CHARACTERIZATION OF THE PATZ1 BTB DOMAIN
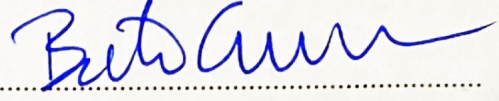
by
SOFIA PIEPOLI

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
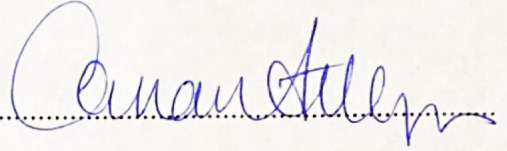
Sabancı University
August 2020

CHARACTERIZATION OF THE PATZ1 BTB DOMAIN
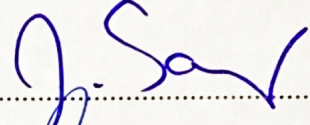
APPROVED BY:

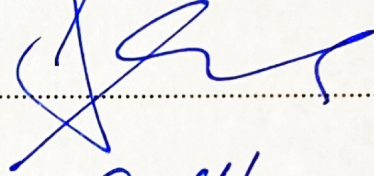Prof. Dr. Batu ERMAN
(Dissertation Supervisor)

Prof. Dr. Canan ATILGAN
(Dissertation Co-Supervisor)
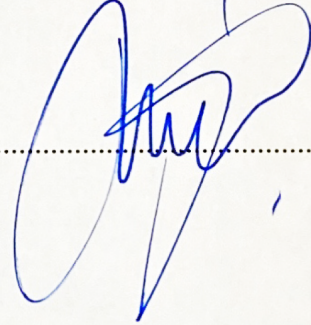
Prof. Dr. Zehra SAYERS

Asst. Prof. Dr. Stefan H. FUSS

Asst. Prof. Dr. Ogün ADEBALİ

Asst. Prof. Dr. Hasan DEMİRCİ

DATE OF APPROVAL: 28/08/2020

ABSTRACT

CHARACTERIZATION OF THE PATZ1 BTB DOMAIN

Sofia Piepoli

Molecular Biology, Genetics and Bioengineering, Ph.D. Thesis, 2020

Thesis Supervisor: Batu Erman

Thesis Co-supervisor: Canan Atılgan

Keywords: BTB domain, PATZ1 transcription factor, Protein Crystallization, Dimerization, Molecular Dynamics

The BTB domain is a conserved, multi-functional protein interaction domain engaged in the coordination of various large protein complexes, including ion channels, degradation signaling complexes and DNA transcription factors. The ZBTB protein family of transcription regulators is characterized by an N-terminal BTB domain followed by a variable number of Zinc Fingers and it is involved in cellular proliferation, immune cell lineage differentiation, and cancer development. Among the 49 human ZBTB proteins, PATZ1 stands out with a unique extra 31-amino acids long sequence in the center of the BTB domain. The central loop of PATZ1 is an intrinsically disordered region (IDR) conserved in mammals but absent in fish and has high binding potential. PATZ1 is a ubiquitous transcription factor that plays a role in numerous cancer types and interferes with the p53 tumor suppressor. In this thesis work, the determination of the unique structural features of the PATZ1 BTB domain was primarily accomplished with the solution of its crystal structure as a homodimer in both mouse and zebrafish (PDB IDs: 6GUV and 6GUW). Although similar in the 3D fold and the dimerization interface to the BCL6 BTB domain, the new crystal structures suggest that the PATZ1 co-repressors binding site contains a divergent sequence. The molecular dynamics of the new protein structures compared to other previously known structures allowed us to classify ZBTBs family members. Finally, the formation of stable BTB heterodimers with selected ZBTB proteins was computationally demonstrated.

ÖZET

PATZ1 BTB ALANININ KARAKTERIZASYONU

Sofia Piepoli

Moleküler Biyoloji, Genetik ve Biyomühendislik, Doktora Tezi, 2020

Tez Danışmanı: Batu Erman

İkinci Tez Danışman: Canan Atılgan

Anahtar Kelimeler: BTB alanı, PATZ1 transkripsiyon faktörü, Protein kristalizasyon, Dimerizasyon, Moleküler Dinamikler

BTB bölgesi, iyon kanalları, degradasyon sinyal kompleksleri ve DNA transkripsiyon sistemi dahil olmak üzere çeşitli büyük protein komplekslerinin koordinasyonunda görev alan, çok fonksiyonlu ve korunmuş bir protein etkileşim alanıdır. ZBTB proteinleri bir N-terminal BTB alanına ve ardından değişken sayıda Çinko Parmaklarına sahiptir ve transkripsiyon faktörü olarak çalışır. Hücresel proliferasyon, immün hücre soy farklılaşması ve kanser gelişiminin düzenlenmesinde rol alan 49 insan ZBTB proteininin dizi ve yapı açısında oldukça benzer BTB bölgesi bulunur. Bunların arasında PATZ1 proteininin BTB bölgesinin merkezinde, eşsiz ve işlevi bilinmeyen 31 amino asit uzunluğundaki sekans ile öne çıkmaktadır. PATZ1'in merkezi sarımı memelilerde bulunmaktadır, ancak balıklarda yoktur. Her hücrede ifade edilen PATZ1, çok sayıda kanser türünde rol oynayan ve p53 tümör baskılayıcısına müdahale eden bir transkripsiyon faktörüdür. Bu tez çalışmasında, PATZ1 BTB alanının eşsiz yapısal özelliklerinin belirlenmesi, öncelikle hem fare hem de zebra balıklarında (PDB ID'leri: 6GUV ve 6GUW) bir homodimer olarak kristal yapısının çözümü ile gerçekleştirilmiştir. Merkezi sarım, nükleer co-repressor NCOR1'e bağlanma açısından test edildi. ZBTB'lerin dinamiklere dayalı bir sınıflandırmasını oluşturmak için, yeni protein yapılarının moleküler dinamikleri daha önce bilinen diğer proteinlerle karşılaştırarak incelendi. Son olarak, ZBTB proteinlerini ile BTB dimerizasyon potansiyeli, bağışıklık sistemi hücrelerininin mRNA ifade verileri kullanılarak hesaplamalı olarak araştırıldı.

To my beloved Matilde and Irene:
may your wonder for learning never end

ACKNOLEDGEMENTS

During the years of my PhD education, I felt grateful for the uncountable learning opportunities and for the truly invaluable people that I have met along the way. The times of a global pandemic crisis that we lived in 2020 kept us apart but did not cut our bridges. Before all, I want to thank my thesis advisors: Batu Erman, for showing me his natural teaching passion, open mind and unconditional support; and Canan Atılgan, for her research enthusiasm, attentive guidance and sincere encouragement. I look at their collaborative work and friendship as an inspirational model for colleagues' relationship in academia. With them, I thank the past and current members of Erman Lab and the MIDST lab because being part of two groups at the same time fully enriched my student experience. In particular, Emre Deniz and Haleh Abdizadeh for all their precious teaching, and Ronay Çetin, Gökşin Liu, Furkan Güçlu and Melike Gezen for being the closest fellow students with whom I shared this long experience of research and life.

I feel enormously thankful for the time spent in UK in 2016-17 from which I have the best memories. I want to thank Erika Mancini and Aaron Alt for their heartfelt mentorship and all their dedicated work; for the people in their lab to be so genuine at work as well as in front of a beer; for taking me to the synchrotron, the biggest laboratory a biologist can imagine! Out of academia, I want to thank Emre Yıldız, my drum teacher, because I was lucky to take music lessons from such a talented and passionate artist, but also because I'll treasure the profound life lessons that we've got to study together in long drives to school and witty coffee breaks chats. I also cannot avoid thanking my big family in Italy. Despite the distance I feel their strength at every step. Finally, I want to thank my life mate Ozan, because he has always supported me and kept me together in the last seven years as in the last seven weeks. He shows me his unmeasurable love making sure I have no other worries than work. In our journey together we share both hardship and achievements. Without him this all thesis wouldn't be here because "of this book, he is the spine". A Gallehault indeed Sabanci University where we first met!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	Amino acids
Amp	Ampicillin
AU	Absorbance Units
bp	Base pair
BTB	Broad complex, tramtrack and bric-a-brac
°C	Celsius degrees
Chl	Chloramphenicol
CIP	Calf Intestinal Alkaline Phosphatase
Cl	Chlorine
CMV	Cytomegalovirus
Da	Dalton
DMEM	Dulbecco's Modified Eagle Medium
EGFP	Enhanced GFP
FBS	Fetal Bovine Serum
FW	Forward
GFP	Green Fluorescent Protein
GST	Glutathione S-transferase
h	Hour
HF	High Fidelity (enzyme)
IMAC	Immobilized Metal Affinity Chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-d-1-thiogalactopyranoside
K	Potassium or Kelvin
Kan	Kanamycin
kDa	kilo Dalton
μ	micro
MD	Molecular Dynamics
min	Minute
ml	milli liter
MW	Molecular Weight
NCOR	Nuclear Co-Repressor
NPT	number of particles, pressure, temperature
NR	Not Retained
ns	nano seconds
NVT	number of particles, volume, temperature
o.n.	Overnight
OD	Optical Density
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pg	pico gram
pI	isoelectric point
r.t.	Room Temperature
RE	Restriction enzymes
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation

rpm	Revolution per minute
RV	Reverse
SASA	Solvent Accessible Surface Area
SAXS	Small Angle X-ray Scattering
SD	Superdex
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-Exclusion Chromatography
sec	second
Tm	Annealing Temperature
UK	United Kingdom
VdW	Van der Walls non-bonded interactions
WB	Western Blot
x g	relative centrifugal force unit
ZBTB	ZF and BTB domain proteins
ZF	Zinc Finger

1. INTRODUCTION

1.1 The BTB domain

The BTB (Broad-Complex, Tramtrack and Bric-a-brac) or POZ (Poxvirus and Zinc finger) domain is a protein-protein interaction motif with a conserved alpha/beta-topology fold. The BTB term was first used to describe a new protein domain found in Drosophila transcription regulatory protein encoding genes[1] that were found to be evolutionary related to mammalian transcription factors containing Zinc Finger (ZF) domains. Generally found in the protein N-terminus, the BTB domain plays a variety of roles comprising the assembly of voltage gated ion channels subunits by oligomerization, the maintenance of cytoskeleton stability, the recruitment of transcription co-repressors and chromatin remodeling factors as well as the coordination of E3-ubiquitin ligase complexes[2].

Roughly 1% (~200) of the proteins encoded in the human genome contain a BTB domain that is characterized by well conserved core secondary structures (β1-α1-B1-B2-A1-A2-B3-A3-β2-A4-A5) and a 3D fold that is strikingly similar. The presence of β1-α1 and β2 is common but optional. These elements do not belong to the core structure of the BTB domain but contribute to BTB dimerization. The specific characteristics of the most abundant BTB domain-containing proteins can be categorized in six classes:

- BTB, BACK and Kelch domain (BBK)
- BTB-T1
- BTB-Skp1
- BTB-Elongin C
- BTB-SLX4
- BTB and ZF (ZBTB)

In BBK proteins, the BACK domain is located immediately after the C-terminal of the BTB domain and it is mostly composed by in-tandem α-helices[3]. It is followed by a variable number of the Kelch motif repeats, each characterized by four β-sheets, that assemble in a circular propeller shape. These proteins interact with cytoskeleton elements to coordinate ubiquitinylated proteins targeted for degradation[4].

BTB-T1 is a class of proteins involved in the formation of the potassium channels tetramerization domain (KCTD)[5]. Reported crystal structures and electron microscopy images show however, mostly pentameric assembly[6] along a unique interaction surface, confirming the protein-binding nature of this domain.

The BTB domain of the Skp1 class contains two extra α-helices at the C-terminal of the core BTB structure and it is the farthest related group of proteins in the broad BTB/POZ family[2]. It is found in the E3 ubiquitin ligase SCF complex (Skp, Cullin1, F-box). The Skp1 domain is responsible of the coordination between the Cullin domain and the F-box, substrate recognition domain[7]. The E3 ligase catalyzes the ubiquitinoylation of proteins targeted for cellular degradation[8].

Elongin C BTB domain participates in the E3 ubiquitin ligase SCF complex that include Cullin2 domain[9]. Structurally, this BTB domain is the shortest, containing only the core elements and lacking the last A5 helix.

Another category of BTB domain containing proteins include SLX4 endonuclease (BTBD12) involved in genome stability and DNA repair[10]. These proteins contain additional UBZ4, MLR, SAP and CCD domains. The BTB domain is characterized by the secondary structure assumed by the loop connecting A2 and B3 that turns in an extra β-strand that completes the β-sheet formed by B1-B2 and B3 (PDB entries 4ZOU-4UYI).

Finally, the BTB domain of ZBTB proteins is characterized by the association to ZF domains in the C-terminus that can specifically bind DNA target sequences. Another, albeit rarer group of BTB domain containing proteins are characterized by a C-terminal DNA binding domain, the basic leucin zipper (bZip) found in BACH1 and BACH2[11]. The C-terminal DNA binding domains appoint these proteins to the function of DNA

transcription regulator leaving the BTB domain to mediate oligomerization by self-association to recruit to the transcription machinery co-factors by protein-protein interactions[12-15].

Figure 1.1 Protein sequence alignment of selected human ZBTB proteins and their predicted secondary structure. The sequences of the human PATZ1 and cKrox BTB domains with their unique extra region between the A2 helix and B3 strand are shown above. Arrows and rods identify predicted conserved β-strand and α-helical regions. The eight BTB domains with solved structures are annotated on the left with their common names in addition to the ZBTB nomenclature. Shading, asterisks, colons and periods identify conserved residues according to the Clustal format. A consensus sequence is shown at the bottom, with the three predicted degron residues involved in BTB domain stability indicated by arrows.

The human genome encodes for 49 members of the ZBTB family (Fig. 1.1), all of which contain an N-terminal BTB domain and a variable number of ZF motifs at their C-terminus. Two members of this family contain an additional motif in the form of an AT hook (PATZ1/ZBTB19 and PATZ2/ZBTB24). Most of the available BTB domain

structures of the ZBTB family are homodimers formed by the assembly of two identical monomers [16-31]. In the case of Myc-interacting zinc finger protein 1 (MIZ1), the structures of both a homodimer [24] and a homo-tetramer [32] have been reported. Before this work, the structures of eight different proteins belonging to the ZBTB family were available in the Protein Data Bank (PDB): LRF/ Pokémon (ZBTB7A); PLZF (ZBTB16); MIZ1 (ZBTB17); BCL6 (ZBTB27); MYNN (ZBTB31); FAZF (ZBTB32); KAISO (ZBTB33); HKR3/ TZAP (ZBTB48).

1.2 PATZ1 Transcription factor

The POZ/BTB and AT-hook-containing Zinc Finger protein 1 (PATZ1), is a transcription factor also known as MAZR (MAZ-Related Factor), ZNF278 (Zinc Finger protein 278), ZBTB19 (Zinc Finger and BTB protein 19) or ZSG (Zinc Finger Sarcoma Gene). It was firstly identified through its interaction with BACH2 (BTB And CNC Homology 2) in a Yeast Two Hybrid (Y2H) experiment[33]. It was discovered that the two transcription factors, PATZ1 and BACH2 associate through their BTB domains and that their DNA binding domains target the same promoter of the Fibroblast growth factor FGF4, important in embryonic development. Structurally, PATZ1 belongs to the POZ/ZBTB family of transcription factors[34]. Proteins belonging to this family have a fundamental role in the development of the immune system[35] but they also have been implicated in many biological processes, including transcriptional regulation and development[36] whilst their dysfunction in vertebrates has been linked to tumorigenesis.

A.
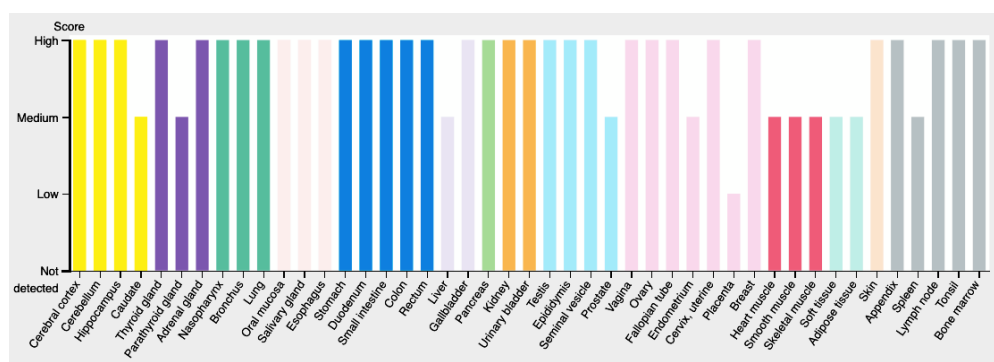
B.
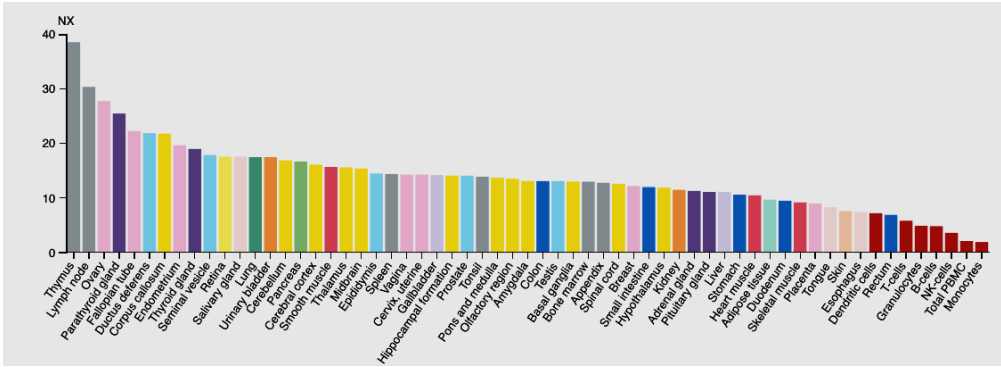
Figure 1.2 Overview of immunohistochemistry (A) and RNA-seq data (B) for PATZ1 protein and RNA expression in normal tissues. Human Protein Atlas available from https://www.proteinatlas.org/.

According to the Human Protein Atlas[37], PATZ1 is a ubiquitous nuclear expression protein detected in all tissues with low specificity (Fig. 1.2). One important function of PATZ1 is however specific of the early stages of T-lymphocyte differentiation where it negatively regulates *CD8* gene expression[38]. PATZ1 has been shown to participate in thymocyte development and CD4, CD8 and T_{reg} lineage choice by repressing the expression of ThPOK (ZBTB7B/ ZBTB15/ cKrox), another BTB domain-containing transcription factor[39-41] and the expression of the FOXP3 transcription factor[42].

The functions of PATZ1 are however not limited to lymphocytes. In mouse mast cells, PATZ1 (MAZR) was shown to interact synergistically with the mi-transcription factor (MITF) through the Zinc Finger C-terminal domain for the transcriptional activation of the gene encoding mouse mast cell protease 6 (mMCP-6)[43]. Mast cell proteases are strored in mast cell granules and released during the inflammatory and allergic responses[44]. An early embryonic role for PATZ1 has been suggested, as Patz1-/- mice are embryonic lethal or born at non-Mendelian frequency and are small in size depending on the genetic background[39]. PATZ1 also negatively regulates induced pluripotent stem cell (iPSC) generation[45, 46]. This function may be related to its interaction with the p53 tumor suppressor as demonstrated by various studies[47-49]. The role of PATZ1 as a regulator of stem cell development is shown to be correlated with the implication of PATZ1 in different aspects of cellular processes including tumorigenesis as it is overexpressed in Glioblastoma brain tumors[50] as well as in pediatric glioma[51].

1.3 Patz1 coding sequence

Isoform	Transcript		bp	aa	kDa	
	Human	**Mouse**				
A	hPATZ1-203	mPatz1-201	3638	641	69	4 exons
B	hPATZ1-204	mPatz1-202	3711	537	58	5 exons
C	hPATZ1-202	mPatz1-205	3895	687	74	5 exons
short	hPATZ1-201	mPatz1-203	2516	537	58	3 exons

Table 1.1 Summary of the 4 alternative splice protein coding isoforms of Patz1 gene sequence in human and mouse.

The human Patz1 gene sequence is located on chromosome 22q12.2 and over the 6 splice variants, 4 are protein-coding (Table 1.1). For this thesis work, all the experimental data referring to the mammalian PATZ1, derive exclusively from the mouse homolog of Patz1 isoform A (641 aa). This isoform of PATZ1 contains an N-terminal BTB domain (1-159), an AT-hook DNA binding motif (KRGRGRPRKANLL; 264-276) and 7 Zinc Finger domains (ZF) (details in Appendix D). The BTB domain is encoded by a single exon (exon1). The N-terminal BTB domain is required for PATZ1 dimerization. The region between ZF6 and ZF7 has been identified for the binding of PATZ1 to p53[48] (Fig. 1.3). All Zinc Fingers (ZF) are C2H2 Krüppel type DNA binding domains characterized by two cysteine (C2) ad two histidine (H2) residues coordinating a Zinc (Zn^{2+}) ion. For each ZF, the alpha helix is the recognition site for the target DNA. The residues -1, 3, 6 of each ZF helix contact directly the nucleotides on the target DNA[52] .
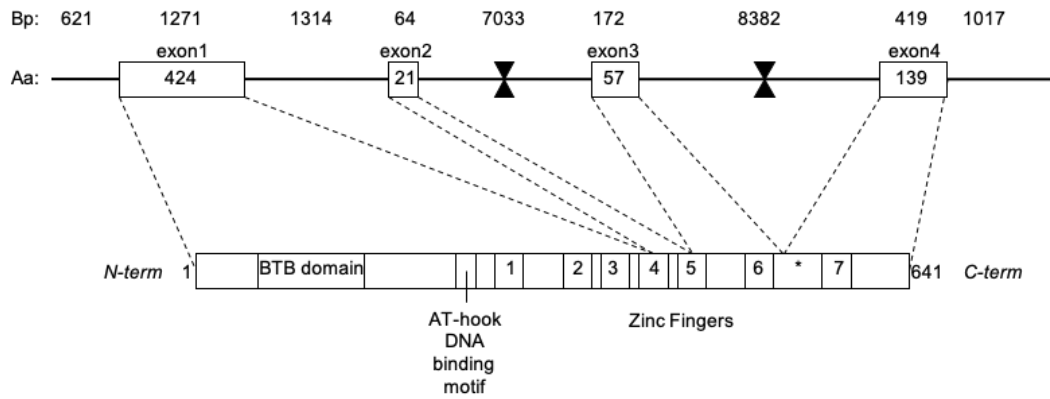
Figure 1.3 Schematic of Patz1 (isoform A) gene sequence and protein architecture. The N-terminal BTB domain is followed by an AT-hook DNA binding motif and 7 ZF domains. *The region between ZF6 and ZF7 contains a pseudo ZF sequence and the binding site for p53.

1.4 PATZ1 interactions

1.4.1 p53

The tumor suppressor p53 and PATZ1 are both transcription factors involved in cancer pathways. When PATZ1 proteins bind to p53, they inhibit it from DNA binding. The p53 association can also inhibit the PATZ1 protein from binding the PATZ1 DNA target site[48]. The binding site for p53 on PATZ1 was identified on the region between ZF6 and ZF7. Point mutations and deletions were tested in this region rich in negatively charged amino acids that resulted in the lost binding between PATZ1 and p53. Their interaction and regulatory action on each other function on gene transcription regulation could be important in cancer development.

1.4.2 NCOR1

The homodimerization of the BTB domains of ZBTB27 (B-Cell Lymphoma 6/BCL6) and ZBTB16 (Promyelocytic leukemia zinc finger/PLZF) creates a charged lateral groove that binds nuclear receptor co-repressors such as NCOR1, NCOR2 (SMRT) and BCOR[53-57]. NCOR1 and SMRT are structurally disordered proteins that share 45% identity[58] and contain a conserved 17-amino acids long BCL6 Binding Domain (BBD). These co-

repressors are components of large complexes containing histone deacetylases[59] contributing to transcriptional silencing. It is not known if co-repressor binding is a generalizable feature of BTB homodimers, as MIZ1, FAZF and LRF BTB homodimers do not interact with these co-repressors[20, 24]. Nevertheless, the PATZ1 BTB domain has been shown to interact with NCOR1, however it is not known if the interaction is mediated by an interface similar to that of BCL6 and PLZF[60]. Co-repressor binding to the BCL6 BTB domain requires dimerization because the interaction interface (lateral groove) is formed by residues in both monomers. 23 residues from each BCL6 monomer contribute to this interface[18]. Additionally, four residues, when mutated (L19S, N23H and L25S/R26L), interfere with co-repressor binding by preventing homodimerization of the BTB domain[21, 53, 58]. Because the PATZ1 BTB domain has also been shown to bind to NCOR1, a similar lateral groove may be mediating this interaction[60]. In fact, when the residues corresponding to L19S, N23H and L25S/R26L in BCL6 were mutated in the PATZ1 BTB domain (L27S, Q33S and R34L), it also failed to bind NCOR1[60].

For this study, two Myc-tagged plasmids encoding NCOR1 or NCOR2 (SMRT) partial cDNA were donated by Dr. V. Bardwell lab (Fig. 1.4). The protein sequences encoded in these plasmids include 1043 residues for NCOR1 (1019-2061) and 315 for SMRT (1368-1682). The protein indices refer to the NCBI-GenBank database entries NP_006302 for NCOR1 and NP_006303 for SMRT. Two large protein fragments (266 aa and 85 aa) from these same plasmids have been purified as small proteins that include the BBD short sequence (17 aa) at the immediate N-terminus[57] (Table 1.2). These fragments were shown to bind BCL6 and PLZF in previous studies[53, 54, 57]. The BBD of SMRT was later crystallized, with a BCL6 BTB domain homodimer[18] (PDB entry: 1R2B).
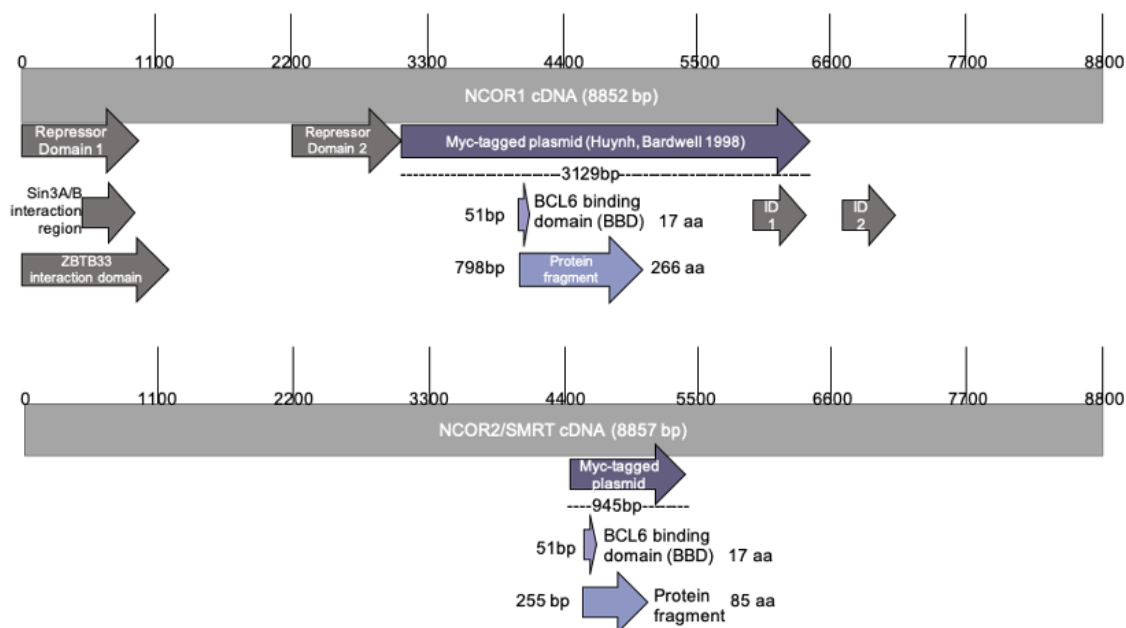
Figure 1.4 Schematic illustration of the NCOR1 and SMRT sequence. The DNA length in base pairs (bp) and the relative translated protein size in number of amino acids (aa) is reported. ID1-2 are annotated interaction domains.

	Myc-tagged plasmid		BBD		Protein fragment	
	NCOR1	SMRT	NCOR1	SMRT	NCOR1	SMRT
Gene Position	3055-6183	4453-5397	4054-4104	4561-4611	4051-4848	4558-4812
nt count	3129 bp	945 bp	51 bp	51 bp	798 bp	255 bp
Protein Residue number	1019-2061	1368-1682	1352-1368	1414-1430	1351-1616	1413-1497
aa count	1043 aa	315 aa	17 aa	17 aa	266 aa	85 aa

Table 1.2 NCOR1 and SMRT sequence annotations. The gene positions and the protein residue numbers with relative nucleotide (nt) and amino acids (aa) counts are given for the Myc-tagged plasmids, the BCL6 binding domain (BBD) and the protein fragments described above. Notice that the BBD is included in both of the larger constructs.

1.4.3 DNA target

ChIP-Seq data for PATZ1 indicates a CG-rich target sequence on DNA[61] as it was already shown as the target of the homologous MAZ protein on the promoter of *c-myc* gene[33]. The sequence motif found is CCCCKCCC (K being G or T in IUPAC code). Through a ZF-target prediction tool[62, 63], it is possible to predict a DNA sequence that would be the binding target of a specific sequence of Zinc Fingers in which each ZF binds 3 nucleotides on DNA. Applying random forest regression method, ChIP-Seq results match the prediction for the binding sequence predicted for ZF2, ZF3, ZF4.
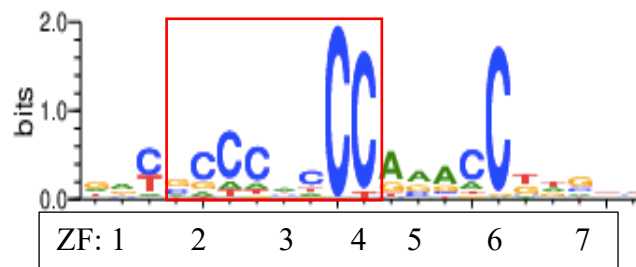
Figure 1.5 Predicted DNA target site for PATZ1 ZF domains. Available from http://zf.princeton.edu/index.php.

1.4.4 SUMO proteins

Small ubiquitin-like modifiers (SUMOs) are proteins that are used to post-translationally modify other proteins by covalent conjugation in a process called SUMOylation[64]. Targeted proteins can be directed to proteasomal degradation or recruited to promyelocytic leukemia (PML) nuclear bodies (NB) that are organelle-like spherical structures into the nucleus, made up of mainly of PML proteins and other components like SP100[65]. A SUMOylation target site consists of a Lysine acceptor, often preceded by a hydrophobic residue. The binding mechanism involves the cleavage of the last 2 residues of the SUMO protein upon synthesis, VY in the case of 95-residue long SUMO2. The c-terminal amino acid now being a Glycine can bind a Lysine residue on other proteins as a reversible post-translational modification through a covalent isopeptide bond. Differently from other ubiquitins, SUMO2 does not target proteins for degradation

but rather for processes including nuclear transport, DNA replication and repair, transcription co-repressor binding, mitosis and signal transduction, transcriptional regulation, apoptosis, and protein stability[66]. PATZ1 could also be recruited to NB and involved in these processes. Evidence of PATZ1 involvement in these interactions may be suggested by the co-localization of PATZ1 with RNF4[67] in nuclear dots. RNF4 is an E3 ubiquitin-protein ligase that interacts with SUMO2/3 to polyubiquitinate PML-NB components for degradation in proteasomes[68]. Other ZBTB proteins like BCL6 and PLZF also localize in nuclear speckles[53, 55, 67, 69, 70]. BCL6 in particular is known to be targeted by another E3 ubiquitin ligase, FBXO11, for proteasomal degradation.

1.5 Gene fusion

Gene fusions are generated from aberrant rearrangements of spatially close genes on the same or on different chromosomes. The hybrid genes formed with parts of two genes are usually dysfunctional but may acquire dominant characteristics and are linked to cancer. On chromosome 22 (Fig. 1.6), specific inversions translocate the EWSR1 gene with the PATZ1 gene and have been observed in various sarcomas. This translocation is not the only translocation of EWSR1. EWSR1 commonly translocates with ETS family proteins in Ewing's sarcoma. When EWSR1 fuses with PATZ1, it potentially encodes two fusion proteins EWSR1-PATZ1 and PATZ1-EWSR1, the former of which will contain the ZF domain, retaining the ability to bind DNA and the latter of which will contain the BTB domain potentially heterodimerizing with other BTB domain proteins[71-76]. PATZ1 was also reported to participate in a gene fusion event with MN1[77] another gene that can be disrupted by translocation on chromosome 22 causing meningioma.
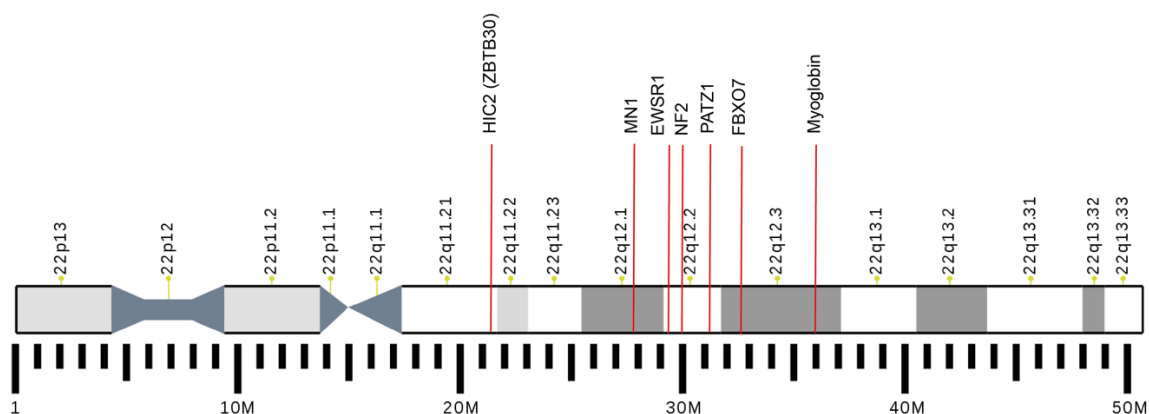
Figure 1.6 Schematic representation of the map of Chromosome 22 with gene locations. Fusion events have been reported to occur between adjacent Patz1 and EWSR1 and Patz1 and MN1 genes.

1.6 PATZ1 previous structures

Already available structure for human PATZ1 are NMR solutions for position 286-338 (2EPP), 350-384 (2EPR), 355-436 (2YT9), 380-411 (2EPQ), 408-445 (2EPS). These structures correspond to ZF1, ZF2, ZF3-4-5, ZF3, ZF4, respectively. No structure including the BTB domain of PATZ1 was reported before this study. While the BTB domain is similar in sequence to other BTB domains, there are some unique features that make it an interesting target for crystallography studies.

1.7 Unique properties of the PATZ1 BTB domain

Although overall fold of the BTB domain is highly conserved, the PATZ1 BTB domain contains a unique feature. The long glycine- and alanine-rich central loop between A2 and B3[33] will be referred to as the central loop (Fig. 1.1). This 31-aa long sequence is conserved in all vertebrate PATZ1 proteins, but it is absent in fish and amphibians. By sequence analysis, the mammalian A2/B3 loop sequence does not have any propensity to fold into a stable secondary structure and could be an evolutionarily acquired insertion sequence encoding an intrinsically disordered loop (IDL)[78]. Many unstructured proteins can fold or become functionally active upon post-translational modifications (PTM) or protein-protein interactions[79]. Phosphorylation by protein kinases is a PTM targeting

tyrosine, threonine or serine residues. It is a strong mechanism in inducing signal transduction and conformational changes of proteins. As many phosphorylation sites are found in long intrinsically disordered regions (IDR)[80], the folding of the PATZ1 disordered central loop could be also regulated by phosphorylation. In particular, threonine 91 found in this large loop could be the hypothetical PATZ1 phosphorylation site. Interestingly, T91A is the only aa mutation that distinguishes human from mouse PATZ1. If this hypothesis is correct, the mouse PATZ1 central loop lacks a phosphorylation site present in humans that could potentially affect its 3D structure. The T91 residue is conserved in primates and bats and it is alanine in other mammal species (Fig. 1.8).

1.8 PATZ1 in human and other species

PATZ1 is a transcription factor present in all vertebrates (Fig. 1.8). Despite 96 million years of divergent evolution[81], mouse and human share 99% of their genes with only about 300 genes found unique to each species[82]. Environment, metabolic rates, size and lifespan represent the main triggers of differentiation between the mouse and human. Also the immune system[83] and the mechanisms of tumorigenesis[84] can show significant differences. The PATZ1 protein sequences in human and mouse (isoform A) are 98.9% identical and 99.4% similar (EMBOSS Needle pairwise alignment). When the expression profile of these genes is compared for the immune systems of mouse and humans, PATZ1 has a 0.844 level of conservation of expression (COE)[85] (Available at http://rstats.immgen.org/comparative/comparative_search.php). For this reason, in thesis human and mouse PATZ1 are used interchangeably, unless differently indicated, although specific differences can have functional significance on the BTB domain or the ZF region in the two different species.

```
                                    BTB domain
PATZ1_HUMAN    1 MERVNDASCGPSGCYTYQVSRHSTEMLHNLNQQRKNGGRFCDVLLRVGDE    50
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE   1 MERVNDASCGPSGCYTYQVSRHSTEMLHNLNQQRKNGGRFCDVLLRVGDE    50


PATZ1_HUMAN   51 SFPAHRAVLAACSEYFESVFSAQLGDGGAADGGPADVGGATAAPGGGAGG   100
                 |||||||||||||||||||||||||||||||||||||||||.|||||||||
```

```
Q9JMG9_MOUSE     51 SFPAHRAVLAACSEYFESVFSAQLGDGGAADGGPADVGGAAAAPGGGAGG    100


PATZ1_HUMAN     101 SRELEMHTISSKVFGDILDFAYTSRIVVRLESFPELMTAAKFLLMRSVIE    150
                    |||||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    101 SRELEMHTISSKVFGDILDFAYTSRIVVRLESFPELMTAAKFLLMRSVIE    150


PATZ1_HUMAN     151 ICQEVIKQSNVQILVPPARADIMLFRPPGTSDLGFPLDMTNGAALAANSN    200
                    |||||||||||||||||||||||||||||||||||||||||:||||||
Q9JMG9_MOUSE    151 ICQEVIKQSNVQILVPPARADIMLFRPPGTSDLGFPLDMTNGAAMAANSN    200


PATZ1_HUMAN     201 GIAGSMQPEEEAARAAGAAIAGQASLPVLPGVDRLPMVAGPLSPQLLTSP    250
                    |||||||||||||.||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    201 GIAGSMQPEEEAARATGAAIAGQASLPVLPGVDRLPMVAGPLSPQLLTSP    250

                                        AT-hook                       ZF1
PATZ1_HUMAN     251 FPSVASSAPPLTGKRGRGRPRKANLLDSMFGSPGGLREAGILPCGLCGKV    300
                    ||:|||||||||.||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    251 FPNVASSAPPLTSKRGRGRPRKANLLDSMFGSPGGLREAGILPCGLCGKV    300


PATZ1_HUMAN     301 FTDANRLRQHEAQHGVTSLQLGYIDLPPPRLGENGLPISEDPDGPRKRSR    350
                    |||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    301 FTDANRLRQHEAQHGVTSLQLGYIDLPPPRLGENGLPISEDPDGPRKRSR    350

                          ZF2                                ZF3
PATZ1_HUMAN     351 TRKQVACEICGKIFRDVYHLNRHKLSHSGEKPYSCPVCGLRFKRKDRMSY    400
                    ||||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    351 TRKQVACEICGKIFRDVYHLNRHKLSHSGEKPYSCPVCGLRFKRKDRMSY    400

                                   ZF4                      ZF5
PATZ1_HUMAN     401 HVRSHDGSVGKPYICQSCGKGFSRPDHLNGHIKQVHTSERPHKCQTCNAS    450
                    |||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    401 HVRSHDGSVGKPYICQSCGKGFSRPDHLNGHIKQVHTSERPHKCQTCNAS    450

                                         ZF6
PATZ1_HUMAN     451 FATRDRLRSHLACHEDKVPCQVCGKYLRAAYMADHLKKHSEGPSNFCSIC    500
                    ||||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    451 FATRDRLRSHLACHEDKVPCQVCGKYLRAAYMADHLKKHSEGPSNFCSIC    500


PATZ1_HUMAN     501 NREGQKCSHQDPIESSDSYGDLSDASDLKTPEKQSANGSFSCDMAVPKNK    550
                    |||||||||||.||||||||||||||||||||||||||||||:||||||
Q9JMG9_MOUSE    501 NREGQKCSHQDLIESSDSYGDLSDASDLKTPEKQSANGSFSCDVAVPKNK    550

                                   ZF7
PATZ1_HUMAN     551 MESDGEKKYPCPECGSFFRSKSYLNKHIQKVHVRALGGPLGDLGPALGSP    600
                    |||||||||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    551 MESDGEKKYPCPECGSFFRSKSYLNKHIQKVHVRALGGPLGDLGPALGSP    600


PATZ1_HUMAN     601 FSPQQNMSLLESFGFQIVQSAFASSLVDPEVDQQPMGPEGK     641
                    |||||||||||||||||||||||||||||||||||||||||
Q9JMG9_MOUSE    601 FSPQQNMSLLESFGFQIVQSAFASSLVDPEVDQQPMGPEGK     641
```

14

Figure 1.7 Sequence alignment of PATZ1 protein (isoform A) in human and mouse. The N-terminal BTB domain is highlighted in yellow, the AT-hook motif in red and the ZF in blue. Clustal W annotation system is used to compare pairs of aligned amino acids. Amino acid differences between the two proteins (7/641) are highlighted in green. Only one amino acid difference is located within the BTB domain.

The mammalian PATZ1 protein is predicted to contain a 31-amino-acid A2/B3 loop (residues 75–105) that is partially conserved in cKrox (ZBTB15), which sets PATZ1 apart from the other ZBTB family members. This large loop replaces a shorter amino-acid stretch that forms a β-strand (B3) in other ZBTB family proteins, as described in detail by Stogios et al. (2005)[2]. This loop is glycine and alanine-rich and is predicted to be partially disordered; however, a short β-strand is predicted for the last six residues. The single difference between the human and mouse PATZ1 BTB domains (T91A) is found within this large loop. This sequence is predicted to be partially disordered or flexible. Sequence alignment shows that while the length of the A2/B3 loop is conserved in all PATZ1 orthologues, it is conspicuously absent in those from fish and amphibians (Fig. 1.8).

Figure 1.8 Protein sequence alignment of PATZ1 BTB domains from selected vertebrate species. The amino acid composition is very conserved apart from the unique central sequence that is conserved only in mammals and is missing in fish and amphibians. Primates and bats contain residue Thr 91 that is Ala for other mammals.

1.9 BTB homo and hetero dimerization

The eight deposited BTB domain crystal structures of ZBTB proteins show that the dimeric interface is formed by a specular (symmetric) interaction of the two BTB monomers with domain swap. The dimerization interface is quite conserved among all BTBs and consists of mostly hydrophobic residues and two absolutely conserved charged residues (charged pocket) flanking B1 and B2 (Fig.1.9).

```
                  β1               α1                    B1      B2      A1          A2
ZBTB7A-LRF>    --GPIGIPFPDHSSDILSGLNEQRTQ-GLLCDVVILVEGREFPTHRSVLAACSQYFKKLFTSGAVVD---------
ZBTB16-PLZF>   -MGMIQLQNPSHPTGLLCKANQMRLA-GTLCDVVIMVDSQEFHAHRTVLACTSKMFEILFHRN-------------
ZBTB17-MIZ1>   ------MDFPQHSQHVLEQLNQQRQL-GLLCDCTFVVDGVHFKAHKAVLAACSEYFKMLFVDQ-------------
ZBTB27-BCL6>   --ADSCIQFTRHASDVLLNLNRLRSR-DILTDVVIVVSREQFRAHKTVLMACSGLFYSIFTDQLKCN---------
ZBTB31-MYNN>   ---------SHHCEHLLERLNKQREA-GFLCDCTIVIGEFQFKAHRNVLASFSYFGAIYRST-------------
ZBTB32-FAZF>   ---PIRLPSPYGSDRLVQLAARLRPA---LCDTLITVGSQEFPAHSLVLAGVSQQLGR------------------
ZBTB33-KAISO>  SRKLISATDIQYSGSLLNSLNEQRGH-GLFCDVTVIVEDRKFRAHKNILSASSTYFHQLFSVA-------------
ZBTB48-HKR3>   -------SFVQHSVRVLQELNKQREK-GQYCDATLDVGGLVFKAHWSVLACCSHFFQSLYGDG-------------
ZBTB19-PATZ1>  -SGCYTYQVSRHSTEMLHNLNQQRKNGGRFCDVLLRVGDESFPAHRAVLAACSPYFESVFSAQLGPGGAADGGPAD
zebfishPATZ1>  ----YTYQVSKHSAEMLHNLNGQRKDGGRFCDVILRVGEESFPAHKAVLAACSEYFESVFSCQTEDD---------
ZBTB15-cKrox>  --DLIGIPFPDHSSELLSCLNEQRQL-GHLCDLTIRTQGLEYRTHRAVLAACSHYFKKLFTEGGGGAVMGAGGSGT
ZBTB24-PATZ2>  -SGQLVVHSDAHSDTVLASFEDQRKK-GFLCDITLIVENVHFRAHKALLAASSEYFSMMFAEEGR----I------
CONSENSUS>     S+GPIT+QFPQHS++LL+NLNQQR+AGGLLCDVTIVVGGEEFPAH+AVLAACSEYFESLFSS+S+D+G+VDQ++++

                  β*        B3           A3       β2        A4            A5
ZBTB7A-LRF>    -----------QQNVYEIDFVS---AEALTALMDFAYTATLT-VSTANVGDILSAARLLEIPAVSHVCADLLDR--
ZBTB16-PLZF>   -----------SQHYTLDFLS--PKTFQQILEYAYTATLQ-AKAEDLDDLLYAAEILEIEYLEEQCLKMLETIQ
ZBTB17-MIZ1>   -----------KDVVHLDISN---AAGLGQVLEFMYTAKLS-LSPENVDDVLAVATFLQMQDIITACHALKSLA-
ZBTB27-BCL6>   -----------LSVINLDPEIN---PEGFCILLDFMYTSRLN-LREGNIMAVMATAMYLQMEHVVDTCRKFIKASE
ZBTB31-MYNN>   -----------SENNVFLDQSQ-VKADGFQKLLEFIYTGTLN-LDSWNVKEIHQAADYLKVERVVTKCKIKME---
ZBTB32-FAZF>   -----------RGQWALGEGIS---PSTFAQLLNFVYGESVE-LQPGELRPLQEAARALGVQSLEEACWRAR----
ZBTB33-KAISO>  -----------GQVVELSFIR---AEIFAEILNYIYSSKIVRVRSDLLDELIKSGQLLGVKFIAELG--------
ZBTB48-HKR3>   -----------SGGSVVLPAGF---AEIFGLLLDFFYTGHLA-LTSGNRDQVLLAARELRVPEAVELCQSFKP---
ZBTB19-PATZ1>  VGGATAAPGGGAGGSRELEMHT-ISSKVFGDILDFAYTSRIV-VRLESFPELMTAAKFLLMRSVIEICQEVIK---
zebfishPATZ1>  -----------GQGKELEMHT-ISPKVFRDILDFAYTSKIV-VRLECFPELMTAAKFLLMRSVIEICQEVI----
ZBTB15-cKrox>  ATGG--------AGAGVCELDFVGPEALGALLEFAYTATLT-TSSANMPAVLQAARLLEIPCVIAACMEILQ---
ZBTB24-PATZ2>  -GQ------------SIYMLEGMVADTFGILLEFIYTGYLH-ASPKSTEQILATAQFLKVYDLVKAYTDFQNNHS
CONSENSUS>     VGGATAAPGGGAGGS+VLELDF+ISAEGFGQLLDFAYTS+LVRLRSEN+DELL+AARFL++ESV+E+CQ++I++++
```

Figure 1.9 Sequence alignment of the BTB domain of 11 human and 1 zebrafish ZBTB proteins. The yellow shaded residues are involved in the dimer interface. Positively and negatively charged residues involved in the dimer interface are indicated in blue and red respectively. The interface residues of cKrox and PATZ2 have been predicted from a structural model built by homology. The characteristic secondary structures are indicated with arrows and cylinders. The consensus sequence is indicated in the last line. Human PATZ1 contains an A2-B3 extra loop with a hypothetical additional β-strand (β*), and is partially conserved in cKrox.

