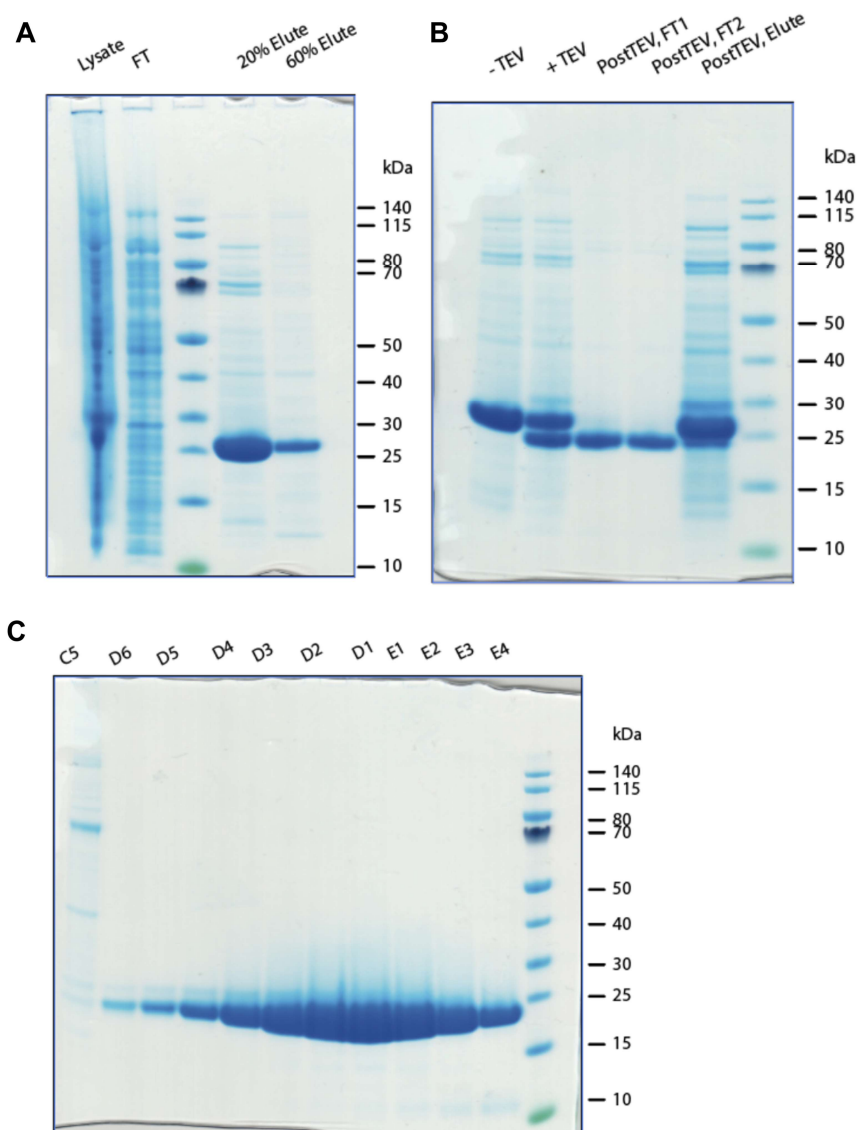


Figure 75| PHD-Br domain purification. **A.** Elution of the protein with 20% and 60% Elution buffer after affinity chromatography (HisTrap 5mL) **B.** Cut of the N-terminal His₆ tag with TEV protease. **C.** Pure fractions after gel filtration chromatography (Superdex 65 16/60 GL)

Thermal shift assay. For the experimental validation we followed the same protocol as Srinivasan and collaborators²⁷⁹. The paper published in 2014 identified new low nanomolar binders for the dihydrofolate reductase and the protein tyrosine phosphatase (PTP). The authors screened 20'000 compounds using a structure based virtual screening approach and they tested the top 1% ligands classified on their binding energies by differential scanning fluorimetry.

The 272 positive compounds together with 8 BRD inhibitors (BIC1¹⁶⁸, NP1¹⁶⁹, Ischemin¹⁷⁰, JQ1²⁸⁰, I-BET762¹⁰⁷, I-BET151¹⁷¹, PF-1¹⁷², RVX-208¹⁷³) were selected and tested by DSF (Differential Scanning

Fluorimetry). The compounds were prepared by the Biomolecular Screening Facility. The ECO-550 dispenser (LABCYTE) for nanoliter pipetting was used to prepare 384-well PCR plates (in triplicates) with 25 nL of compound in each well at 10 mM in 100% DMSO. In each well 21 μ L of KAP1 PHD-Br domain were added at an optimized concentration of 0.2 mg/mL together with 4 μ L of SYPRO™ Orange (ThermoFisher). The final compounds concentration was 10 μ M in 0.1% DMSO.

The melting temperature (T_m) of KAP1 PHD-Br domain without addition of any compound averaged over 24 repetitions was 47.5 °C. Out of 272 tested ligands, 13 compounds produced a relevant shift of the melting curve. The positive compounds are listed in **Table 3** and their structures are reported in **Appendix 8**. The compound F2031-0348 produced the most relevant shift changing the T_m value from 47.5 °C to 59.9 °C.

Compounds	T_m (°C) from Boltzmann fit
EN300-61307	50.9
Z30804950	50.0
F0016-1115	49.7
F1091-0020	49.9
F3139-1743	49.7
F3351-0405	50.2
F3385-6166	50.7
26547668	50.2
5674870	50.0
6572417	49.8
6942873	49.7
7804007	49.7
F2031-0348	59.9
KAP1 PHD-Br	47.5

Table 3| List of the compounds that bind KAP1 PHD-Br domain. In the table are reported the catalog name of each compound together with the corresponding measured melting temperature value upon addition of KAP1. The compound that produced the most relevant change in the T_m value was the ligand F2031-0348 that raised the KAP1 melting temperature of 12°C.

The software used to fit the fluorescence data and to obtained the T_m values for each PHD-Bromodomain ligand complex is the Protein Thermal Shift™ software (ThermoFisher, v1.3). The software was developed for the analysis of the protein melting curves directly from real-time PCR files. For each condition the software generates a unique melting curve shape associated with a specific slope, signal to noise ratio and a

temperature melt range. The software selects automatically a region of the melting curve data to analyze. The algorithm looks for the region of the curve with the steepest slope increment and uses these data as the region of analysis to fit with the Boltzmann equation and to return a T_m value (**Fig. 76**).

$$F(T) = F(\text{pre}) + \frac{[F(\text{post}) - F(\text{pre})]}{1 + e^{(T_m - T)/C}}$$

Where the $F(\text{pre})$ is the fluorescence at the start of the region of analysis or before the melting and $F(\text{post})$ is the fluorescence at the end of the region of analysis or after the melting, and T_m is the melting temperature, and C is the enthalpy of the reaction.

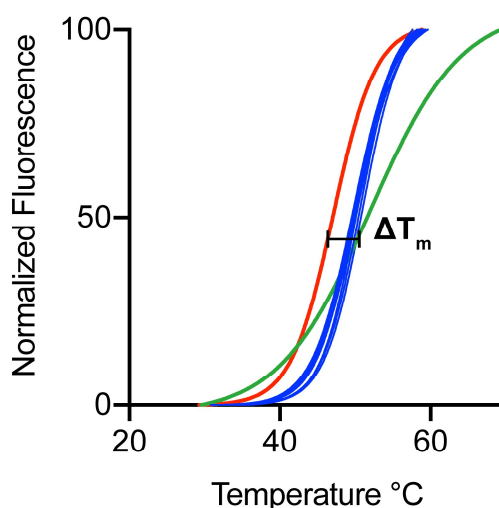


Figure 76| Boltzmann fit of the melt curve data. The graph shows the Boltzmann fit of the melt curve data obtained for KAP1 PHD-Br domain without addition of any compound (T_m 47.5 °C) in red and with the 13 positive compounds listed in **Table 3** in blue. These compounds raise the T_m value of ~ 4°C. The ligand F2031-0348 shown in green produces a thermal shift of ~ 12°C and represents the most promising compound of the screen.

These compounds present chemical groups common to the classical BRD inhibitors such as benzyle, quinoline, carboxamide, benzamide, pyridine. Four of them (Z30804950, F1091-0020, F3139-1743, 6572417) were found in both the screenings (receptor in the “open” and in the “close” conformations). Finally the orientation of the F2031-0348 ligand in the binding pocket and its interaction with the receptor were analysed. The ligand presents two main binding poses. In the first one the pyrazolopyrimidine group enters the binding cavity interacting with Leu100, Ala101, Ser104 and Ser107 in the ZA loop, Pro96 and Leu97 in the last portion of the αZ helix, Leu152, Thr153 and Lys156 in the BC loop and Phe149 in the αB helix (**Fig. 77A**). In this orientation the pyridimidine group forms a hydrogen bond with Ser107. In the second representative binding pose the chlorophenyl group enter the binding pocket and the ligand interacts with Leu100, Ala101,

Ser104 and Ser107 Leu115 and Asp116 in the ZA loop, Pro96 and Leu97 in the last portion of the α Z helix, Leu152, Thr153, Asp155 and Lys156 in the BC loop and Phe149 in the α B helix (**Fig. 77B**). In this orientation the pyrazolo group forms a hydrogen bond with Ser104.

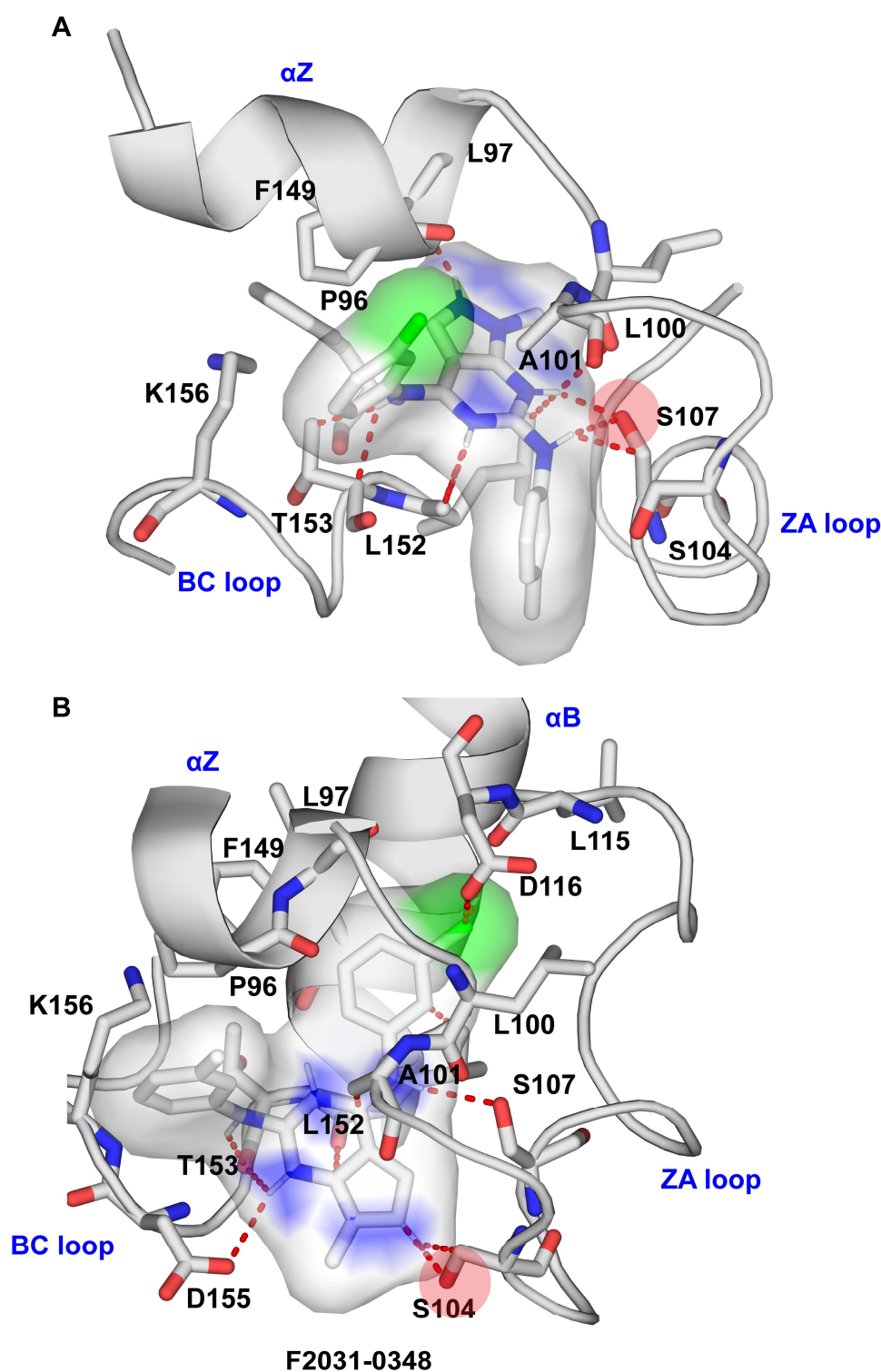


Figure 77| F2031-0348 binding poses. **A.** The pyrazolopyrimidine group enters the binding cavity interacting with Leu100, Ala101, Ser104 and Ser107 in the ZA loop, Pro96 and Leu97 in the last portion of the αZ helix, Leu152, Thr153 and Lys156 in the BC loop and Phe149 in the αB helix. The pyridimidine group forms a hydrogen bond with Ser107 **B.** the chlorophenyl group enter the binding pocket and the ligand interacts with Leu100, Ala101, Ser104 and Ser107 Leu115 and Asp116 in the ZA loop, Pro96 and Leu97 in the last portion of the αZ helix, Leu152, Thr153, Asp155 and Lys156 in the BC loop and Phe149 in the αB helix. In this orientation the pyrazolo group forms a hydrogen bond with Ser104.

4.6 Conclusions

We screened 13'433 compounds against the structure of KAP1 Bromodomain using a structure based virtual screening approach. The compounds are part of a bigger chemical library, The Chemically Diverse Collection (CDC) available at the Biomolecular Screening Facility (BSF), EPFL and they cover 85% of the chemical diversity of the entire library. The ligands were prepared with Open Babel¹⁶⁶ and MGLTools package (MGLTools-1.5.6) .

The structure of KAP1 Bromodomain was solved in 2008 by NMR (PDB: 2RO1⁶²). At first, we compared its structure with the other BRDs (PDBs: 6HDN¹⁶¹, 4WHU (*to be published*), 6HDQ¹⁶¹, 6MO9¹⁶², 5H1T¹⁶³, 5VOM¹⁶⁴, 6HDO¹⁶¹) in order to identify a possible binding pocket (**Fig. 70**). A putative binding site was identified between the ZA loop and the BC loop as in the other BRD structures. We analyzed this pocket and we observed the conservation of the hydrophobic characteristics common to the other BRDs and essential for the accommodation of the charged neutralized acetyl lysine and for the binding of inhibitors (**Fig. 68**). The main difference between KAP1 and the other structures is given by the position of the ZA loop. The ZA loop is conventionally close to the BC loop forming the hydrophobic pocket in which the acetyl lysine or the acetyl lysine mimetic compounds are recognized. In the NMR of KAP1, the ZA loop is shorter and it has a more “open” conformation compared to the other BRDs (**Fig. 67A**). The presence of this unique conformation can have different explanations: either the “open” conformation of the ZA loop together with the amino acids substitutions could preclude the formation of the hydrophobic cavity with the BC loop preventing the recognition of the acetylated peptides or given the flexibility of this protein region, the “open” conformation is a transition state that upon binding of an inhibitor could change dynamically assuming the same conformation as the other BRD structures and leading to the formation of a binding pocket together with the BC loop. This second hypothesis is supported by the high degree of flexibility of KAP1 ZA loop shown in the ¹H-¹⁵N heteronuclear NOEs plot⁶². Given the flexibility of the ZA loop, we decide to proceed with the screening of two protein conformations: the original NMR structure (PDB: 2RO1⁶²) in which the ZA loop is found in an “open” conformation and a new “close” conformation produce by homology modeling using Swiss-Model¹³⁶ based on the crystal structure of TRIM24 (PDB: 5H1T¹⁶³) (**Fig. 67B**). The structures were protonated and prepared for docking using AutoDockTools¹⁶⁵ (ADT4.2). The grid was centered in the middle of the four helices bundle and the dimensions were chosen to maximize the searching space between the ZA loop and the BC loop.

The docking was performed using Autock Vina¹⁵¹. The compounds were classified based on the value of their binding energy. Specifically the best pose for each compound showing the lowest binding energy was retrieved. We selected 136 compounds for each KAP1 Bromodomain conformation (with binding energies between 10 and 8 kcal/mol) for a total of 272 ligands that were binding the hydrophobic pocket between the

ZA loop and BC loop (**Fig. 72**). The compounds were packed on one side to the last portion of the α Z helix and on the other side to the α B helix showing a unique binding mode for KAP1. Among them, 18 compounds were common to both the “open” and the “close” conformation. The selected compounds were retrieved and they were analyzed with Scaffold Hunter¹⁶⁷, classified based on their scaffold and the most populated clusters were analyzed (**Fig. 73** and **Fig. 74**). The ligands included in these clusters were characterized by the presence of 3-5 aromatic rings and chemical groups common to the BRDs inhibitors such as quinoline, benzene, benzamide, pyridine, carboxamide.