Due to this high homology, BTB proteins are theoretically expected to bind not only as homodimer pairs but potentially also as heterodimers. In human cells not all the BTB containing proteins are expressed at the same level or at all, so the probability of all possible dimers to form in real conditions is presumably low. Although homodimerization seems to be favored, heterodimeric interactions between couples of ZBTB proteins have also been documented (Table 1.3). The crystal structure of a tethered "forced" heterodimer between the BTB domains of MIZ1 and B-cell lymphoma 6 protein (BCL6) suggests that heterodimers can use the same interface as homodimers[86]. Together with BCL6, MIZ1 seems to be a promiscuous member of the ZBTB family, making more heterodimers than any other BTB domain. MIZ1 functionally interacts with BCL6 in germinal center B cells[87], ZBTB4[88], ZBTB36[89], ZBTB49[90] and NAC1 (nucleus

accumbens-associated-1[86]). PATZ1 can also form heterodimers with other BTB domain containing proteins such as PATZ2[91], BACH1 and BACH2[33].

The mechanism controlling homodimerization versus heterodimerization of BTB domains has not been elucidated. Co-translational dimerization, a mechanism often required in protein complex assembly, may be at play[92]. Recently, a dimerization quality control mechanism for BTB proteins was proposed, where homodimer stability would exceed that of heterodimers because of the structural masking of destabilizing residues[93, 94]. According to this model, the preferential exposure of three "degron" residues on BTB heterodimers results in targeting by ubiquitin ligases and a shorter half-life. Whether this mechanism is universally shared by all BTB domain-containing proteins including members of the ZBTB family remains unclear.

	ZBTB pair interaction		Bibliography reference	Article results showing interaction
?	ZBTB4	ZBTB17/MIZ1	[88]	Fig.6 F-G
?	ZBTB5	ZBTB17/MIZ1	[95]	Fig 6
*	ZBTB7A/LRF	ZBTB7A/LRF	[96]	
	ZBTB7A/LRF	ZBTB27/BCL6	[97]	Fig 5-6
	ZBTB7A/LRF	ZBTB48/HKR3	[98]	Fig 5A
	ZBTB15/cKrox	ZBTB7A/LRF	[99]	Fig 5
*	ZBTB16/PLZF	ZBTB16/PLZF	[100]	Fig 3
?	ZBTB16/PLZF	ZBTB27/BCL6	[69]	Fig 1-2-4
	ZBTB16/PLZF	ZBTB32/FAZF	[101]	Fig 7
*	ZBTB17/MIZ1	ZBTB27/BCL6	[87]	Fig 2
	ZBTB17/MIZ1	ZBTB36/ZBTB7C	[89]	Fig 1
?	ZBTB17/MIZ1	ZBTB49	[90]	Fig. 4
	ZBTB21/ZNF295	ZBTB14/ZFP161	[102]	Fig 4
	ZBTB21/ZNF295	ZBTB21	[102]	Fig 4
?	ZBTB27/BCL6	ZBTB28/BCL6B/BAZF	[103]	Fig 4
	ZBTB29/HIC1	ZBTB29/HIC1	[104]	Fig 1
	ZBTB29/HIC1	ZBTB30/HIC2	[105]	Fig 6-7
	ZBTB30/HIC2	ZBTB30/HIC2	[105]	Fig 6
	ZBTB33/KAISO	ZBTB35	[106]	Fig 1

Table 1.3 Literature reports of heterodimers with references. The question marks indicate that the studied interactions are suggested or shown to be not exclusively through BTB domains but involve other parts of the proteins (ZF). The asterisk indicates that a BTB dimer crystal structure exist confirming the interaction.

2. AIM OF THE STUDY

The aim of this thesis project is to clarify the mechanism of potential interactions of the BTB domain of PATZ1 in the context of different protein complexes, through the study of its structural characteristics. First, the crystal structures of the homodimer of both mouse PATZ1 BTB domain (99% identical to human) and its zebrafish orthologue were obtained and analyzed in atomic details to determine the ideal state of self-association. The details of nuclear co-repressors binding to PATZ1 BTB domain were computationally predicted. A comparative analysis of the Molecular Dynamics simulations with ten other BTB domains was used to understand the impact that the diversity within this family may have on their functions. Additionally, to set the conditions for the study of functional interactions with other BTB domains of the ZBTB family, the potential heterodimers formation for PATZ1 was computationally explored in different cell types.

3. MATERIALS AND METHODS

3.1   Materials

3.1.1 Chemicals

All the chemicals used in this project are listed in the Appendix A.

3.1.2 Equipment

All the equipment used in this project are listed in the Appendix B.

3.1.3 Buffers and solutions

3.1.3.1 Bacterial Transformation

Calcium Chloride (CaCl$_2$) Working Solution: CaCl$_2$ 2M (exothermic reaction), PIPES 0.5M at pH 7.00 adjusted with NaOH pellet and Glycerol 15% (from 88% stock solution) were mixed in pure water ddH$_2$O, filter sterilized (0.2µm) and stored at 4°C.

3.1.3.2 Mammalian Cell Culture

Polyethyleneimine (PEI): 100 mg PEI was dissolved in 90 ml of pure water ddH$_2$O. pH was adjusted to 7.00 with NaOH 5M and the solution completed to 100 ml with pure water ddH$_2$O. The solution was filter sterilized and store at -20°C.

Trypan blue dye (0.4% w/v): 40 µg of trypan blue was dissolved in 10 ml Phosphate-buffer saline (PBS).

Hypotonic Lysis Buffer: 10 mM HEPES, 2 mM MgCl$_2$, 0.1 mM EDTA, 10 mM KCl, 0.5% NP40 mixed with pure water ddH$_2$O.

Hypertonic Lysis Buffer: 50 mM HEPES, 20 mM MgCl$_2$, 0.1 mM EDTA, 50 mM KCl, 400 mM NaCl, 10% Glycerol mixed with pure water ddH$_2$O.

4X Laemmli Buffer: 200 mM Tris-HCl (pH 6.8), 400 mM DTT, SDS 8%, 0.4% Bromophenol Blue, Glycerol 40%.

3.1.3.3 Gel Electrophoresis Buffers and Solutions

Agarose gel (1%): 1 g of agarose dissolved in 100 ml of 0.5X TBE buffer by heating. 0.2 µl of Ethidium Bromide is added to the solution for nucleic acid detection. The gel solution is poured in a casting tray and a comb is inserted on top to create wells for the loading of the samples. The gel is left to cool down and solidify.

5X Tris-Borate-EDTA (TBE) Buffer: 54 g Tris Base, 27.5 g Boric Acid, 20 ml 05 M EDTA (pH 8.0) completed to 1 L with pure water ddH$_2$O.

SDS Separating Gel:

For 10 ml (2 gel):	10%	14%
ddH$_2$O	3.8 ml	2.6 ml
Acrylamide 30 %	3.4 ml	4.6 ml
Tris 1.5 M, pH 8.8	2.6 ml	2.6 ml
SDS 10%	100 µl	100 µl
Ammonium Persulfate (APS) 10%	100 µl	100 µl
TEMED	10 µl	10 µl

Table 3.1 Composition of SDS Separating Gel at different concentration of Acrylamide.

SDS Stacking Gel (4%): For 5 ml (2 gel); 2.975 ml ddH$_2$O, 0.67 ml Acrylamide 30 %, 1.25 ml Tris 0.5 M pH 6.8, 50 µl SDS 10%, 50 µl APS 10%, 5 µl TEMED.

10X Tris-Glycine Buffer: 25 mM Tris Base and 200 mM Glycine. pH is adjusted to 8.3 with HCl 6N and the solution completed with pure water ddH$_2$O.

10X Running Buffer: 100 ml of 10X Tris Glicine (pH 8.3) and 5 ml of SDS 20% mixed and completed to 1 L with pure water ddH$_2$O.

1X Transfer Buffer (1 L): 100 ml of 10X Tris Glicine (pH 8.3), 200 ml methanol and 1.88 ml of SDS 20% mixed and completed to 1 L with pure water ddH$_2$O. Stored at 4°C.

1X PBS-T: 0.5 ml of Tween 20 detergent dissolved in 1 L of 1X PBS.

Blocking Buffer (5%): 0.5 g of Skim Milk powder dissolved in 10 ml 1X PBS-T and vortexed.

Enhanced Chemiluminescence (ECL) Reaction Buffer: 25 µl Luminol, 12.5 µl Cumaric acid, 234 µl Tris 1.5 and 1.5 µl of H$_2$O$_2$ 30% are added to 4.728 ml of pure water. Luminol 97% is prepared as 250 mM concentration in dimethyl sulfoxide (DMSO), stored in 25 µl aliquots at −20°C; *p*-coumaric acid is prepared as 90 mM concentration in DMSO, stored in 12.5 µl aliquots at −20°C.

3.1.3.4 Protein purification

IMAC Bacterial cells Lysis Buffer: HEPES 50mM, NaCl 250mM, TCEP 0.5mM (or 0.05% β−Mercaptoethanol), Imidazole 10 mM (from 2M stock), 1 tablet of Protease Inhibitor Cocktail (Roche- cOmplete™, EDTA-free), 1 µl of Turbo DNAse (Ambion-Invitrogen AM2238) or 5 µl of DNAse I were mixed on a rocking shaker or roller mixer at 4°C until the protease inhibitor tablet was dissolved in the buffer. The buffer was then kept on ice.

IMAC Wash Buffer: HEPES 50mM, NaCl 250mM, TCEP 0.5 mM (or 0.05% β−Mercaptoethanol), Imidazole 10mM. Filter sterilized (0.45µm) and stored at 4°C.

IMAC Elution Buffer: HEPES 50mM, NaCl 250mM, TCEP 0.5 mM (or 0.05% β–Mercaptoethanol), Imidazole 100/300/600 or 1000 mM. Filter sterilized (0.45µm) and stored at 4°C.

Gel Filtration Buffer: HEPES 20mM, NaCl 250mM, TCEP 5 mM (or 0.05% β–Mercaptoethanol). Filter sterilized (0.45µm) and stored at 4°C.

3.1.4 Growth Media

3.1.4.1 Bacterial Growth Media

Luria-Bertani Broth (LB) or Terrific/Turbo Broth (TB) were used for liquid bacterial cultures. LB was prepared dissolving 20 g of powder in 1 L of distilled water and autoclaved at 120°C for 15 min.

LB-Agar was prepared for bacterial solid medium. 15 g of Agar powder was dissolved together with 20 g of LB powder in 1 L of distilled water and autoclaved at 120°C for 15 min. Antibiotics were added after the cooling down of the solution and before solidification. The medium was poured onto sterile Petri dishes (~15ml/plate). The plates were stored at 4°C until use.

Antibiotics
For bacterial selections, final working concentration of 100 µg/ml for ampicillin, 50 µg/ml for kanamycin and 34 µg/ml for Chloramphenicol were added to LB/TB liquid culture or to LB-Agar before solidification in Petri dishes. Antibiotics stock solutions are stored in -20°C freezer.

3.1.4.2 Tissue culture Growth Media

DMEM HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% Pen-Strep (100 U/ml Penicillium and 100 µg/mL Streptomycin).

3.1.5 Commercial Molecular Biology Kits

All molecular biology kits used in this project are listed in the Appendix B.

3.1.6 Enzymes

All restriction and DNA modifying enzymes and their corresponding buffers used are from New England BioLabs (NEB) or Fermentas (Thermo Fisher Scientific).

3.1.7 Cell Types

3.1.7.1 Bacterial cells

DH5α

E. coli DH5α chemically competent cells were used for bacterial transformation of recombinant plasmids. This bacterial strain (genotype: F-, Φ80dlacZΔM15, Δ (lacZYA-argF) U169, deoR, recA1, endA1, hsdR17 (rk-, mk+), phoA, supE44, λ-, thi-1, gyrA96, relA1) is optimized for efficient cloning and plasmid isolation. Stored at -80°C.

Rosetta2 DE3 pLysS

This bacterial cell strain derives from E. coli BL21 with reduced cytoplasmic and periplasmic proteolytic degradation activity and was engineered to enhance expression of eukaryotic proteins. The cells contain the prophage from λDE3 that carries a chromosomal copy of the T7 RNA polymerase gene under control of a *lacUV5* promoter, allowing expression of proteins cloned in pET vectors by induction with IPTG. The integrated pRARE plasmid supplies tRNAs for 7 codons that are rare in E. coli and a gene for Chloramphenicol resistance. pLysS indicates that this strain expresses T7 lysozyme which further suppresses the basal expression of T7 RNA polymerase prior to induction, thus stabilizing pET recombinants, cell growth and viability. Rosetta2 DE3 pLysS chemically competent cells are stored at -80°C.

3.1.7.2 Tissue Culture Cell Lines

<u>HEK293T</u>

The HEK293T cell line, derives from human embryonic kidney 293 cells, and contains the SV40 T-antigen. This highly transfectable tissue culture cell line was used in this study for expression of recombinant plasmids carrying eukaryotic proteins.

3.1.8   Vectors and Primers

Vector Name	Use	Antibiotic resistance gene for selection
pCMV-HA	Cloning of mouse and Chimera PATZ1 full length; template for PATZ1 fragments	Amp
pEGFP-C3	Cloning of mouse PATZ1 full length	Kan
pET-47b (+)	Protein purification of mouse and zebrafish PATZ1 fragments	Kan
pET-49b (+)	Cloning of mouse PATZ1 fragments	Kan
pET-52b (+)	Cloning of mouse PATZ1 fragments	Amp
pGEX/3C	Protein purification of GST tagged HRV-3C protease	Amp
pKH110-EF(NLS)-9E10 c-Myc-tVP16-NCOR (human)	Co-IP and cloning of BBD peptides	Amp
pKH109-EF(NLS)-9E10 c-Myc-tVP16-SMRT (human)	Co-IP and cloning of BBD peptides	Amp
pcDNA3.1-Myc-His-B (-) Tag GFP-human-BCL6-BTB	Mutagenesis of BCL6	Amp
pcDNA3.1-Myc-His-B (-) Tag RFP-human-BCL6-BTB	Mutagenesis of BCL6	Amp

Table 3.2 Vector Backbones used in this thesis work.

A list of oligonucleotides synthesized to primer amplification reactions for the cloning experiments in this study is attached in Appendix D.

3.1.9   X-ray Crystallization Facilities

- Prof. Dr. Erika Mancini's Biochemistry lab at School of Life Sciences, University of Sussex, Falmer Brighton, UK;
- University of Sussex Crystallization facility, Falmer Brighton, UK;
- Beamline I04 for Macromolecular crystallography at Diamond Light Source, Harwell Science and Innovation Campus, Didcot Oxfordshire, UK.

3.1.10   Computational resources, Software and Servers

Licensed programs: CLC Main Workbench 7, GraphPad Prism 5, UNICORN 7.1, EndNote X8, Microsoft Office package, Adobe Illustrator.

Open Source Desktop Programs: VMD[107], NAMD[108], Jalview, PyMOL, Python, Tableau Public, Sublime Text, Evernote, Filezilla, MobaX-term, AnyDesk

Online Tools: PRIMO[109], MEME, Swiss Model, SymmDock[110], Z-dock[111], Modeller[112], Multiprot[113], PDBePISA, PROMALS3D[114], NEB Cloner, HHpred[115], Uniprot[116], Genbank, Ensemble, PDB, PRISM-Webserver[117], Expasy, Clustal-W, Emboss Needle, GIFMaker.me, PIC tool[118].

Remote Connection Server: TRUBA (Tübitak Ulakbim), HPC (SUNUM, Sabanci University), TOSUN-HPC (FENS, Sabanci University), KUACC-HPC (Koç University).

3.2   Methods

3.2.1   General Molecular Cloning Methods

3.2.1.1 Bacterial Cell Culture

<u>Bacterial Culture Growth:</u> E. coli DH5α or Rosetta2 (DE3) pLysS strains were grown o.n. at 37°C or 18°C in shaking incubator at 221 or 180 rpm in Luria-Bertani Broth (LB) or Terrific (Turbo) Broth (TB). Bacterial cells were frozen in liquid nitrogen with cryo-protector Glycerol (10%) for storing at -80°C.

<u>Chemically competent Bacterial Cells</u>: E. coli DH5α competent cells were prepared from previously frozen cells stored at -80°C in LB medium and glycerol without any selective antibody. 200 µl were incubated in 40 ml of LB and grown at 37°C o.n. in shaker incubator at 221 rpm. This dense culture was then diluted 1:100 by adding 4 ml to 400 ml of LB in a sterile 2-L flask under the laminal flow biosafety hood. The culture was set to grow at 37°C and Optical Density (OD) Absorbance was measured through a spectrophotometer until OD_{590} reached the value of 0.375. The culture was then transferred in eight 50 ml Falcon tubes, pre-chilled on ice and left on ice for 5-10 min. The cells were then centrifuged for 10 min at 1600 x g at 4°C. After discarding the supernatant, each pellet was resuspended in 10 ml ice-cold $CaCl_2$ solution. The solution was mixed by strongly hitting the tubes to each other on ice. The cells were further centrifuged for 5 min at 1100 x g at 4°Cand the supernatant again discarded. Resuspension of the pellet with 10 ml ice-cold $CaCl_2$ solution was repeated. Then the cells were kept on ice for 30 min. After a last centrifugation at 1100 x g at 4°C for 5 min, the supernatant was discarded, and each pellet resuspended in 2 ml of ice-cold $CaCl_2$ solution and collected in a single tube (16 ml total volume). The cells were finally distributed in 200 µl aliquots in 1.5 ml Eppendorf tubes and immediately frozen in liquid nitrogen for storage at -80°C. Transformation efficiency (TE) was tested for each batch

by transforming the competent cells with 1 µl of 10 pg/µl, 100 pg/µl and 1 ng/µl of pUC19 plasmid, carrying the Ampicillin resistance conferring gene. TE is measured dividing the total number of colonies grown on the transformed plate by the amount of plasmid DNA (in µg) used in the transformation. The appropriate value for TE is about 10^7 colonies/µg.

3.2.1.2 Cloning

Primer design: considerations
Before molecular cloning, it is necessary to follow 3 important steps for the design of constructs that aim at protein purification and crystallization.
1.Research on the protein sequence
2.Prediction of secondary structure elements
3.Choice of the cloning plasmid(s).

1. Research on the protein sequence
Information about the spliced variants of the gene coding the protein of interest is important when planning to PCR-amplify a chosen segment from genomic cDNA. If the construct sequence spans over more than one exon, the risk is to amplify different gene variants.
2.Prediction of secondary structure elements
HHpred[119] or similar bioinformatic tools, allow to predict protein secondary and tertiary structure from sequence based on homology. This step important to prevent cloning a sequence with breaks on the main secondary structures like alpha helices or beta strands. Failing to respect the natural protein conformation may results in misfolding or aggregation during expression and purification.
3.Choice of the cloning plasmid(s).
The cloning plasmid is the vector that contains the designed construct and will be transformed in the expression strain of bacteria or another producing organism. The choice of the cloning plasmid determines the expression and purification methods that will be used to obtain the protein of interest.

<u>Cloning</u>

Recombinant DNA Techniques were used to introduce protein-coding DNA sequencing (insert) into vectors for expression in bacterial or mammalian cells.

For cloning the following protocols were followed according to the origin of the insert:

— Insert from plasmid: 1 μg of miniprep DNA sample from the plasmid containing the sequence of interest is directly digested with cloning enzymes.

— Insert from plasmid + PCR: 2 ng of miniprep DNA sample from the plasmid containing the sequence of interest is used as a template for the amplification by PCR of the gene of interest.

— Insert from genomic DNA: 50 ng of genomic DNA is used as a template for the amplification by PCR of the gene of interest.

Primers were designed to have unique Restriction Sites at both 5'- and 3'-end. The restriction enzymes (RE) were chosen with two characteristics: 1. their recognition sequence is available in the cloning site of the vector plasmids used for cloning; 2. their recognition sequence is absent from the sequence of the insert to avoid the premature digestion of the amplified DNA. At the 5'-ends of both primers, forward (FW) and reverse (RV), 4 extra nucleotides were added to better accommodate the RE on the recognition sequence: TGGA on the FW primer, GGAC on the Rev primer (read in 5'-3' direction). At the 3'-end of the RV primer 2 stop codons (TAATAA) were added after the restriction site. Length of the designed primers was determined by online designing tools that calculate annealing temperature (Tm) for any given oligonucleotide sequence (http://biotools.nubic.northwestern.edu/OligoCalc.html; NEBcloner). Based on the working temperature of the DNA polymerase used during PCR, the length of the designed primers was determined according to the calculated Tm for the annealing sequence only, restriction sites excluded. For example, in the case of Phusion polymerase (NEB), that works at 60°C, the length of the designed primers was adjusted so that the calculated Tm for the annealing sequence was 58-62°C.

<u>PCR</u>

During Polymerase Chain Reaction process (PCR), the enzyme Polymerase elongates the primers sequences by adding nucleotides to the 3'end following the complementary strand template DNA. Phusion and Q5 (High fidelity) polymerases were used in optimal

conditions in order to decrease the error rate in terms of number of mutations introduced per base pair when cloning was a downstream application. Taq Polymerase was used in colony-PCR for diagnostic application only. In colony-PCR, the template DNA was picked from a bacterial colony grown on LB-Agar plate after cloning. The same colony was incubated overnight in liquid culture. If the corresponding PCR results matched the expected size of the amplified region, suggesting successful cloning, DNA was extracted from the liquid culture (miniprep) and analyzed by sequencing to confirm it.

	Taq Polymerase		Phusion Polymerase		Q5-HF	
Template DNA	1 ng		1 ng		2 ng	
FW primer (10µM)	0.5 µl		2.5 µl		1.25 µl	
RV primer (10µM)	0.5 µl		2.5 µl		1.25 µl	
dNTPs (10mM each)	0.5 µl		1 µl		0.5 µl	
Enzyme	0.125 µl		0.5 µl		0.25 µl	
Buffer	10X		5X		5X	
Final Volume	25 µl		50 µl		25 µl	
Initial Denaturation	95°C, 30"		98°C, 30"		98°C, 30"	
Denaturation	95°C, 25"		98°C, 10"		98°C, 10"	
Annealing (temperature depends on the Tm of the primers)	45-68°C, 25"	25 cycles	45-72°C, 20"	35 cycles	50-72°C, 20"	35 cycles
Extension (time depends on the length of sequence to be amplified)	68°C		72°C		72°C	
Final Extension	68°C, 5'		72°C, 10'		72°C, 2'	

Table 3.3 Summary of the PCR setup with 3 different Polymerase enzymes. All enzymes are from New England BioLabs (NEB). These reactions are optimized for PCR-thermocycler.

Restriction Enzyme Digestion

Vector and insert are digested from the same set of enzymes to create compatible ends. Generally, the vector DNA is then treated with CIP (NEB) at 37°C for 1h to avoid recircularization.

DNA	1 µg DNA
Restriction Enzyme(s)	0.5-1 µl
Buffer	10X
Final Volume	30-50 µl

Table 3.4 Restriction enzyme reaction. Reaction takes place in 5-20 min for FastDigest enzymes or 30 min-1h for commonly optimized enzymes.

Agarose gel electrophoresis

The digested sequences are loaded on 1% Agarose gel and ran until clear band separation are visible. 12-18 µl of each sample are loaded to each well. 1-2 µl of a marker mixture of DNA molecules with known length and concentration is also loaded for reference. All DNA molecules are negatively charged and run by electrophoresis in the direction of the current (negative to positive). Generally, 100-120 Volts are applied for 30-60 min. Agarose gel is used as a diagnostic method to verify the efficiency of digestion. DNA fragments are visible only under UV light through to the absorption by Ethidium bromide intercalated between nucleotides.

DNA Extraction from Agarose Gel

The DNA bands of interest (vector and insert) are compared in size with the marker ladder and can be gel extracted and cleaned-up of the Ethidium Bromide and buffer through a purification kit (Macherey-Nagel™ NucleoSpin™) according to manufacturers instructions.

Ligation

The 1:3 (vector:insert) mole ratio for ligation is chosen based on a second agarose gel run with 1, 2 and 4 µl samples of both digested insert and vector. The ligation reaction is carried on in a 20 µl reaction volume with 0.5 µl T4 ligase enzyme (NEB) and its buffer (1:10) for 20 min at r.t.

Transformation of Bacterial Cells for Cloning

Chemically competent cells aliquots of E. coli DH5α (200 µl) were thawed from -80°C to 4°C on ice and 100 pg of recombinant plasmid DNA was added. The cells were kept on ice for 20 min. A heat-shock was induced to the cells by transferring them on a heat-block set to 42°C for 90 sec and then immediately back on ice. After 3 min, 800 µl of LB was added to the cells then left to reacclimate in water bath or thermocycler at 37°C for 45 min. The cells were then centrifuged at max speed for 30 sec and 800 µl of LB medium supernatant was discarded. The pellet was resuspended in the remaining medium and 100 µl of suspension was distributed using glass beads on LB-Agar plates enriched with antibiotics for selection. The plates were then incubation o.n. at 37°C.

Plasmid DNA isolation

colonies were extracted for diagnostic digestion. Miniprep DNA extracted from bacterial colonies yields about 100-200 ng/µl concentrated plasmid DNA. MIDI prep yield about 800-4000 ng/µl concentrated plasmid DNA.

Sequencing

15µl of 100 ng/µl concentrated miniprep DNA sample are sent for Sanger sequencing analysis (MC Lab, South San Francisco, CA).

3.2.2 Mammalian Cell Culture

Mammalian Culture Growth and Maintenance

Complete DMEM was used for growth and maintenance of HEK293T adherent cells. The cells were normally grown in 10 cm sterile tissue culture plates in 10 ml of DMEM medium and incubated in constant conditions of 37°C and 5% CO_2 level. At confluency, reached when the cells coated about 80% of the plate surface, the medium was removed by aspiration, the cells washed with serum-free DMEM and then detached with trypsin treatment. For maintenance, the cells were then suspended in complete DMEM and split to a new 10 cm sterile tissue plate at 1:10 ratio. For transfection experiment preparation, the cells were suspended in fresh complete DMEM and were counted as 10^6 for 10 cm plates and 200000 cells/well for 6-well plates, then incubated at 37°C for 16 h.

Mammalian Cells Transfection

16 hours after seeding the HEK293T cells on 10 cm or 6-well plates, 5 or 3 μg of MIDIprep plasmid DNA were mixed with 550 or 200 μl serum-free phenol-free DMEM and 15 or 9 μg of PEI (1μg/μl) for transfection (1:3 DNA-PEI ratio). The mixture was vortexed and left to settle at r.t for 15 min. After incubation, the mixture was distributed on the cells drop by drop.

Cell Lysis

For co-immunoprecipitation experiments 2 or 3 10 cm plates per experimental condition were seeded. 48 hours after transfection $2x10^7$ cells were counted for each experimental condition. The cells were then centrifuged at 300 x g for 5 min in 15 ml Falcon tubes. After discarding the supernatant by aspiration, the cells' pellet was resuspended in 1 ml of ice-cold PBS (1X) for washing and centrifuged at 16000 x g for 15 min at 4°C. This wash was repeated twice after the last supernatant removal the cell pellet was either stored at -80°C or treated with cell lysis buffers. 500 μl of H1 buffer for cytoplasmic lysis was added to each 10^7 cells pellet. The resuspension was left on ice for 10 min, then centrifuged at max speed for 10 min at 4°C. While the supernatant was transferred in a new tube labelled as "cytoplasmic lysis" and stored at -20°C, the pellet was further treated for nuclear lysis extraction. 75 μl of H2 buffer was added to each 10^7 cells pellet. With a manual homogenizer, the pellet was pushed towards to bottom of the tube with consistent pressure and rotation for 10 times until a sticky whitish pellet was obtained. The smashed cells were kept on ice for 15 min, then centrifuged at max speed for 10 min at 4°C. The supernatant was transferred in a new tube labelled as "nuclear lysis" and use in co-immunoprecipitation experiments.

Co-Immunoprecipitation (Co-IP)

150 μl of nuclear lysate (from $2x10^7$ cells) are mixed with 2.5 μl (1 μg) of mouse monoclonal IgG1 c-myc-antibody (Ab) and diluted with the 150 μl of H2 buffer. The mixture is incubated o.n. at 4°C on rotating tube stand. After the incubation time, Protein G magnetic beads are prepared for Co-IP. 30 μl of the beads' slurry is added to each reaction tube. With the use of a magnetic rack, the beads in the tubes separate from the storage solution (PBS) and are collected on the side of the tube. The liquid is removed, and the beads resuspended in 30 μl of H2 buffer. This wash phase it repeated twice. After removing the liquid for the last time, the mixture of Ab and nuclear lysate samples (~300

μl) is added to the beads and the reaction tubes incubated for 3h at 4°C on rotating tube stand. The mixture is then separated on the magnetic rack. The beads are isolated by removing the supernatant and then washed 3 times with 200 μl of H2 buffer. A final volume of 50 μl buffer is added to the beads. 50 μl of 2X Laemmli are added to the samples (IP) that are then boiled at 95°C for 10 min. 20 μl of each IP sample are then run by electrophoresis on a protein gel at 10% Acrylamide concentration. The proteins are then transferred from the gel to membrane for Western Blot.

Microscopy

After 24 hours from transfection, live cells from 6-well plates were observed under inverted fluorescence microscope (Axio Vert.A1, Zeiss) with bright field and GFP filters.

3.2.3   Protein Expression and Purification

The main steps described in this section are summarized in Fig. 3.2.

Bacterial culture

Clone plasmids containing the recombinant protein of interest are transformed in E. coli expression strain cells, Rosetta2 (DE3) pLysS. These cells carry Chloramphenicol resistance conferring gene. The transformed cells are inoculated on a petri dish containing solid LB-agar in which working concentration of two antibiotics are added. antibiotic would be Chloramphenicol, and the second depends on the antibiotic resistance conferring gene present on the inserted plasmid. The plates are incubated at 37°C o.n. A single colony from this plate is then inoculate in 3 or 25 ml of liquid LB or Turbo Broth in which the same antibiotics have been added. The small or big scale cultures are grown at 37°C o.n. The pre-cultures are then added to fresh medium containing 40% of the antibiotic working concentration and incubated at 37°C until OD600~0.4-0.6 for LB and OD600 ~0.6-1.0 for TB in shaker incubator 180 rpm.

Induction of Protein Expression

The growth curve of the bacterial cells is built by collecting values of Optical Density (OD) Absorbance through a spectrophotometer. For each culture, 1 ml sample is collected at different time points. Bacterial growth curve is characterized by a logistic model that after an exponential phase, slows down over the carrying capacity (limited resources) of

the environment. Thus, the synthesis of the recombinant protein is induced while still on the exponential phase. The expression vectors contain a *lac* gene operator under control of the strong T7 promoter upstream of the recombinant protein sequence. The bacteria growing on normal medium rich on glucose are not able to transcribe the recombinant protein sequence because of the *lac* inhibitor blocking the operator site. Once bacterial growth reaches an optimal value of absorbance at OD_{600}, protein expression can be externally induced with isopropyl β-d-1-thiogalactopyranoside (IPTG). IPTG is a molecular mimic of allolactose that, competing for the binding of the *lac* inhibitor, allows the release of the promoter for transcription and expression of the inserted DNA.

	SMALL scale expression	LARGE scale expression
Culture final volume	50 ml	1 L
Pre-culture (LB + selective antibiotics) Start from single colony or alternatively from several colonies picked up by sterile inoculation loops.	3 ml LB at 37°C o.n.	3 ml LB at 37°C for ~ 8 hours; then, 25 ml LB at 37°C o.n.
Incubate the pre-culture (LB or TB) at 37°C until OD_{600} ~0.4-0.6 for LB or ~0.6-1.0 for TB in shaker incubator at 180 rpm; Medium with 40% antibiotics concentration	50 ml medium in 250ml flask	1 L medium in 2L flask
Induce expression: Reduce the temperature to 18°C, then induce protein expression with →	50 μl of 0.1M IPTG	100 μl of 1M IPTG

Table 3.5 Schematic protocol for small- and large-scale culture for protein expression.

<u>Cell Lysis</u>

After induction, the bacterial culture is kept growing at 18°C o.n. in shaker incubator at 180 rpm. Cells are then harvested by centrifugation (10 min at 4000 rpm) until clear. The supernatant consisting mostly of the growth medium is discarded, and the cell pellet

dissolved into 10 ml or 25 ml of lysis buffer for the small scale (50 ml) or the big scale (1 L) culture, respectively. After the chemical lysis, the cells are also sonicated. The sonicator is setup to run elapse time for 1 min and 30 sec with small probe for 50 ml culture lysate; 4 to 6 min with thick probe for 1 L culture lysate. Pulse is setup as 5 sec ON, 10 sec OFF at amplitude set to 40%. The cell lysate is then centrifugated at 15000 rpm for 45min/1h at 4°C in 30 ml centrifuge tubes to precipitate cell walls and keep soluble proteins in the supernatant.

Protein Purification by IMAC

The Immobilized Metal Affinity Chromatography (IMAC) protein purification method is based on the use of a liquid and solid phase component. The liquid phase component is represented by the protein that was cloned in frame with an N-terminal 6x His tag and an HRV-3C protease cleavage site for fusion-tag removal; the solid phase is represented by agarose beads conjugated with chelating metal ions (Nickel, Cobalt or other) forming a resin stored at 4°C in EtOH, 1:1 dilution. The metal beads attract the imidazolic ring structure of Histidines expressed on the a His tag (Fig. 3.1). The binding capacity of the HisPur™ Cobalt Resin (Thermo Fisher Scientific) is 5-15mg of protein/ml. For IMAC purification, the resin is loaded (2-3ml of resin-solution that results into actual 1-1.5 ml of beads) in Chromatography Columns (~50ml volume) for pull-down (gravity flow protocol). Alternatively, the Batch protocol included in the resin package can be followed.

After equilibration of the resin with pure water to remove EtOH, the resin is firstly washed with Buffer IMAC-A at low concentration of Imidazole (10mM). This concentration is sufficient to blandly coat the resin. The supernatant from the centrifugation of the induced bacterial culture containing the soluble proteins is then loaded over the resin. The complex of His tagged proteins and the beads is allowed to form by gentle rocking incubation of the column at 4°C for 10 min. After the incubation time, the supernatant is left to flow through the column. This fraction contains all the soluble material of the cell lysate that does not bind to the capacity of the resin beads and can be stored for analysis. The flow-through of the not-retained fraction (NR) is collected in a 50 ml falcon tube and stores at 4°C. 10 μl sample are taken for SDS gel analysis. The resin and the retained proteins are washed with 10ml of buffer IMAC-A for 3times to remove non-specific binding and the flow-through discarded. Then, higher concentration of Imidazole is added to elute the protein.

Imidazole: Histidine:

Figure 3.1 Chemical structure of Imidazole and Histidine amino acid. The similarity of the characteristic imidazolic ring in the two compounds is exploit in imidazole elution of isolated His tagged- proteins.

The imidazole molecules added with the elution buffer compete for the binding to the resin beads and the protein is released. Different concentrations of Imidazole can be added in an increasing gradient (IMAC-B: 100mM, 300mM, 600mM, 1mM…) and each eluate has to be analyzed in order to find the optimal concentration of the buffer at which the protein of interest is being completely released. The elution with buffer IMAC-B occurs in 3 steps of 3 to 5 ml each for a total of 7 to 12 ml elution volume collected in a 15 ml falcon. 20 µl sample are taken for SDS gel analysis. The eluted proteins are stored at 4°C.

Protein Concentration

The sample containing the eluted purified protein (7-12 ml) needs to be concentrated down to 5ml (or less) for following application like SEC. Concentration process is facilitated by concentrators spun at high centrifuge speed. The concentrator is tube divided in two compartments by a filtering membrane. An adequate cut-off for the porosity of the membrane is chosen based on the size of the protein to concentrate. 10000MWCO (10000 Da Molecular Weight cut-off) is used for protein size higher than10 kDa. After 1 min spin, the protein initial concentration is checked by Nanodrop using the buffer passed through the membrane as blank. Depending on the purity of the protein it can take minutes to hours to concentrate. The concentration value in the final volume is compared with the initial value to make sure that any protein amount was lost in the process. In that case, the flow-through can be analyzed instead.

The protein A280 concentration analysis was performed on NanoDrop 2000 (Thermo Scientific) with specific extinction coefficient (E 1%) for the protein of interest is entered. This value can be retrieved from *Swissprot-Protparam* tool. After entering the sequence of protein of interest the tool calculates some protein characteristics such as residue count, MW, aa percentage composition, ext. coeff. (E 0.1%).

SDS Gel Electrophoresis

The analysis of the collected nuclear lysate, cell pellet, not-retained fraction and eluates consists in running the samples on an SDS denaturing gel and check for expected size of the purified protein in molecular weight. The purified proteins are treated with SDS loading buffer (2X or 4X Laemmli, Dye) and denatured at 95°C for 5-10 min. 10 µl of each sample are loaded in each gel well. The first lane is loaded with the protein marker (tri-colour pre-stained ladder, stored at 4°C). Comassie blue is then used for gel staining and protein bands visualization.

Protein Purification by SEC

The eluate containing the protein at higher concentration is then used for further purification (size exclusion gel filtration) in FPLC (AKTA pure by GE).

Depending on the method setup, the protein is injected to the AKTA through 0.5, 1, 5ml loop or combinations. The gel filtration column must be previously equilibrated with Gel filtration (GeFi) buffer. For size exclusion chromatography, there are various choices of columns according to the size of the protein to be purified. SD Superdex 75 10/300 (30cm) or 16/600 HiLoad (60cm) by GE can separate protein in a size range of 3000 to 70000 Da and loaded with different volumes. Superdex 200 10/300or 16/600 can separate proteins in the size range of 10000-600000Da.

The protein separation gives a chromatogram. When the column is calibrated with known size proteins, it is possible to correlate the elution point (ml) where the protein is detected by UV and the size of the molecule.

His-tag cleavage

The HRV (Human Rhinovirus) 3C Protease recognizes the pre-scission site LEVLFQ|GP and cuts between Q and G, leaving the last two residues (GP) at the N-terminus of the recombinant protein sequence. The pre-scission site sequence is encoded on the vector backbone. To cleave the His tag from the recombinant proteins expressed in 1L culture 100 µl of 3C protease is used with a 16 hours incubation (o.n.) in rolling mixer at 4°C (cold room). The enzyme is stored at -20°C in 200 µl aliquots at 2mg/ml concentration.

Protein Flash Freezing

Pure proteins can be aliquoted and directly frozen in liquid nitrogen for storage at -80°C. Thawing procedure must also be rapid.
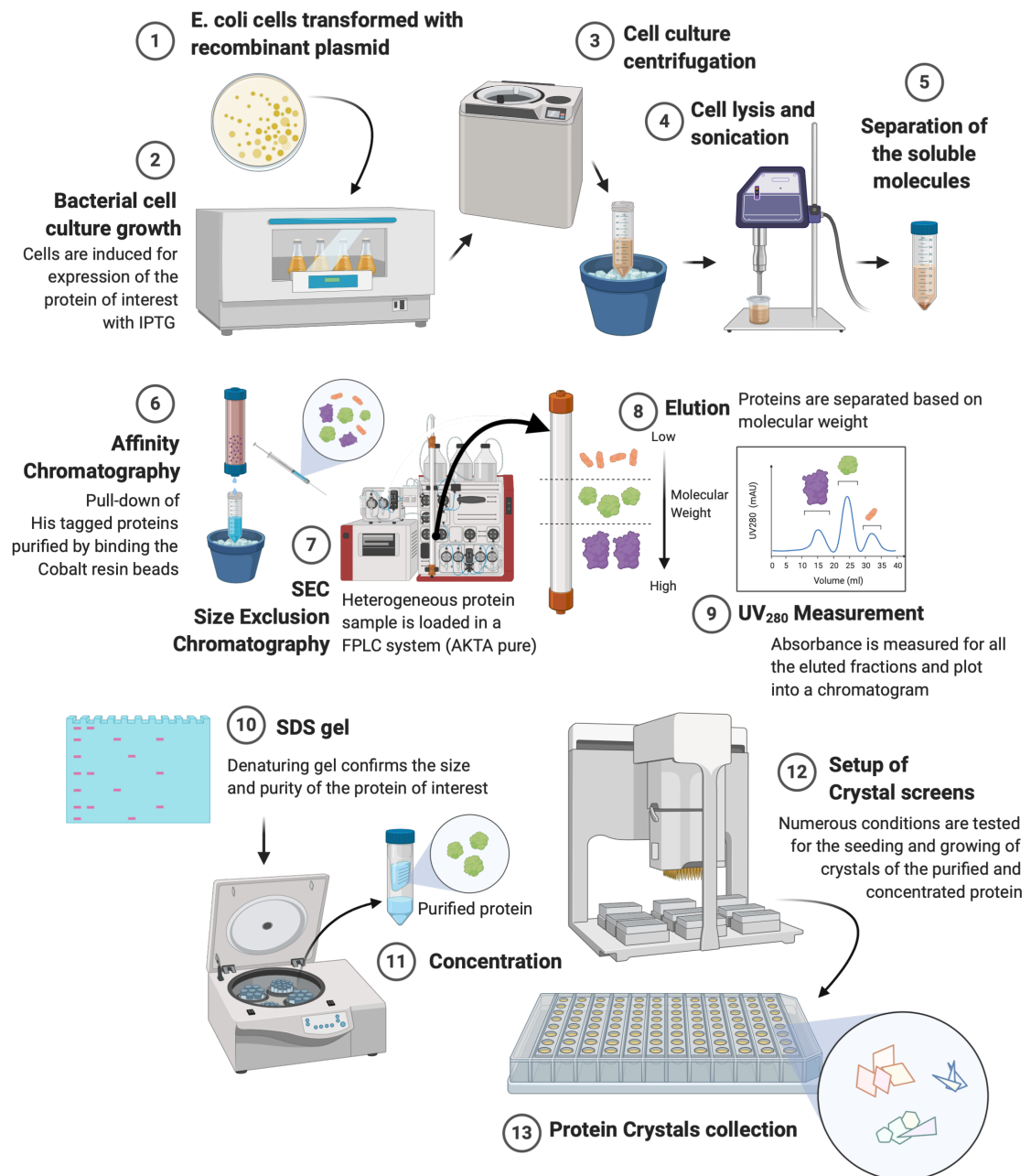
Figure 3.2 Schematic summary of the protein purification methods from bacterial cells to protein crystals. Created with BioRender.com

3.2.4 X-ray Crystallization

3.2.4.1 Theory

X-ray crystallography is a technique that allows to magnify macromolecules like proteins or nucleic acids in order to deduce their atomic structure. A covalent bond between two carbon atoms measures approximately 1.5 Å (0.15 nm) and a hydrogen bond is about twice as long. Because of this size scale, protein structures can be "seen" only under a specific light source. According to the properties of light, the behavior of light particles (photons) can be described in waves terminology. The wavelength (λ) is defined as the distance between two consecutive wave crests (nm) and is measured dividing the wave speed (c = 3 x 10^8 m/s in the case of light) by its frequency (ν, namely the number of waves in the unit of time (s). Photons have also energy (E) properties. These properties are expressed by the Einstein-Planck's equation:

$$\text{E} = h \cdot \frac{c}{\lambda} = h \cdot \nu \tag{3.1}$$

where h is the Planck's constant (4.1357 $\times$ 10^{-15} eV $\cdot$ s).

The energy of a photon is inversely proportional to its wavelength and directly proportional to the frequency. While the visible spectrum spans in a limited fraction of the electromagnetic field, between 400 (violet) and 700 nm (red), the atomistic details of the molecules of life range in the Angstrom (10^{-1} nm) scale and can be appreciated only with the appropriate wavelength of light. As seen in the previous equation (3.1), such a small wavelength will obviously have a very high energy. X-rays have short enough wavelength (10^{-2} to 10 nm) and can penetrate organic matter (energy in the order of 10^3 eV). The X-rays used in protein crystallography range between 0.5-1.6 Å. Such homogenous and tunable light source is generally provided only by synchrotron experimental beamlines.

This source of light is nevertheless as efficient as harmful to experimental samples of single molecules. For this reason, this light source is used experimentally on frozen crystallized samples in which every molecule to be analyzed is present in multiple copies (in average there are 10^{15} or more molecules per crystal), all in the same repeated spatial configuration. Thus, X-rays can penetrate the sample and scatter on a diffractor giving enough information about the disposition of atoms in the crystal before damaging it. Ideal crystal size must range between 0.2-1 mm to balance between good signal and absorption of X-rays. Enough many molecules in the crystal, allow the collection of sufficient signals from X-ray diffraction to reconstruct the source image of the space arrangement of the atoms. When an X-ray photon hits an electron in the sample, the electron starts vibrating and emits a photon of the same energy and wavelength of the X-ray photon in random direction. This is called coherent scattering. If the X-ray photon hits the nucleus of an atom, the nucleus vibrates much less because is much more massive than the photon, resulting in the emission of photons of lower energy, the so-called incoherent scattering and molecule damage. X-ray crystallography sees electrons and the structural results represent the average configuration found at the peak of the distribution of electron density[120].