The 272 positive compounds together with 8 BRD inhibitors (BIC1¹⁶⁸, NP1¹⁶⁹, Ischemin¹⁷⁰, JQ1²⁸⁰, I-BET762¹⁰⁷, I-BET151¹⁷¹, PF-1¹⁷², RVX-208¹⁷³) were selected and tested by DSF (Differential Scanning Fluorimetry). The compounds were ordered from the Biomolecular Screening Facility. The melting temperature (T_m) of KAP1 PHD-Br domain without addition of any compound averaged over 24 repetitions was 47.5 °C. Out of 272 tested ligands, 13 compounds produced a relevant shift of the melting curve (**Fig. 76**). Since there are no available data on KAP1 ligands, we decided to select for further tests only compounds that produced a shift in the melting curve ≥ 3 °C based on the change of the same amplitude produced by RVX-208 on the melting temperature of BRD2 (51°C to 54 °C on a thermal shift assay). Moreover, in the same study one of the most potent poli-BRDs inhibitor JQ1, produced a thermal shift of ~ 8 °C raising the melting temperature to 59 °C. The 13 selected compounds presented chemical groups common to the classical BRD inhibitors such as benzyle, quinoline, carboxamide, benzamide, pyridine. Four of them (Z30804950, F1091-0020, F3139-1743, 6572417) were found in both the screenings (receptor in the “open” and in the “close” conformations). In particular the compound F2031-0348 produced the most relevant thermal shift changing the T_m value of 12 degree from 47.5 °C to 59.9 °C (**Fig. 76**). Finally we evaluated the orientation of the F2031-0348 ligand in the binding pocket and its interaction with the receptor (**Fig. 77**). The ligand presents two main binding poses. In the first one the pyrazolopyrimidine group enters the binding cavity packing between the ZA loop, the α Z helix and the α B helix. In this orientation the pyridimidine group forms a hydrogen bond with Ser107. In the second representative binding pose the chlorophenyl group enters the binding pocket packing against the ZA loop, the α Z helix and the α B helix. In this orientation the pyrazolo group forms a hydrogen bond with Ser104.

Several protein-ligand complexes have been proved to exhibit multipose binding modes as observed for the KAP1-F2031-0348 complex. Among these multipose binding complexes, we find trypsin bound to benzo [b] thiophen-3-ylmethanamine²⁸¹, HIV type 1 protease in complex with 3-4-amino-pyrrolidine based inhibitor²⁸² and SH3 domain bound to a proline-rich decapeptide²⁸³. The benzo [b] thiophen-3-ylmethanamine shows two different orientations for the amine group within the trypsin binding pocket, the 3-4-amino-pyrrolidine based inhibitor shows two completely different binding modes for which the only similarity is represented by the interaction of the pyrrolidine nitrogen atom with the catalytic aspartates of the HIV type 1 protease and the

ability of the SH3 domain to recognize peptides in two distinct modes has been proved to play a crucial role in the specificity of the pathways involving this domain. In the case of KAP1-F2031-0348 complex, NMR studies and eventually crystal structures will be able to confirm the existence of multiple ligand binding modes possibly to be related to different protein functions thought in vitro assays.

Moreover, we must repeat the thermal shift assay experiments for the thirteen compounds identified in the first round of selection. The compounds have been already ordered and the experiment will be repeated with a ligand concentration of 10 μ M and 100 μ M, in triplicate. If our observations will be confirmed, especially for the compound F2031-0348, we will performed NMR studies on the KAP1- F2031-0348 complex to identify the residues involved in the binding of the drug. We will retrieve the K_d of the complex with direct methods such as ITC or Biacore (Biacore 8K from GE healthcare). We will attempt the crystallization of the receptor-ligand complex guided by the information available for the other BRDs in complex with inhibitors. We will screen more compounds using a ligand based virtual screening approach to enrich the pool of molecules presenting the same scaffolds as the positive hits found in the first screening and eventually we will synthesize even more potent ligands based on these scaffolds. Finally in collaboration with Professor Trono Laboratory of Virology and Genetics, EPFL, we will study the effect of the drugs *in vitro* and *in vivo* to better understand how the compounds regulate the transcriptional activity of KAP1 and how eventually they effect the interaction with chromatin remodelling enzymes and KRAB ZFPs.

Chapter 5 Conclusions and Future Perspectives

KAP1 (KRAB domain associated protein 1), also known as TRIM28 or TIF1 β , has been identified almost twenty years ago as a transcriptional co-repressor factor because of its interaction with the KRAB-ZFPs⁴². Several studies have demonstrated that KAP1 is involved in the control and in the regulation of many biological pathways and cell activities such as DNA damage response⁵¹, retroviral replication²⁸⁴, cytokine production²⁸⁵ and immune responses²⁸⁶, embryo development³³, cell proliferation²⁸⁷ and differentiation²⁸⁸. Moreover, KAP1 plays a fundamental role in regulating gene expression in mammalian cells¹⁰³, by recruiting different transcription factors and altering chromatin state¹⁰⁴. While doing so, KAP1 acts both as a platform for macromolecular interactions and as an E3 SUMO ligase¹⁹⁸. Given the emerging fundamental role of KAP1 in the development and differentiation of both normal and cancer cells, understanding how this protein controls the expression of the genetic information at the molecular level will have broad implications for both basic biology and translational research.

Structural information. At the time this work was performed, the X-ray structure of KAP1 RBCC domain was not available. In fact only recently the structure of the N-terminal RBCC domain was published by Stoll and collaborators (PDB: 6QAJ⁵⁹, July 2019) and by Lim and collaborators (PDB: 6H3A⁶⁰, June 2019). Moreover the NMR structure of the KAP1 RING domain has been recently solved (PDB: 6I9H⁶⁴, April 2019). This study showed that KAP1 RING domain does not form dimers observed in other TRIM family members (TRIM 25 and TRIM 32) fundamental for the binding of the E2 ubiquitin (E2-Ub) conjugation enzyme. However since KAP1 is known to be involved in the ubiquitination of different substrates including p53²⁸⁹, the authors speculate that the “inactive” KAP1 RING domain could form heterodimers with an “active” RING partners such as TRIM24 or TRIM33 stabilizing the E2-Ub conjugate. They suggest that KAP1 E3 ligase activity is suppressed while other TRIM family members are active and only when proteins turnover is required this activity is unleashed. Moreover the structure of the first B-box domain (PDB: 6O5K⁵⁸, June 2019) revealed a novel dimer interface for the formation of high-order assemblies with other TRIM proteins.

The RBCC domain and high oligomeric states. Stoll and collaborators⁵⁹ describe the RBCC domain as an antiparallel dimer in which the interface for the interaction with the KRAB domain of the ZFPs is represented by the coil-coiled domain. Furthermore, the authors discussed the oligomeric state of the RING domain that do not form dimers as observed for TRIM5 α , TRIM25, TRIM32 and BIRC-family E3 ligases^{187,290,291}. These results are in agreement with our SAXS data according to which the RBCC domain is a symmetric antiparallel dimer and the RING domain can act as an additional independent catalytic domain for the autoSUMOylation of KAP1 or can act synergistically with the PHD domain. This last finding is supported by our SAXS based KAP1 asymmetric model in which one PHD-Br domain is in close proximity with the RING domain.

Therefore, we concluded that KAP1 has two separate SUMO E3 ligase active sites which might work independently or synergistically. The role of the RING domain as a second catalytic site is supported by the observation that deletions of the PHD-Br domain module do not abolish the SUMOylation of IRF7 (Interferon regulatory factor 7)⁶⁸ while the deletion of the RING domain completely abolishes its SUMOylation¹⁹⁸. Moreover, KAP1 has been proved to interact with the E2-conjugating enzyme Ubc9 mainly through the RING domain¹⁹⁸, pointing to a catalytic activity of the latter. Furthermore, the authors observed by mutagenesis experiments that high-order oligomeric states of the RBCC domain are mediated by the B-box1 and not by the RING or the B-box 2 domains. In contrast to what suggested in this study, we never observed high-order oligomers when working with KAP1 full-length that was always observed as a dimer alone in solution and in complex with binding partners. Similarly, Lim and collaborators⁶⁰ solved the structure of the RBCC domain in complex with SMARCAD1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1). The RBCC domain showed an antiparallel organization in which the surface for the interaction with SMARCAD1 is the central region of the coiled-coil domain. The same results have been obtained by Sun and collaborators⁵⁸ that reconstructed the KRAB-RBCC complex by small-angle X-ray scattering experiments. The authors identified the central portion of the coiled-coil domain as the region of interaction with the KRAB-ZFPs and they observed high-oligomeric states mediated by the B-box 2 domain.

Comparison of our RBCC model with the other X-ray structures. To further compare our RBCC models with the recently published X-ray structures (PDBs: 6QAJ⁵⁹ and 6H3A⁶⁰) we fitted the crystallographic structures to our SAXS data using NOLB-NMA¹³³ protocol. Before the fitting, we added the N-terminal His₆ tag present in our construct and with which the SAXS data have been collected. The discrepancy between the theoretical scattering curve of the X-ray structures and the experimental SAXS data (χ^2 value) was calculated using Pepsi-SAXS¹³⁴. Initially the χ^2 value showed a poor fitting to the SAXS data (initial χ^2 values: 70 for 6QAJ⁵⁹ and 56 for 6H3A⁶⁰) however after NOLB-NMA we obtained a good fit with χ^2 values of respectively 2 and 0.8 (**Fig. 78**). The initial discrepancy between the experimental SAXS curve and the scattering curve calculated for the X-ray structures depends on the intrinsic rigidity of the crystallographic structures. In the X-ray structures this rigidity of the macromolecules is due to the hydrogen bonds or other interactions happening between the sub-chains or between atoms belonging to different unit cells and is produced by the crystal packing. Consequently, the relaxation of the side chains through normal mode analysis was sufficient to obtain a good fit of the two RBCC crystal structures to our SAXS data.

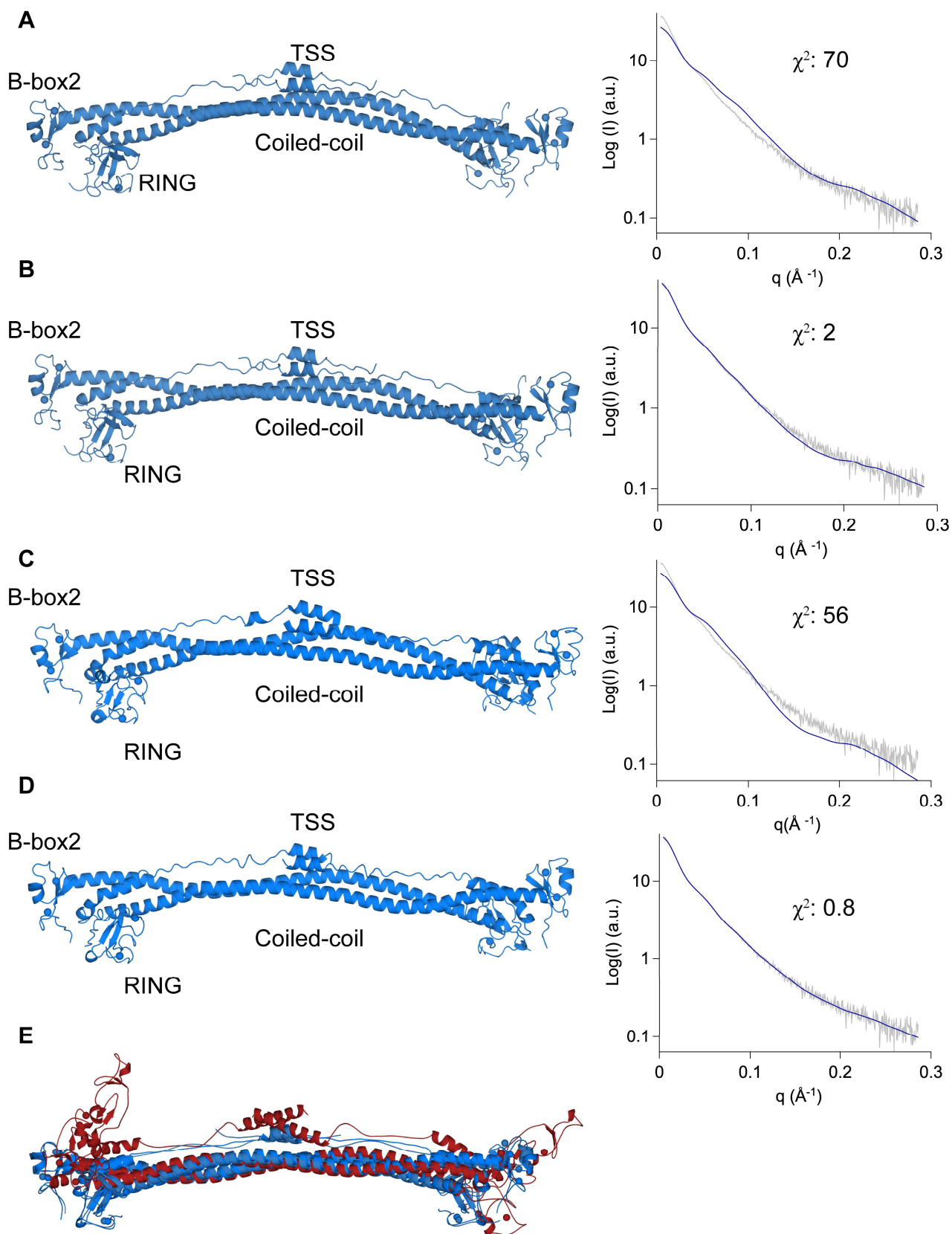


Figure 78| Fitting of the RBCC X-ray structures to the SAXS data. **A.** RBCC structure by Stoll and collaborators (PDB 6QAJ⁵⁹ and fit to the SAXS data (χ^2 : 70) **B.** Fit of the 6QAJ⁵⁹ structure to the SAXS data through NOLB-NMA¹³³ (χ^2 : 2) **C.** RBCC structure by Lim and collaborators (PDB: 6H3A⁶⁰) **D.** Fit of the 6H3A⁶⁰ structure to the SAXS data

(χ^2 : 0.8). Both structures have been modified with the addition of the N-terminal His₆Tag before the fitting **E**. Superimposition of the RBCC model obtain after NOLB-NMA in red, 6QAJ⁵⁹ and 6H3A⁶⁰ structures in blue fitted to the SAXS data.

KAP1 and binding partners. We characterized the interaction between KAP1 and HP1 α , using for the first time full-length purified proteins. HP1 α is a dimer with two globular domains linked by a flexible linker that could be found in solution in the close or in the open state as shown by our SAXS data and others²¹⁵. Previous data using the HP1 Chromo Shadow domain (CSD) and the KAP1 HP1 binding domain (HP1BD) showed that the CSD forms dimer that bind only one KAP1 peptide²¹⁶. According to our MALS data, the interaction between the two proteins is maintained when using full-length proteins but only one of the two KAP1 FL binding sites is available for the interaction. The observed stoichiometry 1:1 (1 KAP1 FL dimer bound to 1 HP1 α FL dimer) suggests that the asymmetric structure of KAP1 modulates the interaction with binding partners. This observation was confirmed within the context of chromatin fibers where KAP1 dimers recruit HP1 as dimers and not as tetramers, to the trimethylated nucleosomes such that two Chromo domains (CD) per complex can reach for the histone tails. This observation confirmed again the accessibility of only one HP1BD bound by a dimer of the HP1 Chromo Shadow domain. This interaction is more durable than the one in the absence of KAP1, allowing us to speculate that in cells, the binding of KAP1 stabilizes HP1 dimerization in the chromatin facilitating heterochromatin formation and spreading.

The future of the structural biology studies for KAP1. SAXS experiments have provided us with the global architecture of full-length KAP1 thus our next aim is to solve the atomic structure of the full-length protein using X-ray crystallography. We will attempt to crystallize KAP1 in complex with small binders specifically selected to interact with the protein and to ideally stabilize its flexible regions, in particular the long loop connecting the N-terminal and C-terminal domains. In collaboration with Andreas Plückthun Laboratory, University of Zurich, we are selecting DARPins (Designed Ankyrin Repeat Proteins), engineered antibody mimetic proteins proven to be valuable tools in the crystallization of many difficult proteins^{189,190,191}. The DARPins will be purified in our laboratory and subsequently complexed with KAP1 FL. One of the most promising avenues we are pursuing in order to restrict the flexibility of full-length KAP1, which is likely the main reason of poorly diffracting crystals, is co-crystallization with binding partners such HP1 that binds the HP1BD with nM affinity. HP1 acts as a linker to connect KAP1 to nucleosomes. While the size of the FL KAP1 dimer complexed with HP1 remain on the lower size limit for cryo-EM studies complexing KAP1-HP1 with nucleosome would increase size and dimensions, and likely better stabilize the assembly for EM work. Beat Fierz's laboratory has expertise producing nucleosomes with different PTMs at their histone tails. Thus, one possible avenue to be explored is the reconstitution of the KAP1-HP1-di-nucleosome complex, engineered as to have the HP1 CSD binding to KAP1 and the HP1 CD to the H3K9me3 modification (**Fig.79**). Since the HP1-di-nucleosome complex has been already structurally characterized by cryo-EM⁷⁶, its structure and electron density map could be useful in the three-dimensional reconstruction of the ternary

complex. We will also study how KAP1 influences the recruitment of different isoforms of HP1 to the nucleosomes by smTIRFM and how this interaction leads to different levels of chromatin compaction, by comparing the k_{on} and dwell times ($\tau_{off,1}$ and $\tau_{off,2}$) values of the three HP1 isoforms in the presence and absence of KAP1. Moreover, the binding of full-length HP1 to full-length KAP1 can be also compared to the binding to a minimal 15-residue long peptide containing the KAP1 HP1BD (aa 483-497), which can be also used to solve the atomistic structure of the CDS of HP1 (HP1 α aa 110-180 and HP1 β aa 108-175) bound to the KAP1-HP1BD peptide.

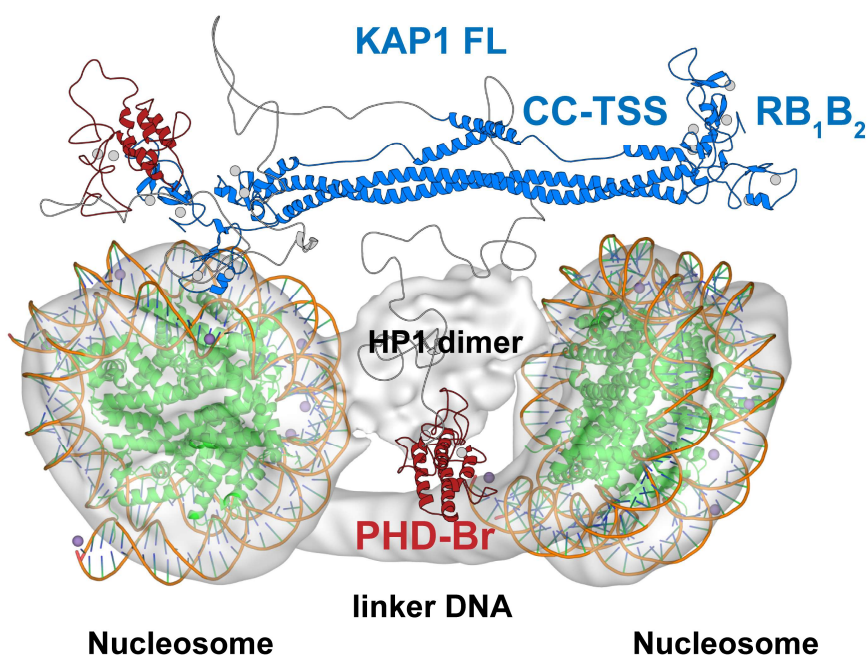


Figure 79| Schematic representation of the KAP1-HP1-di-nucleosome complex. The HP1-di-nucleosome complex has been already structurally characterized by cryo-EM⁷⁶. Here we present a schematic representation of the complex between KAP1 FL (RBCC domain shown in blue and PHD-Br domain in red) and the HP1 dimer-di-nucleosome. HP1 binds KAP1 through the CSD and at the same time contacts the H3K9me3 histone modifications through the CD.

Studies on the post-translational modifications of KAP1. KAP1 can be phosphorylated, SUMOylated and acetylated. Moreover, it is able to act as an E3 SUMO or ubiquitin ligase. There is currently no information on how these PTMs affect the structure of KAP1 and lead to the formation or disruption of the different KAP1 assemblies that in turn regulate cellular development and differentiation. Ser824 is the first target, due to the implication of the phosphorylation at this site in DNA damage response and the de-SUMOylation of KAP1, leading to the disruption of the interaction with SETDB1. Another candidate residue is Ser473, whose phosphorylation affects cell cycle progression and leads to KAP1 translocation to the cytoplasm, probably because HP1 binding at the HP1BD is compromised^{104,51,53}. Other candidate sites are Tyr449, Tyr458 and

Tyr517 which are phosphorylated by Src family kinases (SFKs). The phosphorylation of these tyrosine residues also disrupts the interaction between KAP1 and HP1²⁹². As SFKs regulate several oncogenic processes, including cell growth and differentiation, this result implicates KAP1 in SFK-mediated oncogenic transformation. It would be extremely interesting to study the effect of PTMs at the binding interface with HP1. In addition, mutants defective in HP1 binding (VV681-682EE²⁰²) are not SUMOylated as well as the wild type (wt) *in vivo*⁶⁸. One future avenue is to compare phosphomimetic mutants binding affinities to HP1 to those of wt KAP1 and to test them in SUMOylation assays. Phosphomimetic mutants could be used for structural studies, for assembling complexes and for cellular functional experiments done in collaboration with the Trono's laboratory. We expect that the information resulting from the proposed experiments will help rationalizing the previous findings.

Identification of novel small molecules for KAP1 Bromodomain. We screened 13'433 compounds against the structure of KAP1 Bromodomain using a structure based virtual screening approach. The compounds are part of a bigger chemical library, the Chemically Diverse Collection (CDC) available at the Biomolecular Screening Facility (BSF), and they cover 85% of the chemical diversity of the entire library. We were able to identify for the first time new small molecule ligands for KAP1 C-terminal domain. The C-terminal tandem PHD-Br domain of KAP1 shows a unique function acting as an E3 SUMO ligase promoting both the auto-SUMOylation⁶⁸ of the protein and the SUMOylation of other substrates^{67,237}. The auto-SUMOylation of the C-terminal domain is necessary for the binding of KAP1 to the chromatin remodeling enzymes such as SETDB1, NuRD complex, HDAC and N-CoR inducing the deposition of PTMs and consequently the creation of an heterochromatin environment.^{197,34,35}. Currently, there are no inhibitors for KAP1 able to regulate its transcriptional activity. The inhibition of the SUMOylation activity of KAP1 *in vitro* is achieved through the mutation of the Cys651 that coordinates the zinc ions in the PHD domain⁶². The mutation of this residue into alanine prevents the correct folding of the C-terminal domain and consequently the binding of the SUMO protein.

Future perspectives for KAP1 drug discovery studies. Once confirmed, the results of the current screening will be used to screen again the chemical diversity library in a ligand-based approach. This will provide us with an enriched library of small molecules presenting the same chemical characteristics as the ligands identified in the first screening to be tested *in vitro* and *in vivo*. Specifically, we will retrieve the K_d of KAP1 in complex with the ligands through direct methods such as ITC (if the compounds will be soluble in the same buffer as KAP1 allowing us to remove the DMSO) or by Biacore (Biacore 8K from GE healthcare). NMR studies can be performed to identify the KAP1 residues involved in the ligand binding. Eventually we will attempt the crystallization of the protein-ligand complexes guided by the information available for the other BRDs in complex with inhibitors.

Furthermore, in collaboration with Professor Trono Laboratory of Virology and Genetics, EPFL, we will study the effect of the drugs *in cell* to better understand how the compounds regulate the transcriptional activity of KAP1 and how eventually they effect the interaction with chromatin remodelling enzymes and the KRAB ZFPs. Specifically we can perform a transcriptional repression activity assay in which the repression of a reporter-gene is measured. A Gal4-KRAB fusion protein can be used to recruit KAP1 to specific reporter-promoters regions with five Gal4 sites inducing the repression of the reporter-gene. According to the hypothesis that the small ligands will act as KAP1 inhibitors, we will measure an increased expression of the reporter-gene (**Fig. 80**).

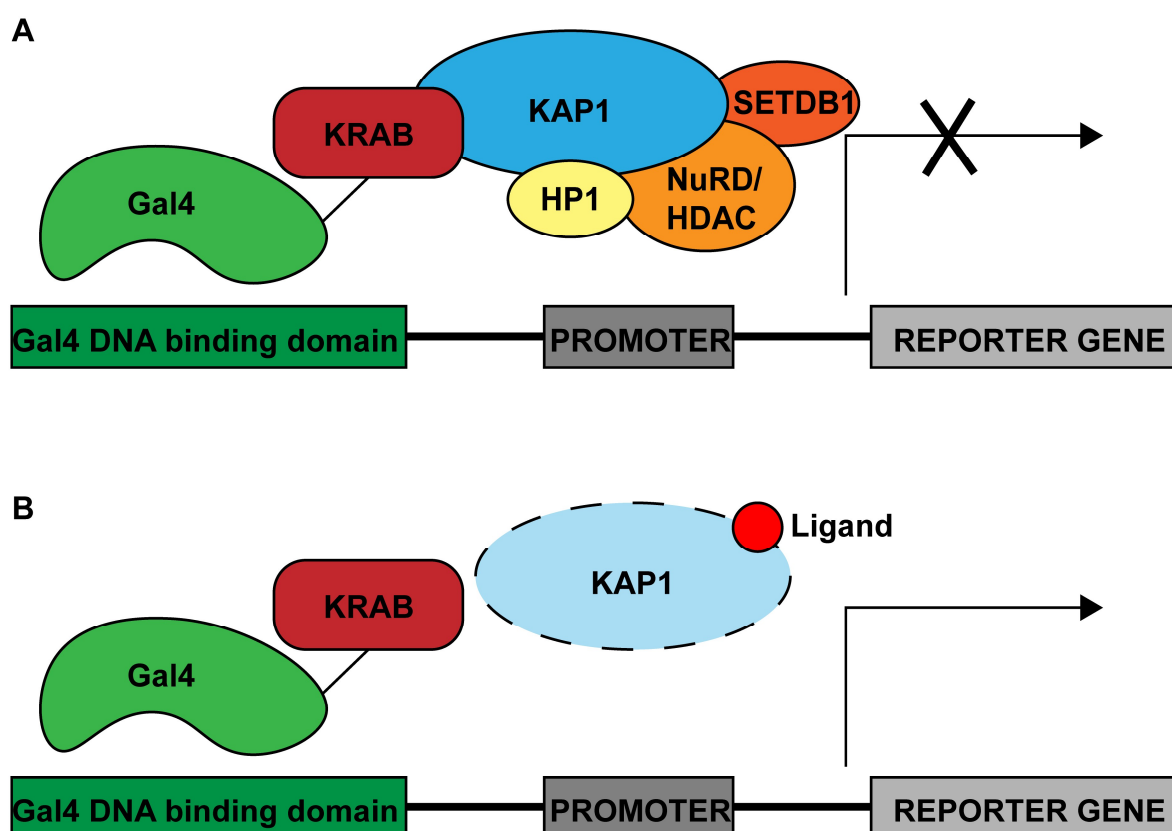


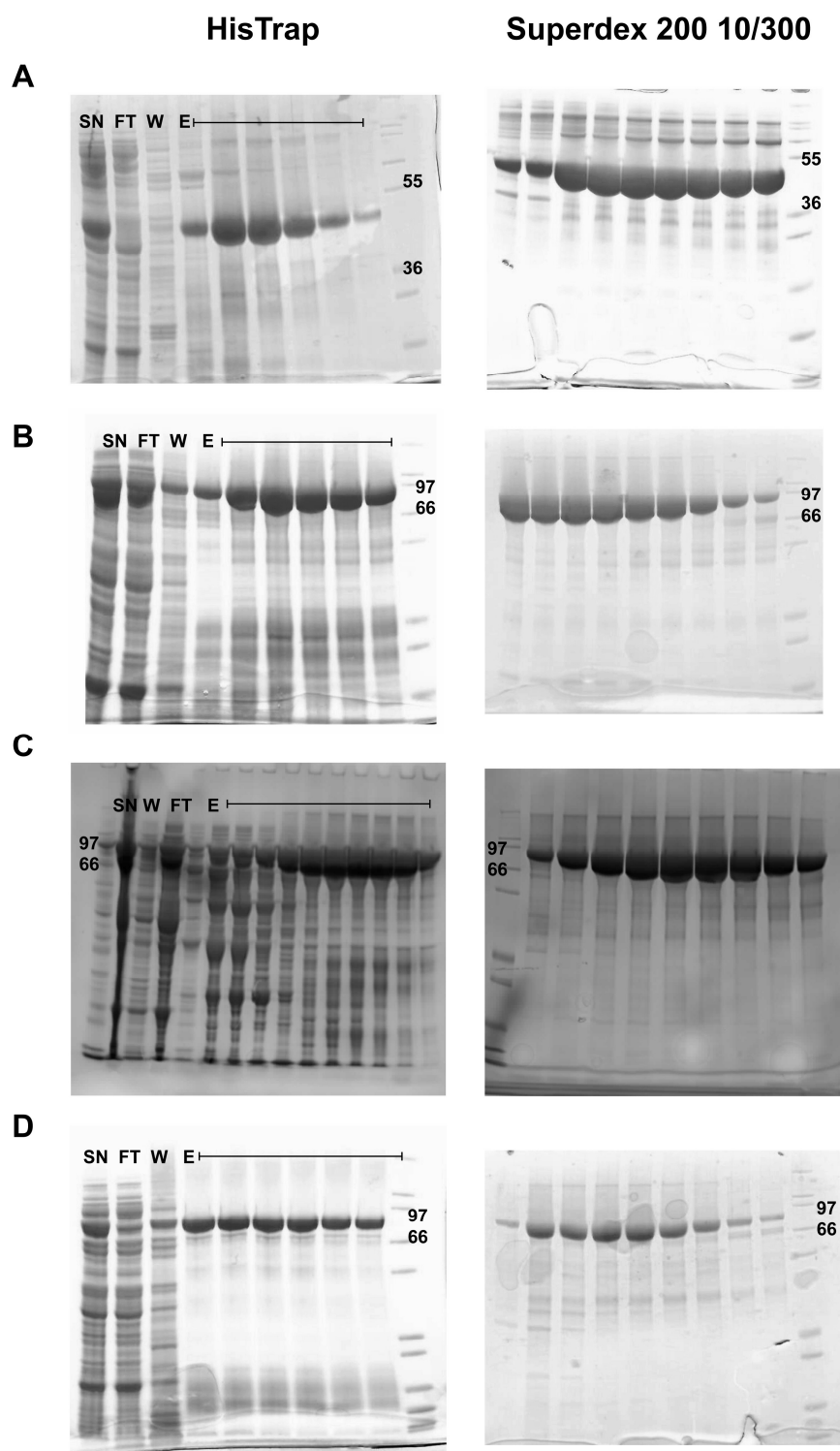
Figure 80| Schematic representation of a transcriptional repression activity assay. A. The Gal4-KRAB fusion protein recruits KAP1 that in turn binds the chromatin remodelling enzymes inducing the deposition of PTMs on the chromatin inducing gene repression **B.** Upon binding of a small molecule acting as an inhibitor, KAP1 is not recruited anymore at the reporter-promoter, it does not induce the formation of the heterochromatin environment and consequently the expression level of the reporter gene is increased.

The importance of IM in the treatment of large and flexible systems. Methods that integrate a large amount of experimental data with computational analysis are commonly known as integrative modeling (IM) and they are today emerging as a powerful strategy to model the architecture of large macromolecular assemblies. IM methods have been successfully applied to reveal the structures of several large important complexes such as the yeast nuclear pore complex (NPC) of 50 MDa involved in the selective transport of

proteins and ribonucleoparticles between the nucleus and the cytoplasm⁹², the proliferating cell nuclear antigen (PCNA) bound to ubiquitin⁹³ and the complex PCNA, Rad9-Hus1-Rad1 (9-1-1) and flap endonuclease 1 (FEN1) specialized in DNA repair⁹⁴. The experimental data to complement the computational approaches can be the low-resolution structural data describing the volume, the shape or the arrangement of the single subunits of the large macromolecular assemblies. These are for example medium or low-resolution cryo-electron microscopy maps and small-angle X-ray scattering (SAXS) data. Moreover, spatial restraints extracted from NMR, electron paramagnetic resonance or cross-link/mass-spectrometry experiments can be used to define the contacts between subunits and to facilitate the modeling process of large assemblies. As for KAP1, the IM approach was a powerful resource to understand the structure of the full-length protein not suitable to be solved by classical high-resolution structural biology techniques such as X-ray crystallography and cryo-EM. The combination of experimental data, molecular modeling and SAXS data fitting allowed us to characterize the structure and the interaction with binding partners of such a dynamic and flexible macromolecule.