3.2.4.2 X-ray Crystallization data collection

Protein structure determination requires large amounts (in the mg scale) of pure and concentrated samples. The proteins expressed from large scale bacterial cultures and purified with the described methods of affinity and size exclusion chromatography were used to setup crystallization trays, microscopic chambers where the crystals seeding and growing conditions are assayed. The trays are 96-well flat-bottom plates in which each well is divided into three compartments, two drop wells and a reservoir (Fig. 3.3).
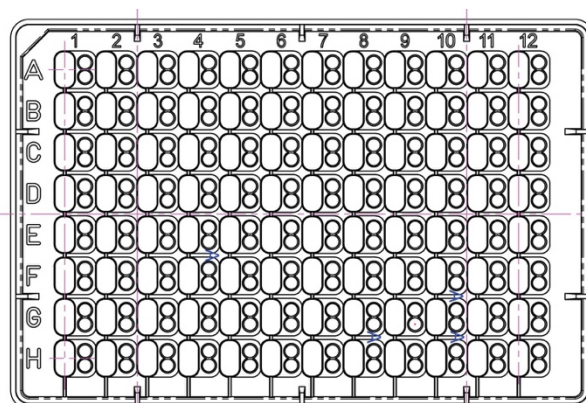
Figure 3.3 Details of a plate for crystallization conditions screening (Hampton Research).

In the sitting-drop vapor-diffusion method, the reservoir is a mixture of chemicals that represent the environmental conditions for the crystal growth, commercially available as crystallization screens. The chemical precipitant volume is first dispensed in the reservoir and in the drop wells. Then, a nano volume of the protein sample (0.2 µl) fills the drop wells, often at different concentration values in each well. The pipetting is performed by a software-controlled robotic needle. The plate is then sealed with UV compatible cover slips and incubated at a constant temperature. The crystal trays are periodically checked under a microscope. When (if) crystals are identified, the screening plate cover is removed, and the crystals individually fished out and mounted on a size-suitable cryo-loop. The crystals are quickly coated with cryoprotectant (i.e. glycerol), then frozen and stored in liquid nitrogen. The samples are transported to the synchrotron light source were the sample storage unit of the experimental beamline is loaded with the experimental pucks, with a single protein crystal for each sample holder.

At this point, the experimental hatch where the X-ray beam is released needs to be safely closed to protect the researchers from radiations. The continuous sample-loading robotic arm and the orientation of each crystal in front of the beam can be controlled remotely through the microscopy information transmitted to connected computers. It is important that the crystal is oriented along its symmetry axes and that multiple diffraction images are collected for each orientation. In this phase to protect the crystal integrity, cooling gas is continuously directed onto it. Explanatory visual reproductions of the equipment used as the setup for X-ray crystallization experiments are available at:
https://www.diamond.ac.uk/Instruments/Mx/I04.html

3.2.4.3 From diffraction data to Structure Determination

Crystallization diffraction data are analyzed by continuously developing software that reorient the multiple collected images to compute the electron density map relative to the crystallized protein. Through molecular replacement it is possible to use a structurally homologous protein as search template. The template is rotated and translated in all possible orientations to fit into the new map. Fitting the backbone and reconstructing the amino acid side chains according to the sequence of the protein of interest is a long process often requiring numerous refinements phases, reliable validation programs and especially high level of expertise. The final structure file contains general remarks about the protein, chain and atoms identifiers, the 3D spatial coordinates for each atom and a temperature value (B-factor). Higher B-factor values indicate a broader distribution of the electron density obtained from that atom due to vibrations or uneven position alignment in the crystal molecules.

3.2.5   Molecular Dynamics Simulations

<u>MD simulations</u>
The average conformation of a protein forced in the crystal packing is rarely the representation of the native protein structure. In order to actually describe the protein, Molecular Dynamics simulate the native environment of the protein in terms of solvent and temperature and apply the laws of statistical mechanics to the motion of atoms to calculate their trajectory as a function of time. Statistical mechanics connects the microscopic description of the system behavior to its macroscopic properties of stability,

flexibility and functional conformations. For the microscopic description it is necessary to determine a potential energy function that defines the force on every particle of the system.

$$\vec{F}(\vec{r}) = -\nabla U(\vec{r}) \qquad (3.2)$$

This equation is built on the summation of the bonded and not bonded interactions terms.

$$U(\vec{r}) = \sum U_{bonded}(\vec{r}) + \sum U_{unbonded}(\vec{r}) \qquad (3.3)$$

The bonded interaction term includes harmonic vibrational motion between pairs of covalently bonded atoms, but also the angles that include three consecutively bonded atoms and the torsion angles between two geometrical planes formed by connected atoms. The non-bonded interactions include long range VdW (Lennard-Jones) and electrostatic interactions (Coulomb).

$$U(\vec{r}) = \underbrace{\sum_{bonds} k_i (r_i - r_0)^2}_{U_{covalent\ bonds}} + \underbrace{\sum_{angles} k_i (\theta_i - \theta_0)^2}_{U_{angles}} + \underbrace{\sum_{torsion} k_i [1 + \cos(n_i \phi_i + \partial_i)]}_{U_{torsion\ angles}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{unbonded}} \qquad (3.4)$$

All energy terms are parametrized in the Force Field in which Molecular Dynamics simulations are conducted. To solve the equation of motion, only the initial spatial coordinates for all atoms in the system are needed, since the initial velocities are randomly assigned from the Maxwell-Boltzmann distribution at a given temperature.

The PDB file provides spatial coordinates (x, y, z), a B-factor and annotations of the residue type and number and chain for each atom of the protein (hydrogens excluded). Starting from the PDB, a PSF file needs to be generated to add bonds and angles between atoms based on a topology file proper of a specific Force Field. For each MD simulation

run for this thesis work, the protein was centered in a solvent box built according to the dimensions of the protein and padded with a 10 Å thick layer of water in every dimension. The solvent was modelled explicitly using TIP3 water molecules. Salt was added to ionize the solvent with 0.15M KCl. All simulations in this work were performed using the CHARMM27 force field[121] in NAMD[108]. Periodic boundary conditions were applied in which long-range electrostatic interactions were treated using the particle-mesh-Ewald method[122] and the cut-off distance was set to 12 Å. All simulations are run for at least 100 ns at constant temperature of 310 K.

Trajectory analysis

The preliminary analysis of the simulation trajectories included the calculation of Root Mean Square Deviation (RMSD) for the protein backbone and the Root Mean Square Fluctuations for the Carbon-alpha atoms of each protein residue. The calculations were conducted with the RMSD Trajectory tool and the Timeline programs within VMD tools. Correlation matrix and modes were obtained by python scripts. Only the first mode was considered for each simulation. Graphics were prepared with Graph Pad Prism 5 and Matplotlib.

SASA

Solvent accessible surface area (SASA) is the surface of a molecule accessible to solvent measured in square Angstroms ($Å^2$). It is calculated as the surface covered by a probe sphere rolling over a molecule and it is used to define exposed and buried area of the molecule. The probe sphere has radius equal to 1.4 Å that mimics the size of a water molecule. In this study, the SASA calculation on selected regions of the proteins analyzed by MD simulations was implemented on VMD by editing the sasa.tcl script available at: https://www.ks.uiuc.edu/Research/vmd/mailing_list/vmd-l/att-18670/sasa.tcl

4. RESULTS

4.1 PATZ1 protein: in-cell analysis

4.1.1 Subcellular localization of PATZ1

To identify the sub-cellular localization of the PATZ1 protein, fusion constructs expressing fluorescent protein tagged PATZ1 were generated. The full-length mouse Patz1 gene sequence was extracted from the pCMV-HA plasmid by EcoRI enzymatic digestion and inserted into the pEGFP-C3 vector to generate a fusion protein sequence of Patz1 N-terminally tagged with EGFP (enhanced green fluorescent protein). The new recombinant protein contains the EGFP sequence (744 bp) in frame with 40 bp followed by the Patz1 sequence (1926 bp), thus expressing an N-terminally EGFP tagged PATZ1. The construct was inserted into HEK293-T cells by PEI transfection. After incubation at 37°C for 16h, the cells were analyzed under fluorescence microscopy (Fig. 4.1).

The images clearly show that EGFP-PATZ1 is expressed and confirm its nuclear localization. The PATZ1 sequence contains two nuclear localization signals (NLS), one located within the AT-hook DNA binding motif and one between ZF1 and ZF2. It was already shown for a shorter isoform that PATZ1 localizes inside the nucleus within defined spots where it co-localizes with RNF4[67]. Because RNF4 is involved in the ubiquitination of SUMOylated proteins from the PML nuclear bodies (NB), PATZ1 could also be recruited to nuclear compartments involved in these activities.
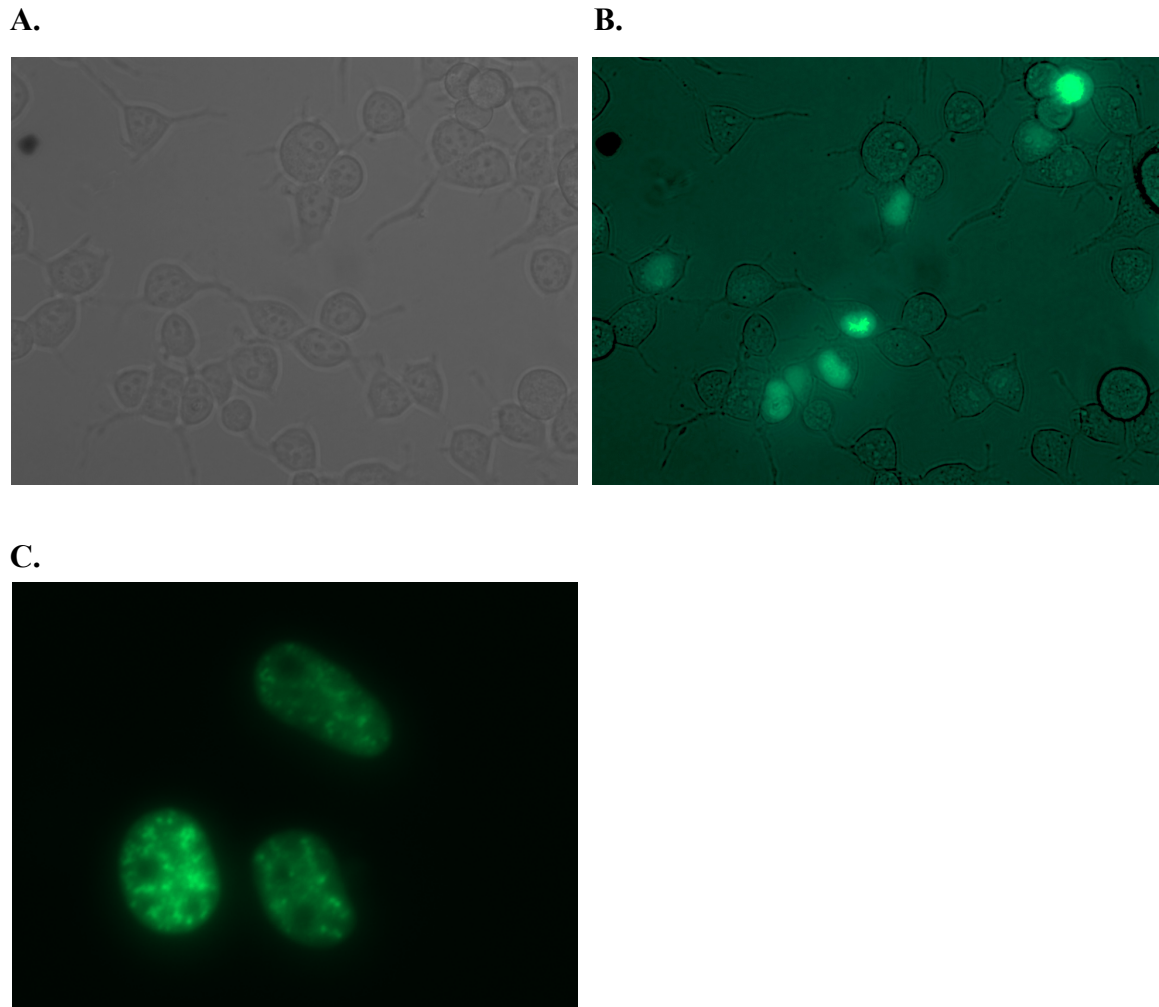
Figure 4.1 EGFP-PATZ1 nuclear localization under inverted fluorescence microscope (Zeiss). Images from the six-well plates are shown at 40X magnification. HEK-293T (A, in bright field) show PATZ1 nuclear expression (B, GFP excited). HeLa cells (C) shows EGFP-PATZ1 localization in nuclear speckles.

In fact, by using the GPS-SUMO web server tool[123, 124] (Available from http://sumosp.biocuckoo.org/online.php), it is possible to identify SUMO interaction motifs (SIM) along the PATZ1 sequence, within the BTB domain (Table 4.1). Consisting of a hydrophobic core[125], SIMs are potential targets for non-covalent interactions of SUMO proteins. These interactions are suggested to create a protein scaffold for the structure of the PML-NB[126]. While we have not addressed the domain requirements for specific localization and we have not definitively demonstrated that the speckles formed by GFP-PATZ1 are indeed PML-NB, we conclude that PATZ1 proteins accumulate in nuclear speckles, similar or identical to PML-NB, a characteristic common to BCL6 and PLZF[53, 55, 67, 69, 70].

Position	SIM	Score	P-Value
44 - 48	GGRFCDV LLRVG DESFPAH	18.389	0.412
126 - 130	DFAYTSR IVVRL ESFPELM	26.015	0.173
148 - 152	KFLLMRS VIEIC QEVIKQS	42.584	0.055
161 - 165	EVIKQSN VQILV PPARADI	29.823	0.197

Table 4.1 Prediction of SUMO Interaction Motifs (SIM) within PATZ1 BTB domain. The underlined sequences were selected by GPS-SUMO without stringency threshold.

4.1.2 The Patz1-Chimera sequence

Excluding the 31-residue long sequence forming the central loop that is present in mammals but absent in fish homologs, the BTB domain of mouse and zebrafish PATZ1 share over 75% aa sequence identity and over 80% similarity. In pursuing the aim of studying the role of the central loop region, the zebrafish BTB domain can be considered as a mutant version of the mouse PATZ1. In fact, functional experiments can be carried on to compare mouse and fish BTB proteins. To this end, we generated chimeric PATZ1 protein expression constructs encoding the BTB domain of zebrafish PATZ1 and the AT-hook and zinc finger domains of murine PATZ1. The experiments with this chimeric protein are ongoing but information about its generation are presented here.

The chimeric PATZ1 expression plasmid was constructed to include not only the BTB domain but the full architecture of the PATZ1 protein for comparative cellular studies. The Chimera-PATZ1 sequence was constructed to contain the BTB sequence of the zebrafish PATZ1 fused to the human PATZ1 C-terminal Zinc fingers. The fusion point was designed to have two complementary hanging tail sequences from the N-terminal (head) and C-terminal (body) parts of the Chimera structure. The construct was assembled in three steps (Table 4.2).

1. For the first step a PCR was set to amplify the zebrafish PATZ1 BTB sequence (~400 bp) already cloned into pET-47b (+) plasmid from zebrafish genomic cDNA. The FW primer (FW1) carries the MfeI restriction site. The RV primer (RV1) was designed to add a 3'-end tail to the amplified sequence.

2. With a second PCR the mouse PATZ1 ZF sequence (~1450 bp) was amplified from the pCMV-HA plasmid. In this case the FW primer (FW2) was design to add a 5'-end tail to the amplified sequence. The RV primer (RV2) carries an extra MfeI restriction site and the TGA stop codon.

3. In the third PCR, the products of the first and second amplification were run on Agarose gel, gel extracted, purified and used as template. The product sequence (~1850 bp) of this third step is the fusion of the zebrafish PATZ1 BTB domain (head) and mouse PATZ1 ZF (body) sequences.

The final fused sequence was also gel extracted and ligated into an empty pCMV-HA plasmid digested with EcoRI enzyme (Fig. 4.2). The EcoRI digested site on the vector backbone is compatible with the hanging hands of MfeI contained in the 5' and 3' of the Chimera sequence insert derived from FW1 and RV2. Although the new fusion vector lost the original EcoRI and MfeI restriction sites, this cloning step of using two different but compatible enzymes was necessary because the zebrafish PATZ1 BTB sequence contained an EcoRI cut site that complicated the cloning procedure.

Reaction	FW primer	RV primer	Template	Product size	
PCR 1	**FW1**	RV1	pET47-ZFPB	400bp	Head
PCR 2	FW2	**RV2**	pCMV-HA-PATZ1	1450bp	Body
PCR 3	**FW1**	**RV2**	Head & Body	1850bp	Chimera

Table 4.2 3-steps cloning strategy for Chimera Patz1 construct.
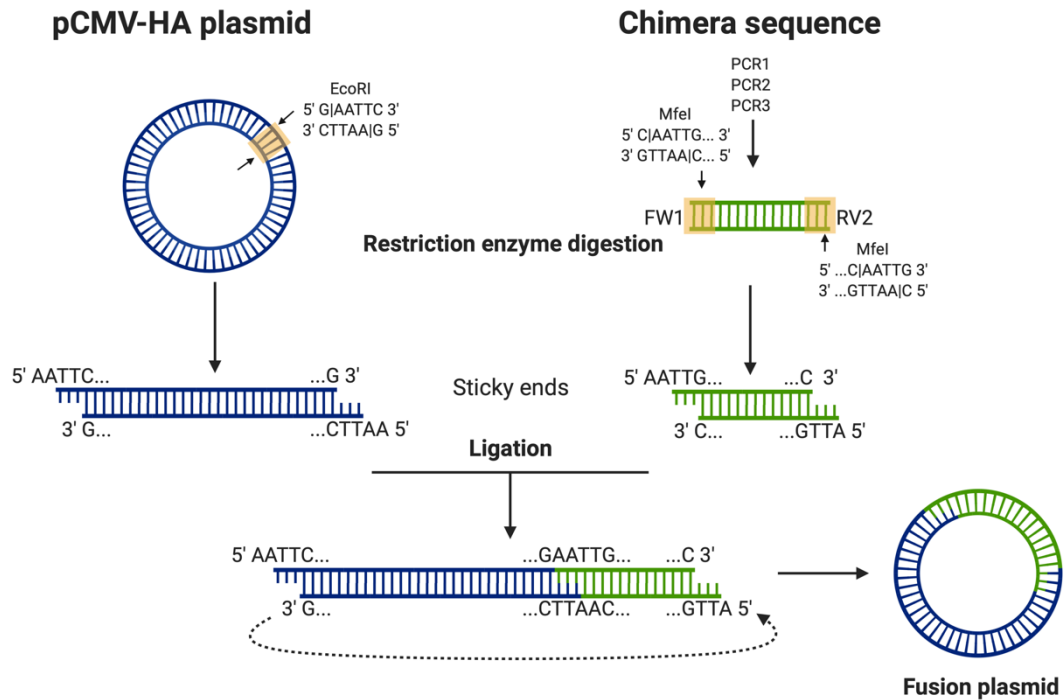
Figure 4.2 Cloning of PATZ1 Chimera Sequence with compatible restriction sites strategy. Created with BioRender.com

The PATZ1 Chimera construct is designed to bind the mammalian PATZ1 target DNA in the cell line of choice and differ only within the BTB domain.

The comparison study of the mouse and zebrafish PATZ1-BTB domains could begin with an immunoprecipitation experiment with antibodies targeting the HA epitope tag on the N-terminus of the PATZ1 protein, followed by western blotting the NCOR co-repressor to confirm that PATZ1 binds to NCOR. In the work from Bilic et al., 2006[60], PATZ1 has already been shown to bind to endogenous NCOR1. The binding site was studied by mutating three residues (L27S, Q33S and R34L) in the PATZ1 BTB domain. In a similar experiment for BCL6, the corresponding residues L19S, N23H and L25S/R26L were also mutated and were necessary for co-repressor binding[53]. After these mutations both PATZ1 and BCL6 failed to bind NCOR1 suggesting that a similar interface on their BTB domains may be mediating this interaction.

4.2 PATZ1 protein: structural study

4.2.1 PATZ1 protein fragments expression and purification

In order to obtain structural information about the PATZ1 protein, 18 alternative constructs were designed containing different regions of the Patz1 sequence as candidates for expression, purification and crystallization (Fig. 4.3).
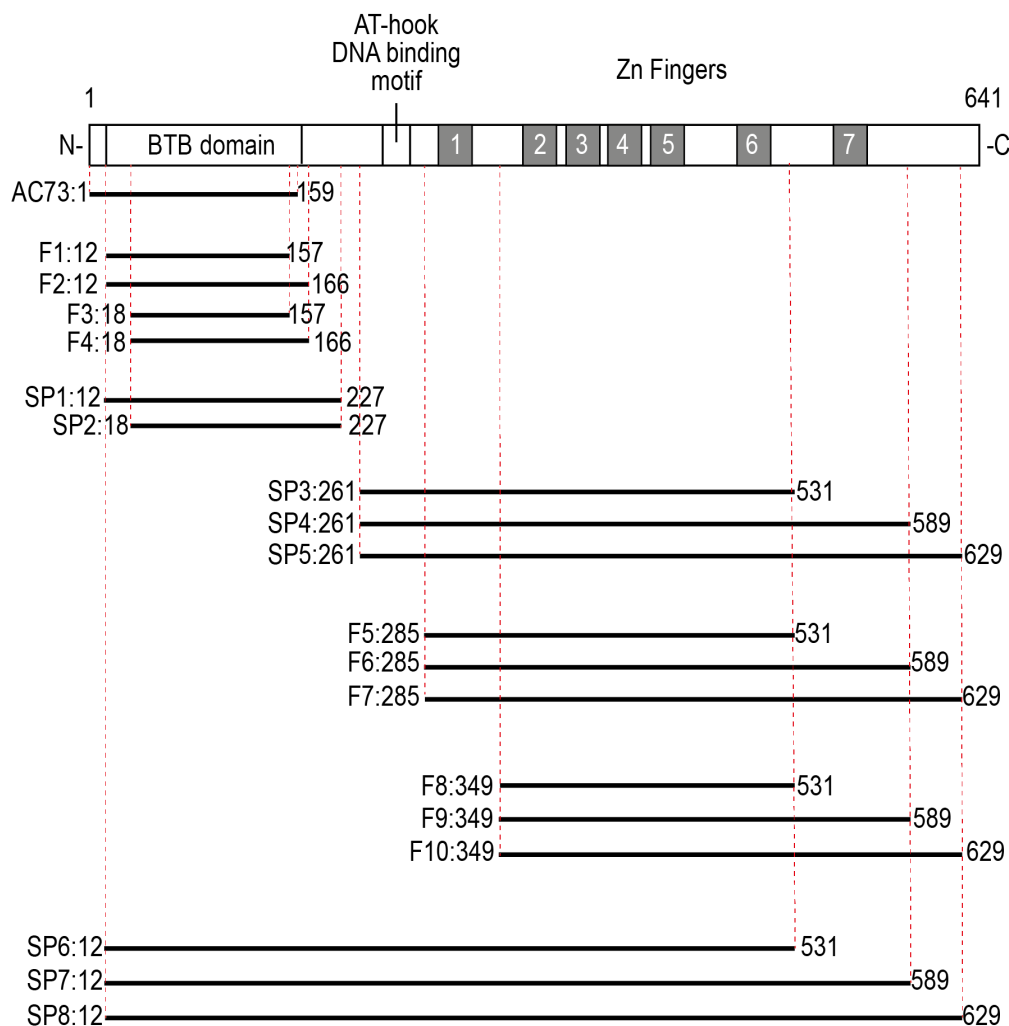
Figure 4.3 Schematic of the 18 fragments of the Patz1 sequence designed for expression and protein crystallization experiments. Numbers correspond to N-and C-terminal residues encoded by the construct.

All sequences were PCR-amplified from a pCMV-HA plasmid construct[48] containing the full-length mouse PATZ1 cDNA (isoform A, encoding 641 aa).

The fragments F1-F10 were designed in the earliest experimental setup. Each sequence was designed and cloned into three different bacterial expression plasmids pET-47b (+),

pET-49b (+) and pET-52b (+) (Novagen), all provided by Prof. Dr. Erika Mancini's Lab at University of Sussex, UK. F1-F4 comprise only the N-terminal BTB domain, while F5-F7 only the C-terminal Zn Finger domains. The fragments SP1-SP8 were designed in a second moment with the intent of including larger portions of Patz1 sequence.

The three cloning vectors are engineered to have the same backbone and cloning site but differ essentially about the fusion tag designed to label the recombinant protein. Each tag requires a different method of protein purification by affinity. In particular, pET47 contains a N-terminal 6xHis•Tag coding sequence, pET49 contains a GST•Tag and pET52 carries a N-terminal Strep•Tag II and a C-terminal His•Tag.
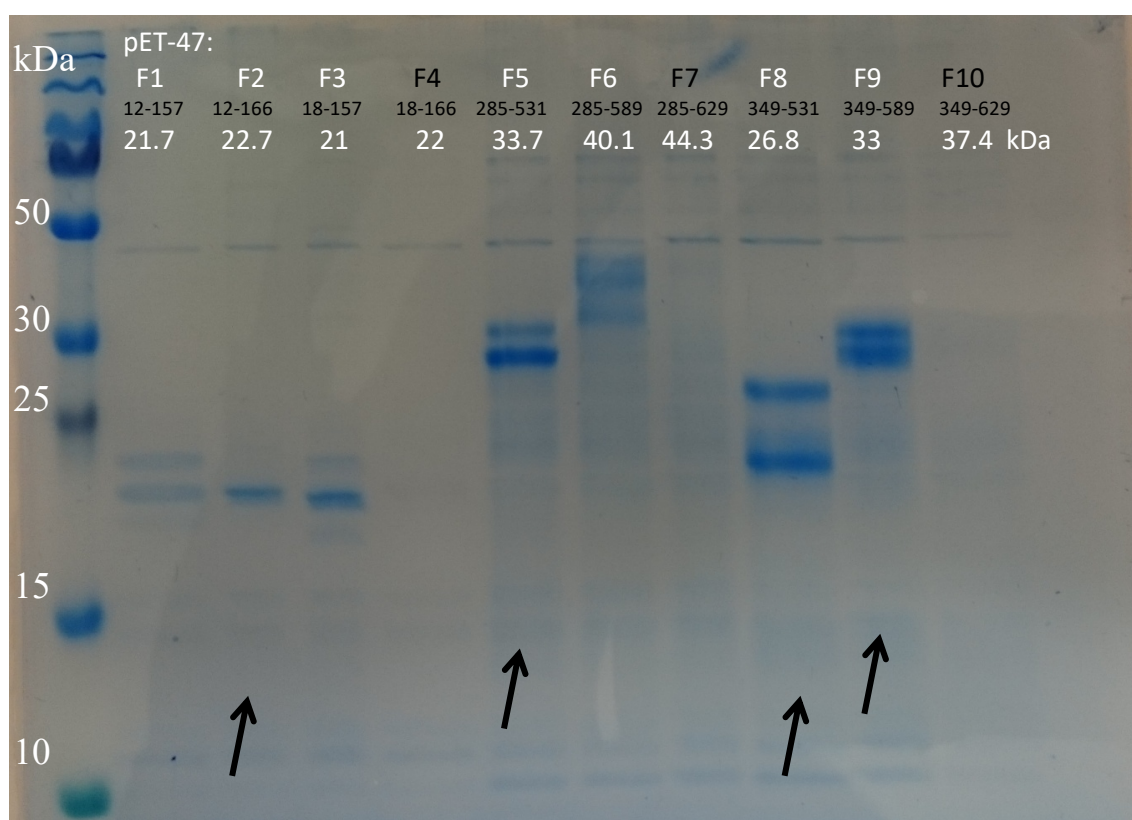
Figure 4.4 SDS gel of small-scale expression assay of fragment F1-F10 in pET-47b (+).

All fragments sequences were inserted within the vectors cloning site between KpnI (then redesigned in XmaI/SmaI) and NotI restriction sites. Initially, all the fragment sequences included an additional 35-residue long C-terminal tail from the vector backbone that was later removed with a new design. The ligated plasmids were firstly transformed in E. coli DH5α cloning strain, and then in Rosetta™(DE3) pLysS expression strain after sequencing. In the first expression test, a small scale 50 ml bacterial culture for each

sequence F1-F10 in pET-47b (+) was grown, purified by affinity and analyzed by SDS gel (Fig. 4.4). Based on the expected protein size (kDa) and the purity of the bands on the SDS gel, F2-F5-F8-F9 were selected for larger scale (1 L) protein expression assay. The proteins were then additionally purified by size-exclusion chromatography (SEC) (Fig. 4.5).
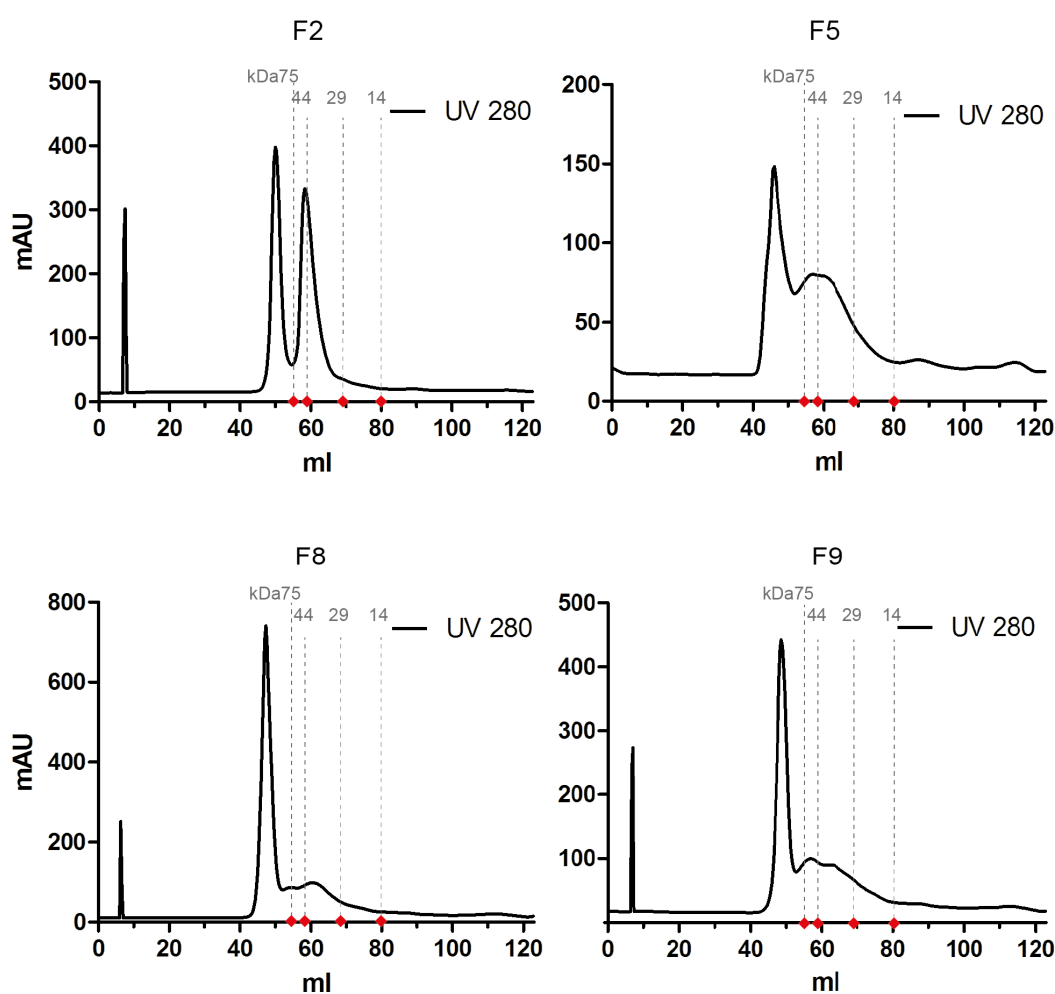
Figure 4.5 The figure shows the SEC profiles for F2-F5-F8-F9. The constructs were purified on AKTA pure (GE) through a HiLoad SD75-1660 calibrated column. Red dots correspond to the elution points of marker proteins used in column calibration (grey dashed lines).

The fragment F2 (Patz1 12-166) covers the region of the BTB domain. Although one peak overlaps with a DNA peak (254 nm), the fractions that were collected after SEC under the latest peak revealed a very pure sample on SDS gel. These fractions eluted at a

point that correspond to a 29-44 kDa protein, about double the size expected for F2. For this reason, this protein appeared to be a dimer. All collected fractions containing the pure protein were pulled together and concentrated. 4 Crystallization trays with two different protein concentrations (16 and 6mg/ml) were set for this protein. Although no crystal was observed to grow in these conditions, the construct remained the most promising for further attempts.

The purification analysis of the fragment F5 (Patz1 285-531) including the first 6 ZFs did not reveal the presence of well separated and pure protein by SEC. Similarly, fragment F8 (Patz1 349-531) and F9 (Patz1 349-589) were collected in a fraction overlapping a significant peak for DNA. Because these proteins cover the ZF region, DNA may be strongly bound and cannot be easily separated by these purification methods.

The fragments SP1-SP8 were all set for cloning in pET-47b (+) (Fig. 4.6). Expression and affinity purification tests were successfully carried on only for SP3-SP4-SP5. These fragments include the AT-hook motif followed by 6 or 7 ZF. SP3, the shortest of these fragments was selected for further purification steps although it did not return clear results.
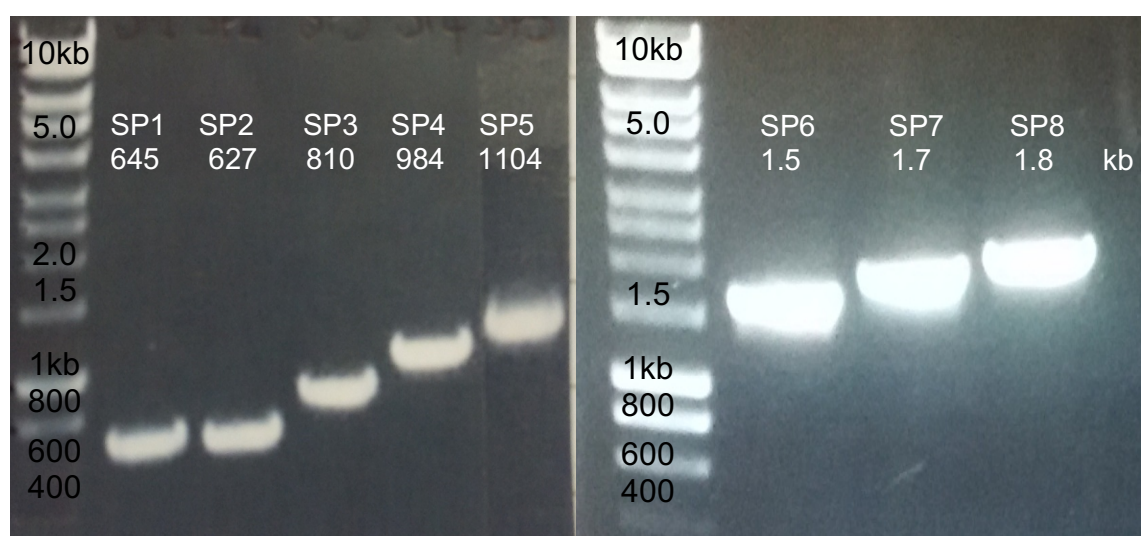
Figure 4.6 PCR amplification of SP1-8 fragments of PATZ1 on agarose gel.

Different PATZ1 fragments were tested for expression and alternative purification methods were attempted. Despite the unremarkable results, some considerations about the purification of PATZ1 fragments that include DNA binding domains could be useful for future work:

- Preliminary experiments show that PATZ1 constructs whose sequence terminates at residue 531, including 5 or 6 Zn Fingers (F5, F8 and SP3), before the structurally unknown region between ZF6 and ZF7, have better expression;
- The protein fragments covering the PATZ1 DNA binding domains are mostly positively charged (15% of residues), with a theoretical isoelectric point (pI) = 9. This information could be useful for an alternative purification method by ion-exchange chromatography;
- The affinity purification test shows that the elution of His tagged proteins pulled down on Co-beads resin (Talon) yields better results when using IMAC buffer with Imidazole concentrated at 600mM;
- The described affinity purification method is not efficient against the high amount of DNA contamination that makes the concentration of the eluted protein very difficult (in 10000MWCO concentrators);
- Desalting methods attempted (250mM to 10mM NaCl) result in protein precipitation and should be avoided in these conditions;
- Ion-Exchange Chromatography with Heparin column and elution with a gradient of salt concentration (200mM to 2M NaCl) is a critical step. Proteins were eluted at ~37% of the gradient (~600mM NaCl);
- Due to the high affinity of these fragments for DNA, a specific target DNA sequence can be introduced in the purification process to stabilize the constructs;
- Different affinity purification methods, including GST tag purification, could be tested.

4.3 Mouse PATZ1 BTB domain

4.3.1 Protein purification

For the new design of F2 fragment (Patz1 12-166), the C-terminal tail from the vector backbone was removed making the new protein about 3 kDa smaller. The construct cloned in pET-47b (+) plasmid was transformed in the bacterial expression strain and incubated in 1 L TB culture. The cell lysate was purified by IMAC. The samples eluted with 100 and 300 mM Imidazole-containing buffers were concentrated together to 2.6 mg/ml in 2.5 ml and loaded on FPLC system for SEC purification (input). In the various purification experiments, PATZ1 BTB domain grown in 1L TB yields ~6mg of pure protein.
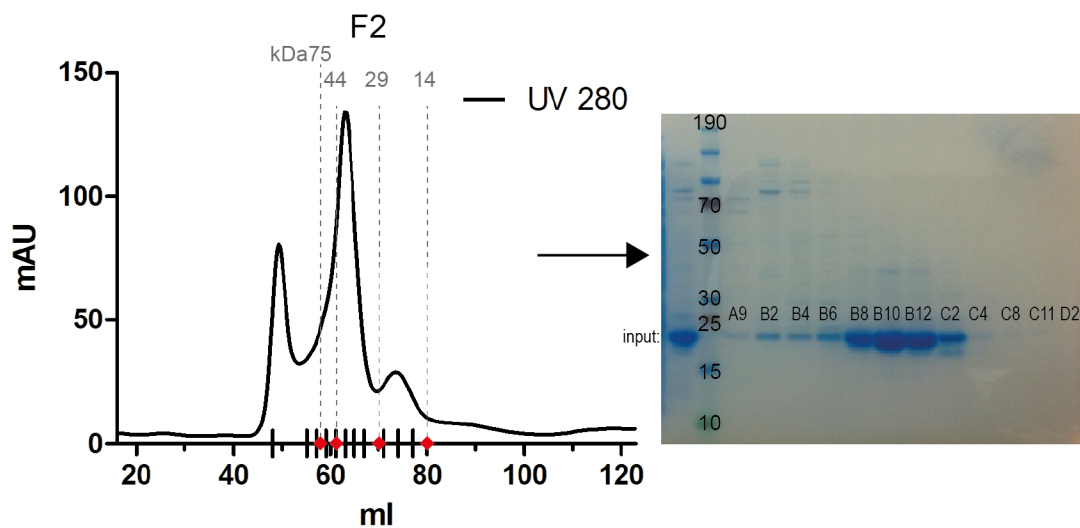
Figure 4.7 SEC purification profile of F2 protein, PATZ1(12-166) and fractions analysis on SDS gel. The BTB domain (19 kDa) elution point on the SD75 16/600 column corresponds to a dimeric conformation of the protein (29-44 kDa). Vertical lines on the X-axis indicate elution points of the fractions collected for the gel. Red dots correspond to the elution points of marker proteins used in column calibration (grey dashed lines).

The fractions collected according to the SEC profile, were then analyzed on SDS gel. The SEC profile (Fig. 4.7) shows three peaks: one overlaps with a DNA peak and correspond

to a very faint protein band at the expected size for F2; the second corresponds to separated molecules in a size range between 29-44 kDa according to the column calibration markers. This protein was considered to be the His tagged BTB domain region in homodimer conformation (~38kDa). The fractions under the third peak do not contain detectable amount of proteins. The samples collected between B10-C2 fractions were pulled together to 9.3 mg/ml concentration.

4.3.2 Protein crystallization

All crystallization experiments were performed using the sitting-drop vapor-diffusion method at the University of Sussex crystallization facility. Thanks to a robotic liquid dispensing system (Phoenix, Art Robbins Instruments) the protein sample was distributed into five 96-well plates, each containing 96 different chemical precipitants to assay the best crystallization conditions specific to this protein. Each well was setup to contain the screen solution and 0.2 µl of the pure protein. The 5 crystallization trays (Screen Index (High throughput-Hampton Research); PACT premier (pH, Anion, Cation); JCSG-plus; Structure Screen 1+2; Morpheus) were sealed and stored at +20°C. Crystals of the mouse PATZ1 BTB domain appeared after 72 h (Fig. 4.8) in 0.1 M MMT (dl-malic acid, MES monohydrate, Tris), 25% (w/v) PEG 1500.

Figure 4.8 Protein crystals under optical microscope.

For data collection, single crystals were briefly immersed in mother liquor supplemented with 20% glycerol prior to flash cooling in liquid nitrogen. For mouse PATZ1 BTB, data were collected to 2.29 Å resolution on beamline I04 at Diamond Light Source, Didcot,

UK. The diffraction data were indexed, integrated, scaled and reduced with xia2[127] and AIMLESS[128]. The space group was P41212 (unit-cell parameters a = b = 43.23, c = 162.95 Å, $\alpha = \beta = \gamma = 90°$), with one molecule in the asymmetric unit. The structure of mouse PATZ1 BTB was solved by molecular replacement with Phaser[129] using the PLZF BTB structure (PDB entry 1BUO;[16] as a search template. The identified solution was then subjected to rounds of manual rebuilding with Coot [130] and refinement with Phenix[131] to give final R_{work} and R_{free} factors of 20.9% and 24.5%, respectively. The final structure was validated with MolProbity[132] and deposited together with the structure factors in the Protein Data Bank as entry 6GUV.

4.3.3 Protein structure

As previously seen in other members of the ZBTB family, the PATZ1 BTB crystal structure reveals a strand-exchange homodimer, organized as a core fold BTB domain preceded by an N-terminal extension (Fig. 10). The characteristic secondary structures of the dimerization interfaces ($\beta 1–\alpha 1$, A1–A2 and $\beta 2–A5$) are conserved. The structure of the mouse PATZ1 BTB domain occupies a volume of 70 nm^3 (~40x60x30 Å).

The murine PATZ1 BTB domain protein was expressed from a construct encoding amino acids 12–166 preceded at the N-terminus by 20 amino acids comprising a His tag and an HRV-3C protease digestion site. The last ten of these amino acids (ALEVLFQGPG) are visible in the structure and fold into a β-strand ($\beta 0$) antiparallel to and interacting with the first N-terminal PATZ1 β-strand ($\beta 1$) (Fig. 4.9).
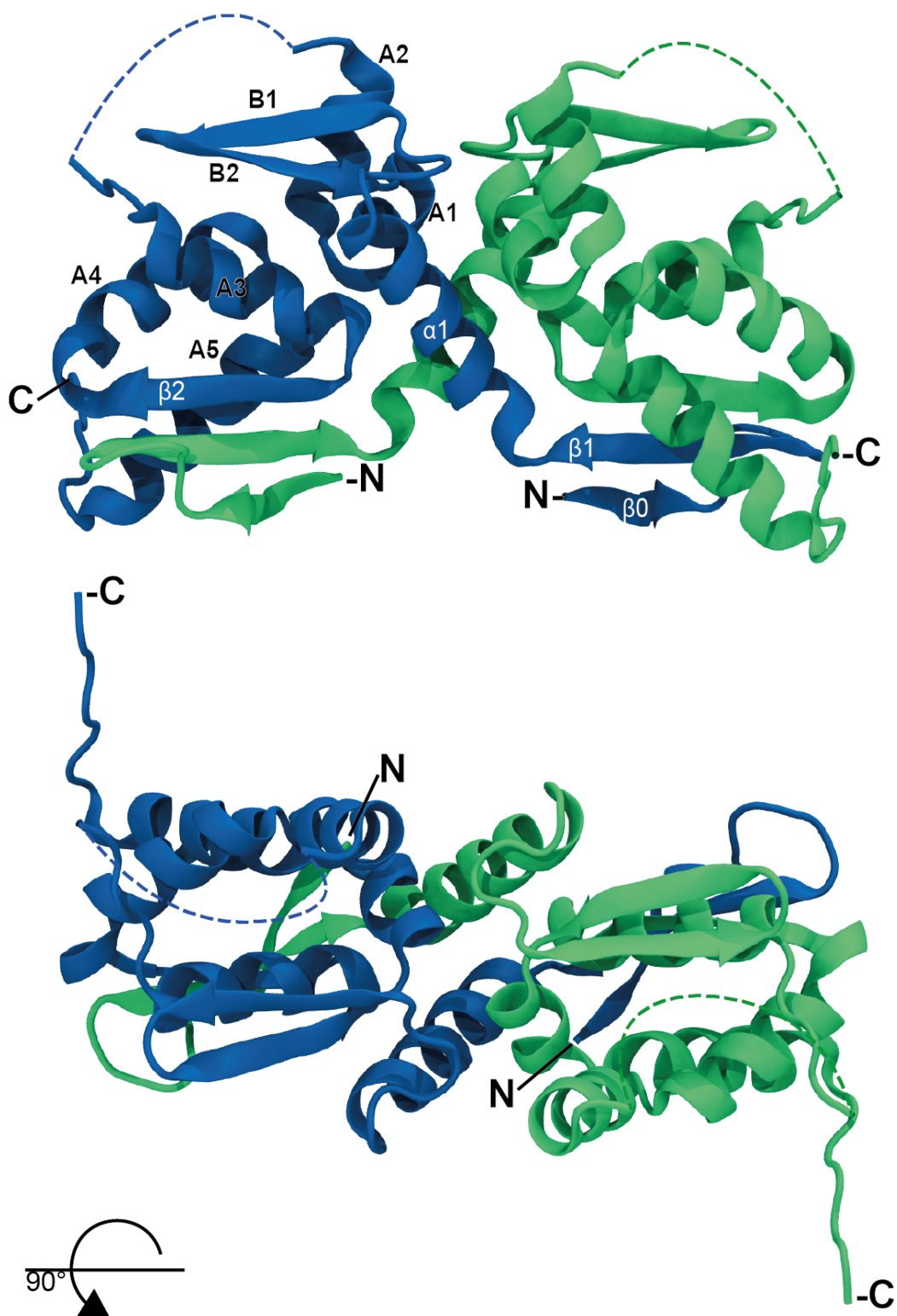
Figure 4.9 Crystal structure of the mouse PATZ1 BTB domain homodimer (front and top view). The two monomers are colored in blue and green. Secondary structure elements are indicated. β0 belongs to the vector backbone. The dashed line represents the region with missing electron density.