Chapter 6 Appendix

6.1 Appendix 1: Gels gallery

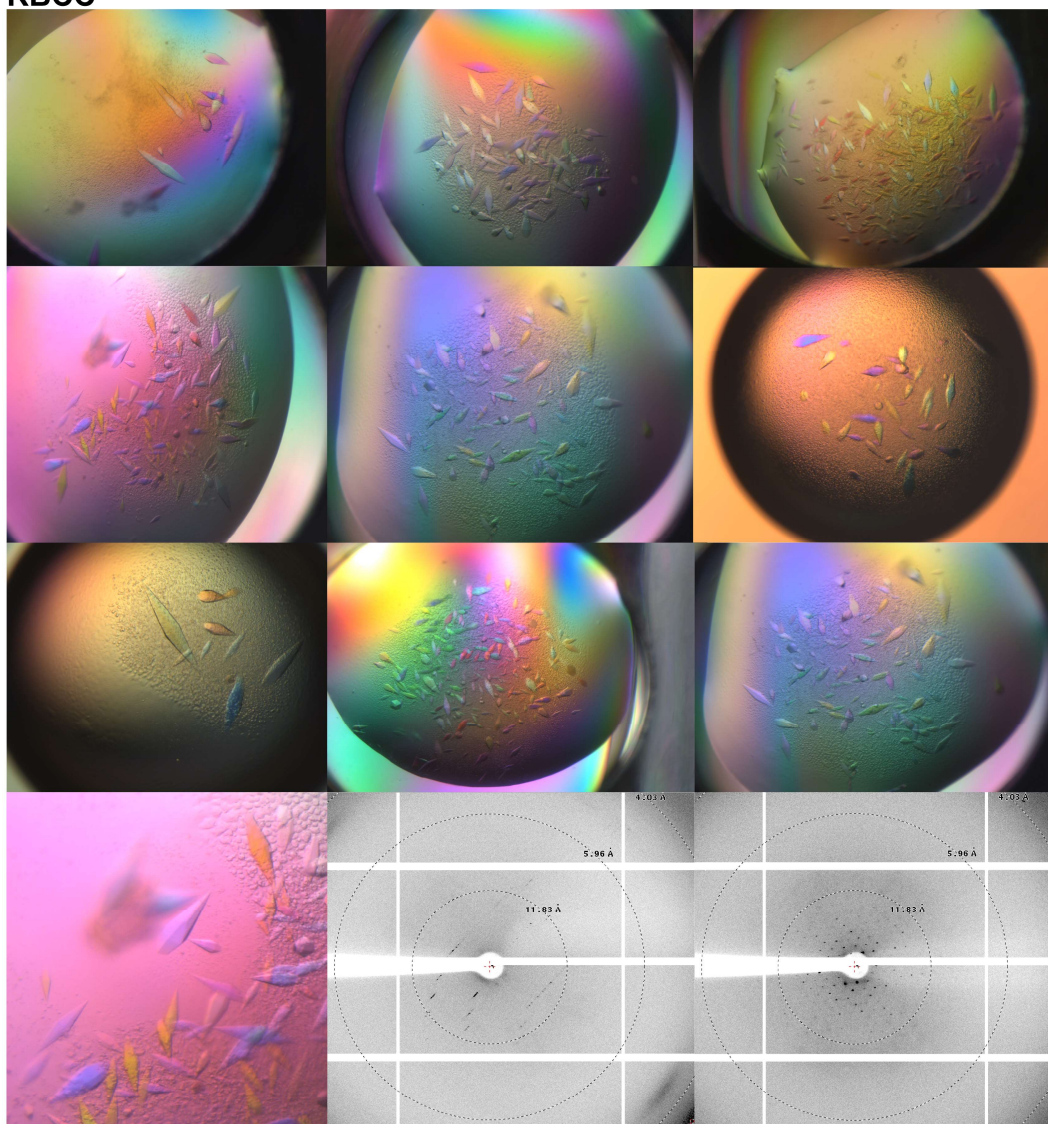


A. RBCC **B.** Δ KAP1 **C.** KAP1 FL **D.** Δ RING.

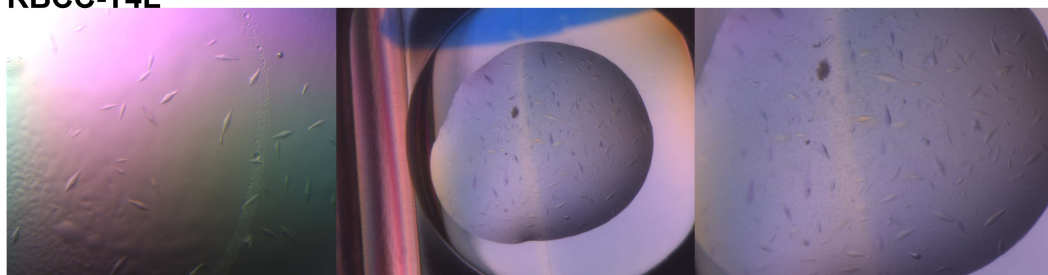
Abbreviations: SN: supernatant, FT: flow-through, W: wash, E: elution.

6.2 Appendix 2: Crystal gallery

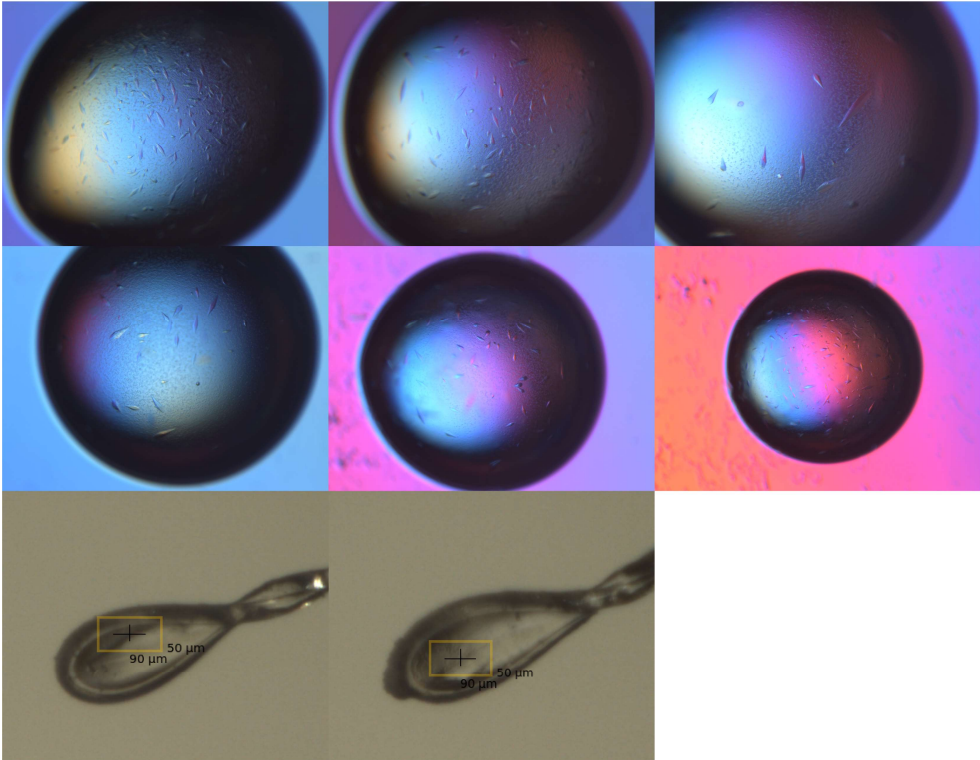
RBCC



RBCC-T4L



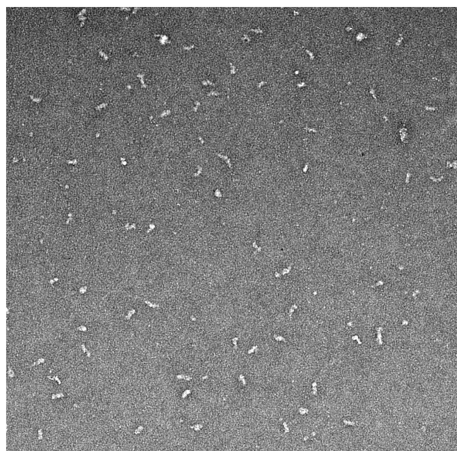
KAP1 FL



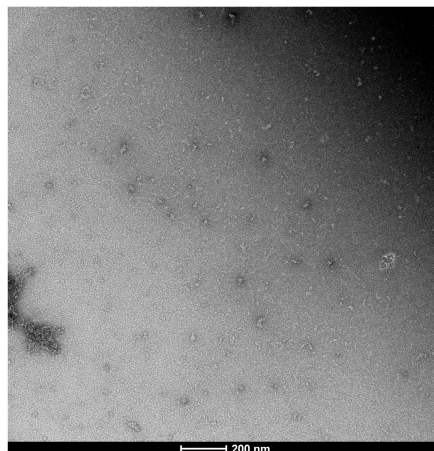
6.3 Appendix 3: Cryo EM gallery

Negative staining

A

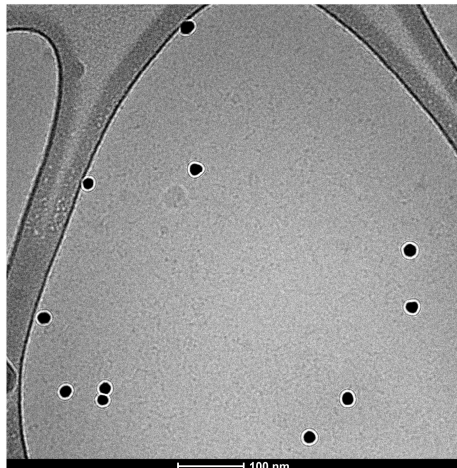


B

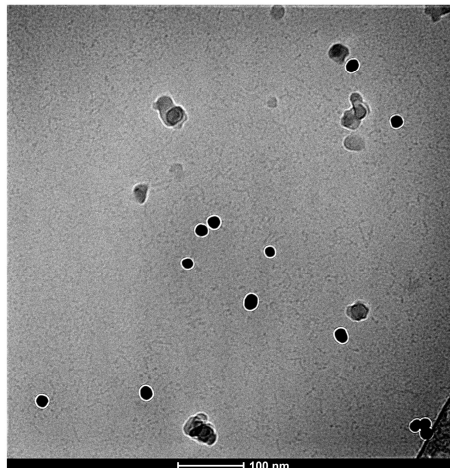


Cryo EM

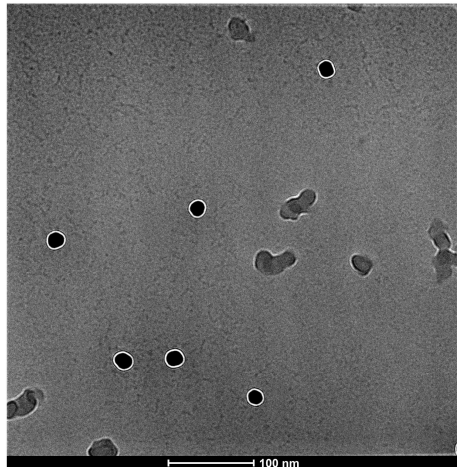
A



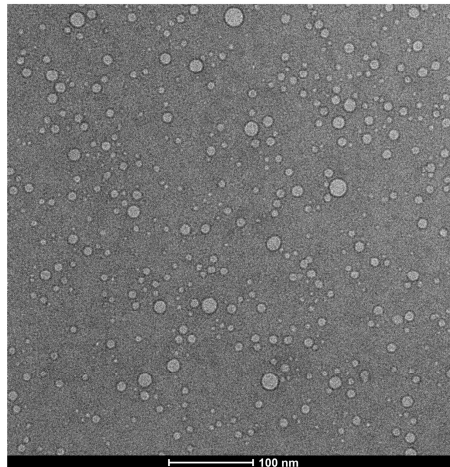
B



C



D



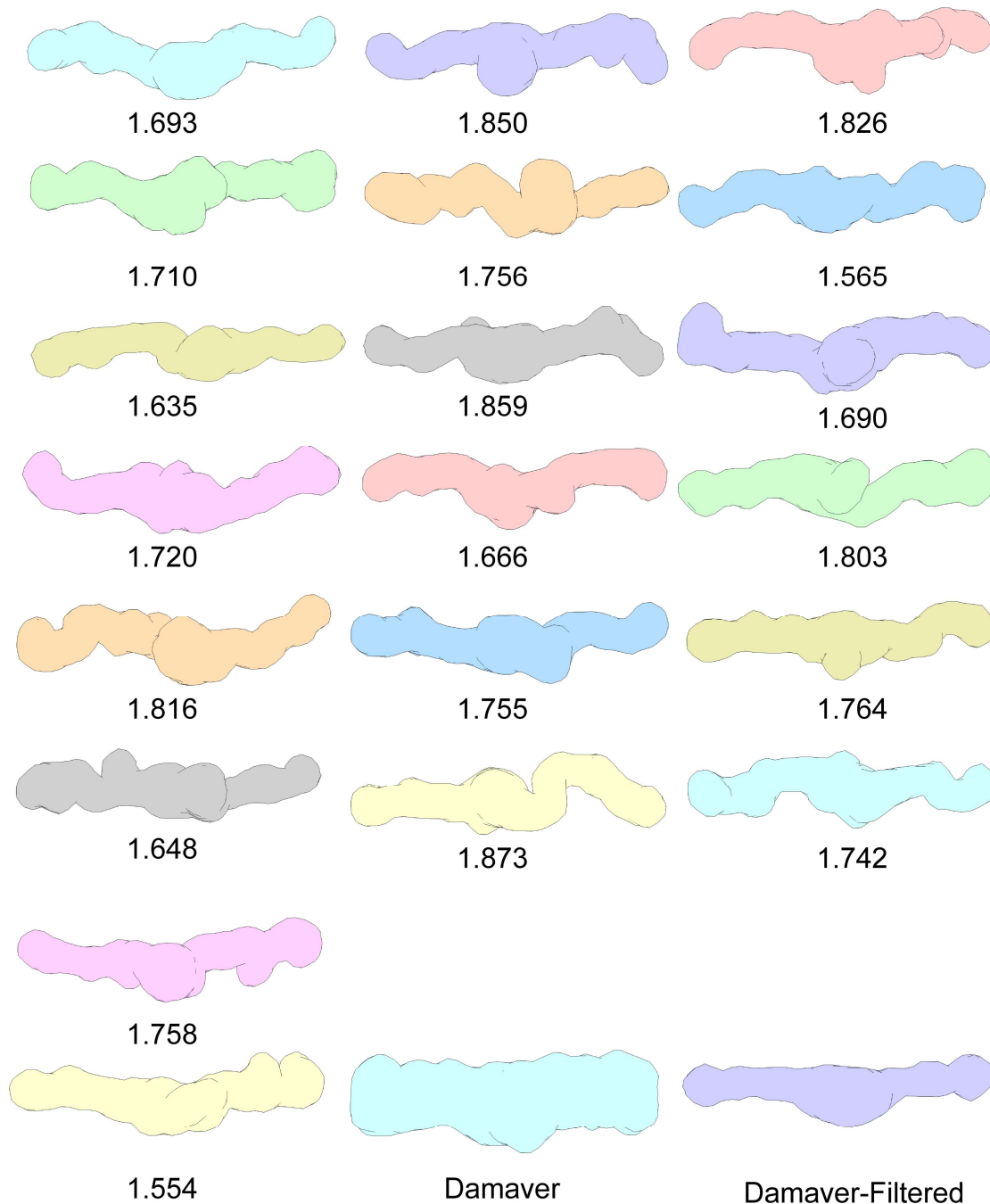
Negative staining **A.** KAP1 crosslinked (**Fig 29**) **B.** KAP1 FL not crosslinked (0.1 mg/mL)

Cryo EM **A.B.C.** Falcon camera 80'000X, 200X dilution (0.6 mg/mL) **D.** Bubbling after high dose exposure

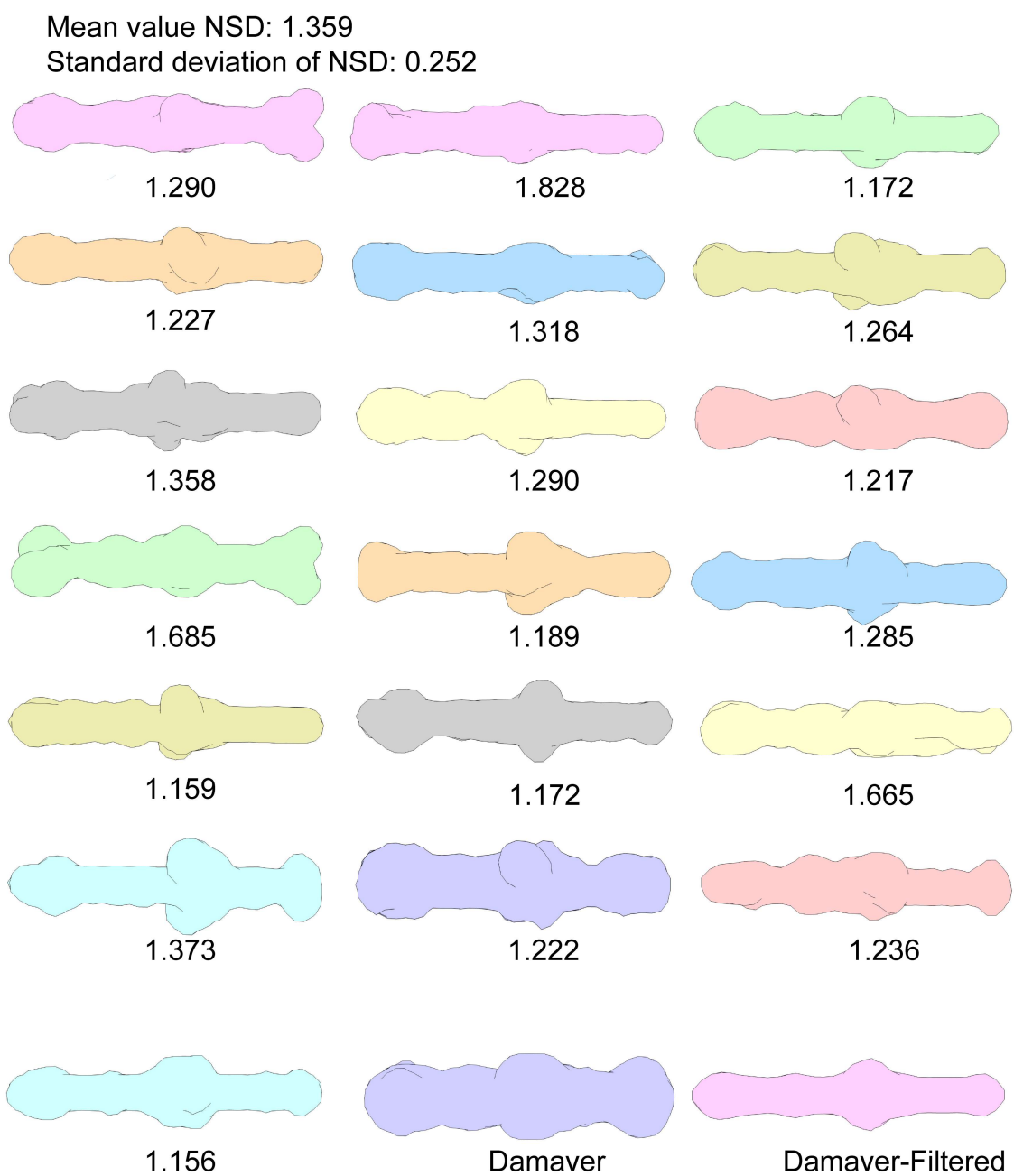
6.4 Appendix 4: *Ab initio* DAMMIN models

Mean value NSD: 1.734

Standard deviation of NSD: 0.091

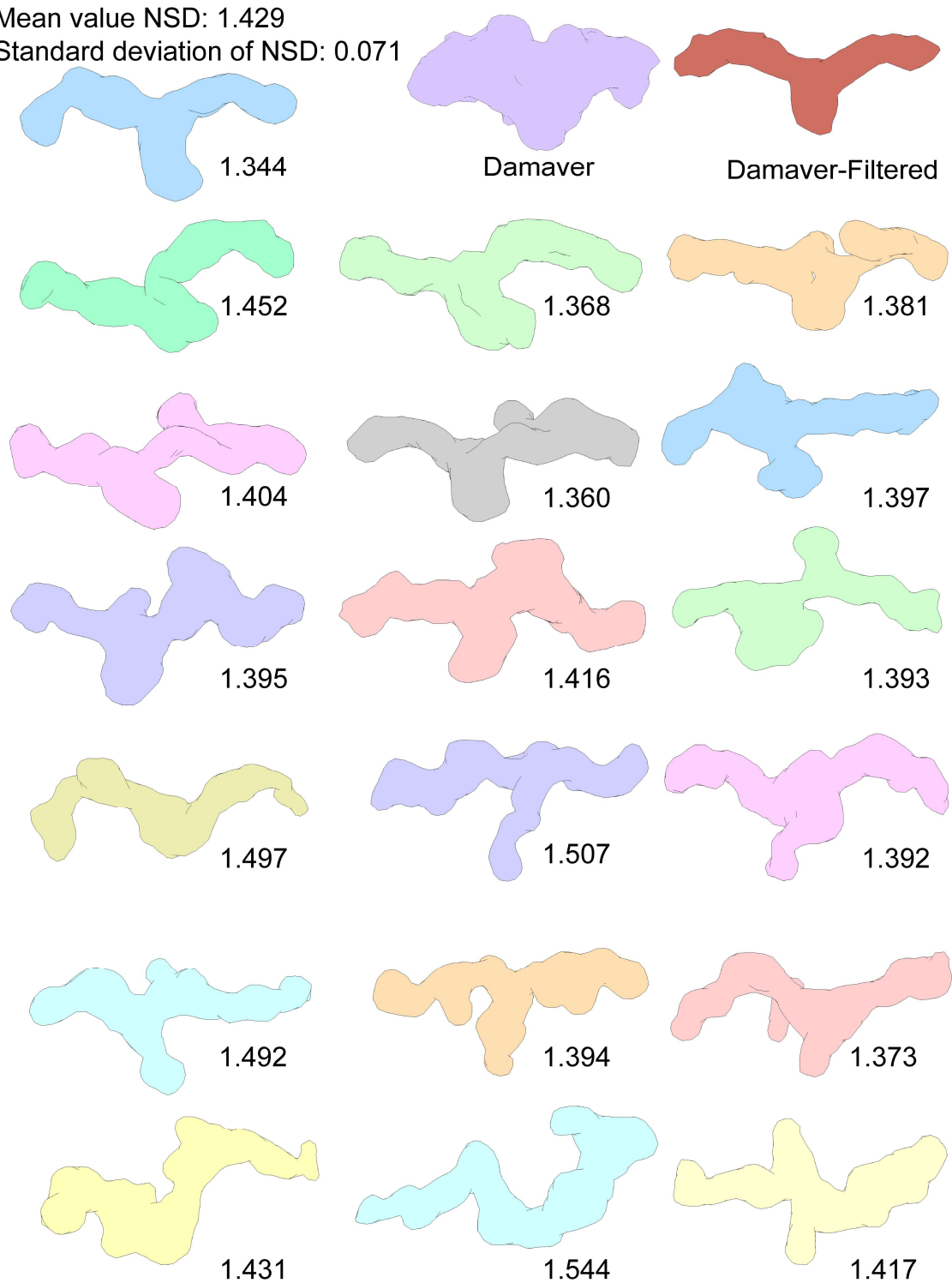


RBCC models generated without 2-fold symmetry (P1)



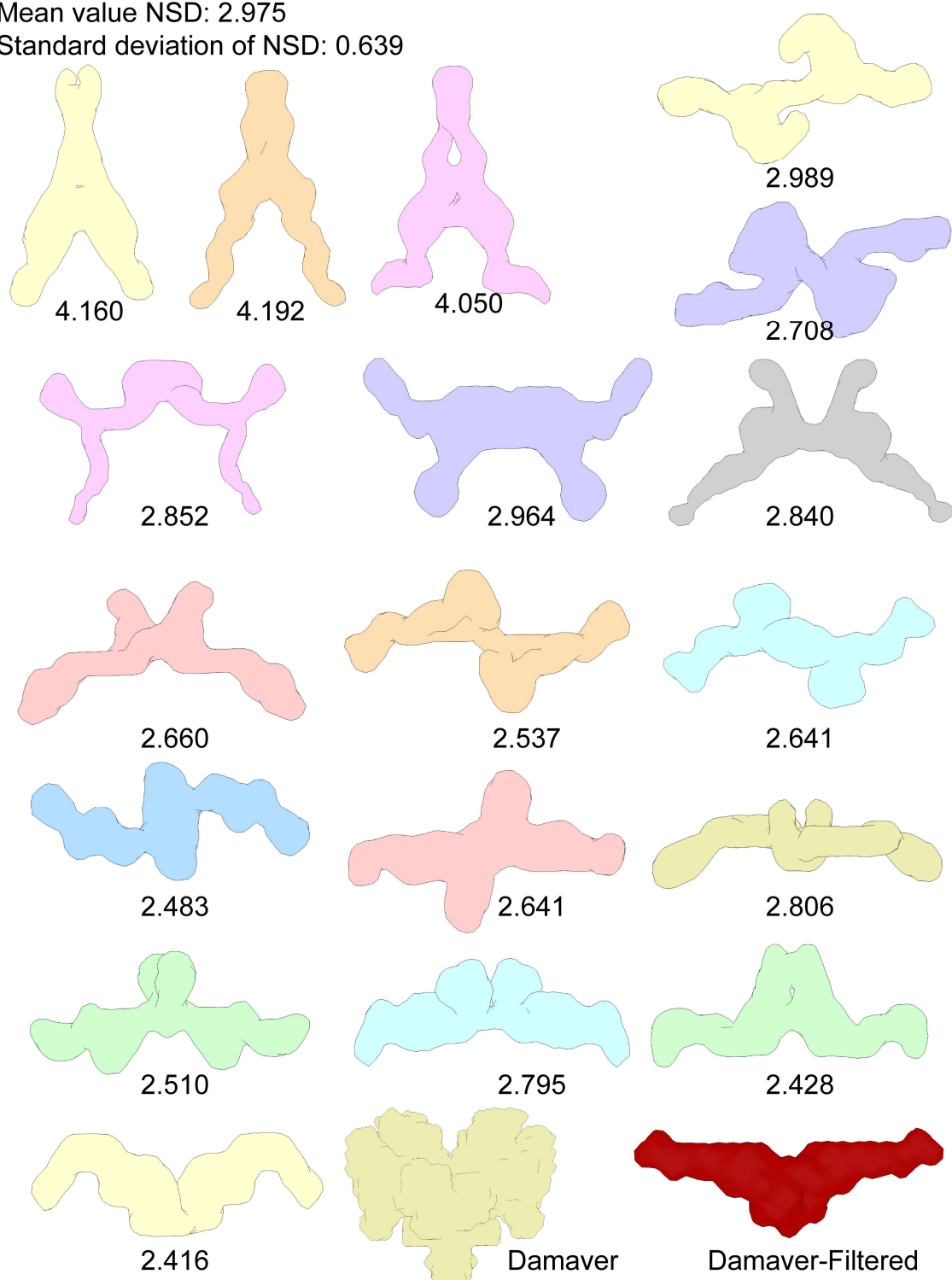
RBCC models generated with 2-fold symmetry (P2)

Mean value NSD: 1.429
Standard deviation of NSD: 0.071



KAP1 FL models generated without 2-fold symmetry (P1)

Mean value NSD: 2.975
Standard deviation of NSD: 0.639



KAP1 FL models generated with 2-fold symmetry (P2)

6.5 Appendix 5: C651A mutant

PCR reactions:

KAP1 C651A	1	2	3
Template 100ng/uL	100ng 1uL	200ng 2uL	300ng 3uL
Primer Fw 100uM	1	1	1
Primer Rv 100uM	1	1	1
10X Buffer	5	5	5
dNTP mix	1	1	1
Quick solution	3	3	3
Pfu TURBO polymerase	1	1	1
H2O	37	36	33
	50uL	50uL	50uL

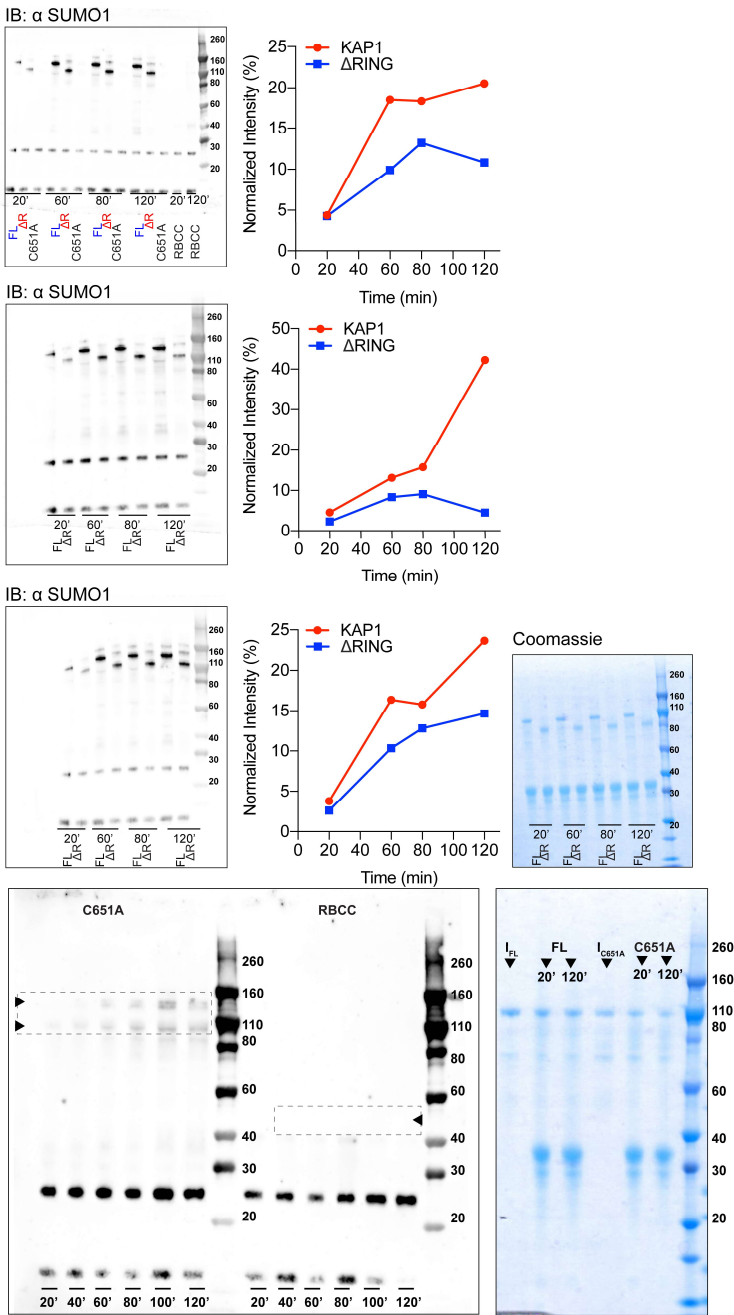
PCR protocol:

1. 95°C for 4 min
2. 95°C for 1 min
3. 60°C for 50 sec
4. 68°C for 18 min
(vector+ KAP1 =7275 bp, 2.5 min/kb)
5. Go to step 2. 18X
6. 68°C for 10 min
7. 4°C HOLD
8. Add Dpn1 37°C for 1 hour
(to digest the parental supercoiled dsDNA)

PCR primers:

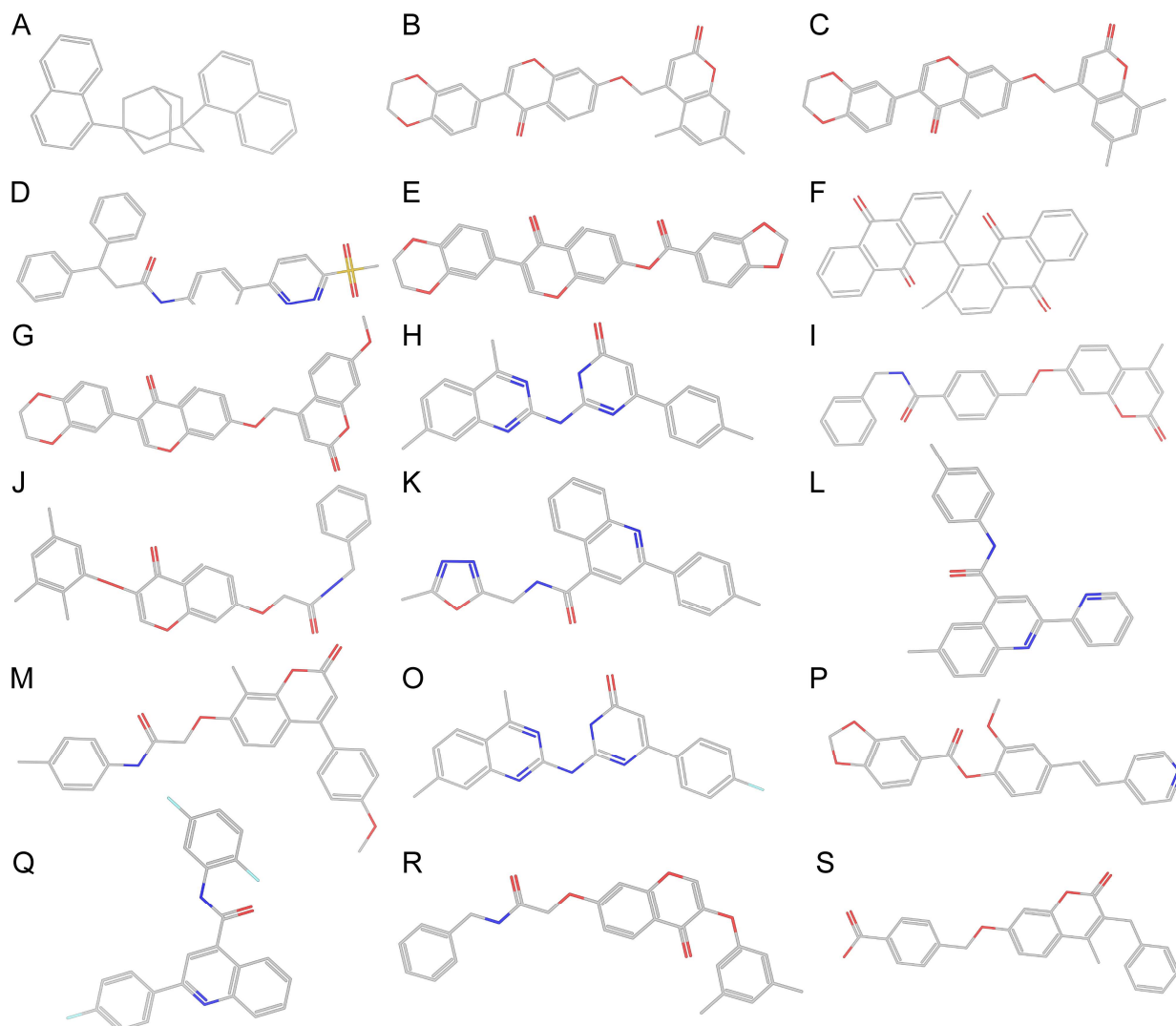
Fw: GGG CCG GCA GGT GAG CGT CCA
GGT GGAAAC
Rv: GTT TCC ACC TGG ACG CTC ACC
TGC CGG CCC