Superposition of this structure with that of BCL6 BTB in complex with co-repressor peptides (PDB entries 1R2B and 3BIM[18, 21] suggests that these extra residues structurally mimic the β-strand-forming residues in both SMRT and BCOR, although their amino-acid sequence is not conserved. Attempts to crystallize the mouse PATZ1 BTB construct following cleavage of the His tag failed, suggesting that the extra N-terminal amino acids are likely to aid the crystallization process in this case.

In the crystal structure no density could be assigned to the residues belonging to a 31-amino-acid A2/B3 loop (residues 75–105), suggesting that these amino acids are partially disordered or flexible. The mammalian PATZ1 A2/B3 loop is only partially conserved in cKrox (ZBTB15), but otherwise unique PATZ1 among the other ZBTB family members. In fact, this large loop in PATZ1 BTB domain replaces a shorter amino-acid stretch that forms a β-strand (B3) in other ZBTB family proteins, as described in detail by Stogios et al. (2005)[2]. This loop is glycine and alanine-rich and is predicted to be partially disordered; however, a short β-strand is predicted for the last six residues. The single difference between the human and mouse PATZ1 BTB domains (T91A) is found within this large loop.

In the crystal unit cell, seven amino- acid stretch from the C-terminus of the BTB domain (160-166) an extends into the region on an adjacent molecule normally occupied by the B3 β-strand in other ZBTB proteins (Fig. 4.10). This crystallization artefact presumably aids crystal packing, as the 'B3 strand mimic' appears to stabilize the β-sheet formed by B1 and B2. To test this hypothesis, the AC73 construct (Patz1 1-159) was crystallized after His tag cleavage. The crystals of this protein diffracted poorly and to lower resolution (3.4 Å), showing the same homodimeric BTB domains in a different crystal packing (data not shown). This suggests that these seven amino acids were in fact important for stabilizing the B1–B2 β-sheet and for crystal packing (hence the better diffraction), yet their absence did not encourage the folding of the A2/B3 loop.
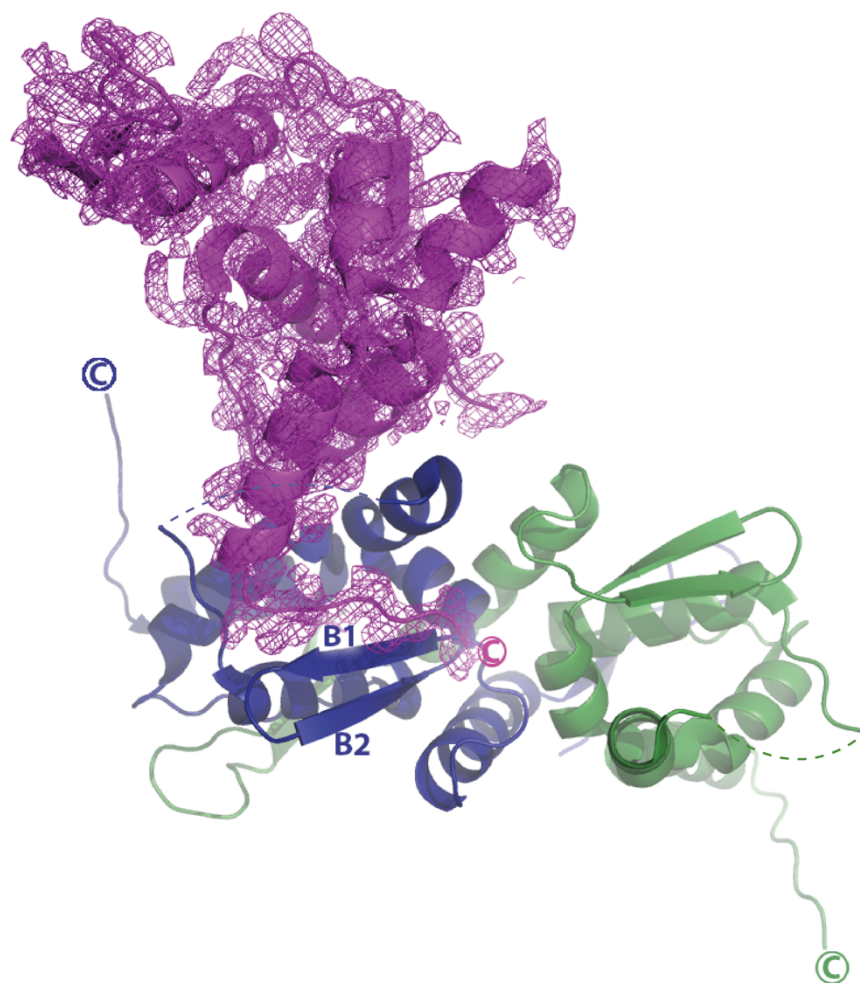
Figure 4.10 Intermolecular artifact within the crystal unit cell. Two PATZ1 BTB monomers forming a homodimer in the asymmetric unit are shown in green and blue respectively (top view). A third monomer from an adjacent molecule (shown in magenta with additional electron density) contributes to the β-sheet formed by B1 and B2. The missing residues of murine PATZ1 A2/B3 loop are indicated by dotted lines.

4.3.4 Protein Modelling and docking

The central loop of the mammalian PATZ1 BTB structure (residues 83–113, numbering referred to the residues in the crystal structure, PDB entry: 6GUV) for one monomer was modelled using the PRIMO suite[109] based on the MODELLER program[112]. Structural information from MEME motif analysis of the loop sequence was added to the template[133]. The obtained model was aligned with the deposited structure in PyMOL (version 1.8; Schrödinger). The coordinates of the loop model were added to the crystal

structure and fragments were joined using the VMD AutoPSF plugin. SymmDock[110] was used to reconstruct the dimer conformation.

The molecular modelling was performed for the human protein sequence, namely by modelling position 99 as a Threonine residue instead of Alanine, from the mouse PATZ1 sequence. The detailed study of the possible structure and dynamic behavior of the A2/B3 loop in human PATZ1 BTB, was addressed by molecular dynamics simulations (MD). MD simulations lasting 200 ns indicated that while the overall dimer forms a stable structure, the A2/B3 loop region is uniquely flexible. The simulations also suggest that a new β-strand could form within this loop (Fig. 4.11) for the predicted B3 residues. Other BTB domains such as LRF and MIZ1 also contain flexible loops in this region[19, 20, 32]. In the case of MIZ1, the A2/B3 region mediates tetramerization of its BTB domain, whilst using size-exclusion chromatography (Fig.4.12) no evidence was found for such stable oligomerization in mouse or zebrafish PATZ1.
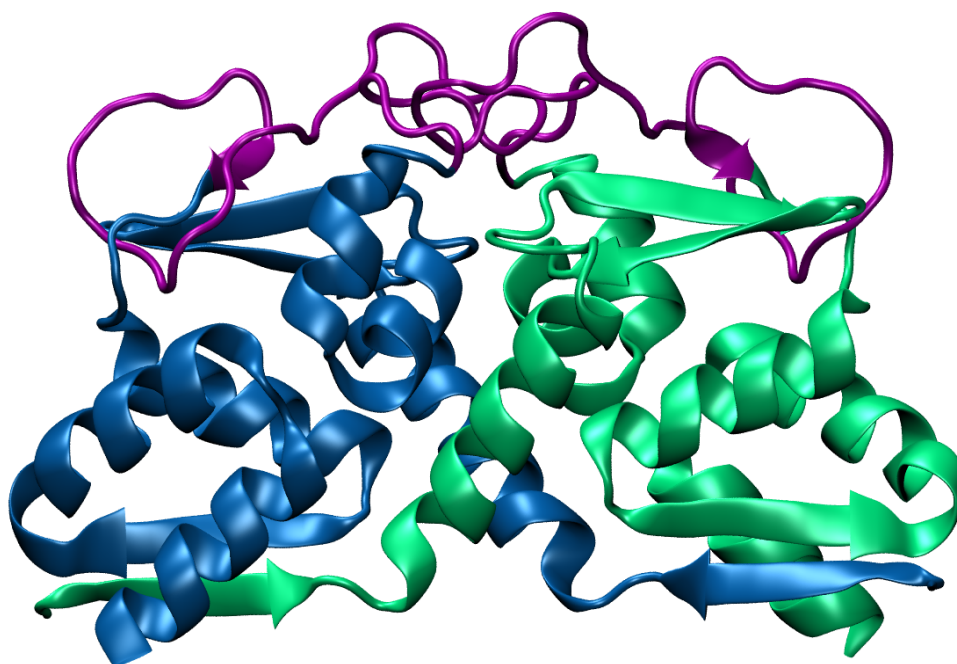
Figure 4.11 Cartoon representation of the mouse PATZ1 BTB domain structure. The missing central loop structure (purple) was computationally modelled.

4.3.5 Tetramerization hypothesis

In order to analyze the hypothesis of an additional stable conformation for the PATZ1 BTB domain besides the dimeric structure, a sequential SEC experiment was set to analytically decompose the possible oligomerization states. The AC73 construct (PATZ1 1-59) was cloned into pET-47b (+) plasmid for His tag affinity purification. Differently from F2 fragment (PATZ1 12-166), AC73 contains the full length BTB domain N-terminal sequence and 7 additional C-terminal residues are excluded from it.

SEC purification profile shows the separation of two main structural conformations for this protein, the most abundant being the BTB dimer under the peak 29-44 kDa. A relatively shorter protein peak of higher size consists of an aggregate conformation of PATZ1 that could indicate the presence of a tetramer form for the BTB domain, by size comparison. For two consecutive SEC experiments (Fig. 4.12), the fractions eluted from the two peaks were collected independently and concentrated for a new run on a different column of different volume. Only the SEC profile from the "tetramer" peaks elution shows again a separation of two main structural conformations, the most abundant being the BTB dimer under the peak 29-44 kDa preceded by a shorter peak of higher size. On the contrary, the fractions eluted under the dimer peaks do not show any additional oligomerization form.

These results suggest that if a tetrameric form is accessible to PATZ1 BTB domain, an equilibrium should exist between this and the dimeric conformation. In fact, while the dimer peak fractions remain stable in a single conformation over sequential purifications, the tetramer peak divides into two at every step with the conformations ratio always moved towards the dimer form.
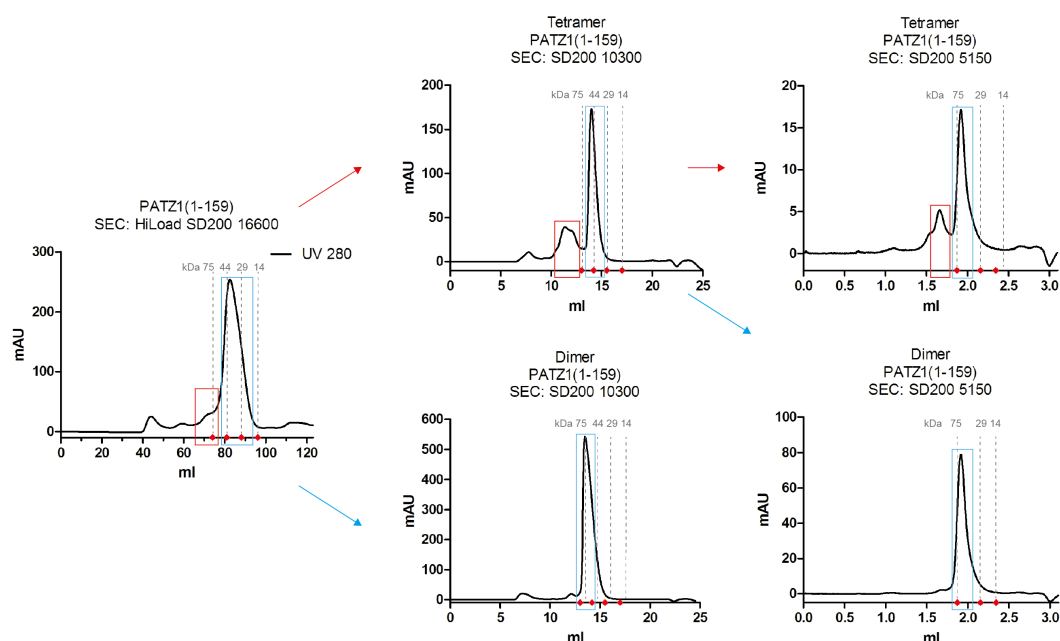
Figure 4.12 Study of the His tagged mouse PATZ1 (1-159) BTB domain by SEC. The red and blue boxes indicate two main structural conformations. A short protein peak of higher size (red box) consists of an aggregate conformation of PATZ1 while the most abundant is the BTB dimer under the peak 29-44 kDa (blue box). The protein sample under the first red and blue peaks in the first graph from left, were collected, re-concentrated and loaded on a second column (left red and blue arrows). Following the red arrow, the new SEC profile deriving from the aggregate conformation shows again a separation of two structural conformations in which the BTB dimer is still the most abundant (blue box) while a protein peak of higher size is still present (red box). Following the first blue arrow, the new SEC profile deriving from the blue peak shows a single peak consisting of the BTB dimer conformation (blue box). The protein samples under the red and the blue peaks are again collected, concentrated and loaded on a third column (right red and blue arrows). Again, the SEC profile deriving from the aggregate conformation separates two peaks of different size (red and blue box) while the SEC profile deriving from the BTB dimer conformation shows a single peak. Following the blue arrows: the protein sample under the first blue dimer peak was collected, re-concentrated and loaded on a second column (left blue arrow). The new SEC profile shows a single peak consisting of the BTB dimer conformation (blue box). The first SEC column used is a HiLoad SD200 16/600 (bed volume 120 ml); the second column is a SD200 Increase 10/300 (bed volume 24 ml); the third column is a SD200 Increase 5/150 (bed volume 3 ml).

4.4 Zebrafish PATZ1 BTB domain

4.4.1 Protein purification

The zebrafish PATZ1 BTB domain construct (1-135) was obtained by amplification of the zebrafish genomic DNA (kind gift of Prof. S. H. Fuss) and cloned in pET-47b (+) vector. It was then transformed in E. coli Rosetta2 bacterial expression strain and incubated in 2 L TB culture. The cell lysate was purified by IMAC. The samples eluted with 300 mM Imidazole-containing buffers were concentrated and loaded on FPLC system for SEC purification (input).
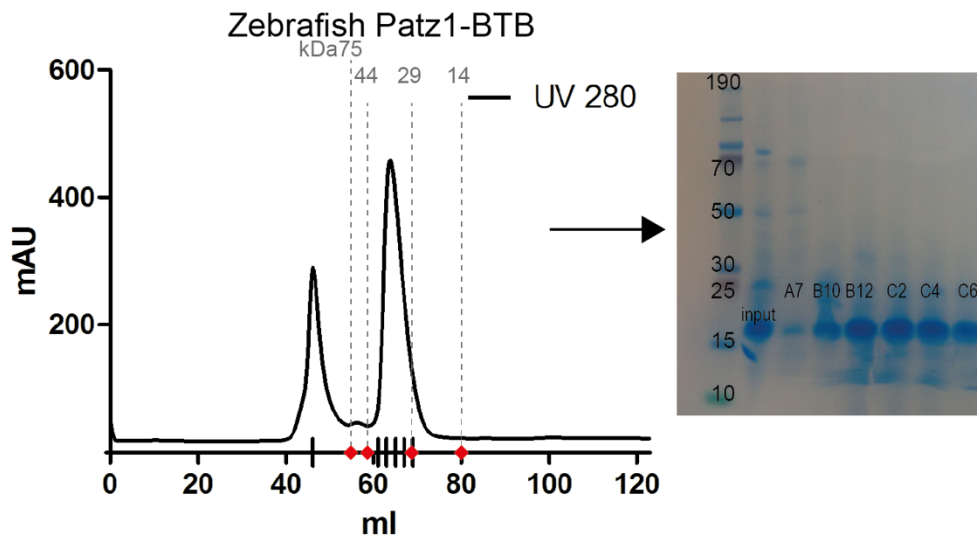
Figure 4.13 SEC purification profile of the zebrafish Patz1- BTB protein (1-135) and fractions analysis on SDS gel. The BTB domain (17.7 kDa) elution point corresponds to a dimeric conformation of the protein (29-44 kDa). Vertical lines on the X-axis indicate elution points of the fractions collected for the gel. Red dots correspond to the elution points of marker proteins.

The fractions collected according to the SEC profile, were then analyzed on SDS gel. The SEC profile shows two peaks (Fig. 4.13): one overlaps with a DNA peak and correspond to a very faint protein band (A7) at the expected size for F2; the second corresponds to separated molecules in a size range between 29-44 kDa according to the column calibration markers. This protein was considered to be the His tagged BTB domain region

66

in homodimer conformation (~35.4 kDa). The samples collected between B12-C4 fractions were pulled together to 8.5 mg/ml concentration.

4.4.2 Protein crystallization

As for the mouse orthologue, also for the zebrafish PATZ1 BTB domain crystallization experiments were performed using the sitting-drop vapor-diffusion method at the University of Sussex crystallization facility. The protein sample was distributed by automated liquid dispensing system into five 96-well plates, each containing 96 different chemical precipitants to assay the best crystallization conditions specific to this protein. Each well was setup to contain the screen solution and 0.2 µl of the pure protein. The 5 crystallization trays (Screen Index (High throughput-Hampton Research); PACT premier (pH, Anion, Cation); JCSG-plus; Structure Screen 1+2; Morpheus) were sealed and stored at +20°C. Crystals of the zebrafish PATZ1 BTB domain appeared after 72 h (Fig. 4.14) in 40%(v/v) PEG 500 MME, 20%(w/v) PEG 20K, 0.1 M Tris pH 8.5, 0.06 M magnesium chloride hexahydrate, 0.06 M calcium chloride dihydrate.
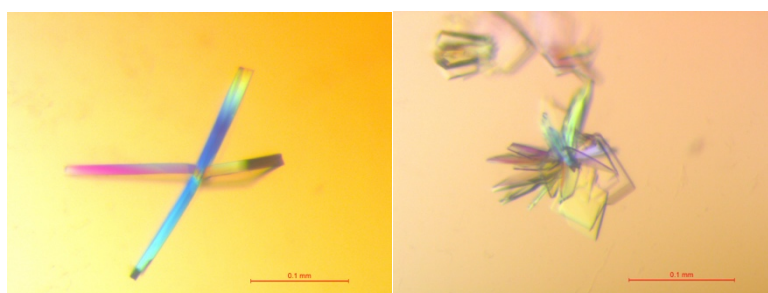
Figure 4.15 Protein crystals under optical microscope.

For data collection, single crystals were briefly immersed in mother liquor supplemented with 20% glycerol prior to flash cooling in liquid nitrogen. For zebrafish PATZ1 BTB, data were collected to 1.8 Å resolution on beamline I04 at Diamond Light Source. The diffraction data were indexed, integrated, scaled and reduced with xia2 and XDS[134]. The space group was P3121 (unit-cell parameters a = b = 43.08, c= 123.64 Å, $\alpha = \beta = \gamma = 90°$), with one molecule in the asymmetric unit. The structure of zebrafish PATZ1 BTB was solved by molecular replacement with Phaser using the mouse PATZ1 BTB structure (PDB entry 6GUV) as the search template. The identified solution was then subjected to rounds of manual rebuilding with Coot and refinement with Phenix to give final R_{work} and

R_{free} factors of 21.5% and 22.4%, respectively. The final structure was validated with MolProbity and deposited together with the structure factors in the Protein Data Bank as entry 6GUW.

4.4.3 Protein structure

The zebrafish PATZ1 BTB domain was expressed from a construct encoding amino acids 1–135 preceded at the N-terminus by 20 amino acids comprising a His tag and an HRV-3C protease digestion site. The 20 amino acids at the N-terminus and the first ten residues of the zebrafish PATZ1 BTB domain are not visible in the electron-density map. In addition, seven amino acids belonging to the A2/B3 loop are also missing from the electron density. The zebrafish BTB domain has the same quaternary structure as the mouse BTB domain: a strand-exchange homodimer (Fig. 4.15). Owing to the absence of the long disordered A2/B3 loop, the zebrafish PATZ1 BTB domain is structurally more similar to other ZBTB proteins than the mammalian PATZ1 BTB domain.

The structure of the zebrafish PATZ1 BTB domain occupies a volume of 50 nm^3 (~35x56x26 Å), slightly smaller than the mouse counterpart.
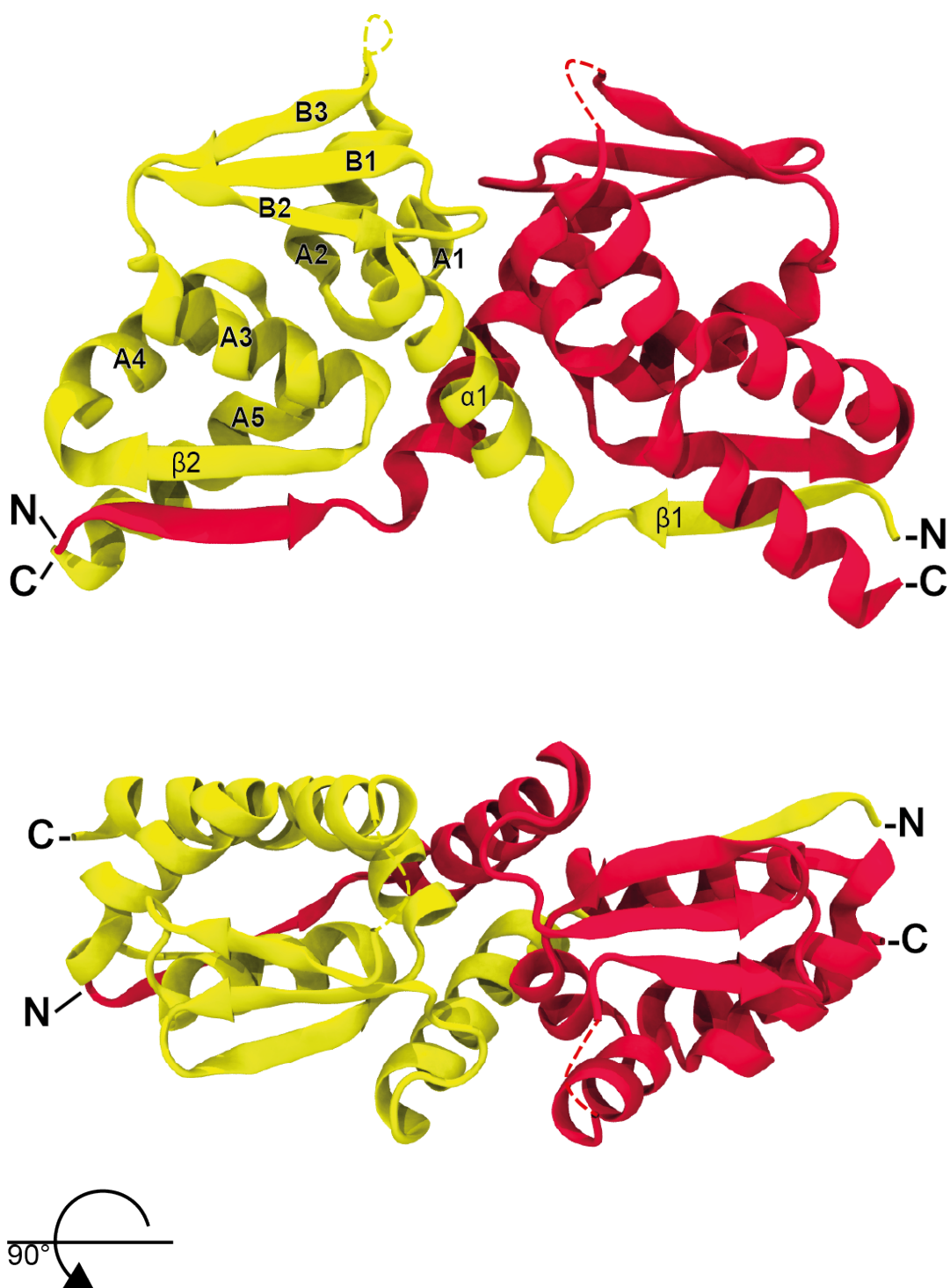
Figure 4.15 Crystal structure of the zebrafish PATZ1 BTB domain homodimer (front and top view). The two monomers are colored in yellow and red. Secondary structure elements are indicated. The dashed line represents the region with missing electron density.

4.4.4 Protein Modelling

The zebrafish PATZ1 BTB domain has a seven-amino-acid sequence between A2 and B3 (70-76) for which no discernible electron density could be found. It was possible to model the coordinates of such short loop simply by using ModLoop[135, 136] (Fig. 4.16).
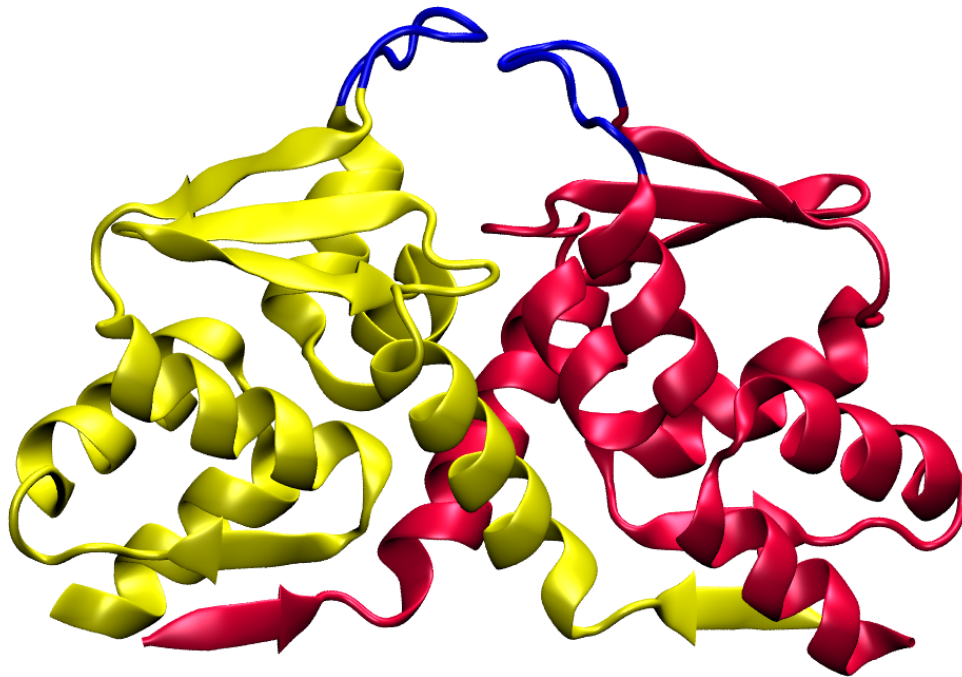
Figure 4.16 Cartoon representation of the zebrafish PATZ1 BTB domain structure. The missing central loop structure (blue) was computationally modelled.

4.5 Comparison of the Mouse and Zebrafish PATZ1 BTB domains

When excluding the central loop, the murine and zebrafish protein sequences share 83.7% identity, whilst their structures can be superimposed with an RMSD of 0.62 Å (Fig. 4.17). Besides their structural homology, the dimerization interfaces of the mouse and zebrafish PATZ1 BTB domains are also very similar. Their crystal structures contain a single structurally corresponding salt bridge (Arg47–Glu75 in the mouse protein and Arg36–Glu64 in the zebrafish protein; the residue numbering refers to the crystal structures), in addition to the residues engaged in inter-chain hydrophobic interactions and hydrogen bonds.
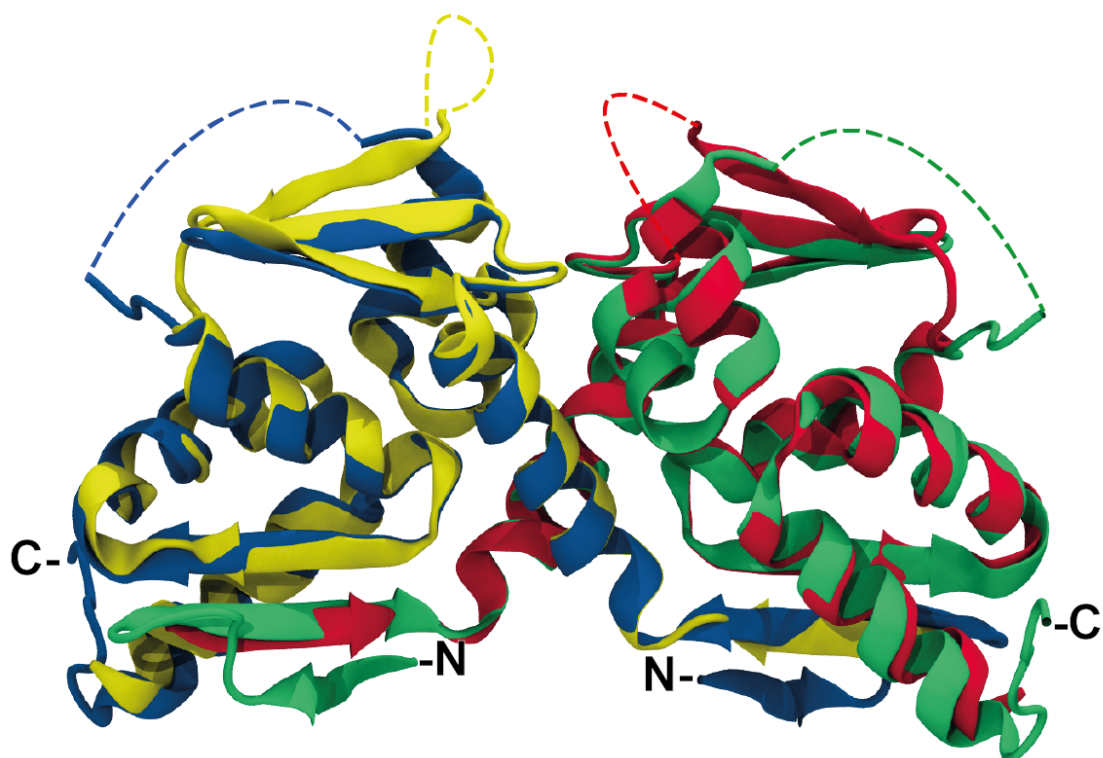
Figure 4.17 Superimposition of the mouse (blue and green) and zebrafish (yellow and red) PATZ1 BTB domains (RMSD of 0.62 Å). Regions with missing electron density are represented by dashed lines.

This analysis confirms the typical BTB dimer interface that includes the β1, α1, A1, A2, A3, β2 and A5 elements (Fig. 4.18). The interfaces forming residues retrieved from PDBePISA[137] include four basic and five acidic residues for the murine protein and four basic and three acidic residues for the zebrafish protein. Interestingly, these charged residues in the crystal structure conformation are not participating in more than one ionic bond. Nevertheless, comparing the number of contact-forming residues in the energy-minimized modelled BTB domains improves the understanding of the dynamics of these interfaces (Fig. 4.19). Using the VMD tools, we found a dramatic increase in the number of charged residues (mostly negative) that participate in interface contacts (marked with asterisks in Table D.3, Appendix D).
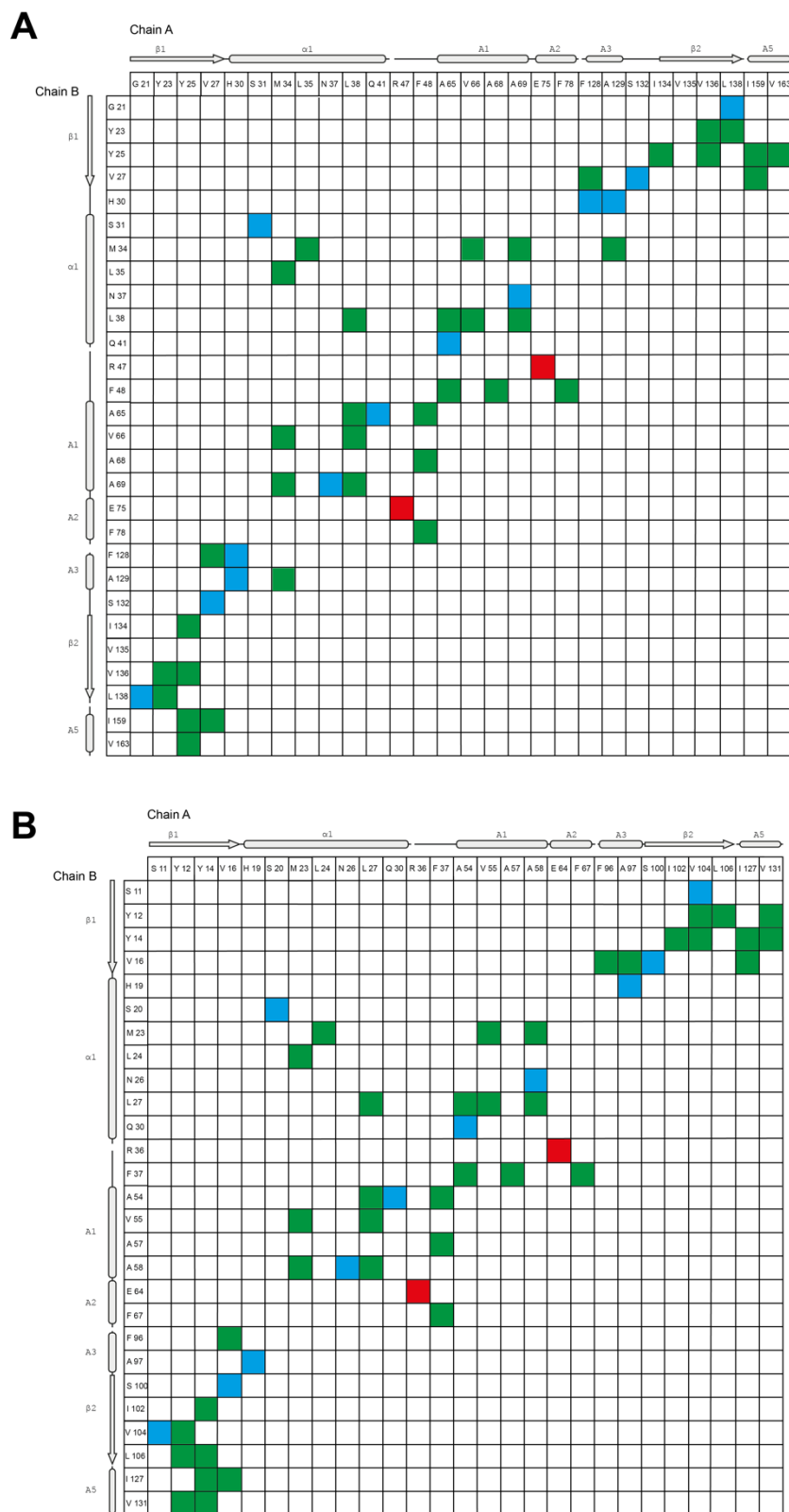
Figure 4.18 Identification of residues involved in homodimer interaction interfaces. Contact maps of inter-chain residue interactions of the mouse (A) and zebrafish (B) PATZ1 BTB domain crystal structures. Hydrophobic interactions (green), hydrogen bonds (blue) and ionic bonds (red) are indicated. The relevant elements of the secondary structures are shown for orientation.

To understand the stability of these contacts, we assessed those that persist above a threshold value (15%) during the lifetime of the 200 ns of MD simulation (Fig.4.19). During the simulation, the A2/B3 loop region significantly contributes to the interface in both models, resulting in flexibility of the salt bridges that form between a single charged amino acid from one monomer and multiple opposite charged amino acids from the opposite monomer. While Arg47 of mouse PATZ1 BTB is only engaged with Glu75 in the crystal structure, MD shows that it can contact a broader number of charged residues.
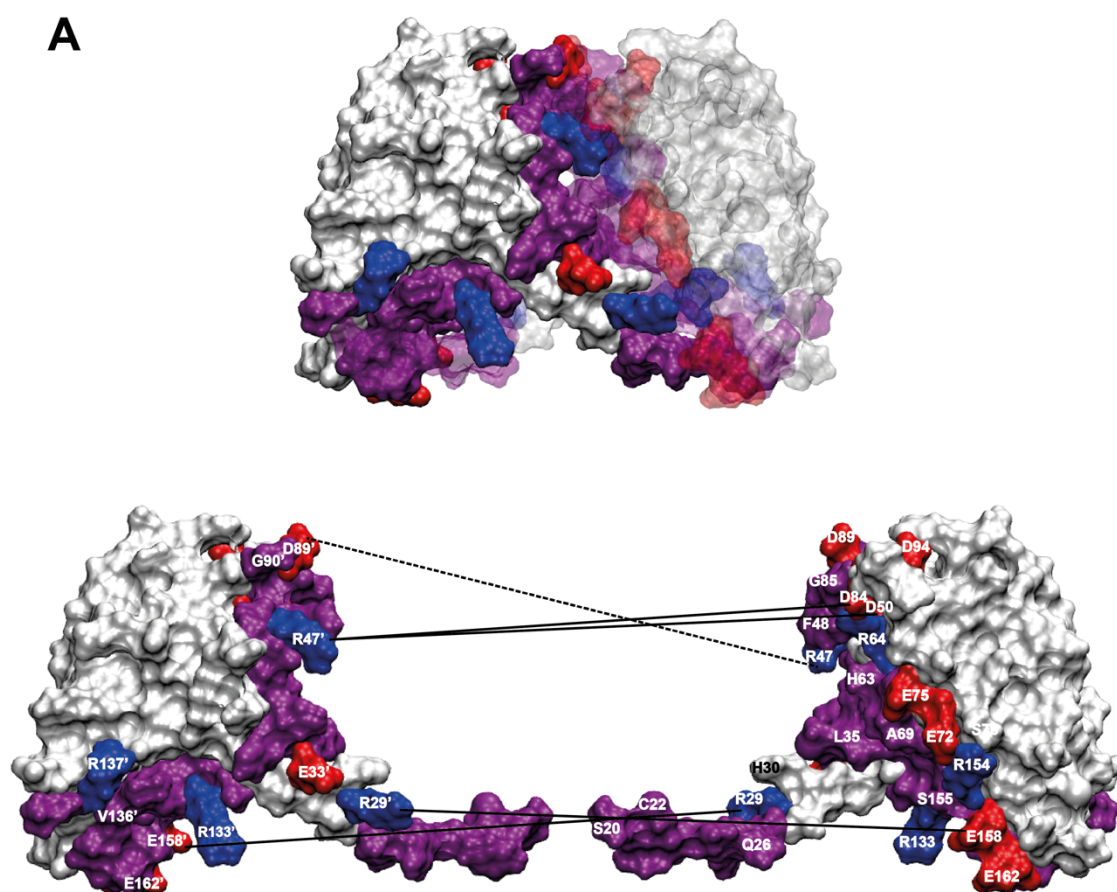
B
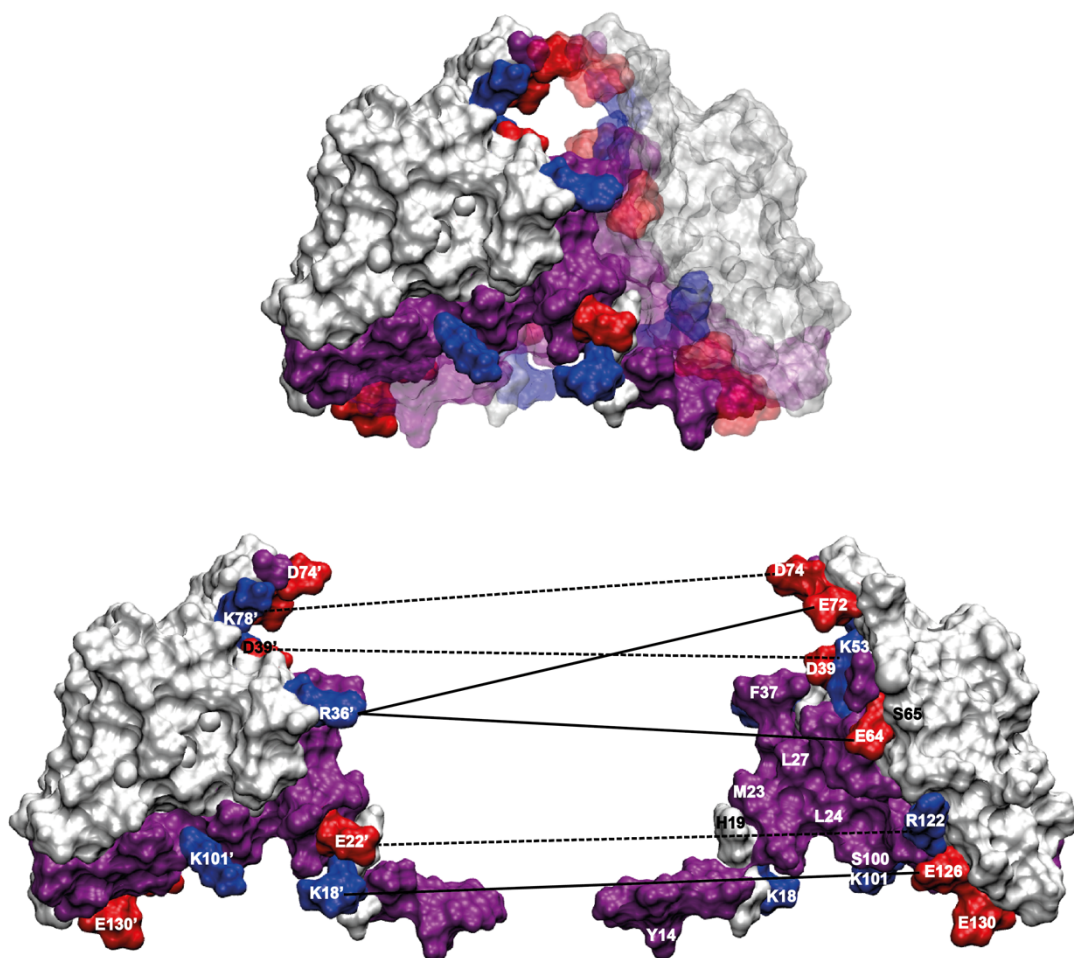
Figure 4.19 A highly charged dimerization interface mediates PATZ1 BTB homodimerization. (A) A split homodimer view in surface representation and completed with the modelled loop highlights the residues involved in the interaction interface. In the top protein the two monomers are distinguished with opaque and transparent representations. Positively (blue) and negatively (red) charged residues are annotated and neutral residues are shown in purple. Inter-chain salt bridges that persist above the threshold are indicated by straight lines and those that do not persist by dotted lines. (B) The zebrafish PATZ1 dimerization interface is also shown as a split homodimer view for comparison. Numbering refers to the crystal structure files.

In fact, at different time points during the MD simulation, it is possible to measure the alternative intermolecular salt bridges established by the residues Arg47 on both monomers, with different negatively charged residues including those from the flexible loop (Asp50, Asp84 and Asp89; Fig. 4.20).
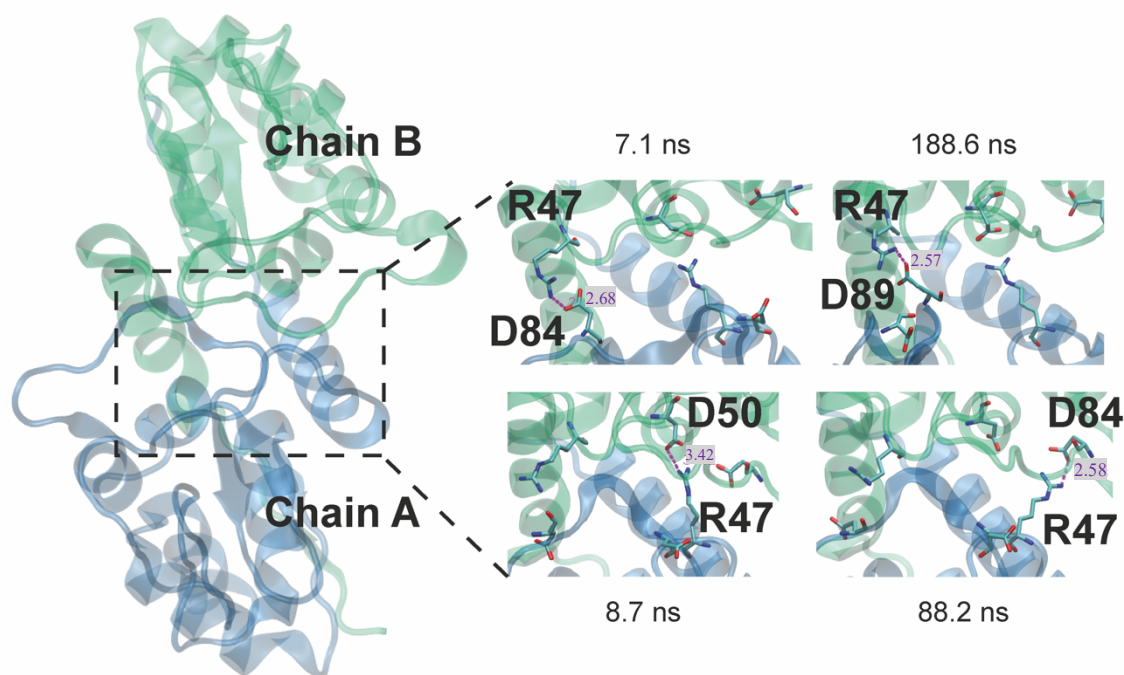
Figure 4.20 Molecular Dynamics (MD) Simulation shows that interface residues rearrange through alternative contacts. A partial dimer interface between the two monomers (Chain A and B) of the mouse PATZ1 BTB domain including the energy minimized modelled central loop is shown in the dashed box (top view). Alternative salt bridges (purple dotted lines, distance in Å) are shown at different time frames of the MD simulation.