6.6 Appendix 6: Uncropped Western Blot images related to Fig. 41



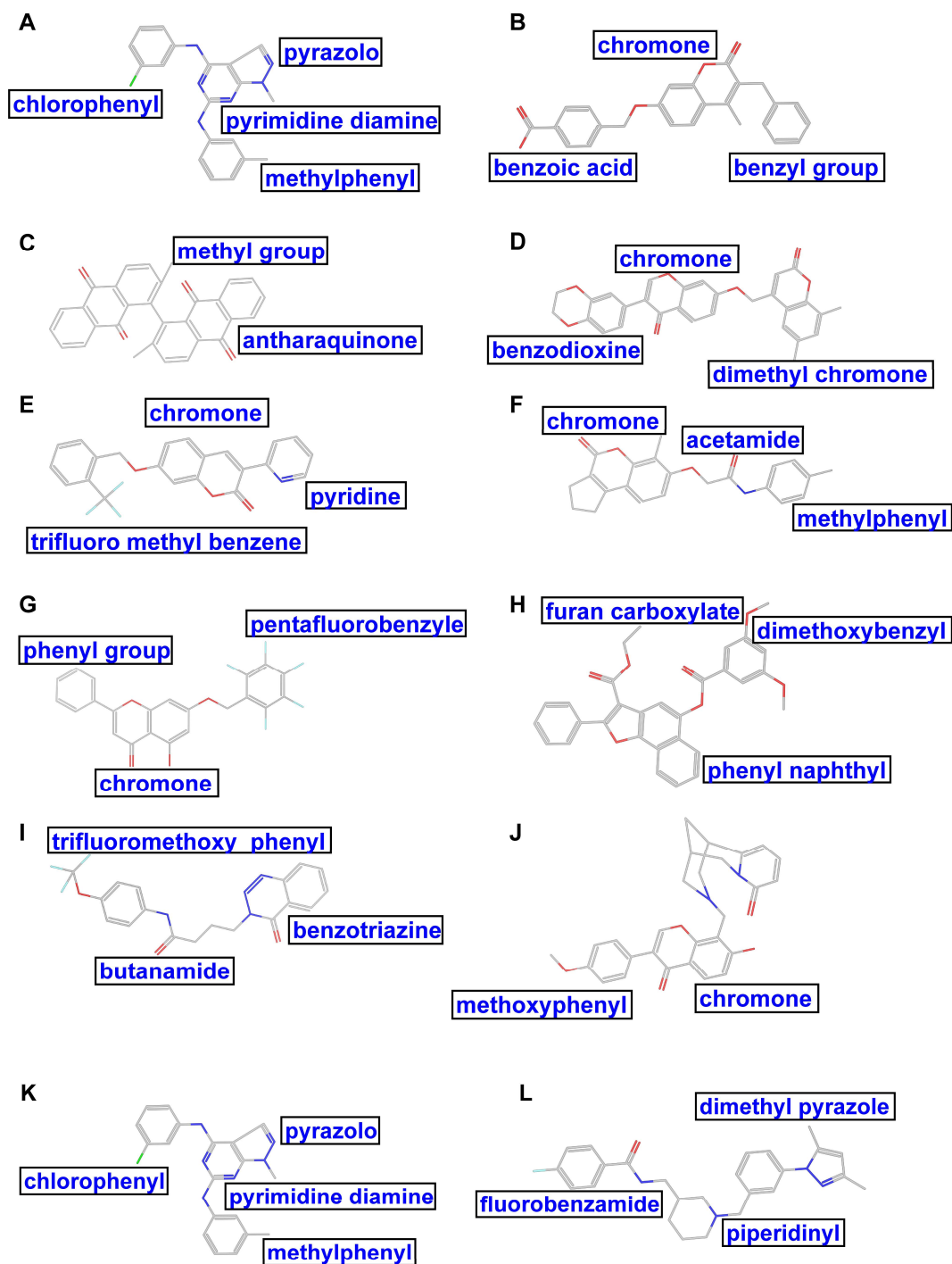
Un-cropped western blots related to Fig 39. The first three panels show the experimental replicas for the auto-SUMOylation assay comparing KAP1 FL and Δ RING. The intensities were normalized with respect to the sum of all considered intensities. The Coomassie stained SDS-PAGE is shown on the right as loading control. The bottom panel shows the auto-SUMOylation assay using the C651A mutant and KAP1 RBCC. As expected, KAP1 RBCC did not get SUMOylated. When comparing the MW shifts of the proteins upon SUMOylation, it is interesting to notice how when the PHD is active, the size change is large (50 kDa, seen for FL and Δ RING) but when the PHD is compromised, the size change is very small (seen for the C651A mutant). On the right, the Coomassie stained SDS-PAGE of the FL and C651A reactions is shown as loading control, where non SUMOylated proteins are indicated by I FL and I C651A. (I=input)

6.7 Appendix 7: 18 common inhibitors found in the open and close ZA loop conformations



- A. 1,3-dinaphthalen-1-yladamantane
 B. 4-(((3-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-4-oxo-4H-chromen-7-yl)oxy)methyl)-5,7-dimethyl-2H-chromen-2-one
 C. 4-(((3-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-4-oxo-4H-chromen-7-yl)oxy)methyl)-6,8-dimethyl-2H-chromen-2-one
 D. N-(4-(6-(methylsulfonyl)pyridazin-3-yl)phenyl)-3,3-diphenylpropanamide
 E. 3-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-4-oxo-4H-chromen-7-yl benzo[d][1,3]dioxole-5-carboxylate
 F. 2,2'-Dimethyl-1,1'-bianthraquinone
 G. 4-(((3-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-4-oxo-4H-chromen-7-yl)oxy)methyl)-7-methoxy-2H-chromen-2-one
 H. 2-[(4,7-Dimethyl-2-quinazolinyl)amino]-6-(4-methylphenyl)-4(1H)-pyrimidinone
 I. N-Benzyl-4-[(4-methyl-2-oxochromen-7-yl)oxymethyl]benzamide
 J. N-benzyl-2-[[4-oxo-3-(2,3,5-trimethylphenoxy)-4H-chromen-7-yl]oxy]acetamide
 K. N-[(5-methyl-1,3,4-oxadiazol-2-yl)methyl]-2-(4-methylphenyl)-4-quinolinecarboxamide
 L. 6-Methyl-2-(pyridin-2-yl)-N-(p-tolyl)quinoline-4-carboxamide
 M. 2-[[4-(4-methoxyphenyl)-8-methyl-2-oxo-2H-chromen-7-yl]oxy]-N-(4-methylphenyl)acetamide
 O. 2-[[4-(7-dimethyl-2-quinazolinyl)amino]-6-(4-fluorophenyl)-4(3H)-pyrimidinone
 P. (E)-2-Methoxy-4-(2-(pyridin-4-yl)vinyl)phenyl benzo[d][1,3]dioxole-5-carboxylate
 Q. N-(2,5-Difluorophenyl)-2-(4-fluorophenyl)quinoline-4-carboxamide
 R. N-benzyl-2-[[3-(3,5-dimethylphenoxy)-4-oxo-4H-chromen-7-yl]oxy]acetamide
 S. 4-[[[(3-benzyl-4-methyl-2-oxo-2H-chromen-7-yl)oxy]methyl]benzoic acid

6.8 Appendix 8: Positive compounds retrieved from thermal shift assay



- A. N-4~-(3-Chlorophenyl)-1-methyl-N-6~-(3-methylphenyl)-1H-pyrazolo[3,4-d]pyrimidine-4,6-diamine
 B. 4-[[[(3-benzyl-4-methyl-2-oxo-2H-chromen-7-yl)oxy]methyl]benzoic acid
 C. 2,2'-Dimethyl-1,1'-bianthraquinone
 D. 4-(((3-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-4-oxo-4H-chromen-7-yl)oxy)methyl)-6,8-dimethyl-2H-chromen-2-one
 E. 3-(2-Pyridinyl)-7-[[[2-(trifluoromethyl)benzyl]oxy]-2H-chromen-2-one
 F. 2-[[[(6-Methyl-4-oxo-1,2,3,4-tetrahydrocyclopenta[c]chromen-7-yl)oxy]-N-(4-methylphenyl)acetamide
 G. 5-Hydroxy-7-[(pentafluorobenzyl)oxy]-2-phenyl-4H-chromen-4-one
 H. Ethyl 5-[(3,5-dimethoxybenzoyl)oxy]-2-phenyl-1-naphtho[1,2-b]furan-3-carboxylate
 I. 4-(4-Oxo-1,2,3-benzotriazin-3(4H)-yl)-N-[4-(trifluoromethoxy)phenyl]butanamide
 J. 11-[[[7-Hydroxy-3-(4-methoxyphenyl)-4-oxo-4H-chromen-8-yl]methyl]-7,11-diazatricyclo[7.3.1.0^{2,7}]trideca-2,4-dien-6-one
 K. N-4~-(3-Chlorophenyl)-1-methyl-N-6~-(3-methylphenyl)-1H-pyrazolo[3,4-d]pyrimidine-4,6-diamine
 L. N-[[[1-[3-(3,5-Dimethyl-1H-pyrazol-1-yl)benzyl]-3-piperidinyl]methyl]-4-fluorobenzamide

6.9 Appendix 9: Table 4-SAXS data

	KAP1 FL	Δ KAP1	RBCC
SASPDDB accession code	SASDEV6	SASDER7	SASDEW6
Data collection			
Instrument		ESRF BM29	
Beam size at sample (μm)		700x700	
Wavelength (\AA)		0.992	
q range (\AA^{-1})		0.0025-0.500	
Detector		Pilatus 1M	
Detector distance (m)		2.867	
Exposure (s per image)		1	
Column		Superose 6 Increase 10/300	
Flow rate (ml/min)		0.5	
Sample volume (μl)		100	
Sample concentration (mg/ml)	15	12	12
Temperature (K)		293	
Structural parameters			
R_g (\AA) Guinier	90	89	83
R_g (\AA) P(r)	95	95	89
R_{gc} (\AA)	35.8	38.8	20.2
D_{max} (\AA)	380	380	370
Porod volume (\AA^3)	710000	853000	381000
Molecular mass determination			
Theoretical MW (kDa)	183	175	92
MALLS MW (kDa)	190	183	88
DATPOROD MW (kDa)	444	533	238
DATVC MW (kDa)	255	290	117
DATMOW MW (kDa)	247	321	145
Data analysis software			
Data reduction		PRIMUS & ScÅtter	
<i>Ab initio</i> modelling		GASBOR	
Homology modelling		Swiss Model/I-tasser	
Computation of model fitting to data		Pepsi-SAXS	
3D graphics representation		Pymol	

6.10 References and Bibliography

1. Lee, T. I. & Young, R. A. Transcriptional Regulation and its Misregulation in Disease. *Cell* **152**, 1237 (2013).
2. Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C. & Gelbart, W. M. Transcription: an overview of gene regulation in eukaryotes. *Introd. Genet. Anal. 7th Ed.* (2000).
3. Roy, A. L. & Singer, D. S. Core Promoters in Transcription: Old Problem, New Insights. *Trends Biochem. Sci.* **40**, 165–171 (2015).
4. Li, J. & Gilmour, D. S. Promoter proximal pausing and the control of gene expression. *Curr. Opin. Genet. Dev.* **21**, 231–235 (2011).
5. Rao, A., States, D. J., Hero, A. O. & Engel, J. D. Understanding distal transcriptional regulation from sequence, expression and interactome perspectives. *J. Bioinform. Comput. Biol.* **8**, 219–246 (2010).
6. Martinez, E. Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.* **50**, 925–947 (2002).
7. Tang, L., Nogales, E. & Ciferri, C. Structure and Function of SWI/SNF Chromatin Remodeling Complexes and Mechanistic Implications for Transcription. *Prog. Biophys. Mol. Biol.* **102**, 122–128 (2010).
8. Li, M. *et al.* Dynamic regulation of transcription factors by nucleosome remodeling. *eLife* **4**,
9. Clarkson, M. J., Wells, J. R., Gibson, F., Saint, R. & Tremethick, D. J. Regions of variant histone His2AvD required for *Drosophila* development. *Nature* **399**, 694–697 (1999).
10. Bernstein, E. & Hake, S. B. The nucleosome: a little variation goes a long way. *Biochem. Cell Biol. Biochim. Biol. Cell.* **84**, 505–517 (2006).
11. Henikoff, S. & Ahmad, K. Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* **21**, 133–153 (2005).

-
12. Pidoux, A. L. & Allshire, R. C. The role of heterochromatin in centromere function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 569–579 (2005).
 13. Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nat. Struct. Mol. Biol.* **11**, 1037–1043 (2004).
 14. Garcia-Ramirez, M., Rocchini, C. & Ausio, J. Modulation of chromatin folding by histone acetylation. *J. Biol. Chem.* **270**, 17923–17928 (1995).
 15. Tse, C., Sera, T., Wolffe, A. P. & Hansen, J. C. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol. Cell. Biol.* **18**, 4629–4638 (1998).
 16. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847 (2006).
 17. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
 18. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574–1577 (2004).
 19. Zhao, T., Heyduk, T., Allis, C. D. & Eissenberg, J. C. Heterochromatin Protein 1 Binds to Nucleosomes and DNA in Vitro. *J. Biol. Chem.* **275**, 28332–28338 (2000).
 20. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
 21. ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
 22. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
 23. Lan, F. *et al.* A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449**, 689–694 (2007).

-
24. Ahn, S.-H. *et al.* Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in *S. cerevisiae*. *Cell* **120**, 25–36 (2005).
 25. Bode, A. M. & Dong, Z. Post-translational modification of p53 in tumorigenesis. *Nat. Rev. Cancer* **4**, 793–805 (2004).
 26. van Attikum, H. & Gasser, S. M. The histone code at DNA breaks: a guide to repair? *Nat. Rev. Mol. Cell Biol.* **6**, 757–765 (2005).
 27. Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Dev. Camb. Engl.* **144**, 2719–2729 (2017).
 28. Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biol.* **4**, 231 (2003).
 29. Witzgall, R., O’Leary, E., Leaf, A., Onaldi, D. & Bonventre, J. V. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4514–4518 (1994).
 30. Collins, T. & Sander, T. L. *The Superfamily of SCAN Domain Containing Zinc Finger Transcription Factors*. (Landes Bioscience, 2013).
 31. Garton, M. *et al.* A structural approach reveals how neighbouring C2H2 zinc fingers influence DNA binding specificity. *Nucleic Acids Res.* **43**, 9147–9157 (2015).
 32. Bellefroid, E. J. *et al.* Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *EMBO J.* **12**, 1363–1374 (1993).
 33. Quenneville, S. *et al.* The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep.* **2**, 766–773 (2012).
 34. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. 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 (2002).

-
35. Underhill, C., Qutob, M. S., Yee, S. P. & Torchia, J. A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *J. Biol. Chem.* **275**, 40463–40470 (2000).
36. Lechner, M. S., Begg, G. E., Speicher, D. W. & Rauscher, F. J. 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 (2000).
37. Schmitges, F. W. *et al.* Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.* **26**, 1742–1752 (2016).
38. Segal, D. J., Crotty, J. W., Bhakta, M. S., Barbas, C. F. & Horton, N. C. Structure of Aart, a Designed Six-finger Zinc Finger Peptide, Bound to DNA. *J. Mol. Biol.* **363**, 405–421 (2006).
39. Friedman, J. R. *et al.* KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**, 2067–2078 (1996).
40. Ozato, K., Shin, D.-M., Chang, T.-H. & Morse, H. C. TRIM family proteins and their emerging roles in innate immunity. *Nat. Rev. Immunol.* **8**, 849–860 (2008).
41. Short, K. M. & Cox, T. C. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J. Biol. Chem.* **281**, 8970–8980 (2006).
42. Peng, H. *et al.* Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J. Mol. Biol.* **295**, 1139–1162 (2000).
43. Moosmann, P., Georgiev, O., Le Douarin, B., Bourquin, J. P. & Schaffner, W. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res.* **24**, 4859–4867 (1996).
44. Okamoto, K., Kitabayashi, I. & Taya, Y. KAP1 dictates p53 response induced by chemotherapeutic agents via Mdm2 interaction. *Biochem. Biophys. Res. Commun.* **351**, 216–222 (2006).

-
45. Mita, P. *et al.* URI Regulates KAP1 Phosphorylation and Transcriptional Repression via PP2A Phosphatase in Prostate Cancer Cells. *J. Biol. Chem.* **291**, 25516–25528 (2016).
46. Cammas, F. *et al.* Cell differentiation induces TIF1beta association with centromeric heterochromatin via an HP1 interaction. *J. Cell Sci.* **115**, 3439–3448 (2002).
47. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23**, 837–848 (2009).
48. Ziv, Y. *et al.* 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 (2006).
49. Li, X. *et al.* SUMOylation of the transcriptional co-repressor KAP1 is regulated by the serine and threonine phosphatase PP1. *Sci. Signal.* **3**, ra32 (2010).
50. O’Geen, H. *et al.* Genome-Wide Analysis of KAP1 Binding Suggests Autoregulation of KRAB-ZNFs. *PLOS Genet.* **3**, e89 (2007).
51. White, D. E. *et al.* The ATM substrate KAP1 controls DNA repair in heterochromatin: Regulation by HP1 proteins and Serine 473/824 phosphorylation. *Mol. Cancer Res. MCR* **10**, 401–414 (2012).
52. Noon, A. T. *et al.* 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nat. Cell Biol.* **12**, 177–184 (2010).
53. Chang, C.-W. *et al.* Phosphorylation at Ser473 regulates heterochromatin protein 1 binding and corepressor function of TIF1beta/KAP1. *BMC Mol. Biol.* **9**, 61 (2008).
54. McNamara, R. P. *et al.* KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. *Mol. Cell* **61**, 39–53 (2016).
55. P-TEFb: Finding its ways to release promoter-proximally paused RNA polymerase II: Transcription: Vol 9, No 2. Available at:
<https://www.tandfonline.com/doi/full/10.1080/21541264.2017.1281864>. (Accessed: 13th August 2019)

-
56. McNamara, R. P., Guzman, C., Reeder, J. E. & D'Orso, I. Genome-wide analysis of KAP1, the 7SK snRNP complex, and RNA polymerase II. *Genomics Data* **7**, 250–255 (2016).
57. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **27**, 1147–1157 (2005).
58. Sun, Y. *et al.* 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 (2019).
59. Stoll, G. A. *et al.* Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. *Proc. Natl. Acad. Sci.* **116**, 15042–15051 (2019).
60. Lim, M. *et al.* A Ubiquitin-Binding Domain that Binds a Structural Fold Distinct from that of Ubiquitin. *Structure* (2019). doi:10.1016/j.str.2019.05.003
61. Venturini, L. *et al.* TIF1gamma, a novel member of the transcriptional intermediary factor 1 family. *Oncogene* **18**, 1209–1217 (1999).
62. Zeng, L. *et al.* Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat. Struct. Mol. Biol.* **15**, 626–633 (2008).
63. Smothers, J. F. & Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol. CB* **10**, 27–30 (2000).
64. Stevens, R. V., Esposito, D. & Rittinger, K. Characterisation of class VI TRIM RING domains: linking RING activity to C-terminal domain identity. *Life Sci. Alliance* **2**, e201900295 (2019).
65. Ragvin, A. *et al.* Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
66. Zhou, Y. & Grummt, I. The PHD Finger/Bromodomain of NoRC Interacts with Acetylated Histone H4K16 and Is Sufficient for rDNA Silencing. *Curr. Biol.* **15**, 1434–1438 (2005).
67. Neo, S. H. *et al.* TRIM28 Is an E3 Ligase for ARF-Mediated NPM1/B23 SUMOylation That Represses Centrosome Amplification. *Mol. Cell. Biol.* **35**, 2851–2863 (2015).