4.6 Structural details of BTB domain and co-repressor binding

The BTB domain of BCL6 is the only example of a solved molecular structure bound to co-repressor peptides (NCOR2/SMRT and BCOR) and it reveals the molecular details of this functional interaction. Co-repressor binding to the BCL6 BTB domain requires BTB dimerization because the interaction interface (lateral groove) with the co-repressor peptide is formed by residues in both monomers. 23 residues from each BCL6 monomer contribute to this interface[18] (Fig. 4.21). Even though BCL6 and PATZ1 are structurally very similar, their corresponding co-repressor binding interface sequences are not conserved (Fig. 4.21-D). The calculated RMSD between the PATZ1 and BCL6 BTB domains individual monomers is equal to 1.56 Å. The structural similarity between BCL6 and PATZ1 was more evident when the flexible PATZ1 loop was excluded (RMSD of 1.23 Å).
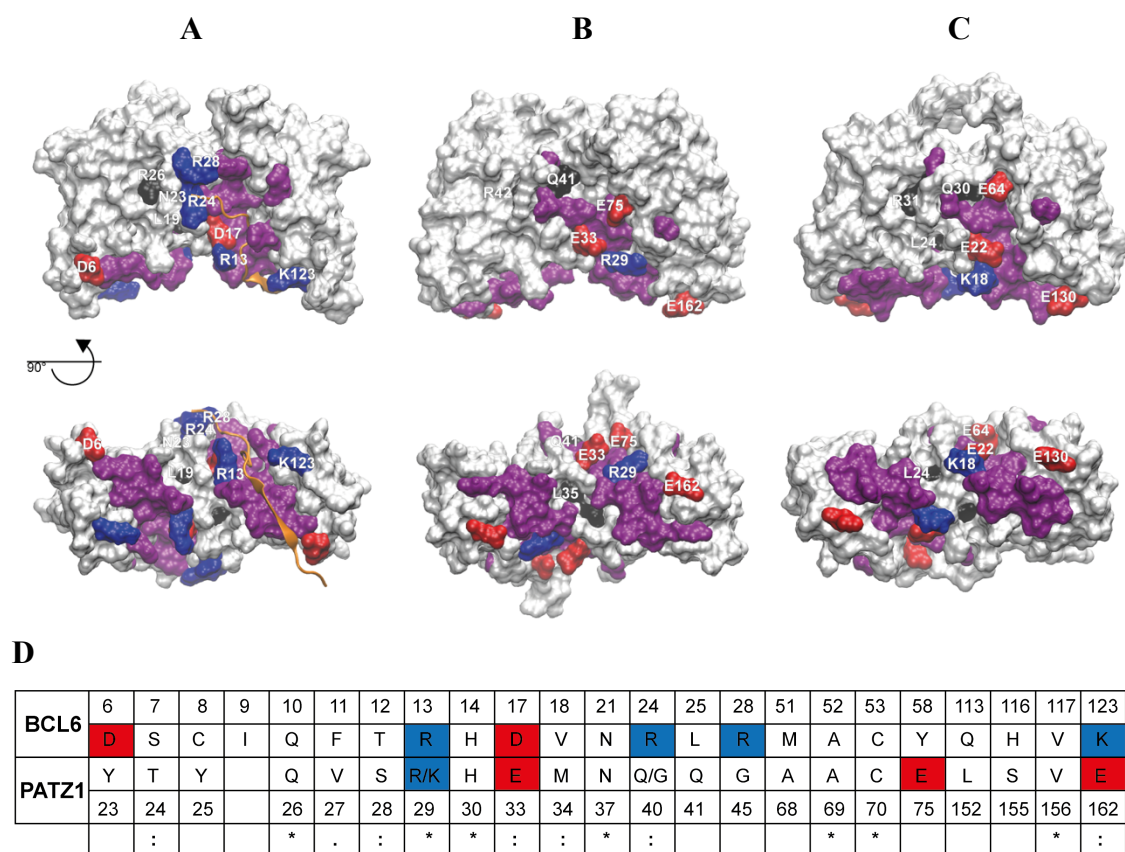
Figure 4.21 Comparison of the lateral grooves of PATZ1 and BCL6 BTB domains. The structures of the BCL6 (PDB entry 1R2B) (A) and the energy-minimized modelled mouse (B) and zebrafish (C) PATZ1 BTB domains are shown in surface representation viewed from the front and bottom. The surface area of the residues in the lateral groove of the BCL6 BTB domain is buried upon formation of the BCL6–SMRT complex[18]. The SMRT peptide binding to the BCL6 lateral groove is shown in cartoon representation in orange. All residues in the lateral groove are labelled for one monomer. Colors indicate residue type: positively charged, blue; negatively charged, red. All other residues in this region are colored purple. The positions of the mutations that affect the binding to the co-repressor peptides in BCL6 and in PATZ1 are indicated in black (references are given in the text, section 4.1.2). A sequence alignment of BCL6 and PATZ1 BTB residues located in the lateral groove region is shown in (D). Residues are numbered according to the BCL6 and mouse PATZ1 structure files, with the charged residues colored as in (A)–(C). Apart from positions 29 and 40, where the two alternatives are indicated, mouse and zebrafish PATZ1 contain the same residues in these structurally corresponding positions. Although the residue conservation for BCL6 and PATZ1 in this region is low, SMRT/NCOR peptides are predicted to bind the BTB domain of PATZ1 in the same region.

Comparison of the surface charge distributions of the two proteins indicates major differences. Specifically, the BCL6 lateral groove contains a high density of positively charged amino acids that interact with the co-repressors (the interaction with SMRT is shown in Fig. 4.21). Surprisingly, the surface of PATZ1 corresponding to the BCL6 lateral groove did not contain as many basic residues. In fact, this region of mouse and

zebrafish PATZ1 is highly conserved (91% identical) and contains more acidic amino acids. The presence of alternatively charged residues in the lateral groove of PATZ1 may indicate that its interaction with corepressors may be through a different mode compared with BCL6. In this regard, the lateral groove of the PATZ1 BTB domain is more similar to that of LRF compared with BCL6[20]. Moreover, the mouse PATZ1 residue Asp50 (Fig. 4.22) is part of the charged pocket that is conserved between BCL6, LRF and PLZF[20]. Residue Asp50 in mouse PATZ1, corresponding to Asp39 in zebrafish and to Asp33 in BCL6, is absolutely conserved in all ZBTB proteins and happens to be the previously mentioned second degron residue (Fig. 1.1). The charged pocket that is formed by the participation of Asp50 residues from both monomers has been suggested as an alternative region for ligand binding[56, 57].
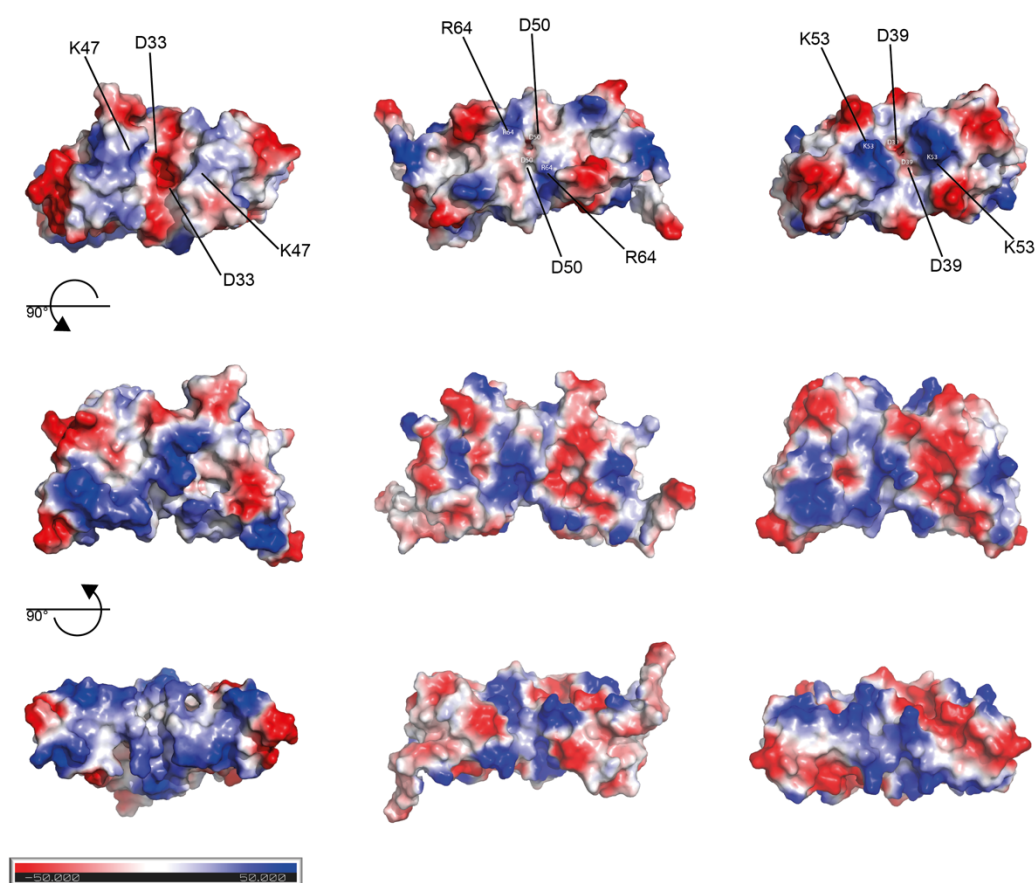
Figure 4.22 Comparison of the electrostatic surface potentials of PATZ1 and BCL6 BTB domains. From the central front view, two 90º rotations show top and bottom views of the electrostatic surface potential of the BTB domain of: (left) BCL6 (PDB entry 3E4U); (center) murine PATZ1 (6GUV); (right) zebrafish PATZ1 (6GUW). In the top view two conserved charged residues (negative and positive) that characterize the BTB domain charged pocket are indicated. Residues are numbered according to the BCL6 and PATZ1

structure files. The colors range from red (negative potential) to blue (positive potential) with white near neutral.

4.6.1 BCL6 Dimerization mutants

According to the study by Huyhn and Bardwell, 1998[53], BCL6 BTB domain fails to bind NCOR1 and SMRT co-repressors when residue 19 Leu is mutated to Ser. This position is found at the dimer interface of BCL6 but is not one of the lateral groove residues (Fig. 4.21). A hypothesis is that this mutation may indirectly affect the binding with the co-repressors because it actually interferes with the formation of the necessary dimer interface. Following this logic, a mutation on the lateral groove sequence that does not belong to the dimerization interface, should affect the co-repressor binding but not the dimer formation. To test this hypothesis, the Q5® Site-Directed Mutagenesis Kit (NEB) was used to generate point mutations on BCL6 BTB domain. First, Leu amino acid 19 was mutated to Ser and Thr amino acid 12 into Gly. The procedure described in the kit includes three steps: exponential amplification (PCR); KLD (phosphorylation, ligation and template digestion) reaction and transformation in bacterial cells. The plasmid with the mutated DNA was then extracted by miniprep and successful mutations confirmed by sequencing. For this experiment, the BCL6-BTB containing plasmids include a GFP or RFP Tag. Future experiments will test these mutations in the context of dimer formation and co-repressor binding between overexpressed fluorescence-tagged proteins in live cell imaging. More details about this experiment are described in section 4.9.3.

4.7 BTB domain stability

The control of the dimer formation could be necessary for the dynamic turnover of BTB-containing protein complexes. Alternatively, giving the highly similar fold of numerous BTB domain containing proteins, mechanisms to allow only functional assemblies could be in place. As a previous study shows for different BTB domain sequences[94], three residues could be subjected to modifications signaling for degradation of intolerable pairs by the E3 ubiquitin ligase complex SCF-FBXL17. These residues are renamed degrons and are shown in crystal structures to be buried when the BTB domain forms a stable homodimer and become exposed otherwise.
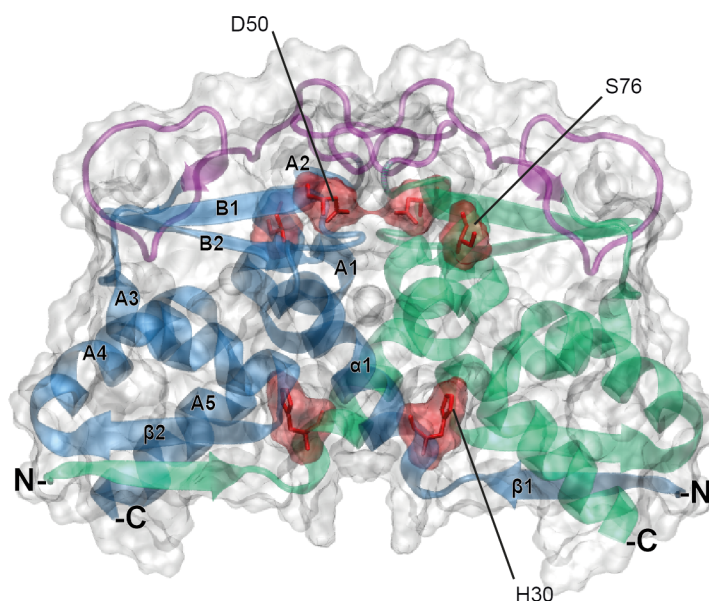
Figure 4.23 Cartoon representation of PATZ1 BTB domain dimer (blue and green) with the modelled A2/B3 central loop in purple. The surface rendering of the protein (transparent) allows the localization of the three buried degrons (in red). Secondary structure elements are indicated.

To determine if these degron residues play a role in PATZ1homodimer stability, their exposure to the solvent was studied in our crystal structure. According to the mentioned study, PATZ1 degrons are residues 30 (His), 50 (Asp) and 76 (Ser) (Fig. 4.23). While the first two degrons are conserved among other BTB domains with exception of the first one in KAISO, the third residue at the end of A2 is not a conserved Ala in all ZBTB proteins (Table 4.3). Additionally, the first two degron residues (annotated in Figs. 1.1 and 4.23) participate in the interaction interface of the PATZ1BTB homodimer structures and are buried.

	BTB degrons		
KEAP1 (4CXI)[94]	HIS 15	ASP 34	ALA 60
LRF (2NN2)	HIS 16	ASP 35	LYS 61
PLZF (1BUO)	HIS 16	ASP 35	ILE 61
MIZ1 (3M52)	HIS 6	ASP 25	MET 51
mPATZ1 (6GUV)	HIS 30	ASP 50	SER 76
ZFPATZ1 (6GUW)	HIS 19	ASP 39	SER 65

BCL6 (1R29)	HIS 14	ASP 33	SER 59
MYNN (2VPK)	HIS 6	ASP 25	ALA 51
KAISO (3M4T)	TYR 14	ASP 33	GLN 59
HKR3 (3B84)	HIS 8	ASP 27	SER 53

Table 4.3 Degron residues positions in aligned BTB domain structures. With the exception of KAISO, the first two degrons are conserved in all proteins.
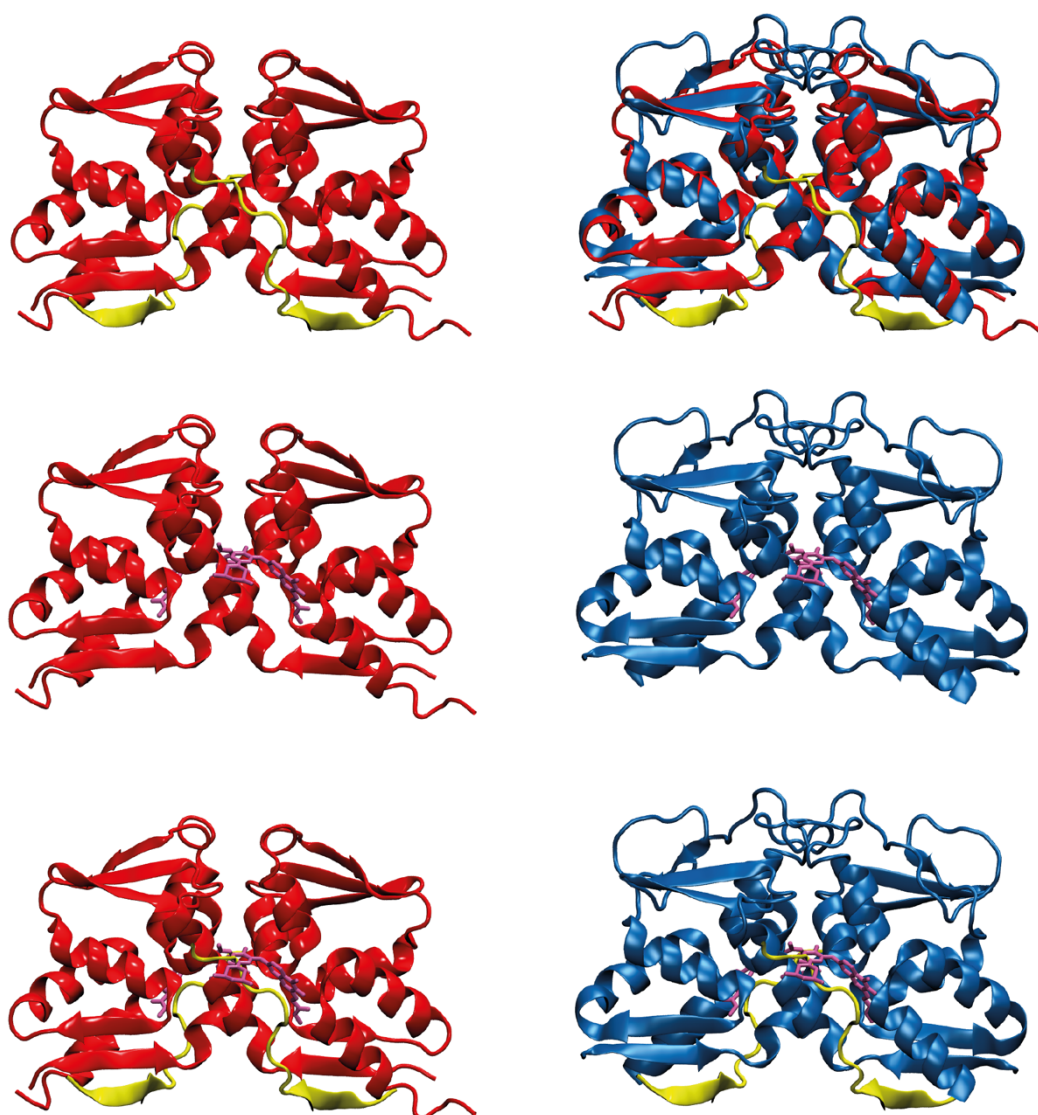
Figure 4.24 Comparison of the binding site of BCOR and BI-3802 drug on BCL6 and PATZ1 BTB domains. BCL6 homodimer is shown in red, BCOR peptides in yellow (from PDB entry 3BIM), PATZ1 homodimer in blue, BI-3802 drug in pink (from PDB entry 5MW2). The top right figure shows the structural overlap of BCL6 and PATZ1 BTB domains.

A different system applied to control and aimed to inhibit the BTB dimer formation was found by the Boehringer Ingelheim pharmaceutical company[29] for BCL6 degradation in B-cell lymphoma disease. The research produced two high affinity drugs BI-3802 and BI-3812 that specifically target the BTB domain of BCL6 and induce its degradation or to inhibit its binding to co-repressors, respectively. The first of these drugs was crystallized on the BTB domain of BCL6 and deposited to the PDB with the code 5MW2. By overlapping structures of BCL6 bound to the degrading drug and to BCOR peptide (3BIM), it becomes clear that the drug targets the co-repressor binding on the BTB domain (Fig. 4.24). In fact, the overlap of the drug and the peptide structures shows major clashes that indicate mutual exclusion of the two BCL6 interactors. A parallel analysis shows the similarity of PATZ1 and BCL6 BTB domains when structurally aligned, although the sequence of the lateral grooves is not conserved (Fig. 4.21). This research could be expanded to further test the specificity of these drugs as well as to design new functional inhibitors based on the structural knowledge of BTB domains.

4.8 The BTB domain in other ZBTB proteins

cKrox

The well optimized expression and purification procedures established and described for mouse and zebrafish PATZ1 BTB domains were applied also to other proteins belonging to the human ZBTB protein family. The first clone was the BTB domain (1-144) of cKrox/ThPOK/ZBTB7b/ZBTB15 in pET-47b (+) vector. After His tag affinity purification, the protein was also further analyzed by SEC (Fig.4.25). The fractions collected according to the SEC profile, were then analyzed on SDS gel. The SEC profile shows two peaks: one overlaps with a DNA peak and correspond to a very faint protein band (A7) at the expected size for cKrox-BTB; the second corresponds to separated molecules in a size range between 29-44 kDa according to the column calibration markers. This protein was considered to be the His tagged BTB domain region in homodimer conformation (~34 kDa). The samples collected between B8-C4 fractions were pulled together to 8.8 mg/ml concentration and 5 crystallization trays were setup for this protein. No crystals were observed for this construct.
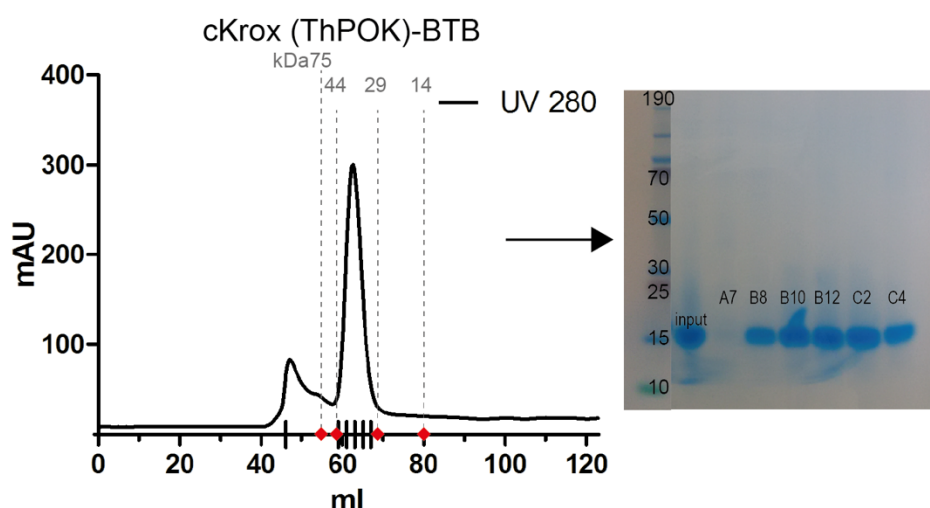
Figure 4.25 SEC purification profile of the cKrox BTB protein and fractions analysis on SDS gel. The BTB domain (17 kDa) elution point corresponds to a dimeric conformation of the protein (29-44 kDa). Vertical lines on the X-axis indicate elution points of the fractions collected for the gel. Red dots correspond to the elution points of marker proteins.

Similarly to mammalian PATZ1 sequence, also the cKrox BTB domain contains a mid-long central loop predicted to be disordered and possibly adding flexibility to the whole domain. This glycine-rich central loop sequence partially aligns with PATZ1 A2/B3 loop (Fig 1.1). Although shorter, this apparently internally disordered region (IDR) also contains a Threonine residue that could act as a phosphorylation site. This post translational modification could affect the stability of the protein conformation as well as its functions.

BCL6 and PATZ2

Two examples of BTB domains of other ZBTB family members that were purified are from human BCL6/ZBTB27 and PATZ2/ZBTB24. The sequences were cloned in pET-47b (+) plasmid, expressed and purified following the same methods described above.

The first crystal structure of the BCL6 BTB domain (5-129)[18] was obtained mutating three cysteine residues (C8Q; C67R; C84N) into more polar amino acids (PDB entry 1R29). Differently from the original crystal structure, the construct sequence was retained as the wild type human BCL6 (1-129, 16.9 kDa). Another group successfully crystallized the wild type BCL6 BTB domain (5-129) (PDB entry 3E4U)[22].

After His tag affinity purification, the protein was analyzed by SEC in Gel filtration buffer (HEPES 20mM, NaCl 250mM, pH 7.4). The SEC profile shows two peaks (Fig. 4.26): one overlaps with a DNA peak measured at UV 254; the second corresponds to separated molecules in a size range between 29-44 kDa according to the column calibration markers. Although this protein was considered to be the His tagged BTB domain region in homodimer conformation (~33.8 kDa), it could not be stored for additional experiments due to precipitation occurring after sample centrifugation or freezing.

While BCL6 is the most studied ZBTB structure, there is no report of a PATZ2 BTB domain protein structure. The sequence (1-133) was designed on the wild type human sequence without modifications. After affinity purification and SEC in the same running buffer, the chromatogram of PATZ2 shows the elution of a single sharp peak (Fig. 4.26), corresponding to a protein size slightly higher than what it was expected for the His tagged construct of PATZ2 BTB domain. The dimer conformation MW of this protein should in fact be equal to 34.2 kDa.
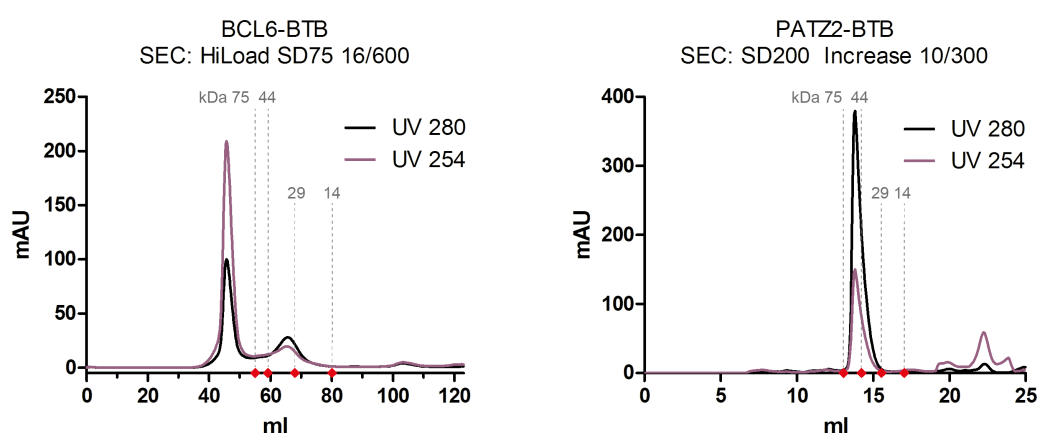
Figure 4.26 SEC purification profile of the BCL6 (left) and PATZ2 (right) BTB proteins. Vertical lines on the X-axis indicate elution points of the fractions collected for the gel. Red dots correspond to the elution points of marker proteins.

Among the limitation of SEC, it has been reported that there exists a dependence between the protein pI and the running buffer pH and NaCl concentration[138]. A protein globular conformation would be affected and the retention time through the column would vary the elution point and the consequent size estimation. In the case of the PATZ2 BTB domain and the buffer conditions used, the low pI of the protein and could cause ion

repulsion from the resin resulting in earlier elution. The comparison of these two elution profiles suggests that protocol optimization for each BTB protein is necessary and that it should be refined according to the residues content and biochemical characteristics (Table 4.4) as well as on available literature.

BTB	Theoretical pI	Amino acids content			Percentage	
		Negative	Positive	Tot	% neg	% pos
BCL6	6.26	14	13	129	10.85	10.08
BCL6 mut	6.82	14	14	129	10.85	10.85
PATZ1	5.32	19	14	159	11.95	8.81
Zebrafish PATZ1	5.37	20	15	135	14.81	11.11
PATZ2	4.74	19	8	133	14.29	6.02

Table 4.4 Protein sequence parameters useful for the choice of expression system and elution buffers.

4.9 Molecular Dynamics of the BTB domains

The availability of existent and newly produced data from the structures of BTB domains in ZBTB proteins allowed an extensive study of the dynamics of these proteins that had never been reported before. Molecular modelling and molecular dynamics (MD) simulations are accurate methods to model the behavior of proteins by performing inexpensive and fast calculations on the molecules' stability and thermodynamically driven conformational changes that can further aid biochemical studies.

For all proteins, a crystal structure was retrieved from the PDB and when more than one entry was available for the same protein, the one with higher resolution was chosen. Only the homodimer assembly annotated in PDBePISA database was considered for the MD simulations. In the simulation environment each atom of the protein is defined by three space coordinates derived from the PDB file and its chemical parameters. Solvent and ions (K^+ and Cl^-) at 15 mM concentration are added to build a simulation box that is set to reproduce the aqueous cellular environment. The structural analysis by molecular simulation includes an initial run of equilibration at constant temperature and constant volume (NVT) only to evaluate the stability of the protein in the simulation environment.

All simulations were then performed at 310 K (37 °C) in isothermal and isobaric conditions (NPT) after equilibration. The simulations are performed by the NAMD software in the context of the Charmm27 force field. The resulting trajectory is the description of the all-atoms motion as function of time in a thermodynamically driven native-like environment.

At least two replicas of 100 ns simulation were performed for every available BTB domain dimer of the proteins in the ZBTB structural family. Quantitative analysis of the trajectory includes the calculation of root mean square deviation (RMSD) and fluctuations (RMSF) for every atom of the protein to measure the molecule dynamics with respect to the initial coordinates. The molecular analysis qualitative focus is then moved on the identification of key residues for the dynamics of the structure, the modes of motion as well as the study of the potential multiple interaction interfaces of BTB domains. The BTB domain is a primary protein-protein interaction domain. The interaction surface for dimerization in the ZBTB family is mostly hydrophobic and involves the N and C termini of the two molecules and the central α–helices. Additionally, there are some conserved solvent exposed residues that suggest the presence of interaction sites with other proteins.

4.9.1 Analysis of the MD simulations of the BTB domain homodimers

The known crystal structures of BTB domains analyzed in this study are listed below (Table 4. 5). The descriptions for each protein are referred to data shown in Fig. 4.27.

Human ZBTB ID	Common name	PDB code	Resolution (Å)	Publication year and ref.
ZBZTB7a	LRF (G)	2NN2	2.10	2006[20]
ZBZTB16	PLZF (B)	1BUO	1.90	1998[16]
ZBTB17	MIZ1 (F)	3M52	2.59	2010[24]
ZBTB19	PATZ1 (J)/ Zebrafish PATZ1 (I)	6GUV/ 6GUW	2.29/ 1.80	2018[139]
ZBTB27	BCL6 (A)/ BCL6+SMRT	1R29/ 1R2B	1.30/ 2.20	2003[18]

ZBTB31	MYNN (D)	2VPK	2.00	2008
ZBTB32	FAZF (H)	3M5B	2.00	2010[24]
ZBTB33	KAISO (C)	3M4T	2.05	2010
ZBTB48	HKR3 (E)	3B84	1.74	2007

Table 4.5 BTB homodimer crystal structures used in this study.

A-B: BCL6 and PLZF

BCL6 and PLZF proteins are two transcription repressor factors involved in cancer biology and known to heterodimerize[69], although their BTB domain is not essential for this interaction. They are the only two proteins whose BTB domain was shown to bind SMRT and NCOR peptides, even if with different affinities[57]. For the simulation of BCL6-BTB, the Mutate Residue plugin from VMD[107] was used to recreate the WT sequence of BCL6 from the 1R29 structure in which three residues were mutated to aid the crystallization process (C8Q; C67R; C84N)[18].

Both proteins present a big cavity from top view within the dimer interface that may be functional for binding to the several interactors in the transcription regulation network these proteins are involved into, like co-repressors and histone deacetylases (HDAC). Both proteins show a stretching motion of the N-terminal β-strand (β1) and the last 2 helices A4-A5 to the sides in the X direction ↔, suggesting that this region could be stabilized upon binding, as it actually occurs in the BCL6-bound form with the SMRT peptide.

C: KAISO

KAISO is a transcriptional regulator that interacts with the NCOR and chromatin remodeling factors complex[140]. In the BTB structure of KAISO the central cavity observed in BCL6 and PLZF is more compact between the three couples of the top β-strands (B1-B2-B3).

D: MYNN

The myoneurin MYNN protein is a transcription regulator expressed in muscle tissues at different stages of development[141]. In the MYNN-BTB structure, each monomer has two types of motion divided in two dynamical parts. The bottom half moves towards outside and up while the top part moves towards inside and down in a dynamical crunch-like movement. The N-terminal β1-strand is absent from MYNN-BTB structure resulting in a more flexible central helix α1. In short, it is an anomalous dimer internally uncorrelated.

E: HKR3

The telomeric zinc finger-associated (TZAP) or Human Krüppel-related 3 (HKR3) protein binds to telomers controlling the length of the chromosomal ends[142]. In the BTB structure simulation analysis for this protein, the dimer interface remains very compact despite a marked motion of the top loop that involves the reshaping of the all structure. This motion is unique and might be a binding motion. As seen in MYNN-BTB, also this protein is missing the N-terminal β1-strand.

F: MIZ1

The myc-interacting zinc finger protein 1 (MIZ1) is a transcription repressor active in embryonic development (gastrulation)[143] and B-cell differentiation[144]. From top view, clearly, MIZ1-BTB presents a cavity between B1-B2 and B3 with B3 strand being very flexible. This cavity may be important to allocate a second BTB dimer for tetramerization. In fact, MIZ1 is the only protein of the ZBTB family that has be shown to tetramerize through the BTB domain (PDB entry 2Q81)[32]. In MIZ1 the B3 strand mediates tetramerization and in the MD analysis, and it has asymmetrical conformation on the two monomers. Its motion is inversely correlated to all other residues in the own monomer but coordinates with the C-term of the other monomer. MIZ1, as well, is missing the N-terminal β1-strand residues.

A BCL6

B PLZF

C KAISO

D MYNN

E HKR3

F MIZ1

G LRF

H FAZF

I zfPATZ1

J hPATZ1
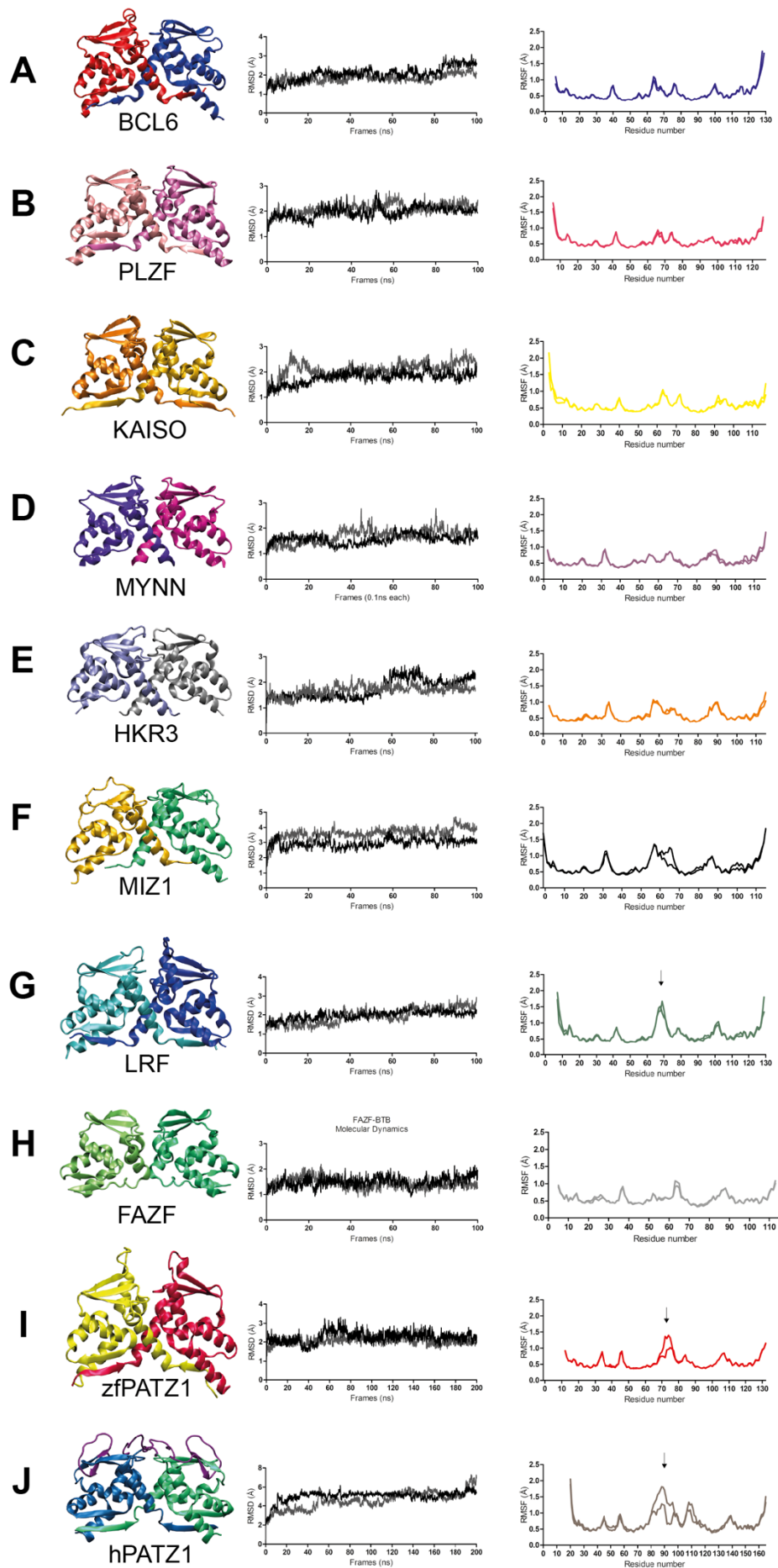
Figure 4.27 Data from the MD simulation of the known crystal structures of ZBTBs. Structure in cartoon representation, RMSD data in replica, RMSF for each residue in both monomers. Arrows on the RMSF graph in part G, H, I, indicate the flexible loops modelled in LRF, zebrafish (zf) and human (h) PATZ1 BTB domains, respectively.

G: LRF

The lymphoma-related factor LRF is involved in oncogenesis and regulates the differentiation of B and T cells[145]. The crystal structures of LRF BTB domain lack electron density for the A2/B3 loop. In order to study the MD of this protein, ModLoop[135] was used to model the coordinates of the missing loop (66-71). Therefore, these residues show high flexibility during the simulation and a significant hinge motion towards the center of the dimer over the cavity residues.

H: FAZF

FAZF is the Fanconi-anemia zinc finger protein, a transcription factor binding the same DNA target of PLZF[101] and important in the hematopoietic stem cell development[146]. The BTB domain of FAZF is very different from the other analyzed structures based on the dimerization interface. In fact, this is the only BTB dimer without strand exchange between the two monomers. The dimer interface is reduced to the residues of a shorter α1 helix while the N-terminal β1-strand, essential in the other dimers for inter-monomer contacts, interacts with the β2-strand at the C-terminal of the same monomer. Because of these unusual characteristics, FAZF-BTB is considered an outlier among the other known ZBTB structures and was excluded from further comparative analyses.

I-J: PATZ1

MD simulations lasting 200 ns were performed for human and zebrafish PATZ1 BTB domains, while 100 ns was the standard maintained for all other protein complexes examined. The simulations of human PATZ1 BTB domain, inclusive of the homology modelled A2/B3 loop, indicated that while the overall dimer forms a stable structure, the A2/B3 loop region is uniquely flexible. During the simulations a new b-strand could form within this loop (Fig. 4.12 and 4.28-J). Interestingly, the b-strand that is generated in the simulations contains the threonine residue that is the only amino acid that differs between the human and mouse BTB domains.

The two BTB monomers are known to homo-dimerize through a specific dimerization interface that includes β1-α1; A1-A2; β2-A5. In the case of human PATZ1, the two A2/B3 loops form an additional central interface between the two monomers with at least ten amino acids contributing to the BTB homodimerization (representing 17.5% of the total interface of 57 amino acids). The two loops freely flap towards and away from the center in a flexible hinge motion around the dimer's axis of simmetry. This motion sampled in a limited 200 ns-long simulation suggests that the central loop may have a gating function and supposedly be involved in protein-protein interactions. A measure of this motion is reported in terms of solvent accessible surface area (SASA) of the residues forming the cavity below the A2/B3 loop (Fig. 4.28).

To ensure the replication of the native environment conditions, zebrafish PATZ1-BTB structure with the modelled central loop was analyzed by MD at both 28 and 37°C to compare the effect of temperature on the dynamics of the Zebrafish PATZ1-BTB domain. The different temperature does not affect the general motion of this protein. In both cases, the top central loops are flexible bringing a large contribution to the overall dynamics and could gate a possible binding site.

4.9.2 Quantitative analysis of the BTB dynamics

The comparative analysis of the MD of ZBTB proteins revealed that in all cases, the BTB structures are stable as homodimers. With the exception of the FAZF BTB domain, they all share a common 3D fold guided by a similar dimerization interface. Nevertheless, a distinctive trait of the BTB dynamics can be highlighted and it is identified in the A2/B3 loop dynamics that can be negligible or very pronounced.

In order to quantify the effect of the A2/B3 loop dynamics in every BTB, the solvent accessible surface area (SASA) of seven residues was considered (Table 4.6). These residues are located below the loop where they form a cleft or the sides of a cavity across the dimerization interface. The analysis of the cumulative SASA values ($Å^2$) for the cavity residues as a function of time gives a measure of the flexibility of the A2/B3 loop that can bury or expose the cavity. The values are calculated during the MD simulation and plotted against the trajectory timeline (ns) (Fig. 4.28).

BTB-containing protein	Cavity residues indexed						
	1	2	3*	4	5*	6	7
BCL6	31	32	33	46	47	48	49
	L	T	D	H	K	T	V
MIZ1	23	24	25	38	39	40	41
	L	C	D	H	K	A	V
LRF	33	34	35	48	49	50	51
	L	C	D	H	R	S	V
zfPATZ1	37	38	39	52	53	54	55
	F	C	D	H	K	A	V
hPATZ1	48	49	50	63	64	65	66
	F	C	D	H	R	A	V
PLZF	33	34	35	48	49	50	51
	L	C	D	H	R	T	V
KAISO	31	32	33	46	47	48	49
	F	C	D	H	K	N	I
MYNN	23	24	25	38	39	40	41
	L	C	D	H	R	N	V
HKR3	25	26	27	40	41	42	43
	Y	C	D	H	W	S	V

Table 4.6 List of the seven cavity-forming residues considered in the SASA calculations of nine ZBTB proteins. Amino acids are indicated with one-letter code and their structure residue number. Although these structurally corresponding residues are extensively conserved among these BTB proteins (residue 3 and 4 are absolutely conserved), the variation of their level of exposure to the solvent gives a measure to distinguish the dynamic behavior of each protein. Residue 3 and 5 are starred because they form the charged pocket of BTB domains[57] (Fig. 4.22).

As a result, in this calculation, BCL6, PLZF, KAISO, MYNN and HKR3 show small variations in the SASA values of their cavity residues that indicates overall a low flexibility of the A2/B3 loop. The variation can be compared to a threshold area, arbitrarily set on 1100 Å^2 and indicated in the SASA graphs with a dashed black line (Fig.4.28). While for BCL6, PLZF, KAISO and MYNN the SASA values follow the threshold trendline, for HKR3 the variation is still low, but the values are below the threshold indicating a more fastened loop over the cavity.

On the contrary, MIZ1, LRF and the PATZ1 cases show higher variation and that arises from the overall higher flexibility as already discussed. For MIZ1, the SASA values are variable but constantly above the threshold, permitting high solvation of the cavity residues. On the other hand, the flexibility of LRF and hPATZ1 loops allows the SASA values to fluctuate above and below the threshold according to the opening and closing hinge motions. On the time frames from the simulation trajectories added for five proteins, the distance between the absolutely conserved Asp (D) residues on the two monomers is indicated as a linear reference (Å).
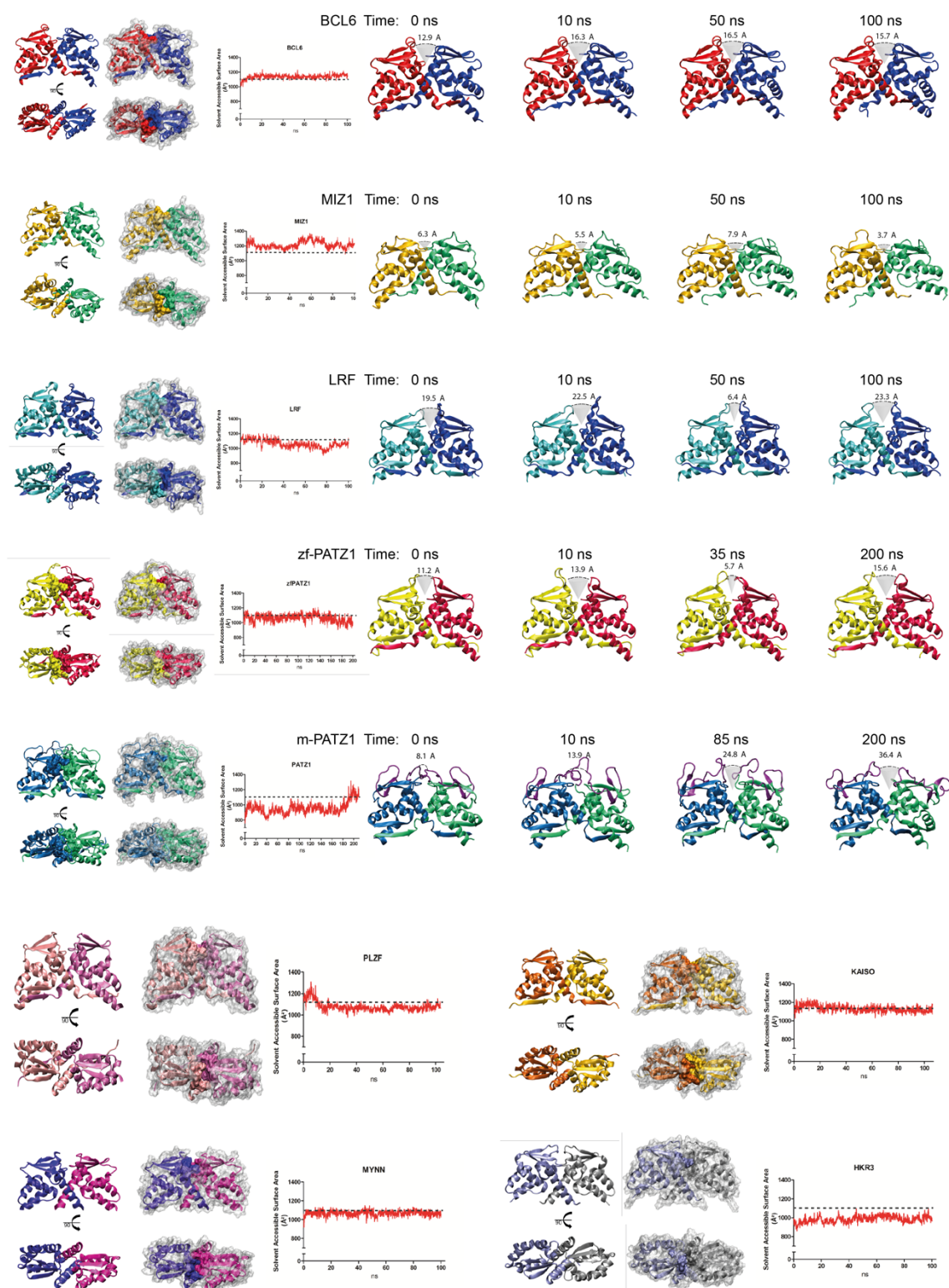
Figure 4.28 SASA measurements for the cavity residues of each BTB domain structure. For five proteins, time frames from the MD simulations are added to show the level of flexibility of the A2/B3 residues and how it affects the solvent accessibility of the cavity residues under it. The change in the SASA values for BCL6 is almost null as for PLZF, KAISO, MYNN and HKR3, likewise and therefore time frames were not represented.

4.9.3 The BTB dimer and co-repressor peptides

The single ZBTB protein in the literature that has been crystallized in a form bound with a co-repressor is BCL6. The co-repressor is represented by only 17 residues aka the BCL6 binding domain (BBD). Two structures in the PDB report BCL6 bound to the BBD of NCOR2/SMRT (PDB entry 1R2B)[18] and to the BBD of BCOR (PDB entry 3BIM)[21]. Based on these two structures, a model exists of BCL6 BTB domain bound to the predicted BBD of NCOR1[58] where the binding modality of SMRT BBD were used to assemble the binding interface of the homologous NCOR1 BBD .

As described in section 4.6, the homodimer conformation of BCL6 is necessary to form the lateral groove that binds SMRT BBD peptide. The structurally corresponding residues that form the lateral groove in BCL6 are not conserved in PATZ1 (Fig. 4.21). However, PATZ1 can bind at least one co-repressor, NCOR1, although the binding site of this interaction is not known. The structural rearrangements observed for PATZ1 BTB dimer during MD simulations (section 4.5), may be the key to understand the binding potential of PATZ1 to the corepressor-peptides.

Firstly, a detailed analysis of the MD simulation for BCL6 BTB domain bound to SMRT was conducted in order to understand the binding modality of the BBD binding to BCL6. In the detailed analysis of the BCL6-SMRT interaction, one mutation was also considered. The mutation L19S on BCL6 was shown to prevent the BTB dimer formation, precluding also NCOR1 and SMRT long fragments from binding[53] (1). The binding rules observed were then transferred from BCL6 to PATZ1 BTB domain docked to a SMRT BBD peptide. The docking of SMRT peptide was repeated on human and zebrafish BTB domain to model the possible effect of the central loop on this interaction (2).

1. The effect of the co-repressor peptide binding to BCL6 can be appreciated by comparing the RMSF graphs from two MD simulations of the same BCL6-BTB protein with and without SMRT BBD (Fig. 4.29). In particular seven residues are remarkably more affected than others, namely their RMSF values averaged over the all simulation (100 ns) are lower in the bound form than in the unbound form: Arg 13, Asp 17, Asn 23, Arg 24 and 28, Cys 53, His 116. Only three of these residues are conserved in PATZ1.

These residues are part of the BCL6 lateral groove (Fig.4.21) and a lower RMSF value indicates less fluctuations for these residues that are stabilized upon binding. Interestingly, in the BCL6-bound form the A2/B3 loop is not affected and remains as rigid as in the free form, demonstrating that in BCL6 the region of this loop does not contribute to co-repressor peptide binding.
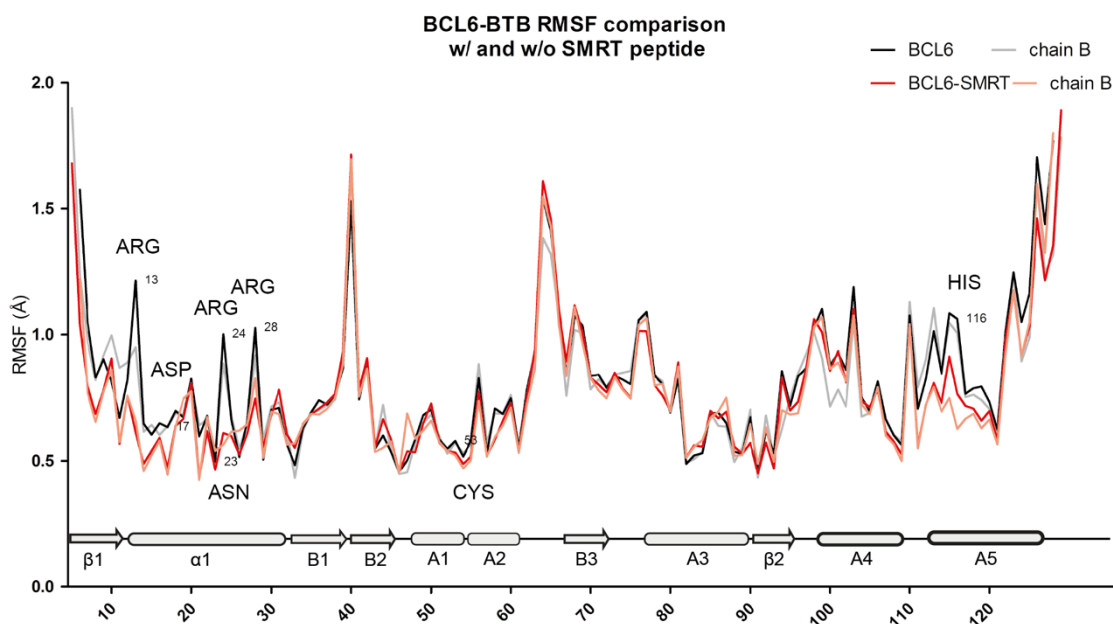
Figure 4.29 Comparison of the RMSF of the residues of BCL6 in bound (red and orange lines) and unbound form (black and gray lines) to SMRT peptide. The two lines for each form represent data coming from the two monomers. The BTB domain secondary structures are added for orientation. The residues with highest difference in RMSF in the bound vs unbound form are indicated.

The L19S mutation in the MD system did not reveal any evident disruption of the BCL6-BTB dimer or the binding with the SMRT peptide (Fig. 4.30). This residue is located at the center of the dimer interface where the two α1-helices overlap. The mutation from a Leu to a Ser amino acid introduces two polar residues in a core that was essentially hydrophobic. However, because residue 19 of BCL6 is not among the residues directly involved in SMRT binding, as described in the original crystal structure[18], the peptide binding does not seem to be affected by its mutation. The limitations of MD sampling

95

(running time) could explain the disagreement of this new observations with the original mutation experiment[53].
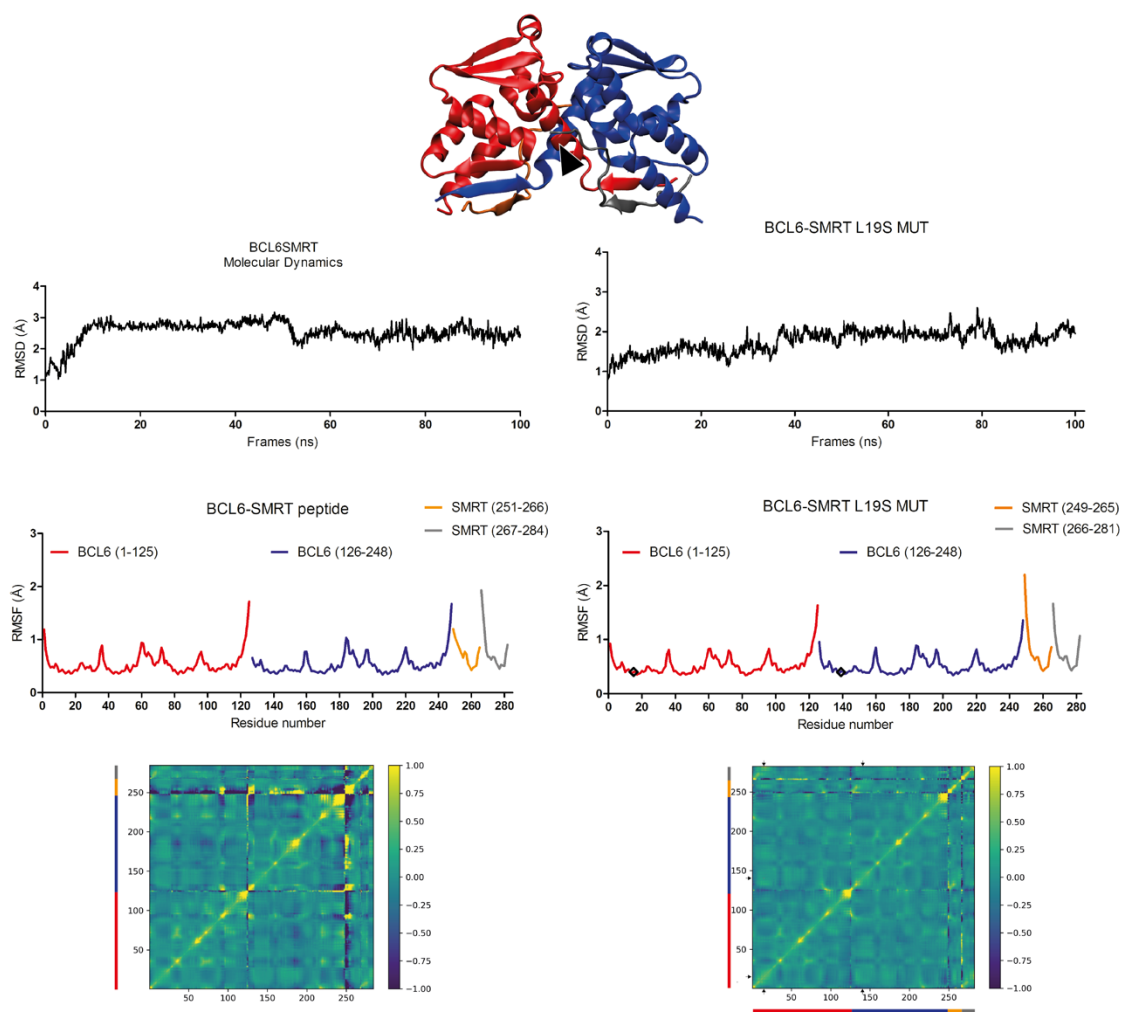
Figure 4.30 MD analysis comparison between BCL6 (left) and its L19S mutant (right) bound to SMRT BBD. The position of the mutated residue on the dimer structure is indicated by a black arrowhead. Following the colors of the protein chains on the protein cartoon (top), red and blue lines on the RMSF graphs and on the axis of the cross correlation matrices, represent the data coming from the residues of the two BCL6 monomers; orange and grey, the data coming from the two SMRT peptides.

2. The SMRT peptides were docked on mouse and zebrafish PATZ1 BTB dimer structures using the Z-Dock online tool[111] where specific residues were selected on the BTB protein to guide the docking. The selection of the binding residues was based on the sequence alignment with BCL6 BTB domain and on the above discussed analysis of the binding details acquired from the BCL6-SMRT MD analysis. In both cases the C-terminal

region of one of the peptides (orange in RMSF graph Fig. 4.31) tend to detach from the BTB domain showing low affinity in the complex without affecting the regular motion of the central A2/B3 loop.
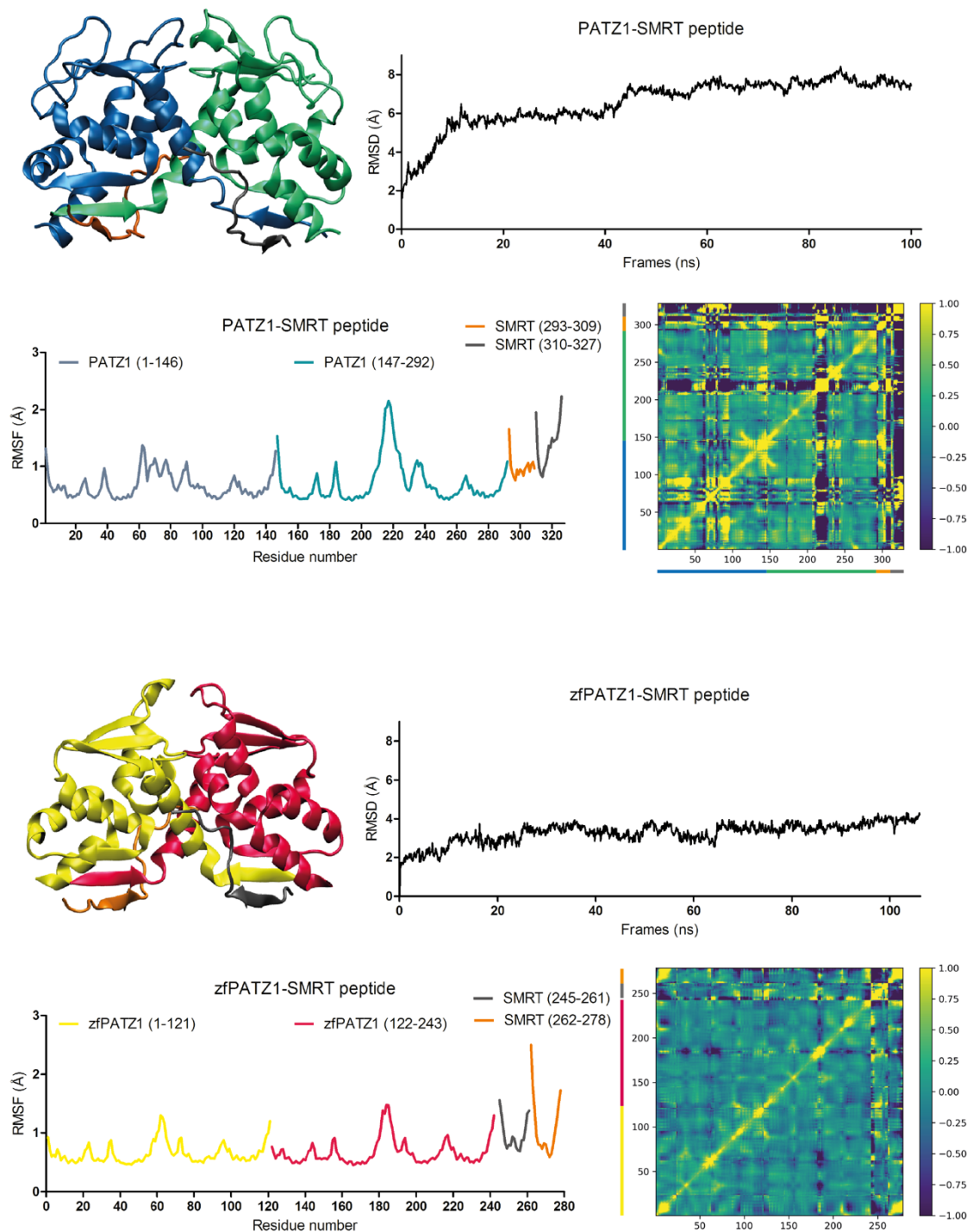
Figure 4.31 MD analysis of the docked structure of the BTB domain of human PATZ1 (top) and zebrafish (zf) PATZ1 (bottom) bound to SMRT BBD. The same color-code used for the protein chains on the protein cartoon (top, left for both panels) is used also for the lines on the RMSF graphs and on the axis of the cross-correlation matrices.

4.9.4 Analysis of the MD simulations of selected BTB domain Heterodimers

Based on literature data, ongoing experimental evidence from B. Erman's lab, previous analysis and the ZBTB available structures, six BTB domain heterodimers were selected and studied by MD simulations. All simulations were run for 100 ns in the same temperature and solvent conditions described before for the homodimers. RMSD, RMSF and cross-correlation analysis for each protein complexed are reported. The RMSF data of each monomer in the new heterodimer complex is compared to the previously calculated RMSF of the same monomer in the homodimer assembly (when available).

PATZ1-BCL6 was the first selected heterodimer for this analysis due to the extensive structure study that had been conducted on these proteins' homodimers. The heterodimer structure was generated with PRISM docking server[117] where the top energy score result was selected and minimized for 0.06 ns (30000 steps). Minimization was started at lower temperature (280 K) conditions in which atomic motions are slower and structure rearrangement smoother. A series of short runs (2 ns) with ramping temperature (280-290-300-310 K) was performed after that to reach the final running temperature of 310 K. This strategy was used because of the artificial assembling of the dimer. The concerns about the effect of the human PATZ1 A2/B3 central loop on this interaction were addressed by taking into consideration also the BTB heterodimer of the zebrafish PATZ1 with BCL6.

BCL6-PATZ1
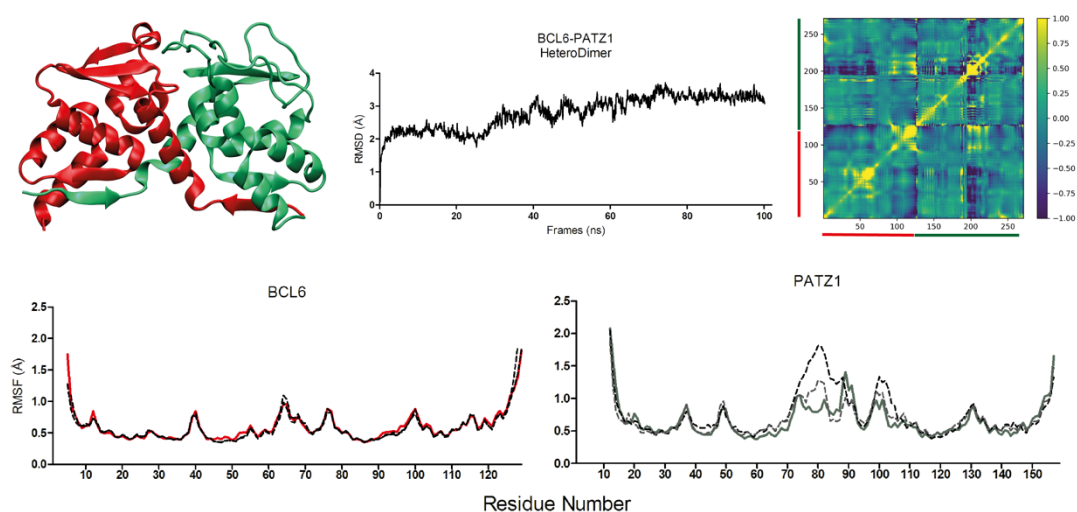
BCL6- zfPATZ1

Figure 4.32 MD analysis of the heterodimer structures of the BTB domain of human PATZ1 (top) and zebrafish (zf) PATZ1 (bottom) with BCL6. The same color-code used for the protein chains on the protein cartoon (top-left for both panels) is used also for the lines on the RMSF graphs and on the axis of the cross-correlation matrices. The dashed lines indicate the RMSF data for the same residues in both monomers during the homodimer simulations.

In both cases (Fig. 4.32), the new interaction of PATZ1-BCL6 BTB heterodimer is not destabilizing. According to the RMSF comparison, the BCL6-BTB monomer is not noticeably affected by the PATZ1-BTB monomer binding. Interestingly, the major differences can be observed on the RMSF of PATZ1 (human and zebrafish), especially on the residues forming the A2/B3 loop. In fact, the simulation shows that the rigidity of BCL6 A2/B3 loop restrains the flexibility of PATZ1 corresponding loop that would normally coordinate with the likewise flexible opposite loop in the homodimer.

The third heterodimer examined was BCL6-MIZ1 BTB domain. This protein is the only so far deposited crystal structure[86] of a BTB heterodimer between ZBTB family members (PDB entry 4U2M). The construct cloned to obtain this crystal structure is a forced heterodimer expressed as a fusion protein of BCL6 (WT) and MIZ1 BTB domain sequences connected by a linker peptide. The electron density from the linker peptide is not reported in the final structure, so the PDB coordinates were used in the simulation files preparation without further modifications.

MIZ1- BCL6

Figure 4.33 MD analysis of the heterodimer structures of the BTB domain of MIZ1 and BCL6. The same color-code used for the protein chains on the protein cartoon is used also for the lines on the RMSF graphs and on the axis of the cross-correlation matrices. The dashed lines indicate the RMSF data for the same residues in both monomers during the homodimer simulations.

Like in the heterodimers with PATZ1, also in this case BCL6 monomer does not seem to be largely affected by the new interaction with the MIZ1 monomer in terms of RMSF of each residue (Fig. 3.33). The only significant difference in the new BCL6 dynamics compared to its homodimer, regards β2 and A4 secondary structure elements. MIZ1-BCL6 heterodimer is in fact asymmetric due to the missing residues of β1 from MIZ1-BTB. The lack of β1−strand, normally contributing to the dimerization interface by forming a sheet with β2 in BCL6, destabilizes this region of BCL6. From the MIZ1 perspective, the fluctuations of B3 as well as the C-terminal elements are reduced in this heterodimer, probably influenced by the β1 of BCL6, this time normally absent in MIZ1 homodimer.

For the next selected BTB heterodimers, PATZ1-PATZ2 and LRF-cKrox, the generation of two new BTB dimer models was necessary (Fig. 4.34). Protein structures of PATZ2 (ZBTB24) and cKrox (ZBTB7b/15) BTB domains are not in fact reported in literature. The models were built by homology in PRIMO and SWISSMODEL suits. No simulation was run for the modelled homodimers, but they were used only in heterodimers simulations with crystal-validated structures.
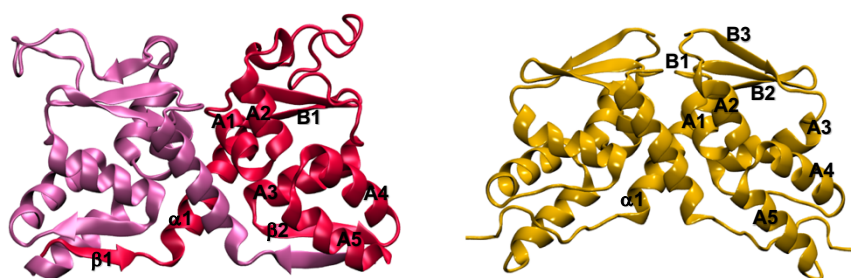
Figure 4.34 Structural model of the BTB domain homodimers of cKrox and PATZ2. On the left, cKrox/ThPOK/ZBTB15/ZBTB7b protein (1-144). As in PATZ1, the A2/A3 central loop is predicted to be partially disordered. The homology modelling was performed on SWISSMODEL. On the right, PATZ2/ZBTB24 protein (1-126). The homology modelling was performed on PRIMO suite and the dimer symmetry reconstructed on Symmdock. On both models, secondary structure elements are indicated for one monomer.

PATZ1-PATZ2 BTB heterodimer emerged from a recently performed experiment carried on in B. Erman's lab to study cellular BTB heterodimer formation. In the experimental setup (ChromoTek), the two BTB monomers were cloned each into a fluorescent-tagged plasmid (tag-GFP and tag-RFP) and co-expressed in BHK cells[147]. The cells were then observed under fluorescence microscope for evidence of GFP and RFP tagged proteins co-localization. The PATZ1-PATZ2 BTB nuclear colocalization has been demonstrated for the first time with this setup and represents an ideal candidate for further confirmation experiments (MSc. Thesis by Liyne Noğay, 2019).

Despite the similar name, PATZ1 and PATZ2 (1-129) BTB domains share only 26.5% identity and 42.4% similarity but among the ZBTB family members, they are the only proteins containing an AT-hook DNA binding motif. Mutations on Patz2 gene product are associated with the Immunodeficiency, Centromeric region instability, Facial anomalies syndrome (ICF) that is a rare autosomal recessive disease[148]

In relation to the concerns about the effect of the human PATZ1 A2/B3 central loop on hetero-dimerization, again both human and zebrafish PATZ1-PATZ2 BTB heterodimers were analyzed. The two heterodimer structures are rather stable during the MD simulation making the docked models solid and suitable for future work (Fig. 4.35).
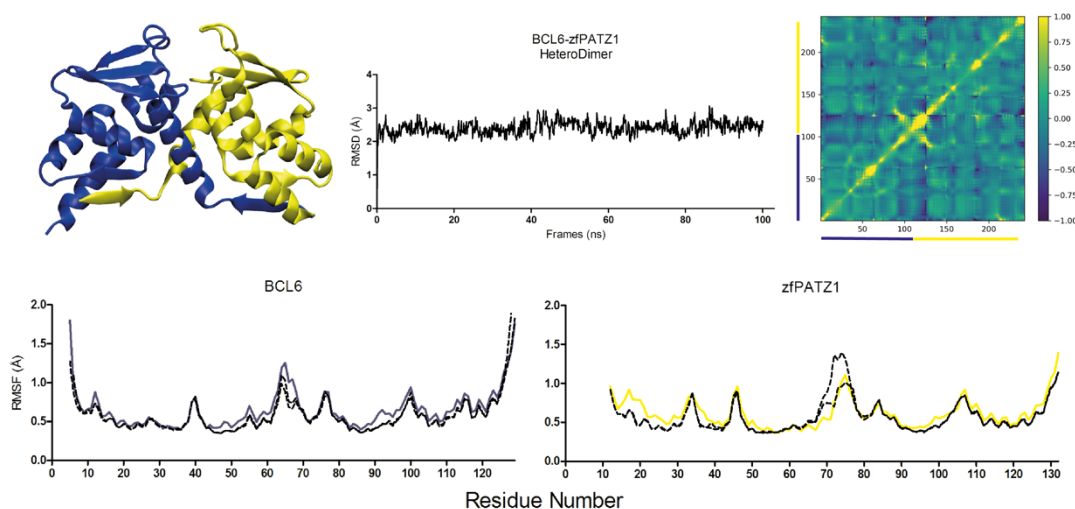
PATZ1-PATZ2
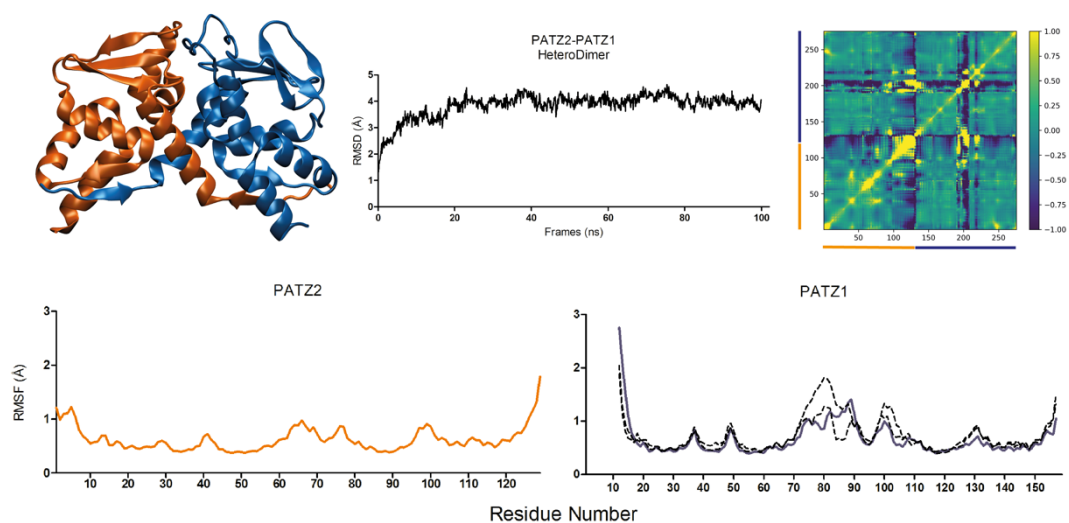
zfPATZ1-PATZ2
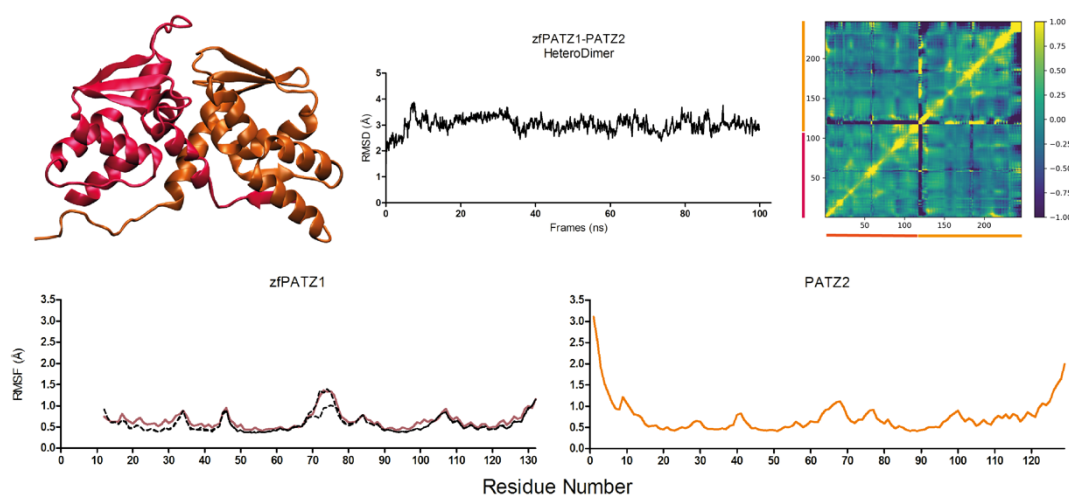
Figure 4.35 MD analysis of the heterodimer structures of the BTB domain of human (top) and zebrafish (bottom) PATZ1 with PATZ2. The same color-code used for the protein chains on the protein cartoon is used also for the lines on the RMSF graphs and on the axis of the cross-correlation matrices. The dashed lines indicate the RMSF data for the same residues in both monomers during the homodimer simulations of the PATZ1 structure models.

The last heterodimer analyzed in this study is the literature documented LRF- cKrox BTB dimer[99]. Despite they are genetically related proteins (ZBTB7a and ZBTB7b) and share 55% sequence identity, they differ about the length of the A2/B3 loop that is longer in cKrox and predicted to be disordered in both of the proteins. The structure was generated in PRISM[117] by docking and minimized as described before. The heterodimer overall dynamics is stable making the docked models solid and suitable for future work (Fig. 4.36).
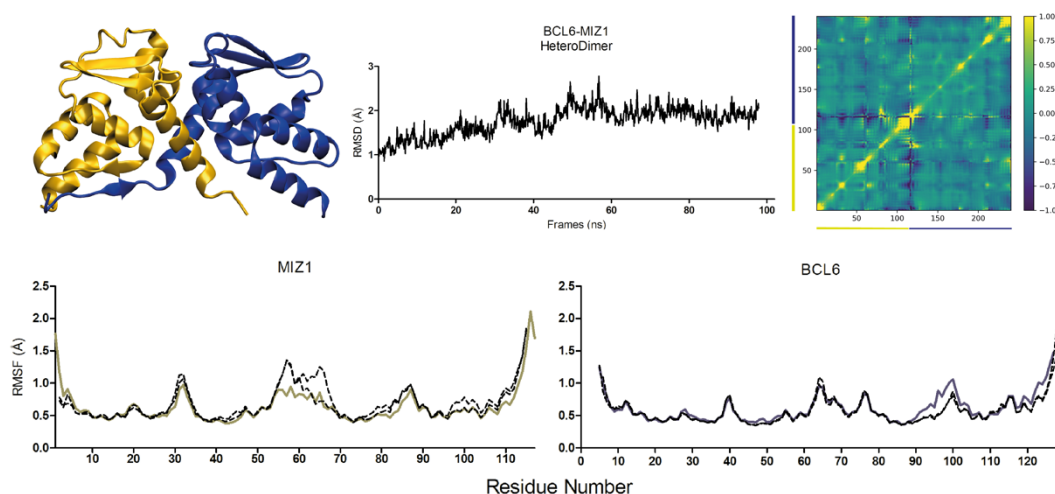
LRF-cKrox
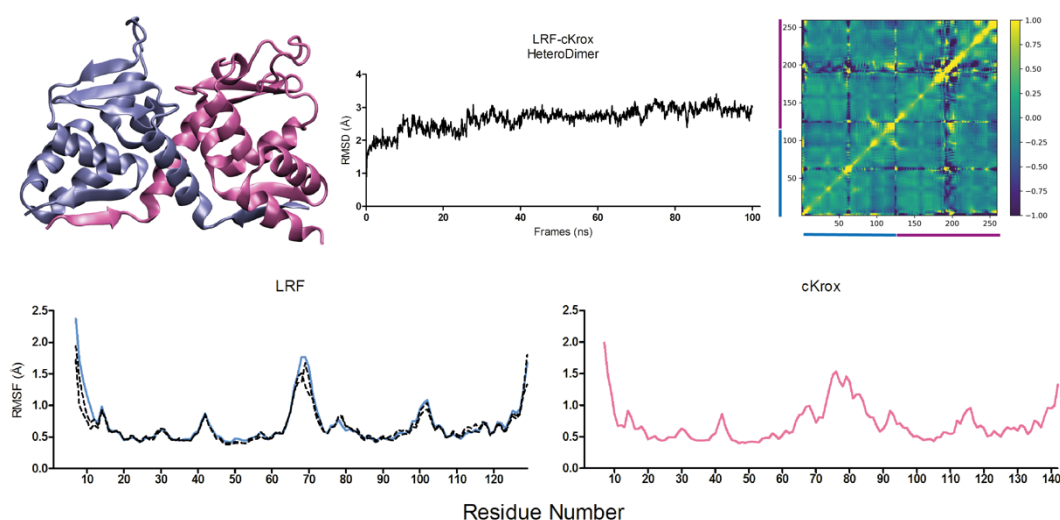
Figure 4.36 MD analysis of the heterodimer structures of the BTB domain of LRF and cKrox. The same color-code used for the protein chains on the protein is used also for the lines on the RMSF graphs and on the axis of the cross-correlation matrices.

4.10 Classification of BTB domains based on MD

The findings revealed by the MD simulation analysis of the ten known crystal structures of BTB domain homodimers, can be summarized in a classification scheme that takes into consideration the dynamically relevant traits of each protein. Certainly, the length of the protein chain and the composition of the secondary structure elements (lack of β1 or substitution of B3 with a more flexible loop) affect the overall dynamics of the BTB domains. However, it emerged from this study that the flexibility of the A2/B3 central loop in the BTB domain can be considered as a distinctive dynamic attribute that may be related with protein function. For reference, a list of the amino acid sequences belonging

to the A2/B3 loop of the known ZBTB structures, including the produced BTB models of PATZ2 and cKrox BTB domains, is shown below (Table 4.7).

ZBTB protein	A2/B3 loop sequence	B3 strand sequence
FAZF	GRR	GQWA
PLZF	FHRNS	QHYT
MYNN	YRSTSE	NNVF
MIZ1	FVDQK	DVVH
KAISO	FSVAG	QVVE
HKR3	YGDGSGGS	VVL
PATZ2	FAEEGEIG	QSIY
LRF	FTSGAVVDQQ	NVYE
BCL6	FTDQLKCNLS	VINL
ZF-PATZ1	FSCQTEDDGQGK	ELEM
PATZ1	FSAQLGDGGAADGGPADVGGATAAPGGGAGGSR	ELEM*
cKrox	FTEGGGGAVMGAGGSGTATGGAGAG	....

Table 4.7 Protein sequence of the A2/B3 loop and the B3 strand of the known ZBTB crystal structures. In red and blue, the content of negatively and positively charged amino acids is indicated. *the asterisk indicates that for human PATZ1 the B3 strand is predicted but not solved as a β-strand in the crystal structure or in the model. In cKrox model the B3 strand does not form.

Charged residues are consistently present in all the A2/B3 loop sequences, with the exception of KAISO whose sequence is largely hydrophobic. The charged residues are mostly negative (in red in Table 4.7) with the exception of FAZF and PLZF sequences that carry only positive charges instead. MYNN, MIZ1 and BCL6 loop sequences are neutral in net charge and overall hydrophilic. The presence of Glycine residues in this loop certainly contributes to its flexibility. According to the size of the A2/B3 loop, it is possible to suggest a classification of the ZBTB proteins. Based only their BTB domain characteristics, the proposed classification is in agreement with the previously discussed MD analysis results.

Class I: FAZF

In addition to the uniqueness of FAZF BTB dimerization interface, the short three-residues long A2/B3 loop places this protein in a separate class of its own.

Class II: PLZF, MYNN, MIZ1, KAISO

These proteins cluster together because of their short A2/B3 loop sequence (5-6 residues), that contains positive charges. SASA values variation for the cavity residues in these proteins was very low (see section 4.9.2), reflecting directly the loop characteristics of low flexibility. The loop of KAISO is the only one lacking any charged residue.

Class III: HKR3, PATZ2, LRF, BCL6, zfPATZ1

These protein class is defined by a longer A2/B3 loop (8-10 residues), that contains negative charges. The length and the moderate Glycine content allow the remarkable flexibility of these loops. LRF and zf-PATZ1 BTB crystal structures lack electron density for these residues.

Class IV: PATZ1, cKrox

Exceptional in the length and amino acid composition, PATZ1 and cKrox loops place these proteins in a separate class. The long alanine and glycine rich loops are predicted to be disordered in both cases.

4.11 Phylogenetic Tree of the human ZBTB family

The phylogenetic tree of the 49 human ZBTB proteins describes this family evolutionary relationships (Fig. 4.37). The leaves of the tree represent the existent human ZBTB protein sequences. The leaves are attached to branches whose length is a measure of the evolutionary time spent for the diversification of the protein on the edge. The tree was built on the alignment of the top ten BLASTP hits for the BTB domain sequence of each ZBTB member. The BLAST search includes homologous sequences from related species. The sequence variation accumulated by each protein during its evolution is visible in the width of the leaves. Additionally, the tree counts also BACH1 and BACH2 BTB sequences even if belonging to a different transcription factor family. Lastly, the tree was rooted on the ZBTB32/FAZF BTB sequence, that, because of its structural characteristics, was considered the outgroup.
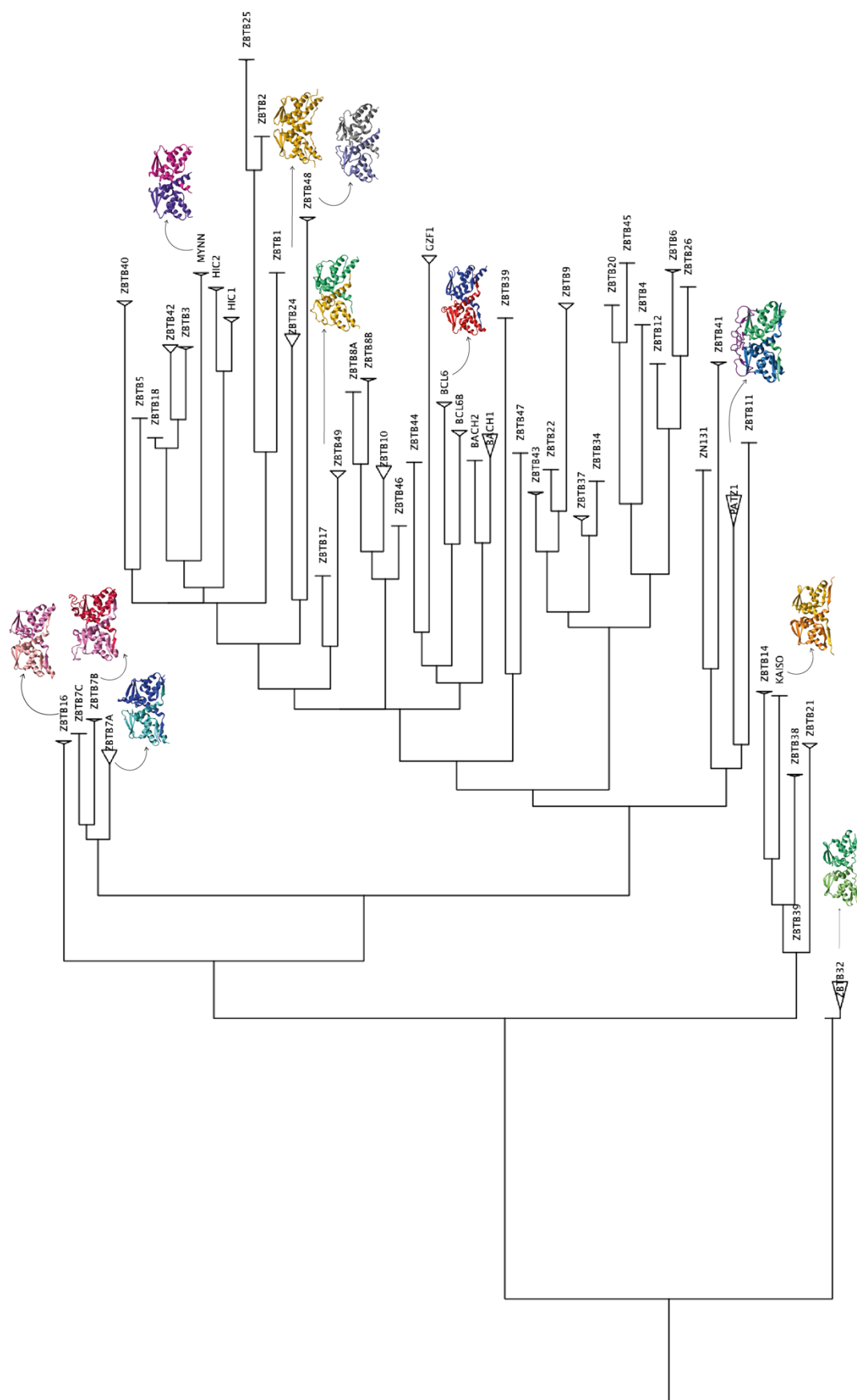
Figure 4.37 Phylogenetic tree of the ZBTB protein family evolution generated in IQ-TREE[149] software. Cartoon representations of the BTB structures used in this study are added next to the corresponding leaves.

4.12 ZBTB proteins Co-expression

The 49 members of the human coded ZBTB protein family are characterized by the same protein architecture with a N-terminal BTB domain followed by various number of ZF domains. The degree of sequence conservation, the similarity of the 3D fold and the dimerization mechanism studied in this thesis work, suggest that ideally all ZBTB proteins can heterodimerize with each of the other family members through their BTB domain. If this is true, a total of 1225 combinations would exist, theoretically.

As seen in the constructed phylogenetic tree (Fig. 4.37), and evident in the known BTB fold, the ZBTB proteins belong to a closely related family. According to an emergent hypothesis, BTB proteins with less divergent sequences have higher probability of dimerization[150]. However, although they might have the binding potential, not all proteins are expressed in the same cells or are active for the same cellular function. In fact, some might have evolved as paralogues and diverged in their functions so that even, if genetically related and highly similar, the proteins may never interact at cellular level. Therefore, the probability that all possible 1225 combinations involve functional interactions could be very low in reality.

In the attempt of providing useful indications about the co-expression of ZBTB pairs, an initial theoretical analysis was conducted on available data retrieved from online databases. Importantly, co-expression of two or more ZBTB genes in the same cell type does not imply occurring interactions. On the contrary, the negative correlation of the expression profile of two genes is sufficient to exclude heterodimer formation, assuming the accuracy of the source data. In particular, when the expression profile data are plotted for a pair of genes, positively correlated data distribute in an ascending linear relation along the main diagonal of the graph. Differently, an interesting subset of uncorrelated data can be visualized as a L-shaped distribution. In this case, for each tissue or cell type, either one of the ZBTB pair is expressed implying that physiologically they are never expressed together and cannot interact.

Two databases were consulted for this analysis regarding the human normal tissue cells from the Human Protein Atlas[37] and the homologues mouse immune system cell lines

from the ImmGen Project[151]. The normal tissue dataset derives from combined sources of RNA transcript expression data in 63 human tissues types (Appendix Fig. D.2). On the other hand, the immune system cells dataset was obtained from the combination of ImmGen Microarray Phase 1 and Phase 2 datasets that include normalized microarray data for 287 cell populations from multiple immune lineages (Fig. 4.38 and Appendix Fig. D.3).

When analyzing the distribution of the expression profile of ZBTB pairs from the normal tissue dataset graphs (Appendix Fig. D.2), the cases of the data from Patz2 (part 3) and Kaiso (part 5) plotted against Patz1, show a clear trend for positive correlation of the gene transcripts' expression. Different is the case Fazf-Patz1 plot (part 4) in which the data points are totally uncorrelated except for one, that come from these genes' expression in testis tissue. For this specific case, although the data refers to RNA rather than protein expression, the limited correlation found between Fazf and Patz1 may suggest that the detection and the study of the interaction between these two ZBTB proteins would be difficult in cell types different from the testis tissue.

A much larger quantity of data was analyzed with the ImmGen dataset. For each cell type in the immune lineages, including B-cells, Dendritic cells, T-cells, Stromal cells, Granulocytes, NK cells, Mast cells, Basophils and Eosinophils, Macrophages, Monocytes and Stem cells, experimental values in replica were averaged. In order to focus on the characterization of PATZ1, like in the previous analysis, the expression data from selected ZBTB genes were all plotted against Patz1. In this case, co-expression data is defined in exact murine cell lines with detailed information available in the downloadable metadata on the ImmGen webpage. Thus, the choice of the working system for the study of the pair interaction of interest becomes more specific.

This second analysis (Fig. 4.38) could be affected by an intrinsic bias of the input data due to the over-populated group of premature and differentiated T-cells that represents the 40% of the all dataset. However, similar positive correlation trends can be observed in Patz2-Patz1 plot. In agreement with previous data, limited correlation between Fazf and Patz1 transcript expression indicates that this interaction should be preferably researched in an isolated group of B-cells. Other cases of L-shaped data distribution are for the pairs of Patz1 with BCL6B, HIC1, ZBTB39, 42 and 46.

Figure 4.38 Extract of the matrix constructed from the ZBTB RNA expression profile data in all Immune System Cells retrieved from ImmGen database. Each plot shows PATZ1 data plot against another ZBTB gene. Each colour represents data from one cell type indicated in the legend.

5. DISCUSSION AND CONCLUSION

Proteins are defined as the motor of life because all cellular processes are driven and coordinated by the activities of their complex networks. Thus, the study of a protein function often relies on the identification of its interaction partners. Protein-protein interactions can be studied by integrating molecular biology, biochemical and computational methods. For this purpose, it is necessary to reach a fundamental understanding of the molecular mechanisms and the structural details that allow such interactions. The structural characterization of the PATZ1 BTB domain achieved in this thesis work is a relevant step forward to model the interaction potential and therefore the function of this protein. As a result of this work, two PATZ1-BTB crystal structures, from mouse and from zebrafish were submitted to the Protein Data Bank (PDB) with entries 6GUV and 6GUW, respectively, and a detailed article with their analysis was recently published[139].

The BTB domain is primarily a protein interaction domain nicknamed Born-To-Bind[12] to highlight its interaction nature. Besides its contribution in the formation of oligomers in voltage gated Potassium channels or the coordination activity in E3-ubiquitin ligase SCF complexes, when connected to Zinc Fingers, BTB proteins are associated in binding transcription co-factors and mediating the recruitment of chromatin remodeling proteins[152]. An example of this association is the complex formed by the ZBTB proteins PLZF and BCL6 binding co-repressors and recruiting Histone deacetylase (HDAC)[104, 153, 154]. The enzyme favors the tighter packaging of chromatin thus inducing the transcription repression of the underlying genes. The role of largely disordered co-repressors proteins NCOR1 and SMRT is to bring together transcription factors and regulatory enzymes and act as a scaffold to support the repression complex[155].

In order to bind co-repressors, the BTB domain of ZBTB transcription factors require dimer formation. The dimer interface creates the lateral groove that BCL6 uses to bind

co-repressors and that is structurally conserved in PATZ1 (Fig. 4.21). However, the discrepancy in charged amino acids in this groove (Fig. 4.22) may indicate altered binding modalities and/or affinities for co-repressors. A known binding partner of PATZ1 is the nuclear co-repressor NCOR1, an interaction that requires the BTB domain[60], though the actual binding modality is not yet clear. To investigate the role of PATZ1-BTB central loop in the binding of NCOR1, the Chimera PATZ1construct was purposely designed. It consists of the mouse PATZ1 protein whose BTB domain was substituted with the zebrafish PATZ1 BTB domain sequence. Hence, in a binding experiment it would be possible to compare the effect of the presence or absence of the central loop on NCOR1 binding.