-
68. Ivanov, A. V. *et al.* PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* **28**, 823–837 (2007).
69. Li, X. *et al.* Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. *J. Biol. Chem.* **282**, 36177–36189 (2007).
70. Hu, C. *et al.* Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response. *J. Biol. Chem.* **287**, 18937–18952 (2012).
71. Cheng, C.-T., Kuo, C.-Y. & Ann, D. K. KAPtain in charge of multiple missions: Emerging roles of KAP1. *World J. Biol. Chem.* **5**, 308–320 (2014).
72. Lai, I.-L. *et al.* Histone deacetylase 10 relieves repression on the melanogenic program by maintaining the deacetylation status of repressors. *J. Biol. Chem.* **285**, 7187–7196 (2010).
73. Eissenberg, J. C. & Elgin, S. C. R. HP1a: a structural chromosomal protein regulating transcription. *Trends Genet. TIG* **30**, 103–110 (2014).
74. Nishibuchi, G. & Nakayama, J. Biochemical and structural properties of heterochromatin protein 1: understanding its role in chromatin assembly. *J. Biochem. (Tokyo)* **156**, 11–20 (2014).
75. Thiru, A. *et al.* Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J.* **23**, 489–499 (2004).
76. Machida, S. *et al.* Structural Basis of Heterochromatin Formation by Human HP1. *Mol. Cell* **69**, 385–397.e8 (2018).
77. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal Structures of Nucleosome Core Particles Containing the ‘601’ Strong Positioning Sequence. *J. Mol. Biol.* **403**, 1–10 (2010).
78. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4405151/>. (Accessed: 7th August 2019)
79. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).

-
80. Kumar, D., Sharma, N. & Giri, R. Therapeutic Interventions of Cancers Using Intrinsically Disordered Proteins as Drug Targets: c-Myc as Model System. *Cancer Inform.* **16**, (2017).
81. Korsak, M. & Kozyreva, T. Beta Amyloid Hallmarks: From Intrinsically Disordered Proteins to Alzheimer's Disease. *Adv. Exp. Med. Biol.* **870**, 401–421 (2015).
82. Tamò, G. E., Abriata, L. A. & Dal Peraro, M. The importance of dynamics in integrative modeling of supramolecular assemblies. *Curr. Opin. Struct. Biol.* **31**, 28–34 (2015).
83. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019).
84. Belsom, A. *et al.* Blind testing cross-linking/mass spectrometry under the auspices of the 11 th critical assessment of methods of protein structure prediction (CASP11). *Wellcome Open Res.* **1**, (2016).
85. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/22106262>. (Accessed: 13th August 2019)
86. Taylor, N. M. I. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
87. Liu, F., Zhang, Z., Csanády, L., Gadsby, D. C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169**, 85-95.e8 (2017).
88. Lander, G. C. *et al.* Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**, 186–191 (2012).
89. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/25760083>. (Accessed: 13th August 2019)

-
90. 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/25953817>. (Accessed: 13th August 2019)
91. Li, K. *et al.* Sub-3 Å apoferritin structure determined with full range of phase shifts using a single position of volta phase plate. *J. Struct. Biol.* **206**, 225–232 (2019).
92. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
93. Tsutakawa, S. E. *et al.* Solution X-ray scattering combined with computational modeling reveals multiple conformations of covalently bound ubiquitin on PCNA. *Proc. Natl. Acad. Sci.* **108**, 17672–17677 (2011).
94. Querol-Audí, J. *et al.* Repair complexes of FEN1 endonuclease, DNA, and Rad9-Hus1-Rad1 are distinguished from their PCNA counterparts by functionally important stability. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8528–8533 (2012).
95. Cerqueira, N. M. F. S. A. *et al.* Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* **582**, 56–67 (2015).
96. Chen, Y. *et al.* 2-(4-Chlorophenyl)-2-oxoethyl 4-benzamidobenzoate derivatives, a novel class of SENP1 inhibitors: Virtual screening, synthesis and biological evaluation. *Bioorg. Med. Chem. Lett.* **22**, 6867–6870 (2012).
97. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
98. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–1107 (2012).
99. ZINC--a free database of commercially available compounds for virtual screening. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/15667143>. (Accessed: 14th August 2019)

-
100. ChemSpider: An Online Chemical Information Resource | Journal of Chemical Education.
Available at: <https://pubs.acs.org/doi/10.1021/ed100697w>. (Accessed: 14th August 2019)
101. Hernández-Santoyo, A., Tenorio-Barajas, A. Y., Altuzar, V., Vivanco-Cid, H. & Mendoza-Barrera, C. Protein-Protein and Protein-Ligand Docking. *Protein Eng. - Technol. Appl.* (2013).
doi:10.5772/56376
102. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
103. Czerwińska, P., Mazurek, S. & Wiznerowicz, M. The complexity of TRIM28 contribution to cancer. *J. Biomed. Sci.* **24**, (2017).
104. Iyengar, S. & Farnham, P. J. KAP1 Protein: An Enigmatic Master Regulator of the Genome. *J. Biol. Chem.* **286**, 26267–26276 (2011).
105. Zhou, Y. & Grummt, I. The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr. Biol. CB* **15**, 1434–1438 (2005).
106. da Motta, L. L. *et al.* The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. *Oncogene* **36**, 122–132 (2017).
107. Interfering with bromodomain epigenome readers as therapeutic option in mucoepidermoid carcinoma. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/30539410>.
(Accessed: 17th June 2019)
108. Andersson, M., Wittgren, B. & Wahlund, K.-G. Accuracy in Multiangle Light Scattering Measurements for Molar Mass and Radius Estimations. Model Calculations and Experiments. *Anal. Chem.* **75**, 4279–4291 (2003).
109. Kwan, T. O. C. *et al.* Selection of Biophysical Methods for Characterisation of Membrane Proteins. *Int. J. Mol. Sci.* **20**, 2605 (2019).

-
110. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).
111. Corrêa, D. & Ramos, C. The use of circular dichroism spectroscopy to study protein folding, form and function. *Afr. J. Biochem. Res.* **3**, 164–173 (2009).
112. Cole, J. L., Lary, J. W., Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **84**, 143–179 (2008).
113. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
114. Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci. Publ. Protein Soc.* **11**, 2067–2079 (2002).
115. RELION: implementation of a Bayesian approach to cryo-EM structure determination. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/23000701>. (Accessed: 17th June 2019)
116. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
117. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
118. Nyman, M. & McQuade, L. Small Angle X-ray Scattering of Group V Polyoxometalates. in 151–170 (2015).
119. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
120. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).

-
121. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
122. Incardona, M.-F. *et al.* EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).
123. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
124. Rambo, R. P. Resolving Individual Components in Protein-RNA Complexes Using Small-Angle X-ray Scattering Experiments. *Methods Enzymol.* **558**, 363–390 (2015).
125. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinforma. Oxf. Engl.* **34**, 1944–1946 (2018).
126. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
127. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
128. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).
129. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
130. Gorba, C. & Tama, F. Normal Mode Flexible Fitting of High-Resolution Structures of Biological Molecules Toward SAXS Data. *Bioinforma. Biol. Insights* **4**, 43–54 (2010).
131. GHYSELS, A. *et al.* Comparative Study of Various Normal Mode Analysis Techniques Based on Partial Hessians. *J. Comput. Chem.* **31**, 994–1007 (2010).

-
132. Panjkovich, A. & Svergun, D. I. Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys. Chem. Chem. Phys. PCCP* **18**, 5707–5719 (2016).
133. Hoffmann, A. & Grudinin, S. NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method. *J. Chem. Theory Comput.* **13**, 2123–2134 (2017).
134. Grudinin, S., Garkavenko, M. & Kazennov, A. Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. Sect. Struct. Biol.* **73**, 449–464 (2017).
135. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
136. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–258 (2014).
137. Huang, A. *et al.* Symmetry and Asymmetry of the RING–RING Dimer of Rad18. *J. Mol. Biol.* **410**, 424–435 (2011).
138. Mrosek, M. *et al.* Structural Analysis of B-Box 2 from MuRF1: Identification of a Novel Self-Association Pattern in a RING-like Fold. *Biochemistry* **47**, 10722–10730 (2008).
139. Li, Y. *et al.* Structural insights into the TRIM family of ubiquitin E3 ligases. *Cell Res.* **24**, 762–765 (2014).
140. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
141. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
142. Huang, Y., Myers, M. P. & Xu, R.-M. Crystal structure of the HP1-EMSY complex reveals an unusual mode of HP1 binding. *Struct. Lond. Engl.* **1993** **14**, 703–712 (2006).

-
143. Greber, B. J. *et al.* The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283–286 (2014).
144. Leaver-Fay, A. *et al.* Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
145. Comparative Protein Structure Modeling Using Modeller. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4186674/>. (Accessed: 25th July 2018)
146. Kilic, S., Bachmann, A. L., Bryan, L. C. & Fierz, B. Multivalency governs HP1 α association dynamics with the silent chromatin state. *Nat. Commun.* **6**, 7313 (2015).
147. Vila-Perelló, M. *et al.* Streamlined expressed protein ligation using split inteins. *J. Am. Chem. Soc.* **135**, 286–292 (2013).
148. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
149. Chambers, J. *et al.* UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J. Cheminformatics* **5**, 3 (2013).
150. Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N. & Sastry, G. N. Virtual screening in drug discovery -- a computational perspective. *Curr. Protein Pept. Sci.* **8**, 329–351 (2007).
151. Forli, S. *et al.* Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).
152. Allen, W. J. *et al.* DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **36**, 1132–1156 (2015).
153. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
154. Taylor, J. S. & Burnett, R. M. DARWIN: a program for docking flexible molecules. *Proteins* **41**, 173–191 (2000).

-
155. Zavodszky, M. I. & Kuhn, L. A. Side-chain flexibility in protein–ligand binding: The minimal rotation hypothesis. *Protein Sci. Publ. Protein Soc.* **14**, 1104–1114 (2005).
156. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **51**, 578–596 (2011).
157. Cross, S. S. J. Improved FlexX Docking Using FlexS-Determined Base Fragment Placement. *J. Chem. Inf. Model.* **45**, 993–1001 (2005).
158. Welch, W., Ruppert, J. & Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **3**, 449–462 (1996).
159. Eisen, M. B., Wiley, D. C., Karplus, M. & Hubbard, R. E. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* **19**, 199–221 (1994).
160. Hamza, A., Wei, N.-N. & Zhan, C.-G. Ligand-Based Virtual Screening Approach Using a New Scoring Function. *J. Chem. Inf. Model.* **52**, 963–974 (2012).
161. Bamborough, P. *et al.* Aiming to Miss a Moving Target: Bromo and Extra Terminal Domain (BET) Selectivity in Constrained ATAD2 Inhibitors. *J. Med. Chem.* **61**, 8321–8336 (2018).
162. Sperandio, D. *et al.* Structure-guided discovery of a novel, potent, and orally bioavailable 3,5-dimethylisoxazole aryl-benzimidazole BET bromodomain inhibitor. *Bioorg. Med. Chem.* **27**, 457–469 (2019).
163. Liu, J. *et al.* The polar warhead of a TRIM24 bromodomain inhibitor rearranges a water-mediated interaction network. *FEBS J.* **284**, 1082–1095 (2017).
164. Millan, D. S. *et al.* Design and Optimization of Benzopiperazines as Potent Inhibitors of BET Bromodomains. *ACS Med. Chem. Lett.* **8**, 847–852 (2017).
165. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

-
166. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
167. Scaffold Hunter: a comprehensive visual analytics framework for drug discovery. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5425364/>. (Accessed: 17th June 2019)
168. Meslamani, J., Smith, S. G., Sanchez, R. & Zhou, M.-M. Structural Features and Inhibitors of Bromodomains. *Drug Discov. Today Technol.* **19**, 3–15 (2016).
169. Discovery of Chemical Inhibitors of Human Bromodomains | Chemical Reviews. Available at: <https://pubs.acs.org/doi/full/10.1021/acs.chemrev.5b00205?src=recsys>. (Accessed: 17th June 2019)
170. Sanchez, R., Meslamani, J. & Zhou, M.-M. The Bromodomain: From Epigenome Reader to Druggable Target. *Biochim. Biophys. Acta* **1839**, 676–685 (2014).
171. Kerscher, B. *et al.* BET Bromodomain Inhibitor iBET151 Impedes Human ILC2 Activation and Prevents Experimental Allergic Lung Inflammation. *Front. Immunol.* **10**, (2019).
172. Pérez-Salvia, M. & Esteller, M. Bromodomain inhibitors and cancer therapy: From structures to applications. *Epigenetics* **12**, 323–339 (2016).
173. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3856850/>. (Accessed: 17th June 2019)
174. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).
175. Cammas, F., Herzog, M., Lerouge, T., Chambon, P. & Losson, R. Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev.* **18**, 2147–2160 (2004).
176. Groner, A. C. *et al.* KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.* **6**, e1000869 (2010).

-
177. Li, X. *et al.* Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. *J. Biol. Chem.* **282**, 36177–36189 (2007).
178. Bunch, H. *et al.* TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. *Nat. Struct. Mol. Biol.* **21**, 876–883 (2014).
179. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23**, 837–848 (2009).
180. Iyengar, S., Ivanov, A. V., Jin, V. X., Rauscher, F. J. & Farnham, P. J. Functional Analysis of KAP1 Genomic Recruitment. *Mol. Cell. Biol.* **31**, 1833–1847 (2011).
181. Goodarzi, A. A., Kurka, T. & Jeggo, P. A. KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nat. Struct. Mol. Biol.* **18**, 831–839 (2011).
182. Koliopoulos, M. G. *et al.* Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. *Nat. Commun.* **9**, 1820 (2018).
183. Sanchez, J. G. *et al.* The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2494–2499 (2014).
184. Goldstone, D. C. *et al.* Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9609–9614 (2014).
185. Wagner, J. M. *et al.* Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5 α . *eLife* **5**, e16309 (2016).
186. Weinert, C., Morger, D., Djekic, A., Grütter, M. G. & Mittl, P. R. E. Crystal structure of TRIM20 C-terminal coiled-coil/B30.2 fragment: implications for the recognition of higher order oligomers. *Sci. Rep.* **5**, 10819 (2015).