In terms of functions, the role of the PATZ1 A2/B3 loop could reside in its disordered nature. Based on the amino acid composition with high net charge (negative) and relatively low hydropathy that does not facilitate the folding process, the central loop of PATZ1 can be classified as an intrinsically disordered region (IDR)[156]. IDRs rich in structure breaker residues Gly and Pro do not spontaneously fold in a single stable structure, but rather remain in a dynamic state and the residues space coordinates cannot be assigned by X-ray crystallography. IDRs evolved with the increasing cell complexity[157] and are thought to have different roles[158] including binding of nucleic acids and proteins and cell signaling. Moreover, IDRs can undergo conformational changes due to post-translational modifications (PTM). In the case of the mammalian PATZ1, the disordered A2/B3 loop contains a threonine suitable for phosphorylation that is conserved in primates and bats while the same position is occupied by alanine in other species (Fig. 1.8). The potential interactions through the central loop, deriving from the induced conformational changes, would be in this case species-specific.

A role for PATZ1 has been demonstrated in various malignancies such as thyroid and testicular cancer[47, 49, 159-162]. The interaction between PATZ1 and the tumor suppressor p53[47-49, 163] mediated by a motif in the zinc-finger DNA-binding domain rather than the BTB domain also links this protein to cancer. The BTB domain could be involved in different forms of cancer development. Chromosome 22 specific inversions that translocate the transcription factor EWSR1 with PATZ1 have been observed in various sarcomas. While EWSR1 translocates with ETS family proteins in Ewing's sarcoma, it potentially encodes two fusion proteins EWSR1-PATZ1 and PATZ1-EWSR1, the latter

of which will contain the BTB domain[71-76]. Similarly, the fusion of PATZ1-MN1 from the same chromosome, could see the participation of the BTB domain of PATZ1 in pediatric meningioma[77]. The dimerization and co-repressor interaction properties of PATZ1 identified in this study may shed light on the mechanism of the development of these tumors.

The gene targets of PATZ1 have not been extensively identified. ChIP-Seq and RNASeq experiments have identified 187 putative targets[48, 164]. Of these targets, roughly half were upregulated, and half were downregulated in the absence of PATZ1 expression. How many of these genes are direct targets and how many require the BTB domain for regulation is not known. The functional conformation of the BTB domain in the context of transcription regulation is also unknown. In this work, the PATZ1 BTB domain was analyzed as a dimer and, although a tetramer conformation was also hypothesized, SEC indicated no other stable oligomer form. In a proposed arrangement for the DNA binding of PATZ1 (Fig. 5.1), if the BTB homodimer is necessary for the formation of the transcription regulation complex with the co-repressors, the ZF of each monomer would bind a palindromic DNA sequence. For heterodimers, the promoter target region of each monomer can be consecutive and their action synergistic (as described for PATZ1-BACH2 interaction)[33]. Or, similar promoter sequences from distant genes could be brought together with the rearrangement of the chromosomal DNA 3D structure. The co-repressor complex would then be expected to bind the BTB domain from the top side, opposite to the ZF (question mark in figure 5.1).
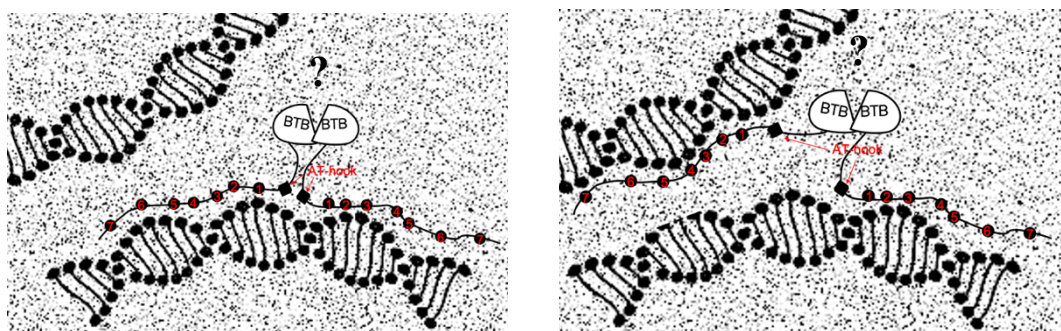
Figure 5.1 Graphic representation of the hypothetical DNA binding arrangements for PATZ1. A BTB domain dimer is shown, followed in each monomer by the AT-hook motif and seven numbered ZF. Note that if the first conformation is preferred (left), the single DNA sequence should be palindromic to allow the specific binding of ZF.

Ideally, the transcription factor function must be governed by a dynamic equilibrium between binding (k_{on}) and releasing (k_{off}) the DNA target sequence[165] in order to control the expression of that gene in response to a specific cell signal. The specificity of the transcription factor for its target should be high as well as transient. To allow the efficient dissociation of the complex, small conformational changes must take place[166], more frequently in the highly disordered regions of the nuclear co-repressors[167]. The same dynamics should apply also to the formation of the BTB dimer and its disassembly. These mechanisms are not well studied as a generalizable feature of all BTB domains. An initial analysis that examined the stability of the BTB dimers found that when exposed, three residues are targeted by ubiquitin ligase for proteasome dependent degradation preferentially in heterodimers.[94] Two of the three degron residues are conserved in ZBTB proteins. Consistent with the stability of homodimers, PATZ1 homodimer structures bury two of these three degron residues in the protein globular structure (Fig. 4.23). Post translational modifications occurring over these or different exposed residues may be the key to understand the mechanisms of BTB domain turnover. A common feature between PATZ1, PLZF and BCL6 is that their BTB domain-mediated localization to nuclear speckles.[55, 67, 70] While PATZ1 interacts with the nuclear speckle-resident ubiquitin ligase RNF4, whether potential post-translational modifications owing to this interaction affect its stability is not known.[168]

A more detailed analysis of the central loop was necessary to characterize further the BTB domain of PATZ1. Because of the absence of electron density in the crystal structure, a molecular model was built by homology for visualization. The homology modelling included a sequence motif search along the central loop and inferred structural information from many proteins. The final result was an unstructured loop. The description of a possible behavior of the PATZ1-BTB model was studied by MD that made clear that the high flexibility of the central A2/B3 loop did not interfere with the overall stability of the BTB domain. The MD simulation shows that the stability of the core structure is maintained by alternative contacts that form between several charged residues at the homodimer interface (Table D.3, Appendix D). Because a single charged residue can contact more than one oppositely charged residue (Fig. 4.20), the dimer interface may be more resilient to the disruption of single contacts. So, energetically, the

presence of alternative salt bridges may be necessary to accommodate the flexibility of the central loop whilst retaining the stability of the dimerization interface.

Focusing on the understanding of the uncommon feature of PATZ1 long central loop, a comparative analysis was performed on the MD of all available BTB domains from ZBTB proteins. The data collected from the MD trajectories was used to calculate the solvent accessible surface area of seven residues located at the dimer interface of each BTB monomer and that create a cavity under the A2/B3 loop. This cavity includes the already observed conserved charged pocket[56, 57]. The SASA calculations are an indirect measurement of the A2/B3 loop flexibility. This study concludes that, despite the conserved fold, the BTB domains can be differentiated through the level of flexibility of their A2/B3 loop. With the available BTB domain structures, four dynamics classes were identified.

Class I	Class II	Class III	Class IV
FAZF (ZBTB32)	PLZF (ZBTB16) MYNN (ZBTB31) MIZ1 (ZBTB17) KAISO (ZBTB33)	HKR3 (ZBTB48) PATZ2 (ZBTB24) LRF (ZBTB7a) BCL6 (ZBTB27) zfPATZ1	PATZ1 (ZBTB19) cKrox (ZBTB7b/15)

Table 5.1 Classification of the analyzed BTB domains based on length and dynamics of the A2/B3 loop.

The flexibility of the A2/B3 loop during the MD simulations was linked to its length and glycine residues composition. These observations could be used as a reference when predicting the classification of other ZBTB proteins, even when no structural or dynamic information is available. The MD simulations were used also to study some examples of heterodimer BTB domains available on the PDB or constructed by docking. When we consider the classification scheme just described, the BTB heterodimers analyzed include proteins from different but consecutive classes. MIZ1-BCL6 (Class II and III); LRF-cKrox, BCL6-PATZ1 and PATZ2-PATZ1 (Class III and IV). Without the peculiarity of the long central loop, PATZ1 would also fall in Class III as its zebrafish homolog. Other documented heterodimers that include the ZBTB proteins in this subset (Table 1.3) are

LRF-BCL6, LRF-HKR3 (Class III); PLZF-BCL6 (Class II and III) and FAZF-PLZF (Class I and II).

Overall, this classification does not directly correlate with the branches of the reconstructed evolutionary tree of the ZBTB family (Fig. 4.37). Not enough specimens among the tree leaves have been examined, but in light of the dynamics information newly obtained, it could be suggested that the same branch host proteins from different classes. In the example of LRF/ZBTB7A and cKrox/ZBTB7B, although derived from a common ancestor, they specialized in their function varying the exposed interaction surface of the A2/B3 loop at their dimerization interface. Or, if proteins of the same class are found on the same branch, they probably evolved under different selective pressures or are activated in different cellular contexts. The accuracy improvement of these predictions will require though a deeper analysis of the ZBTB protein family.

A better comprehension of the different molecular details of these proteins could be the key to identify the rules for ZBTB proteins heterodimerization. It is true that the binding potential of all possible 1225 combinations between pairs of ZBTB proteins could, in principle, be studied with isolated proteins in vitro. But, at least to exclude the not naturally occurring cases, this thesis took in consideration the possible co-expressed pairs of ZBTB gene transcripts as a basis to set the design of heterodimers study. Positive correlated expression does not implicate protein interaction, but a common regulation of their promoters could be suggested. On the other hand, negative correlation would suggest that the pair of transcription factors could directly or indirectly inversely regulate each other's expression. The proposed analysis focuses only on PATZ1 interactions. The current study only scratches the surface of the ZBTB heterodimers complexity. However, an ongoing investigation on the topic is expected to include more comprehensive results.

In reality, the correlation of RNA level with protein abundance is not straightforward. If in principle, the transcript expression drives the protein translation, these two processes are yet subject to dynamic metabolism and undergo active turnover by degradation, silencing (for mRNA), and post-translational modifications (for proteins) that break the theoretical straight correlation. An earlier estimation, accounts for the RNA levels to explain 40% of the variation in protein level[169-171]. Moreover, different gene functions are related to the stability of gene transcripts and translated protein. In fact, differently from

genes with cellular housekeeping roles, transcription factors show a shorter half-life in both RNA and protein level, suggesting that their activity is functional to a specific cell stage or to fast response to external stimuli[169].

A final angle of the structure-based study of the interaction potential of BTB domains, is their being attractive targets for anticancer compounds. Compounds that prevent homodimerization or result in the degradation of BCL6[29] have been shown to have highly effective cytotoxic activity in B-cell lymphomas. Because the BTB domains of ZBTB family proteins all share the same fold, compound specificity requires the targeting of unique features. The residues in the A2/B3 loop of PATZ1, which are unique among the ZBTB proteins, are potentially a specific target for this protein. The structure of the PATZ1 BTB domain reported in this study will aid in the development of therapeutics for those human malignancies that involve PATZ1 and the testing of the specificity of compounds targeting other BTB domains.

APPENDIX A: Chemicals

Chemicals and Media Components	Supplier Company
Acetic acid (glacial)	Sigma, Germany
Acrylamide/Bis-acrylamide (29:1)	Sigma, Germany
Agar	Sigma, Germany
Agarose	Sigma, Germany
Ammonium Persulfate (APS)	Sigma, Germany
Ampicillin Sodium Salt	Sigma, Germany
Anti-c-Myc HRP Antibody from mouse IgG1κ	Roche
Anti-HA HRP Antibody from mouse IgG$_{2b}\kappa$	Roche
b-Mercaptoethanol	Sigma, Germany
Boric Acid	Sigma, Germany
Bromophenol Blue sodium salt	Sigma, Germany
Chloramphenicol	Sigma, Germany
Coomassie Blue Brilliant Blue G	Sigma, Germany
Distilled pure Water	Merck Millipore, USA
DMEM	Gibco, Ireland
DNA Gel Loading Dye, 6X-GeneRuler 1kb Plus	Thermo Fisher Scientific
EDTA	Sigma, Germany
Ethanol, Absolute	IsoLAB, Germany
Ethidium Bromide	Sigma, Germany
Fetal Bovine Serum (FBS)	Gibco, Ireland
Glycerol	Sigma, Germany
Glycine	Sigma, Germany
HEPES	Sigma, Germany
HisPur™ Cobalt Resin	Thermo Fisher Scientific
Hydrochloric Acid	Sigma, Germany
Hydrogen peroxide	TIB-TEK, Turkey

Imidazole	Sigma, Germany
IPTG	Sigma, Germany
Isopropanol	Sigma, Germany
Kanamycin Sulfate	Sigma, Germany
LB Broth	Sigma, Germany
Liquid nitrogen	Microcyl, Koyuncu Gaz, Turkey
Luminol	Sigma, Germany
Magnesium Chloride	Merck Millipore, USA
Methanol	Sigma, Germany
NP40	Sigma, Germany
p-Coumaric Acid	Sigma, Germany
PBS	Gibco, Ireland
Penicillin/Streptomycin	Gibco, Ireland
Pierce, Protein G Magnetic Beads	Thermo Fisher Scientific
PIPES	Sigma, Germany
Potassium Acetate	Sigma, Germany
Potassium Chloride	Sigma, Germany
Protease inhibitor Tablets, Complete	Roche
Protein marker	NEB
SDS	Sigma, Germany
Skim Milk Powder	Sigma, Germany
Sodium Chloride	Amresco
Sodium Hydroxide	Sigma, Germany
TCEP	Sigma, Germany
TEMED	AppliChem
Terrific Broth	Sigma, Germany
Tris Hydrochloride	Amresco
Trizma Base	Sigma, Germany
Trypan Blue Solution	Gibco, Ireland
Trypsin-EDTA	Gibco, Ireland
Tween20	Sigma, Germany

APPENDIX B: Equipment

Equipment	Supplier Company
Centrifuges	
Biofuge Stratos Heraeus	Thermo Scientific
Centrifuge MegaStar	VWR
Sorvall LYNX 6000 Centrifuge	Thermo Fisher Scientific
Tabletop Centrifuge	Eppendorf
Allegra X 5	Beckman Coulter
Bench equipment	
Finnpipette™ F2 GLP set	Thermo Fisher Scientific
Econo-column®	BIO RAD
PES Membrane VivaSpin20	Sartorius
pHmetre, SevenCompact	Mettler Toledo
Sonicators	
VibraCell	Sonics
Q500	QSonica, Sonicators
Spectrophotometers	
Nanodrop	Thermo Fisher Scientific
Ultrospec Spectrophotometer	Amersham Biosciences (GE)
Temperature Control incubators	
Innova 44 shaker incubator	Eppendorf New Brunswick
Thermocycler	MJ Research
Thermomixer	Eppendorf
CO_2 incubator	Binder

Visualization

Luminescent Image Analyzer	Fujifilm
Gel Doc EZ	BIO RAD
Optical and Fluorescence Microscopes	AxioVert.A1, ZEISS

Protein purification and crystallization

AKTA pure, FPLC	GE
Phoenix, liquid dispenser	Art Robbins Instruments

Materials and Preparation Kits	Supplier Company
MIDI prep DNA plasmid extraction kit	Macherey-Nagel
Enzyme purification and GEL extraction	Macherey-Nagel
Mutagenesis Kit	NEB
Agarose gel electrophoresis system	Enduro
SDS protein gel electrophoresis and WB	BIO RAD
Low MW Calibration Kit	GE

Size Exclusion Chromatography (SEC) columns

Superdex 200 Increase 10/300 GL	GE
Superdex 200 Increase 5/150 GL	GE
HiLoad Superdex 75 pg 16/600	GE
HiLoad Superdex 200 pg 16/600	GE

APPENDIX C: Vector maps
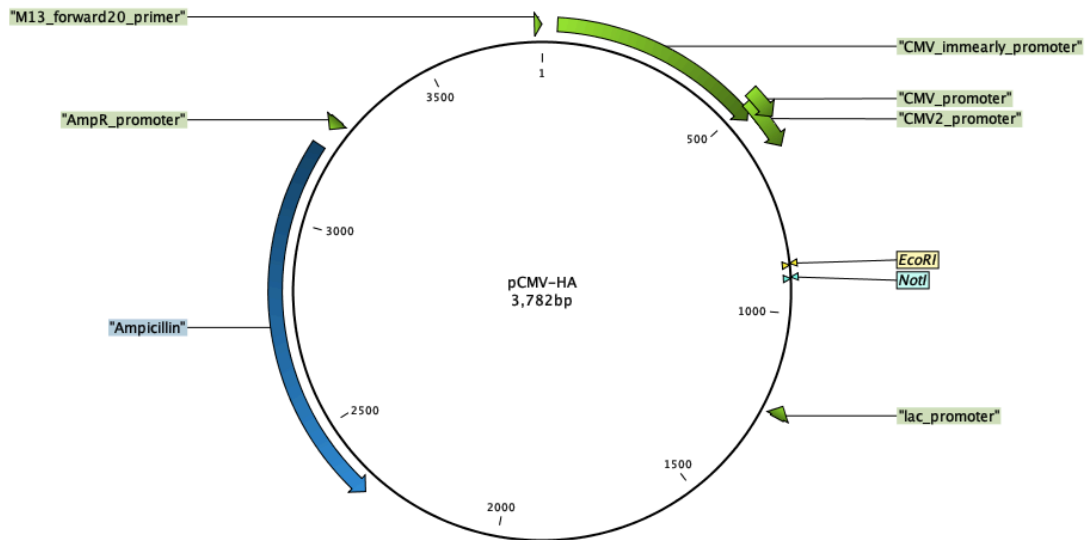
Figure C.1 Map of pCMV-HA plasmid
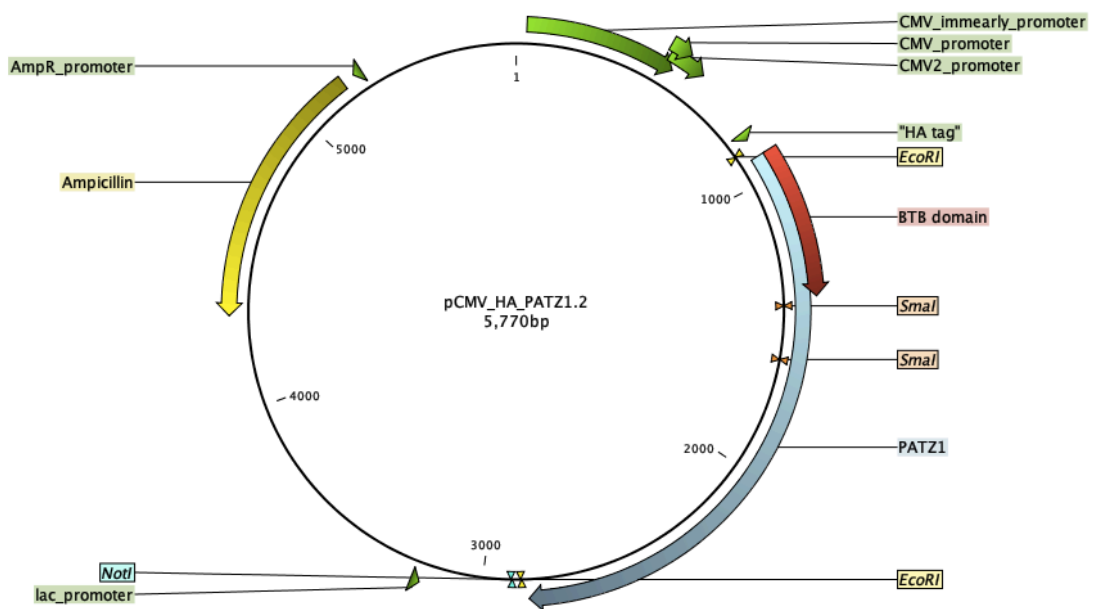
Figure C.2 Map of pCMV-HA mouse PATZ1 plasmid
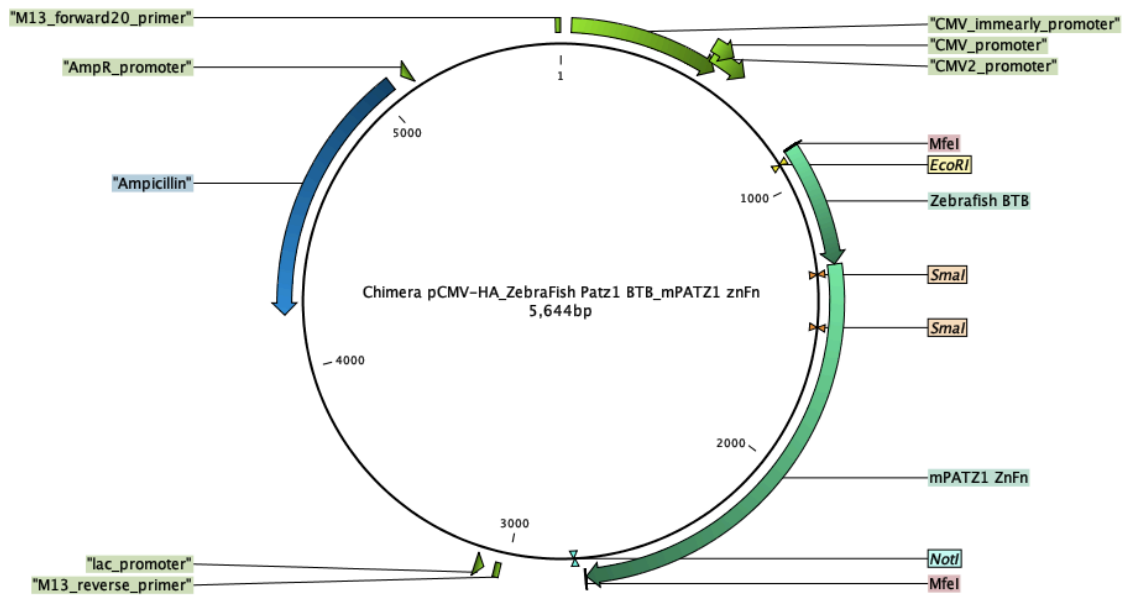
Figure C.3 Map of pCMV-HA-Chimera PATZ1 plasmid
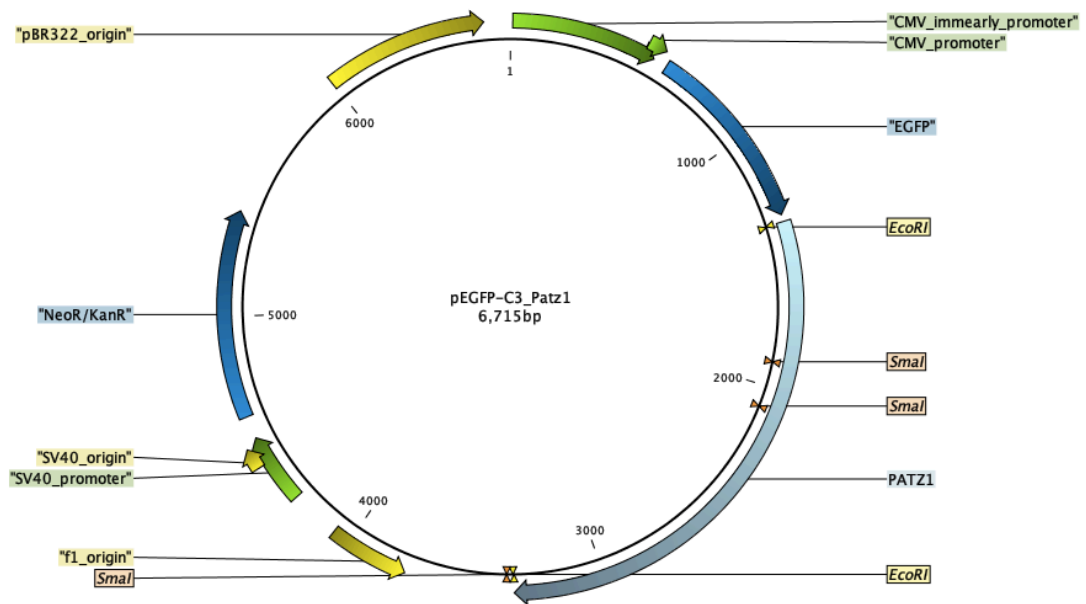
Figure C.4 Map of pEGFP-C3 mouse PATZ1 plasmid
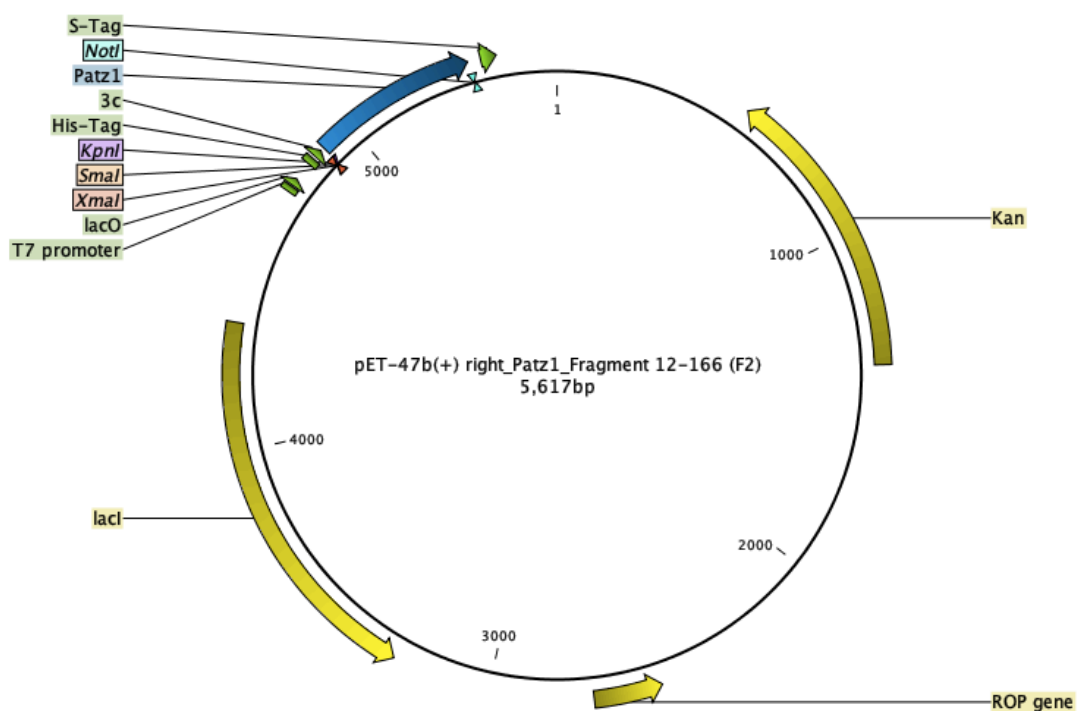
Figure C.5 Map of pET-47b (+) mouse PATZ1 BTB domain plasmid
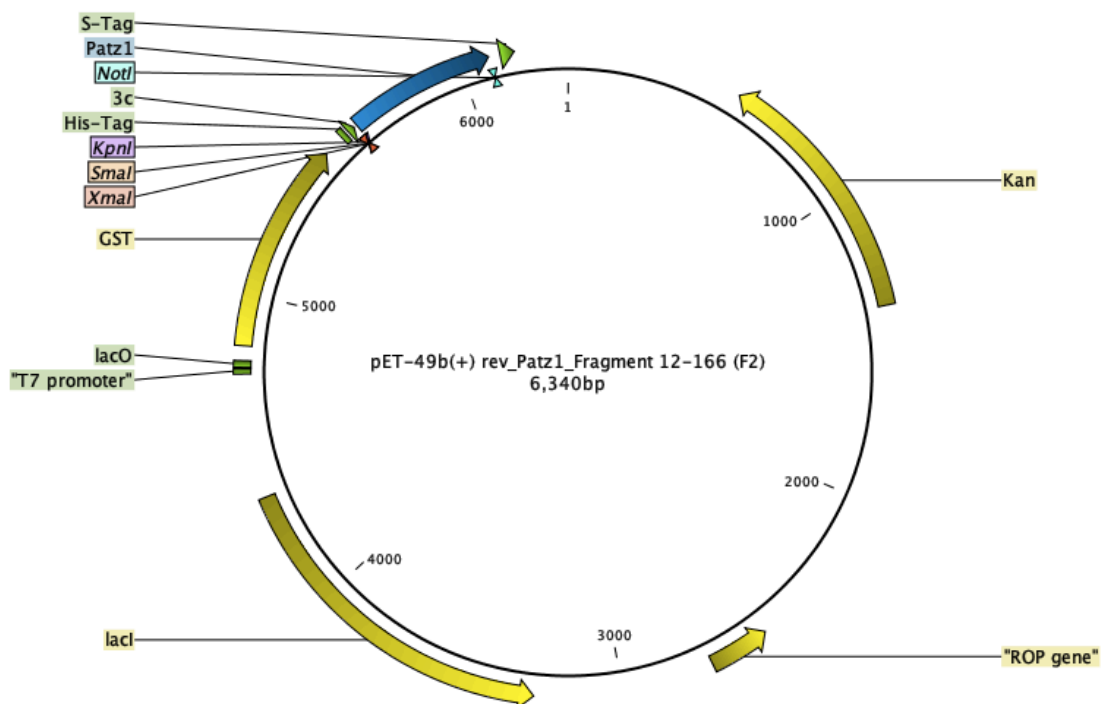
Figure C.6 Map of pET-49b (+) mouse PATZ1 BTB domain plasmid
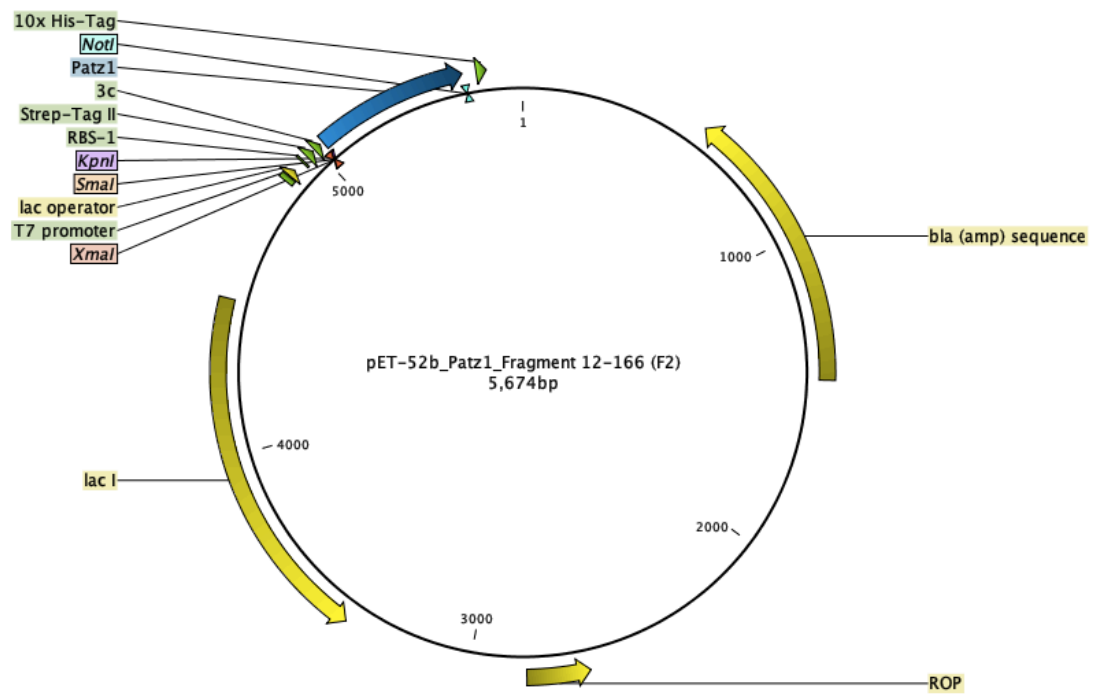
Figure C.7 Map of pET-52b (+) mouse PATZ1 BTB domain plasmid
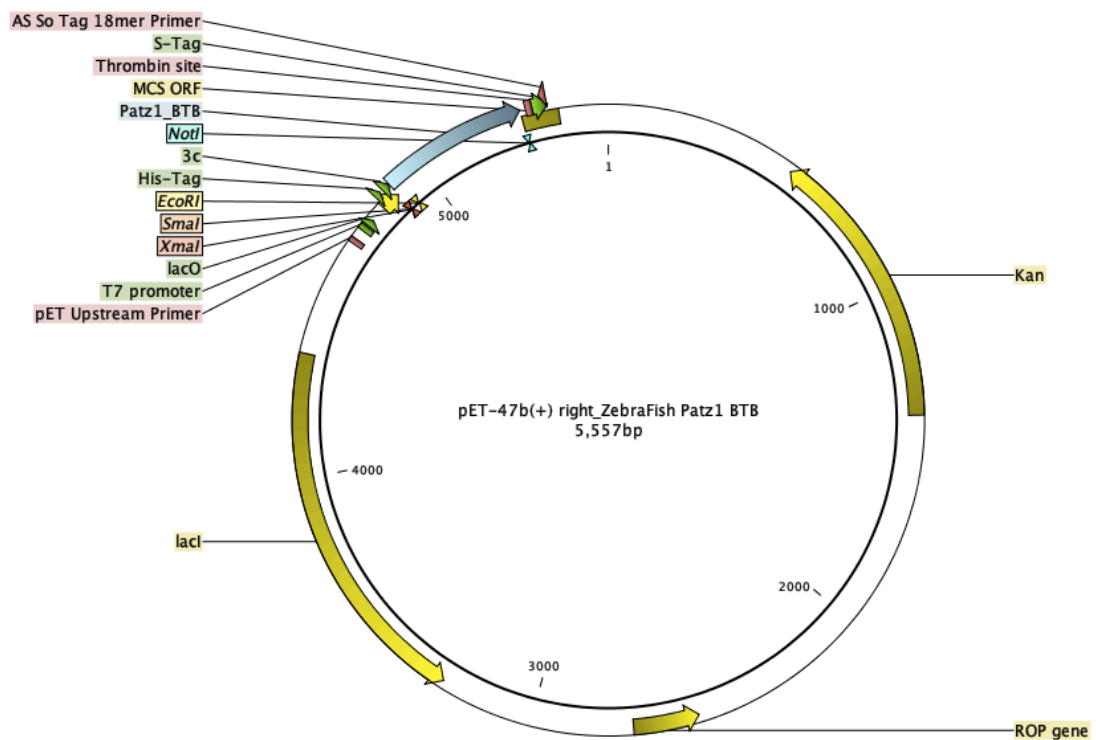
Figure C.8 Map of pET-47b (+) zebrafish PATZ1 BTB domain plasmid
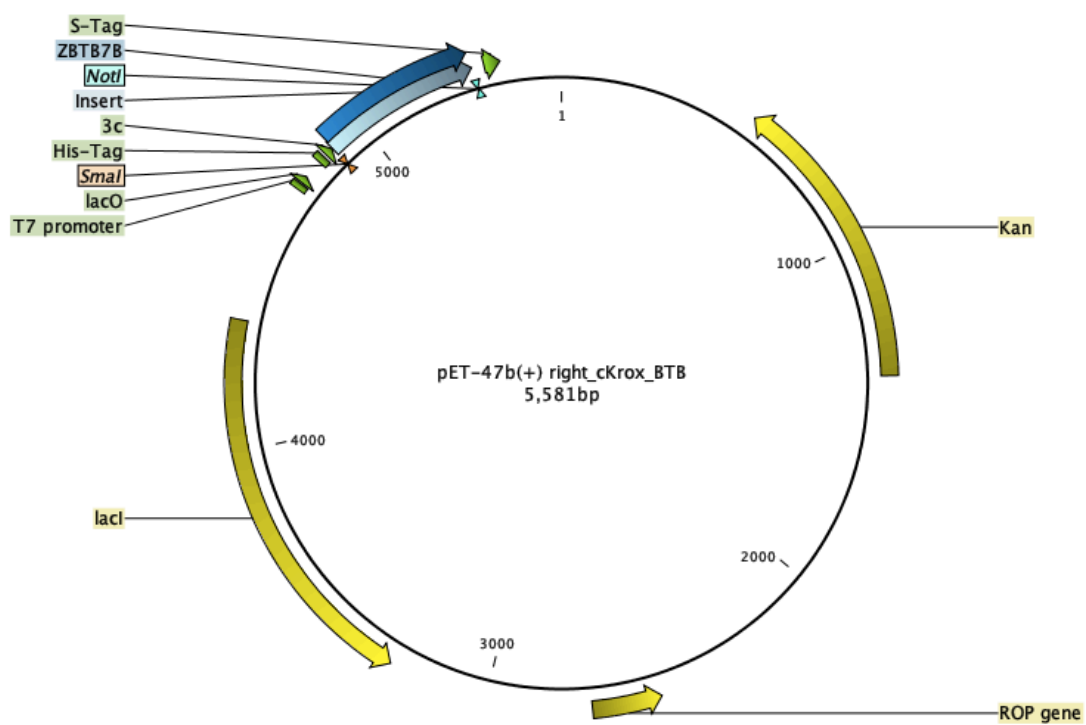
Figure C.9 Map of pET-47b (+) human cKrox BTB domain plasmid
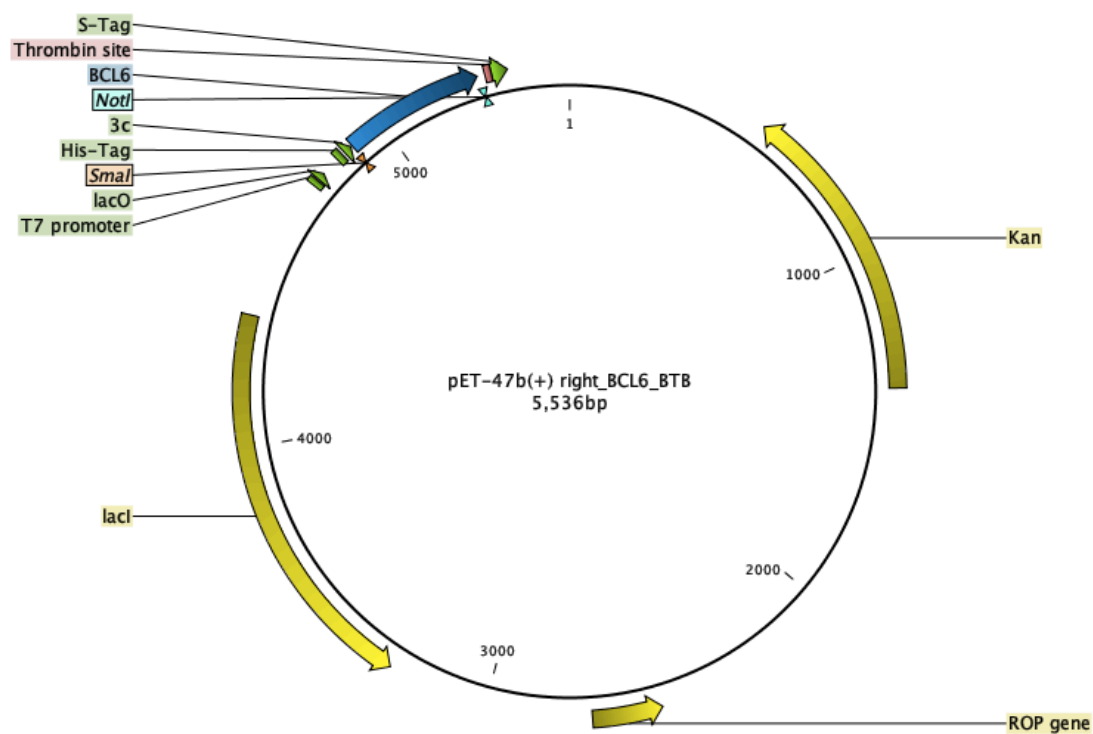
Figure C.10 Map of pET-47b (+) human BCL6 BTB domain plasmid
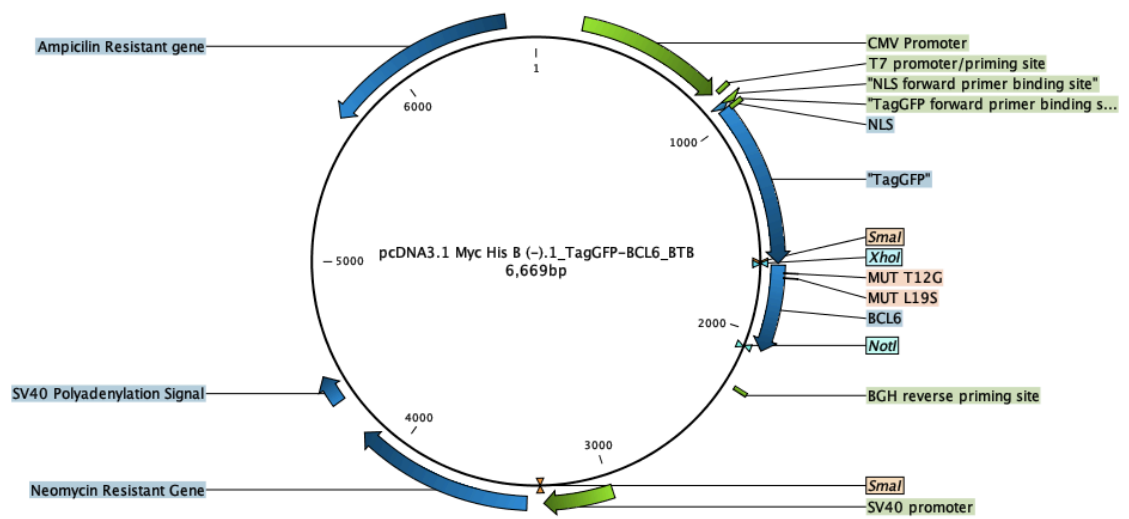
Figure C.11 Map of pcDNA3.1-Myc-His-B (-) Tag GFP-human-BCL6-BTB plasmid

Construct	Patz1 fragment	Bp	Translated AA	MW (kDa)	Abs 0.1 % ext. Coeff.
AC73	1-159	477	159	17.2	0.32
F1	12-157	438	146	15.6	0.39
F2	12-166	465	155	16.6	0.37
F3	18-157	420	140	14.9	0.2
F4	18-166	447	149	15.9	0.19
F5	285-531	741	247	27.6	0.45
F6	285-589	915	305	34	0.46
F7	285-629	1035	345	38.2	0.41
F8	349-531	549	183	20.8	0.52
F9	349-589	723	241	27.2	0.52
F10	349-629	834	281	31.3	0.44
SP1	12-227	648	216	22.6	0.27
SP2	18-227	630	210	21.9	0.14
SP3	261-589	813	271	30.3	0.4
SP4	261-589	987	329	36.6	0.41
SP5	261-629	1107	369	40.8	0.37
SP6	12-531	1560	520	56.2	0.32
SP7	12-589	1734	578	62.5	0.34
SP8	12-629	1854	618	66.7	0.32
HIS Tag (3C digestion site)		57	19	2.2	

Table D.1 List of PATZ1 protein fragments designed for expression and purification.

Protein domain	Amino Acid Sequence	Protein Position index
BTB	SGCYTYQVSRHSTEMLHNLNQQRK NGGRFCDVLLRVGDESFPAHRAVLA ACSEYFESVFSAQLGDGGAADGGPA DVGGATAAPGGGAGGSRELEMHTI SSKVFGDILDFAYTSRIVVRLESFPEL MTAAKFLLMRSVIEICQEVIKQS	12-159
AT-hook	KRGRGRPRKANLL	264-276
ZF 1	LPCGLCGKVFTDANRLRQHEAQH	292-314
ZF 2	VACEICGKIFRDVYHLNRHKLSH	355-377
ZF 3	YSCPVCGLRFKRKDRMSYHVRSH	383-405
ZF 4	YICQSCGKGFSRPDHLNGHIKQVH	413-436
ZF 5	HKCQTCNASFATRDRLRSHLACH	442-464
ZF 6	NFCSICNRGFSSASYLKVHVKTHH	495-518
ZF 7	YPCPECGSFFRSKSYLNKHIQKVH	605-628

Table D.2 Features of PATZ1 protein sequence

BTB Protein Name	PDB ID	Inter-chain Salt Bridges	Frequency in MD (%)	Intra-Monomer Salt Bridges	
				Intra-interface	All protein
mouse PATZ1	6GUV	E75-R47 (crystal only)	0	D50-R64	D127-R42
		D50-R47 (B-A)*	64.84	D124-R137	E111-R54*
		D84-R47*	7.54 (BA)-32.77 (AB)	D84-R64*	E111-R110 (B-B)*
		D89-R47 (A-B)*	11.14	D94-R110*	D127-K43*
		E158-R29*	32.32 (AB)-76.82 (BA)	E72-R154*	E143-K120*
				D50-R110 (A-A)*	D57-R54*
				D89-R64 (A-A)*	D57-K120*
				D89-R110*	D124-K120*
				D89-R47 (B-B)*	E58-R110 (A-A)*

				D127-R133*	E58-K120*
				E33-R29*	E58-K43 (B-B)*
				E75-R64 (A-A)*	E139-K120*
				E139-R137*	D94-R64 (A-A)*
				E143-R137*	
				E158-R154*	
				E162-K165*	
zebrafish PATZ1	6GUW	E64-R36	77.42 (BA)-85.7 (AB)	D92-R105	D95-R31
		D39-K53 (A-B)*	5.44	D33-R36*	E61-R122
		D74-K78*	5.89 (AB)	D39-K53*	E107-K88
		E22-R122 (A-B)*	12.84	D39-K78 (A-A)*	D33-K32*
		E72-R36 (A-B)*	93.06	D73-K53 (A-A)*	D92-R91*
		E126-K18*	19.08 (AB)-28.42 (BA)	D92-K101 (A-A)*	D92-K88*
				D95-K101*	D95-R91*
				E22-K18*	D95-K32*
				E22-R122 (A-A)*	E46-R43*
				E72-K53*	E46-R91 (A-A)*
				E107-R105*	E46-K88*
					E47-R91*
					E47-K88*
					E79-R43*
					E79-K78*
					E81-R43*
					E111-K88*
					E126-R122 (A-A)*

*added with MD simulation of the modelled structures

Residue numbers refer to the crystal structures.