-
187. Koliopoulos, M. G., Esposito, D., Christodoulou, E., Taylor, I. A. & Rittinger, K. Functional role of TRIM E3 ligase oligomerization and regulation of catalytic activity. *EMBO J.* **35**, 1204–1218 (2016).
188. Herquel, B. *et al.* 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 (2011).
189. Gumpena, R., Lountos, G. T. & Waugh, D. S. MBP-binding DARPins facilitate the crystallization of an MBP fusion protein. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74**, 549–557 (2018).
190. Batyuk, A., Wu, Y., Honegger, A., Heberling, M. M. & Plückthun, A. DARPIn-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. *J. Mol. Biol.* **428**, 1574–1588 (2016).
191. Plückthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
192. Volkov, V. V., Svergun, D. I. & IUCr. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* (2003). doi:10.1107/S0021889803000268
193. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
194. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
195. Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5 α | eLife. Available at: <https://elifesciences.org/articles/16309>. (Accessed: 24th July 2019)

-
196. Ortega, A., Amorós, D. & García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **101**, 892–898 (2011).
197. Schultz, D. C., Friedman, J. R. & Rauscher, F. J. 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-2 α subunit of NuRD. *Genes Dev.* **15**, 428–443 (2001).
198. Liang, Q. *et al.* TRIM28 is a SUMO E3 Ligase and Negative Regulator of Interferon Regulatory Factor 7. *J. Immunol. Baltim. Md 1950* **187**, 4754–4763 (2011).
199. Depleting Trim28 in adult mice is well tolerated and reduces levels of α -synuclein and tau | eLife. Available at: <https://elifesciences.org/articles/36768>. (Accessed: 2nd July 2019)
200. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* **8**, 1006–1016 (2007).
201. Shen, T. H., Lin, H.-K., Scaglioni, P. P., Yung, T. M. & Pandolfi, P. P. The mechanisms of PML-nuclear body formation. *Mol. Cell* **24**, 331–339 (2006).
202. Ryan, R. F. *et al.* 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 (1999).
203. Sripathy, S. P., Stevens, J. & Schultz, D. C. 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 (2006).
204. Bryan, L. C. *et al.* Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. *Nucleic Acids Res.* **45**, 10504–10517 (2017). - Cerca con Google. Available at: <https://www.google.com/search?q=Bryan%2C+L.+C.+et+al.+Single->

-
- molecule+kinetic+analysis+of+HP1-
chromatin+binding+reveals+a+dynamic+network+of+histone+modification+and+DNA+interact
ions.+Nucleic+Acids+Res.+45%2C+10504%E2%80%9310517+(2017).&aq=Bryan%2C+L.+C.
+et+al.+Single-molecule+kinetic+analysis+of+HP1-
chromatin+binding+reveals+a+dynamic+network+of+histone+modification+and+DNA+interact
ions.+Nucleic+Acids+Res.+45%2C+10504%E2%80%9310517+(2017).&aq=chrome..69i57.12
54j0j7&sourceid=chrome&ie=UTF-8. (Accessed: 5th July 2019)
205. Kilic, S., Bachmann, A. L., Bryan, L. C. & Fierz, B. Multivalency governs HP1 α association dynamics with the silent chromatin state. *Nat. Commun.* **6**, 7313 (2015).
 206. Koliopoulos, M. G. *et al.* Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. *Nat. Commun.* **9**, 1820 (2018).
 207. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
 208. D’Cruz, A. A. *et al.* Identification of a second binding site on the TRIM25 B30.2 domain. *Biochem. J.* **475**, 429–440 (2018).
 209. Koliopoulos, M. G., Esposito, D., Christodoulou, E., Taylor, I. A. & Rittinger, K. Functional role of TRIM E3 ligase oligomerization and regulation of catalytic activity. *EMBO J.* **35**, 1204–1218 (2016).
 210. Wagner, J. M. *et al.* Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5 α . *eLife* **5**, (2016).
 211. Stevens, R. V., Esposito, D. & Rittinger, K. Characterisation of class VI TRIM RING domains: linking RING activity to C-terminal domain identity. *Life Sci. Alliance* **2**, (2019).
 212. Liang, Q. *et al.* Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J. Immunol. Baltim. Md 1950* **187**, 4754–4763 (2011).

-
213. Rousseaux, M. W. *et al.* Depleting Trim28 in adult mice is well tolerated and reduces levels of α -synuclein and tau. *eLife* **7**, (2018).
214. Ivanov, A. V. *et al.* PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* **28**, 823–837 (2007).
215. Larson, A. G. *et al.* Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
216. Lechner, M. S., Begg, G. E., Speicher, D. W. & Rauscher, F. J. 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 (2000).
217. Yang, B. *et al.* MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines. *Cancer Res.* **67**, 9954–9962 (2007).
218. Celic, I. *et al.* The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation. *Curr. Biol. CB* **16**, 1280–1289 (2006).
219. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
220. Kouzarides, T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* **19**, 1176–1179 (2000).
221. Gray, S. G. & Teh, B. T. Histone acetylation/deacetylation and cancer: an ‘open’ and ‘shut’ case? *Curr. Mol. Med.* **1**, 401–429 (2001).
222. Li, Y. & Seto, E. HDACs and HDAC Inhibitors in Cancer Development and Therapy. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
223. Gryder, B. E., Sodji, Q. H. & Oyelere, A. K. Targeted cancer therapy: giving histone deacetylase inhibitors all they need to succeed. *Future Med. Chem.* **4**, 505–524 (2012).

-
224. Suraweera, A., O'Byrne, K. J. & Richard, D. J. Combination Therapy With Histone Deacetylase Inhibitors (HDACi) for the Treatment of Cancer: Achieving the Full Therapeutic Potential of HDACi. *Front. Oncol.* **8**, (2018).
225. Shu, S. & Polyak, K. BET Bromodomain Proteins as Cancer Therapeutic Targets. *Cold Spring Harb. Symp. Quant. Biol.* **81**, 123–129 (2016).
226. Klein, K. Bromodomain protein inhibition: a novel therapeutic strategy in rheumatic diseases. *RMD Open* **4**, e000744 (2018).
227. Boehm, D., Conrad, R. J. & Ott, M. Bromodomain Proteins in HIV Infection. *Viruses* **5**, 1571–1586 (2013).
228. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov.* **13**, 337–356 (2014).
229. Josling, G. A., Selvarajah, S. A., Petter, M. & Duffy, M. F. The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes* **3**, 320–343 (2012).
230. Zou, Z. *et al.* Brd4 maintains constitutively active NF- κ B in cancer cells by binding to acetylated RelA. *Oncogene* **33**, 2395–2404 (2014).
231. Taniguchi, Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *Int. J. Mol. Sci.* **17**, (2016).
232. Jackson, M. R., Loll, P. J. & Jorns, M. S. X-Ray Structure of Human Sulfide:Quinone Oxidoreductase: Insights into the Mechanism of Mitochondrial Hydrogen Sulfide Oxidation. *Structure* **27**, 794-805.e4 (2019).
233. Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J.* **19**, 6141–6149 (2000).
234. Zeng, L., Zhang, Q., Gerona-Navarro, G., Moshkina, N. & Zhou, M.-M. Structural Basis of Site-Specific Histone Recognition by the Bromodomains of Human Coactivators PCAF and CBP/p300. *Structure* **16**, 643–652 (2008).

-
235. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
236. Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149**, 214–231 (2012).
237. Yang, Y. *et al.* Acetylated hsp70 and KAP1-mediated Vps34 SUMOylation is required for autophagosome creation in autophagy. *Proc. Natl. Acad. Sci.* **110**, 6841–6846 (2013).
238. Sripathy, S. P., Stevens, J. & Schultz, D. C. 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 (2006).
239. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
240. Borah, J. C. *et al.* A Small Molecule Binding to the Coactivator CREB-Binding Protein Blocks Apoptosis in Cardiomyocytes. *Chem. Biol.* **18**, 531–541 (2011).
241. Mertz, J. A. *et al.* Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc. Natl. Acad. Sci.* **108**, 16669–16674 (2011).
242. Tyler, D. S. *et al.* Click chemistry enables preclinical evaluation of targeted epigenetic therapies. *Science* **356**, 1397–1401 (2017).
243. Zhao, L., Okhovat, J.-P., Hong, E. K., Kim, Y. H. & Wood, G. S. Preclinical Studies Support Combined Inhibition of BET Family Proteins and Histone Deacetylases as Epigenetic Therapy for Cutaneous T-Cell Lymphoma. *Neoplasia* **21**, 82–92 (2019).
244. Zhang, G. *et al.* Down-regulation of NF- κ B Transcriptional Activity in HIV-associated Kidney Disease by BRD4 Inhibition. *J. Biol. Chem.* **287**, 28840–28851 (2012).
245. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).

-
246. Chung, C.-W. *et al.* Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J. Med. Chem.* **54**, 3827–3838 (2011).
247. Mirguet, O. *et al.* From ApoA1 upregulation to BET family bromodomain inhibition: discovery of I-BET151. *Bioorg. Med. Chem. Lett.* **22**, 2963–2967 (2012).
248. Gehling, V. S. *et al.* Discovery, Design, and Optimization of Isoxazole Azepine BET Inhibitors. *ACS Med. Chem. Lett.* **4**, 835–840 (2013).
249. Dawson, M. A. *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529–533 (2011).
250. Seal, J. *et al.* Identification of a novel series of BET family bromodomain inhibitors: binding mode and profile of I-BET151 (GSK1210151A). *Bioorg. Med. Chem. Lett.* **22**, 2968–2972 (2012).
251. Aird, F., Kandela, I., Mantis, C. & Reproducibility Project: Cancer Biology. Replication Study: BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *eLife* **6**, (2017).
252. Andrieu, G., Belkina, A. C. & Denis, G. V. Clinical trials for BET inhibitors run ahead of the science. *Drug Discov. Today Technol.* **19**, 45–50 (2016).
253. Chaidos, A., Caputo, V. & Karadimitris, A. Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: emerging preclinical and clinical evidence. *Ther. Adv. Hematol.* **6**, 128–141 (2015).
254. Long, J. *et al.* The BET Bromodomain Inhibitor I-BET151 Acts Downstream of Smoothened Protein to Abrogate the Growth of Hedgehog Protein-driven Cancers. *J. Biol. Chem.* **289**, 35494–35502 (2014).
255. Potent antimyeloma activity of the novel bromodomain inhibitors I-BET151 and I-BET762 | Blood Journal. Available at: <http://www.bloodjournal.org/content/123/5/697?sso-checked=true>. (Accessed: 10th July 2019)

-
256. Hewings, D. S. *et al.* 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J. Med. Chem.* **54**, 6761–6770 (2011).
257. Taylor, A. M. *et al.* Discovery of Benzotriazolo[4,3-d][1,4]diazepines as Orally Active Inhibitors of BET Bromodomains. *ACS Med. Chem. Lett.* **7**, 145–150 (2016).
258. Liu, Z. *et al.* Drug Discovery Targeting Bromodomain-Containing Protein 4. *J. Med. Chem.* **60**, 4533–4558 (2017).
259. Vidler, L. R. *et al.* Discovery of Novel Small-Molecule Inhibitors of BRD4 Using Structure-Based Virtual Screening. *J. Med. Chem.* **56**, 8073–8088 (2013).
260. Chan, K.-H., Zengerle, M., Testa, A. & Ciulli, A. Impact of Target Warhead and Linkage Vector on Inducing Protein Degradation: Comparison of Bromodomain and Extra-Terminal (BET) Degraders Derived from Triazolodiazepine (JQ1) and Tetrahydroquinoline (I-BET726) BET Inhibitor Scaffolds. *J. Med. Chem.* **61**, 504–513 (2018).
261. Fish, P. V. *et al.* Identification of a Chemical Probe for Bromo and Extra C-Terminal Bromodomain Inhibition through Optimization of a Fragment-Derived Hit. *J. Med. Chem.* **55**, 9831–9837 (2012).
262. Picaud, S. *et al.* PFI-1, a highly selective protein interaction inhibitor, targeting BET Bromodomains. *Cancer Res.* **73**, 3336–3346 (2013).
263. Fu, L. *et al.* Inhibition of BET bromodomains as a therapeutic strategy for cancer drug discovery. *Oncotarget* **6**, 5501–5516 (2015).
264. Picaud, S. *et al.* RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci.* **110**, 19754–19759 (2013).
265. Fedorov, O. *et al.* [1,2,4]Triazolo[4,3-a]phthalazines: Inhibitors of Diverse Bromodomains. *J. Med. Chem.* **57**, 462–476 (2014).

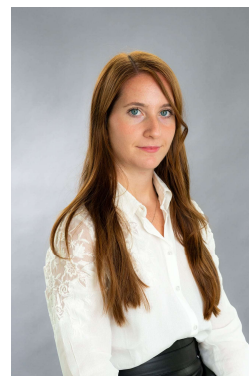
-
266. Pyo, J.-O. *et al.* Protection of Cardiomyocytes from Ischemic/Hypoxic Cell Death via Drbp1 and pMe2GlyDH in Cardio-specific ARC Transgenic Mice. *J. Biol. Chem.* **283**, 30707–30714 (2008).
267. Zhang, G. *et al.* Structure-guided design of potent diazobenzene inhibitors for the BET bromodomains. *J. Med. Chem.* **56**, 9251–9264 (2013).
268. Murphy, T. *et al.* Knockdown of epigenetic transcriptional co-regulator Brd2a disrupts apoptosis and proper formation of hindbrain and midbrain-hindbrain boundary (MHB) region in zebrafish. *Mech. Dev.* **146**, 10–30 (2017).
269. Ito, T. *et al.* Real-Time Imaging of Histone H4K12-Specific Acetylation Determines the Modes of Action of Histone Deacetylase and Bromodomain Inhibitors. *Chem. Biol.* **18**, 495–507 (2011).
270. Zeng, L. *et al.* Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* **127**, 2376–2377 (2005).
271. Sachchidanand *et al.* Target Structure-Based Discovery of Small Molecules that Block Human p53 and CREB Binding Protein Association. *Chem. Biol.* **13**, 81–90 (2006).
272. Picaud, S. *et al.* Promiscuous targeting of bromodomains by bromosporine identifies BET proteins as master regulators of primary transcription response in leukemia. *Sci. Adv.* **2**, e1600760 (2016).
273. Dou, Y. *et al.* Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873–885 (2005).
274. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31**, 35–40 (2006).
275. Identification and Characterization of a Leukocyte-specific Component of the Nuclear Body. Available at: <http://www.jbc.org/content/271/46/29198.full>. (Accessed: 12th July 2019)

-
276. Capili, A. D., Schultz, D. C., Rauscher, F. J. & Borden, K. L. B. Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *EMBO J.* **20**, 165–177 (2001).
277. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91 (2006).
278. Chaikuad, A. *et al.* Structure-Based Identification of Inhibitory Fragments Targeting the p300/CBP-Associated Factor Bromodomain. *J. Med. Chem.* **59**, 1648–1653 (2016).
279. Srinivasan, B., Zhou, H., Kubanek, J. & Skolnick, J. Experimental validation of FINDSITEcomb virtual ligand screening results for eight proteins yields novel nanomolar and micromolar binders. *J. Cheminformatics* **6**, 16 (2014).
280. Sakaguchi, T. *et al.* Bromodomain protein BRD4 inhibitor JQ1 regulates potential prognostic molecules in advanced renal cell carcinoma. *Oncotarget* **9**, 23003–23017 (2018).
281. Atkovska, K., A Samsonov, S., Paszkowski-Rogacz, M. & Pisabarro, M. T. Multipose Binding in Molecular Docking. *Int. J. Mol. Sci.* **15**, 2622–45 (2014).
282. Blum, A., Böttcher, J., Heine, A., Klebe, G. & Diederich, W. E. Structure-Guided Design of C2-Symmetric HIV-1 Protease Inhibitors Based on a Pyrrolidine Scaffold. *J. Med. Chem.* **51**, 2078–2087 (2008).
283. Gushchina, L. V., Gabdulkhakov, A. G., Nikonov, S. V. & Filimonov, V. V. High-resolution crystal structure of spectrin SH3 domain fused with a proline-rich peptide. *J. Biomol. Struct. Dyn.* **29**, 485–495 (2011).
284. Tie, C. H. *et al.* KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Rep.* **19**, (2018).
285. Tanaka, S. *et al.* KAP1 Regulates Regulatory T Cell Function and Proliferation in Both Foxp3-Dependent and -Independent Manners. *Cell Rep.* **23**, 796–807 (2018).

-
286. Kamitani, S. *et al.* KAP1 regulates type I interferon/STAT1-mediated IRF-1 gene expression. *Biochem. Biophys. Res. Commun.* **370**, 366–370 (2008).
287. Addison, J. B. *et al.* KAP1 Promotes Proliferation and Metastatic Progression of Breast Cancer Cells. *Cancer Res.* **75**, 344–355 (2015).
288. Sio, F. R. S. de *et al.* KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood* **119**, 4675–4685 (2012).
289. Wang, C. *et al.* MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation. *EMBO J.* **24**, 3279–3290 (2005).
290. RING dimerization links higher-order assembly of TRIM5 α to synthesis of K63-linked polyubiquitin. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4526822/>. (Accessed: 17th August 2019)
291. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **19**, 876–883 (2012).
292. Kubota, S. *et al.* 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 1 α (HP1 α) with heterochromatin. *J. Biol. Chem.* **288**, 17871–17883 (2013).

Giulia Fonti

Route de Gèneve 103
1026 Denges VD
+41 797690260
giulia.fonti@epfl.ch



18.02.1989
Italian

Professional Summary

I have five years of experience in the application of structural biology techniques. I am able to design and manage research projects of high complexity. While I am able to work autonomously, I thrive in a dynamic and cooperative environment.

Education

Ph.D. – Swiss Federal Institute of Technology (EPFL)	Lausanne	03.2015 - 09.2019
Internship - Swiss Federal Institute of Technology (EPFL)	Lausanne	08.2014 - 02.2015
Master of Science - University of Ferrara, Italy	Ferrara	12.2011 - 07.2014
Bachelor of Science – University of Ferrara, Italy	Ferrara	09.2008 - 11.2011

Technical Expertise

Protein expression and purification:	<i>E. Coli</i> expression system, HPLC chromatography, AKTA purifier
Protein Structural Characterization:	X-ray crystallography, SAXS
Drug discovery:	Autodock Vina, Maestro, DSF, qPCR
Electron Microscopy:	Negative staining and Cryo-EM, use of Tecnai Spirit microscope
Protein Biophysical Characterization:	MALS, AUC, ITC, CD
Molecular Biology Techniques:	Cloning, PCR, Western blot, Mutagenesis, SDS-page
In Silico Techniques:	Homology Modeling, Protein/small ligand Docking
Protein Visualization tools:	VMD, Chimera, Pymol, Coot

IT Skills

Microsoft Excel, Microsoft Word, Microsoft Power Point, Keynote, Adobe illustrator, Prism

Publications

“KAP1 is an antiparallel dimer with a functional asymmetry	<i>Life Science Alliance</i>	08.2019
Assessment of data – assisted prediction by inclusion of crosslinking/mass spectrometry and small angle X – ray scattering data in the 12 th Critical		
Assessment of protein Structure Prediction experiment.	<i>Proteins</i>	03.2018
A rheostat mechanism governs the bifurcation of carbon flux in mycobacteria.	<i>Nat. Commun.</i>	08.2016

Languages

Italian: Mother tongue
English: Proficiency (C1 level, everyday working use)
French: Basic (A2 level)