Table D.3 Inter-chain salt bridges in the dimerization interface of BTB domain of PATZ1

Primer name	Oligonucleotides sequence	Cloning purpose
Patz1_12Fw	TGGACCCGGG TCGGGCTGCTACACCTACCAGG	Patz1 fragments
Patz1_18FW	TGGACCCGGGCAGGTGAGCAGACA CAGTACGGAG	Patz1 fragments
Patz1_157RV	GGACGCGGCCGCTTATTATTTGATT ACTTCCTGGCAGATCTCGATGAC	Patz1 fragments
Patz1_166Rv	GGACGCGGCCGCTTATTAGGGCACG AGGATCTGCACGTTG	Patz1 fragments
Patz1_285FW	TGGACCCGGGGGCTTGAGGGAAGC AGGCATCC	Patz1 fragments
Patz1_531RV	GGACGCGGCCGCTTATTATGGCGTC TTCAGGTCGCTGGC	Patz1 fragments
Patz1_349FW	TGGACCCGGGAGCCGGACCAGGAA GCAAGTGG	Patz1 fragments
Patz1_589RV	GGACGCGGCCGCTTATTAGGGACCC CCAAGGGCACG	Patz1 fragments
Patz1_629RV	GGACGCGGCCGCTTATTA AGGATC CACTAAAGATGATGCAAACGCTG	Patz1 fragments
Patz1_261FW	TGGACCCGGGCTAACTAGCAAGCGA GGCCGGG	Patz1 fragments
Patz1_227RV	GGACGCGGCCGCTTATTAAGGCAGA GAAGCTTGGCCCGC	Patz1 fragments
Patz1_1FW_EcoRI	TGGAGAATTCATGGAGCGGGTCAAC GACGCTTC	EGFP-Patz1
Patz1_1FW	TGGACCCGGGATGGAGCGGGTCAAC GACGCTTC	Patz1 fragments
Patz1_159RV	GGACGCGGCCGCTTATTAGGACTGT TTGATTACTTCCTGGCAGATC	Patz1 fragments
Patz1_Δloop_1FW_ Sma	TGGACCCGGGATGGAGAGGACAGA GCAGCCGTG	Zebrafish Patz1 BTB
Patz1_Δloop_135RV _Not	GGACGCGGCCGCTTATTAGGACTGT TTGATGACCTCCTGGCA	Zebrafish Patz1 BTB
cKrox1_FW	TGGACCCGGGATGGGGAGCCCCGA GGATGAC	cKrox BTB
cKrox1_RV	GGACGCGGCCGCTTATTAACTGCCC TGCAGAAT	cKrox BTB
BCL6_BTBFw:	TGGACCCGGGATGGCCTCGCCGGCT GACAG	BCL6 BTB
BCL6_BTBRv:	GGACGCGGCCGCTTATTCACTGGCC TTAATAAACTTCCGGCAAG	BCL6 BTB
Fw-BCL6-full	TGGACAATTGCCATGGCCTCGCCGG CTGACAG	BCL6 from genomic DNA

Rv-BCL6-full	GGACCAATTGTCAGCAGGCTTTGGG GAGCTCC	BCL6 from genomic DNA
Chimera_FW1:	TGGACAATTGCCATGGAGAGGACAG AGCAGCCGTGGAATTCCTCCTACAC G	Chimera PATZ1
Chimera_RV1:	GGACCGGGCAGGGGGCACGAGGAT CTGCACGTTGGACTGTTTGATGACC TCCTGGCAGATCTCGATGACGGAC	Chimera PATZ1
Chimera_FW2:	TGGAGTCCGTCATCGAGATCTGCCA GGAGGTCATCAAACAGTCCAACGTG CAGATCCTCGTGCCCCCTGCCCG	Chimera PATZ1
Chimera_RV:	GGACCAATTGTCACTTCCCTTCAGG CCCCATGGGCTGCTGGTC	Chimera PATZ1
BCL6_L19S_FW	CAGTGATGTTagcCTCAACCTTAATC GTCTCCG	BCL6 BTB mutagenesis
BCL6_L19S_RV	GCATGGCGGGTGAACTGG	BCL6 BTB mutagenesis
BCL6_T12G_FW	TATCCAGTTCggCCGCCATGCCAGTG ATGTTC	BCL6 BTB mutagenesis
BCL6_T12G_RV	CAGCTGTCAGCCGGCGAG	BCL6 BTB mutagenesis

Table D.4 List of cloning primers

```
              5              10             15             20
H: ATGGAGCGGGTGAACGACGCTTCGTGCGGCCCGTCTGGCTGCTACACATACCAGGTGAGC
M: ATGGAGCGGGTCAACGACGCTTCTTGCGGTCCGTCGGGCTGCTACACCTACCAGGTGAGC
H: -M--E--R--V--N--D--A--S--C--G--P--S--G--C--Y--T--Y--Q--V--S-
M: -M--E--R--V--N--D--A--S--C--G--P--S--G--C--Y--T--Y--Q--V--S-
              25             30             35             40
H: AGACACAGCACGGAGATGCTGCACAACCTGAACCAGCAGCGCAAAAACGGCGGGCGCTTC
M: AGACACAGTACGGAGATGCTGCACAACCTGAACCAACAACGCAAAAACGGCGGGCGCTTT
H: -R--H--S--T--E--M--L--H--N--L--N--Q--Q--R--K--N--G--G--R--F-
M: -R--H--S--T--E--M--L--H--N--L--N--Q--Q--R--K--N--G--G--R--F-
              45             50             55             60
H: TGCGACGTGCTCTTGCGGGTAGGCGACGAGAGCTTCCCAGCGCACCGCGCCGTGCTGGCC
M: TGCGACGTGCTCCTACGAGTAGGCGACGAGAGCTTCCCAGCGCACCGCGCCGTACTGGCT
H: -C--D--V--L--L--R--V--G--D--E--S--F--P--A--H--R--A--V--L--A-
M: -C--D--V--L--L--R--V--G--D--E--S--F--P--A--H--R--A--V--L--A-
              65             70             75             80
H: GCCTGCAGCGAGTACTTTGAGTCGGTGTTCAGCGCCCAGTTGGGCGACGGCGGAGCTGCG
M: GCCTGCAGCGAGTACTTTGAGTCTGTGTTCAGCGCCCAGTTAGGCGACGGCGGAGCTGCA
H: -A--C--S--E--Y--F--E--S--V--F--S--A--Q--L--G--D--G--G--A--A-
M: -A--C--S--E--Y--F--E--S--V--F--S--A--Q--L--G--D--G--G--A--A-
              85             90             95             100
H: GACGGGGGTCCGGCTGATGTAGGGGGCGCGACGGCAGCACCAGGCGGCGGGGCCGGGGGC
M: GATGGTGGTCCTGCTGATGTGGGGAGGCGCGGCGGCGGCTCCAGGCGGCGGAGCTGGAGGC
H: -D--G--G--P--A--D--V--G--G--A--T--A--A--P--G--G--G--A--G--G-
M: -D--G--G--P--A--D--V--G--G--A--A--A--A--P--G--G--G--A--G--G-
              105            110            115            120
H: AGCCGGGAGCTGGAGATGCACACTATCAGCTCCAAGGTATTTGGGGACATTCTGGACTTC
M: AGCCGCGAACTGGAGATGCACACCATCAGTTCCAAAGTGTTCGGAGACATCCTGGACTTC
H: -S--R--E--L--E--M--H--T--I--S--S--K--V--F--G--D--I--L--D--F-
M: -S--R--E--L--E--M--H--T--I--S--S--K--V--F--G--D--I--L--D--F-
              125            130            135            140
H: GCCTACACTTCCCGCATCGTGGTGCGCTTGGAGAGCTTTCCCGAACTCATGACGGCCGCC
M: GCTTATACGTCCCGAATCGTTGTGCGCCTAGAGAGCTTCCCCGAGCTCATGACGGCCGCC
H: -A--Y--T--S--R--I--V--V--R--L--E--S--F--P--E--L--M--T--A--A-
M: -A--Y--T--S--R--I--V--V--R--L--E--S--F--P--E--L--M--T--A--A-
              145            150            155
H: AAGTTCCTGCTGATGAGGTCGGTTATCGAGATCTGCCAGGAAGTCATCAAACAGTCC
M: AAGTTCCTGCTGATGAGGTCGGTCATCGAGATCTGCCAGGAAGTAATCAAACAGTCC
H: -K--F--L--L--M--R--S--V--I--E--I--C--Q--E--V--I--K--Q--S-
M: -K--F--L--L--M--R--S--V--I--E--I--C--Q--E--V--I--K--Q--S-
```

Figure D.1 Comparison of human (H) and mouse (M) PATZ1-BTB including DNA and protein sequence. Apart from a single amino acid difference (T91A), all other nucleotide substitutions are silent.
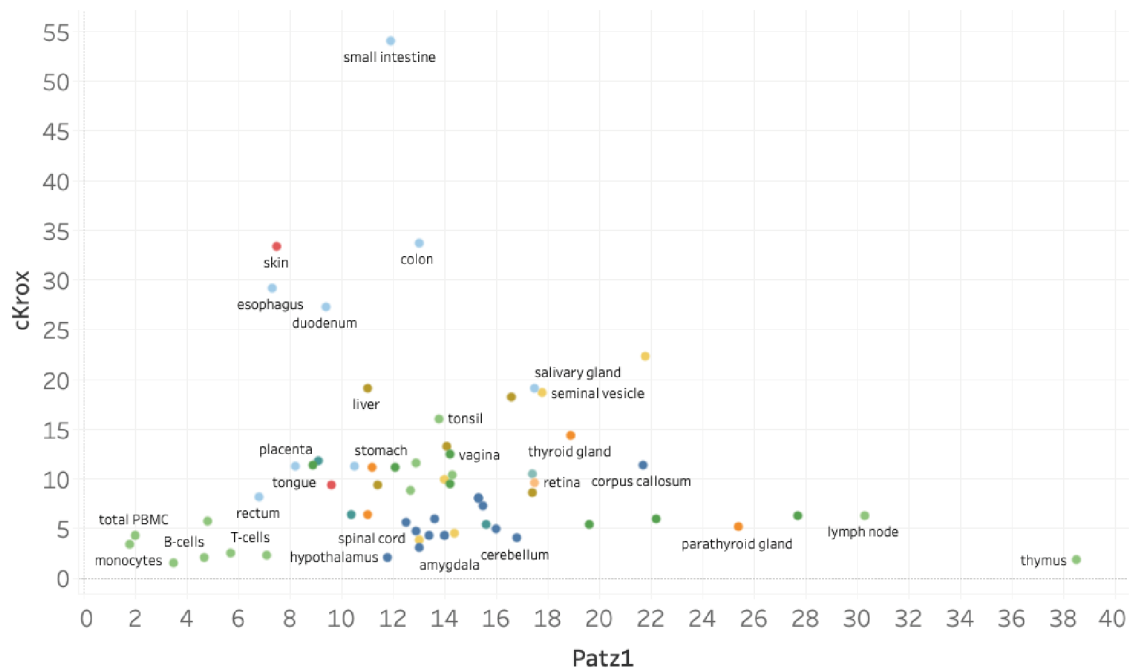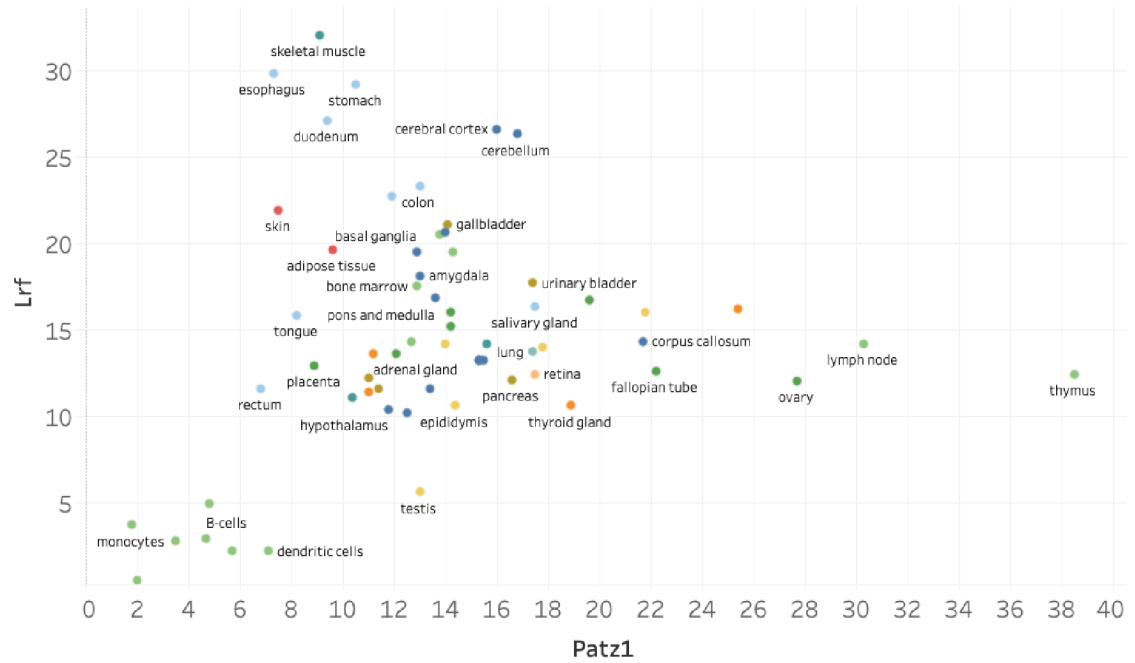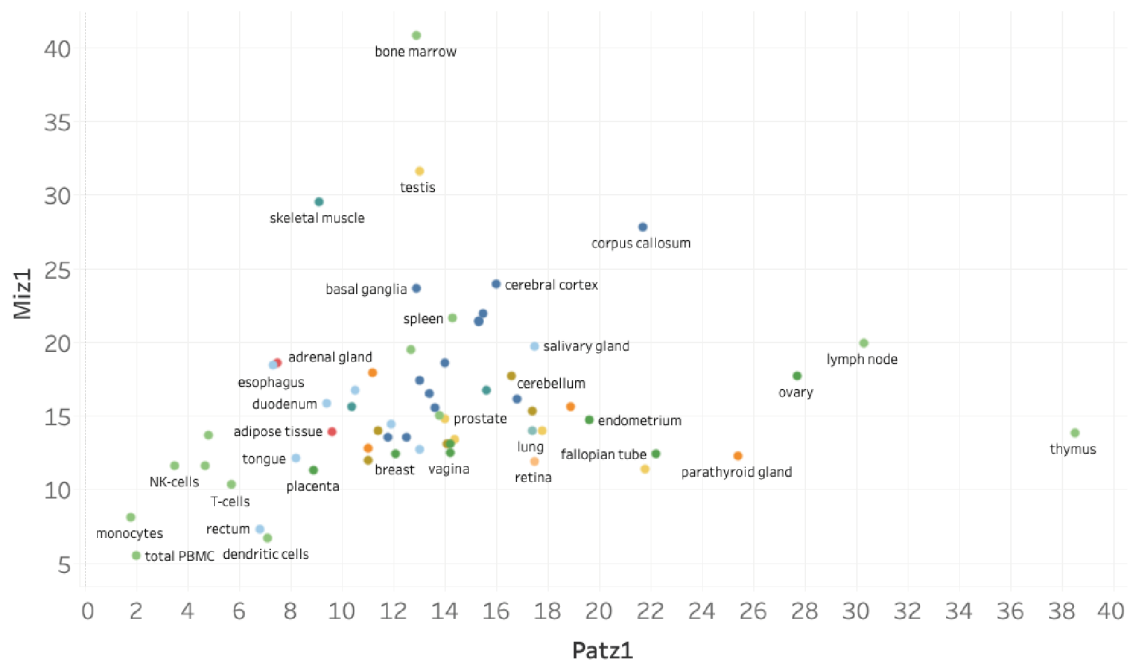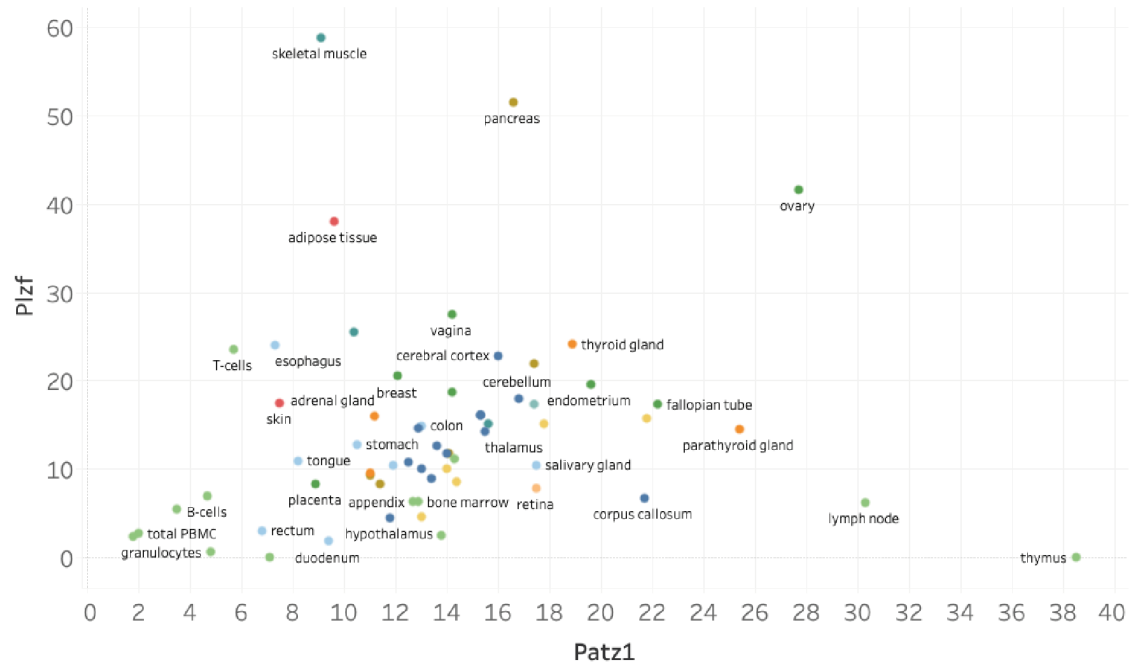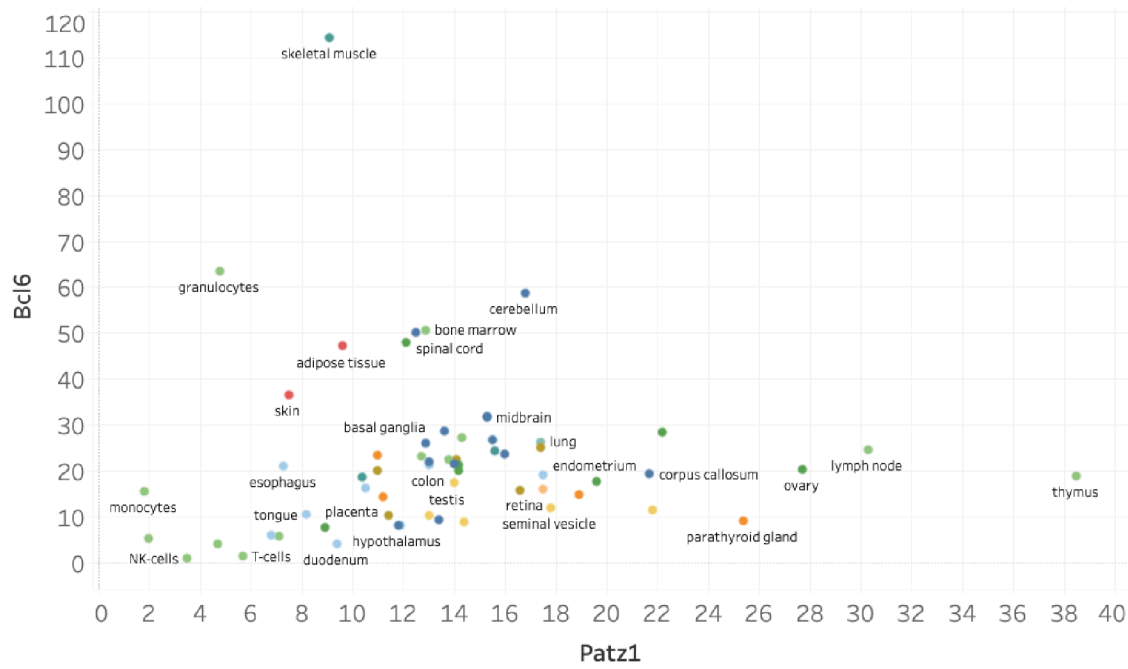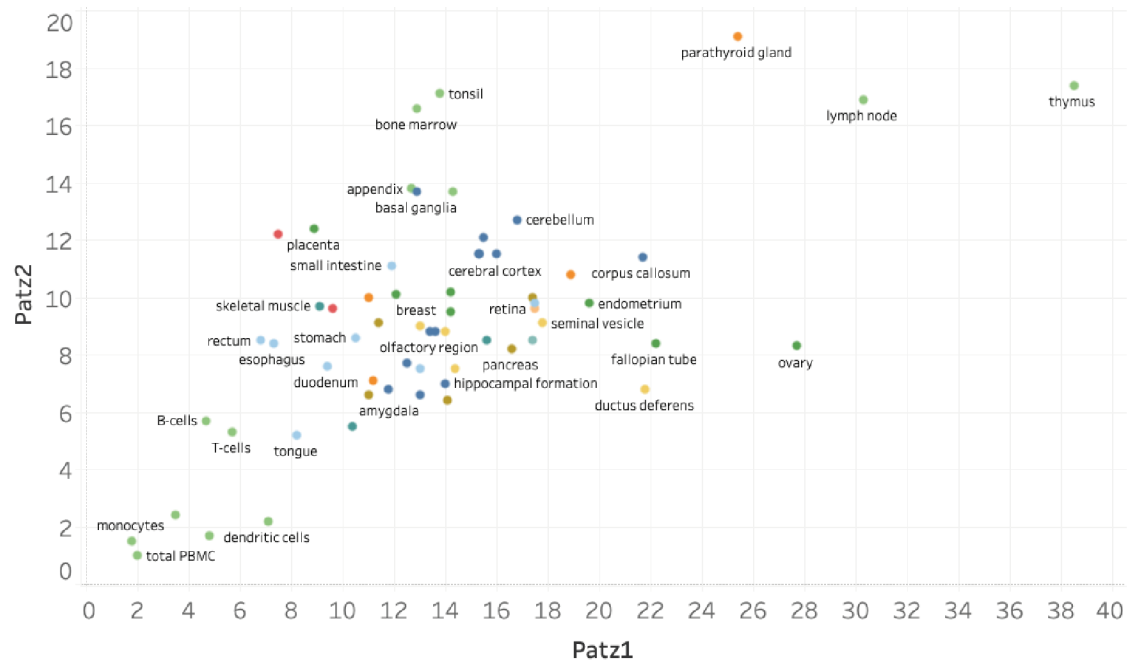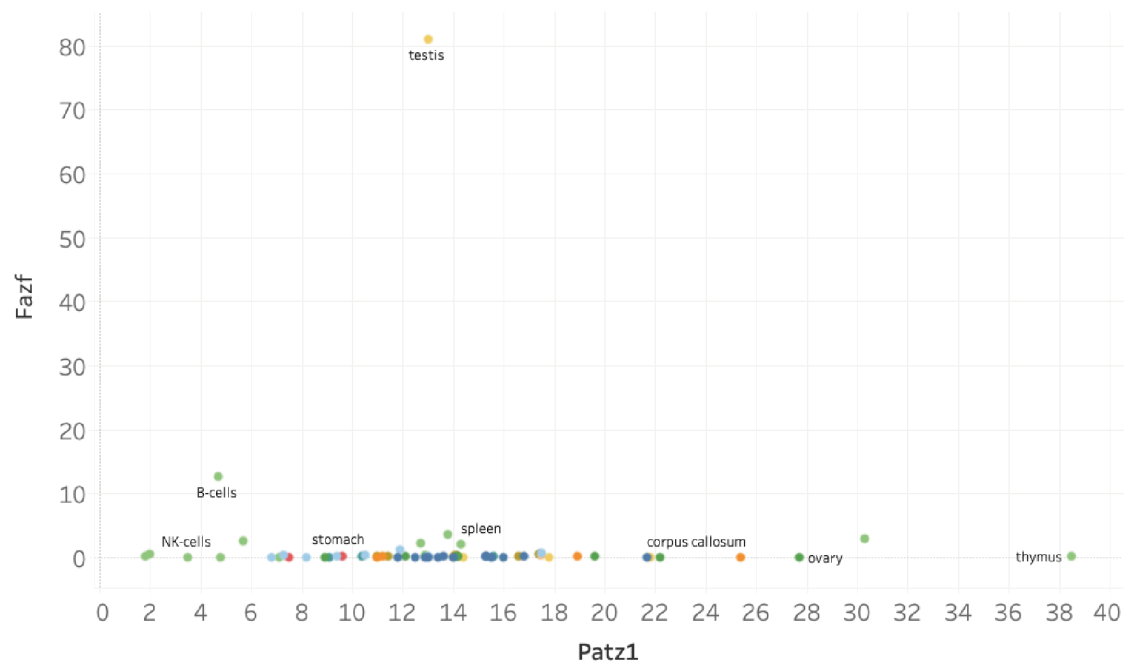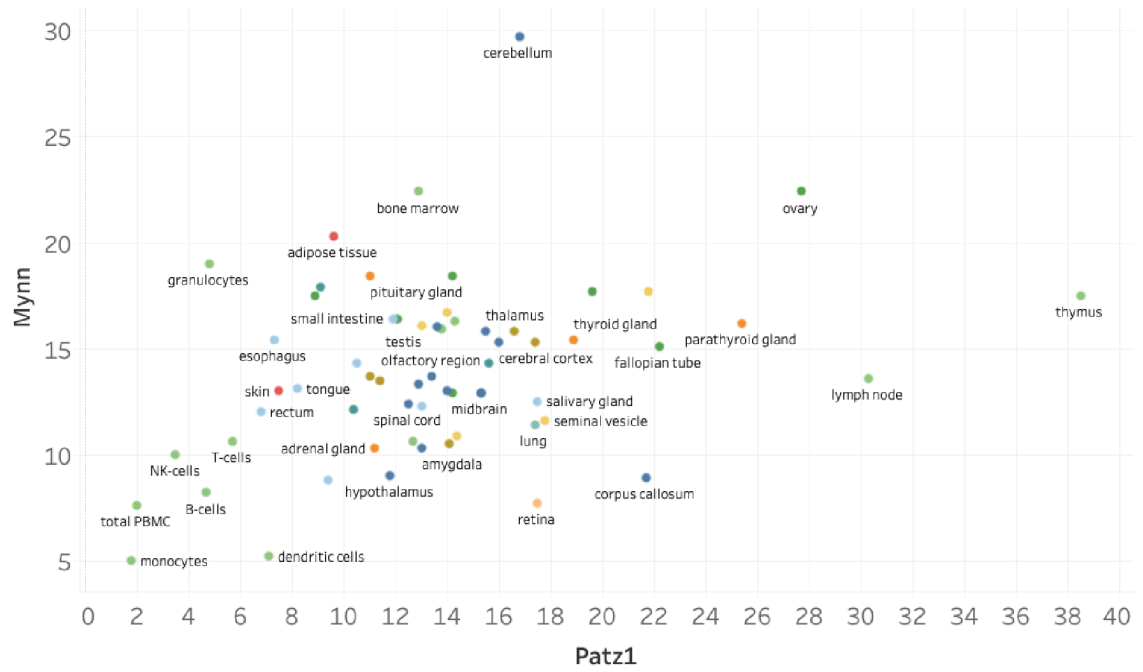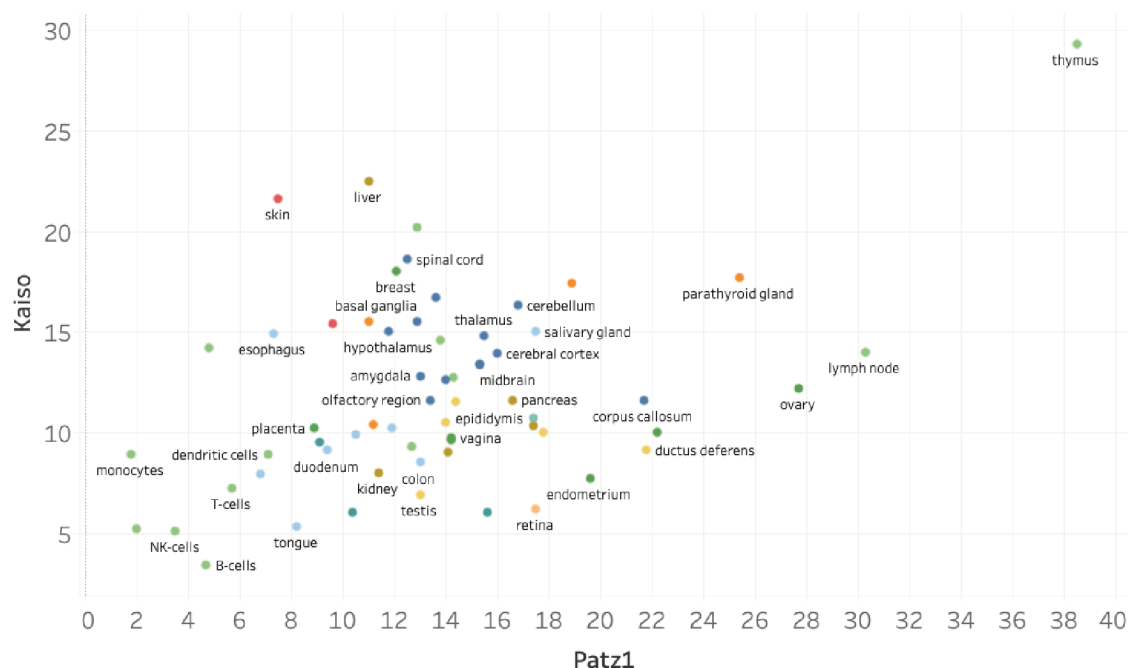
Figure D.2 Normal tissue RNA expression data plotted for ZBTB genes pairs. The normalized data from the ZBTB genes were collected from The Human Protein Atlas-RNA consensus dataset and compared to PATZ1 in nine pairs. Part 1/5: Lrf and cKrox vs Patz1; part 2/5: Plzf and Miz1 vs Patz1; part 3/5: Patz2 and Bcl6 vs Patz1; part 4/5: Mynn and Fazf vs Patz1; part 5/5: Kaiso vs Patz1 and legend.
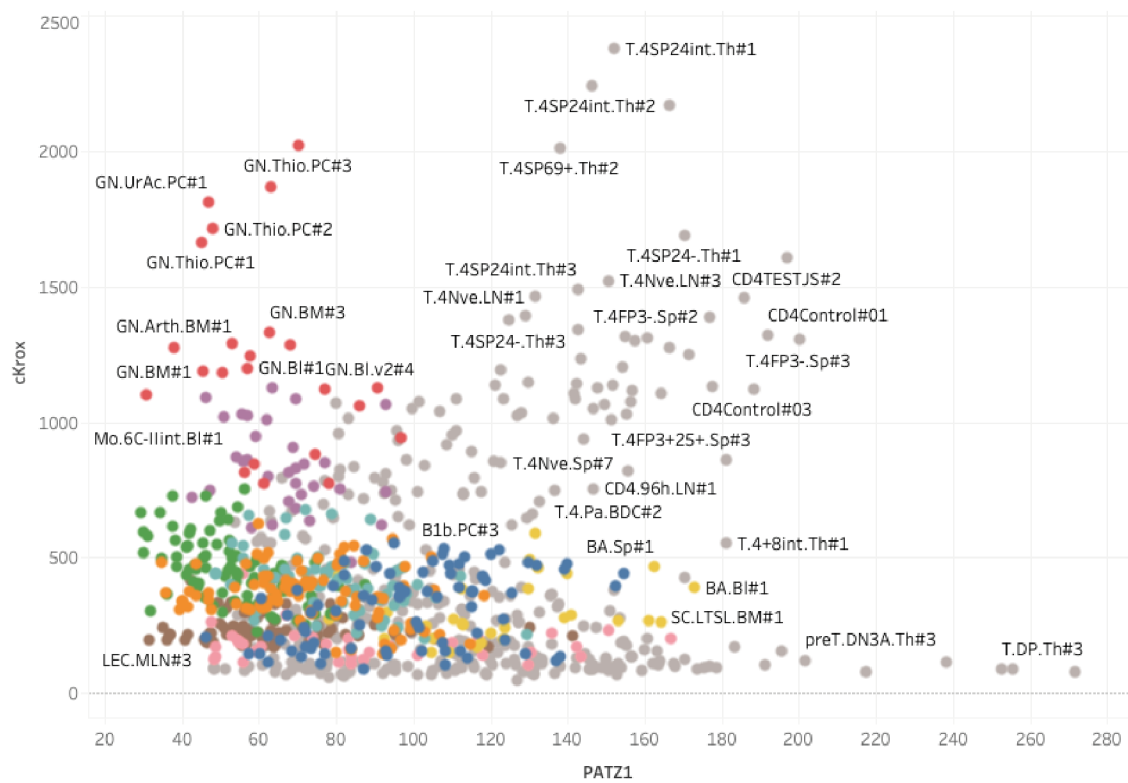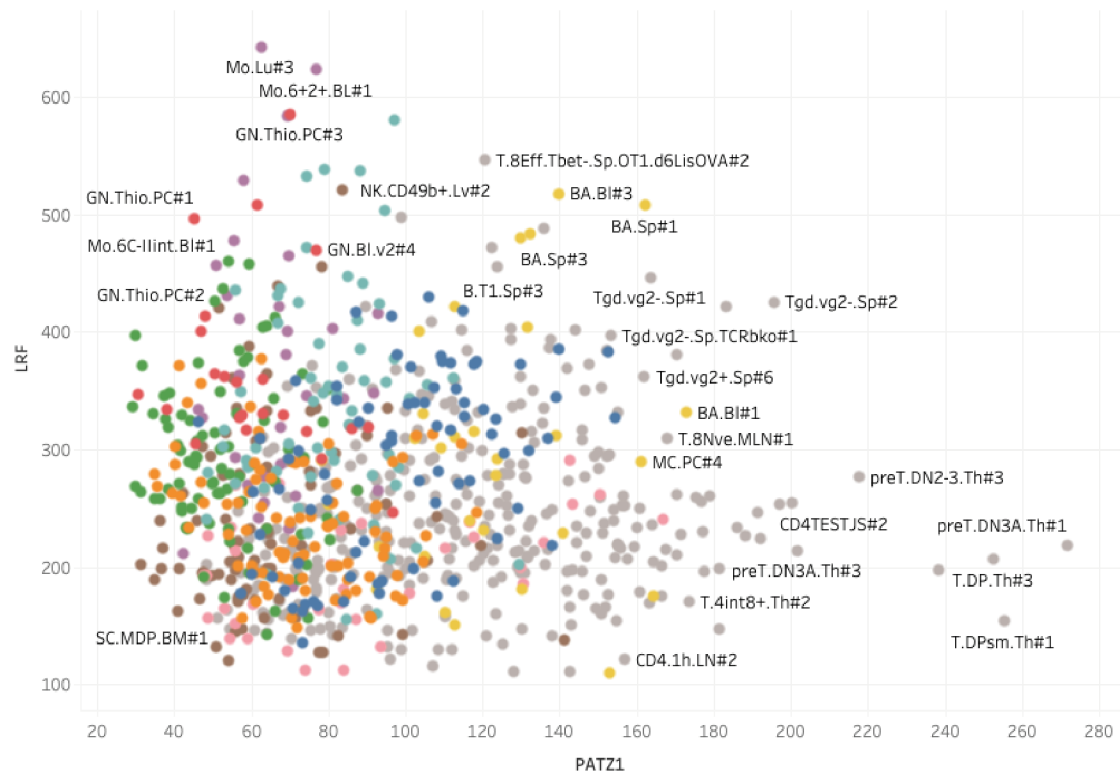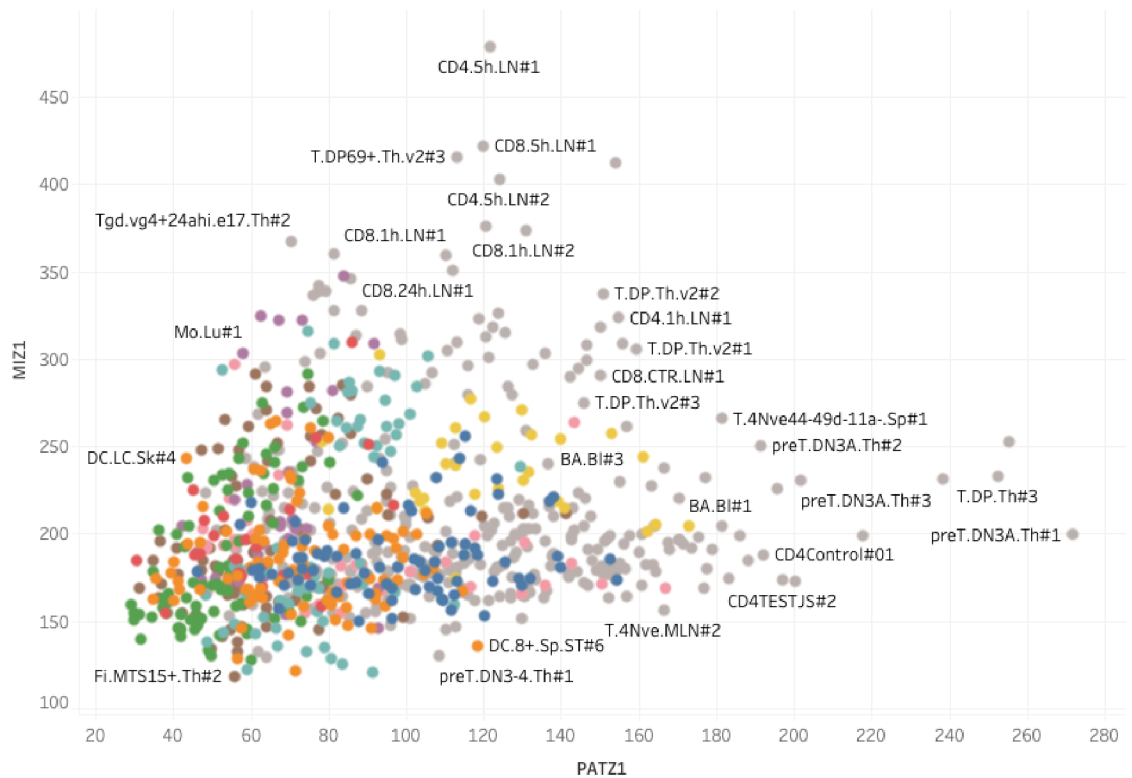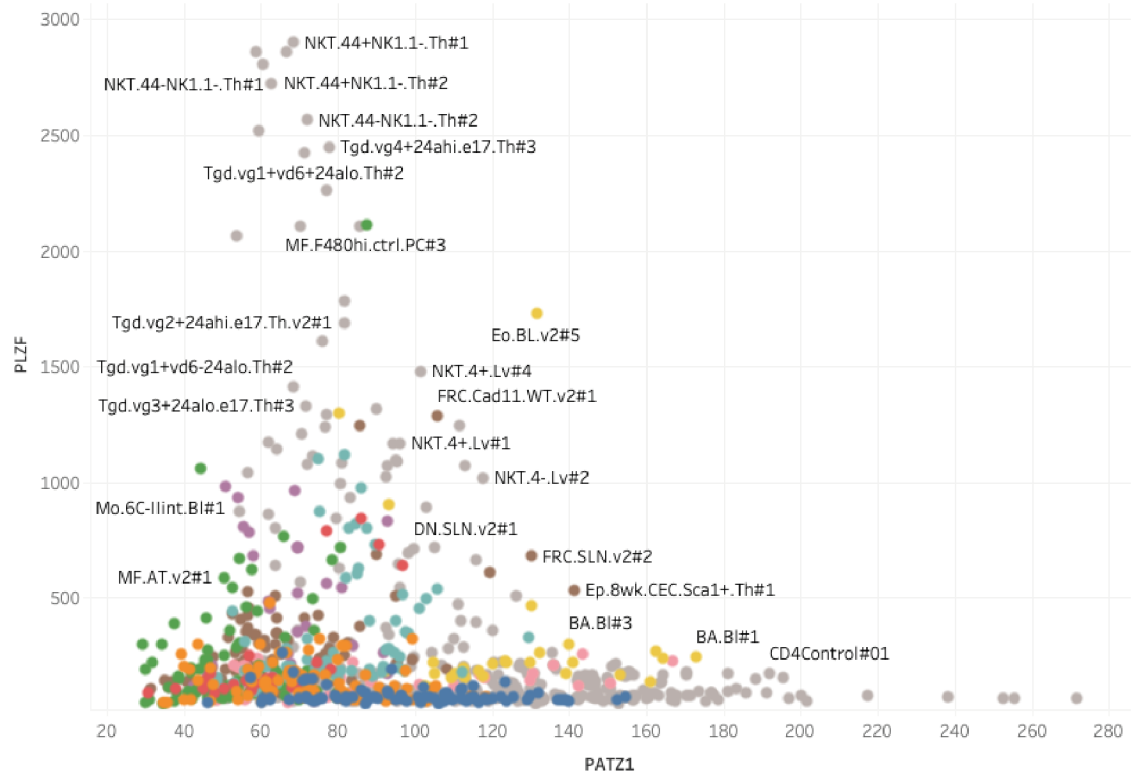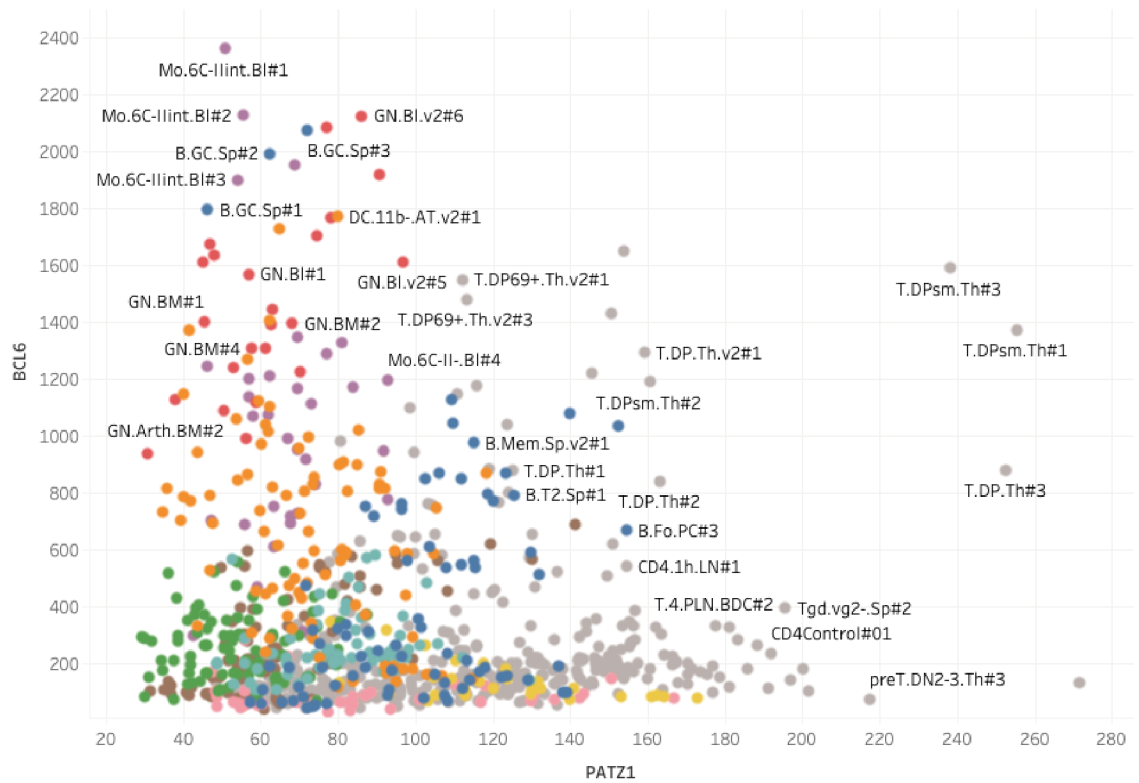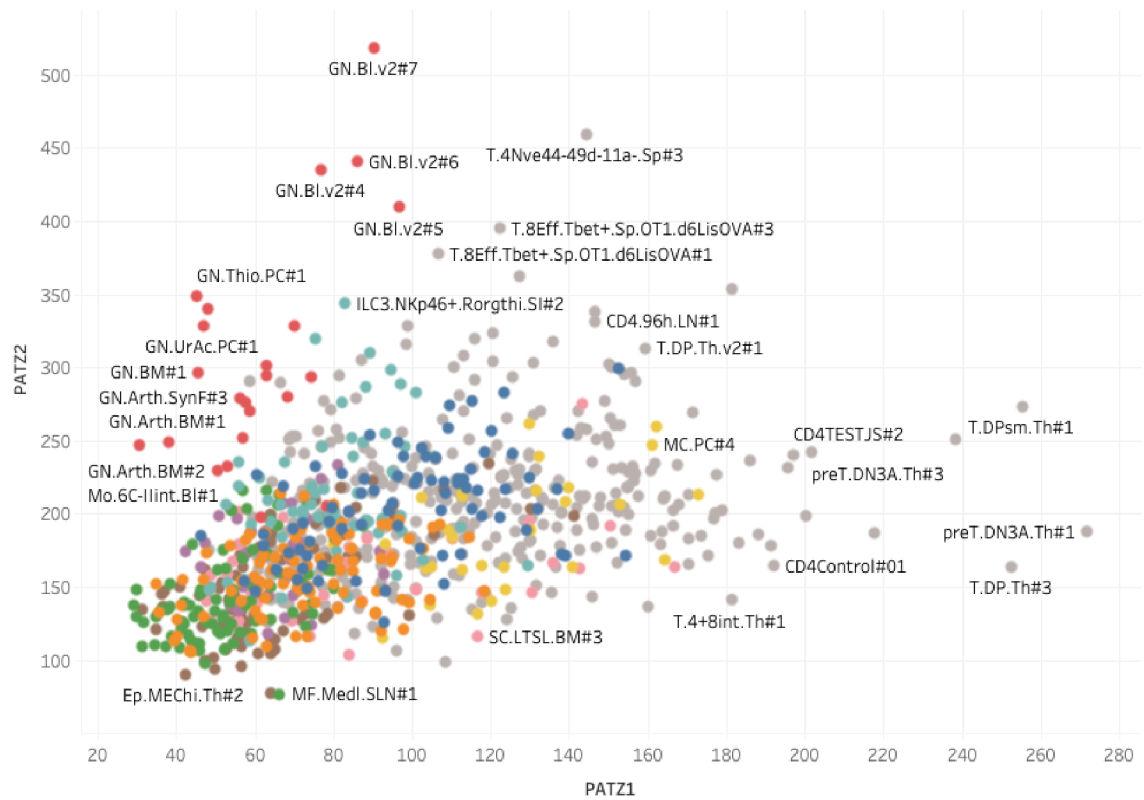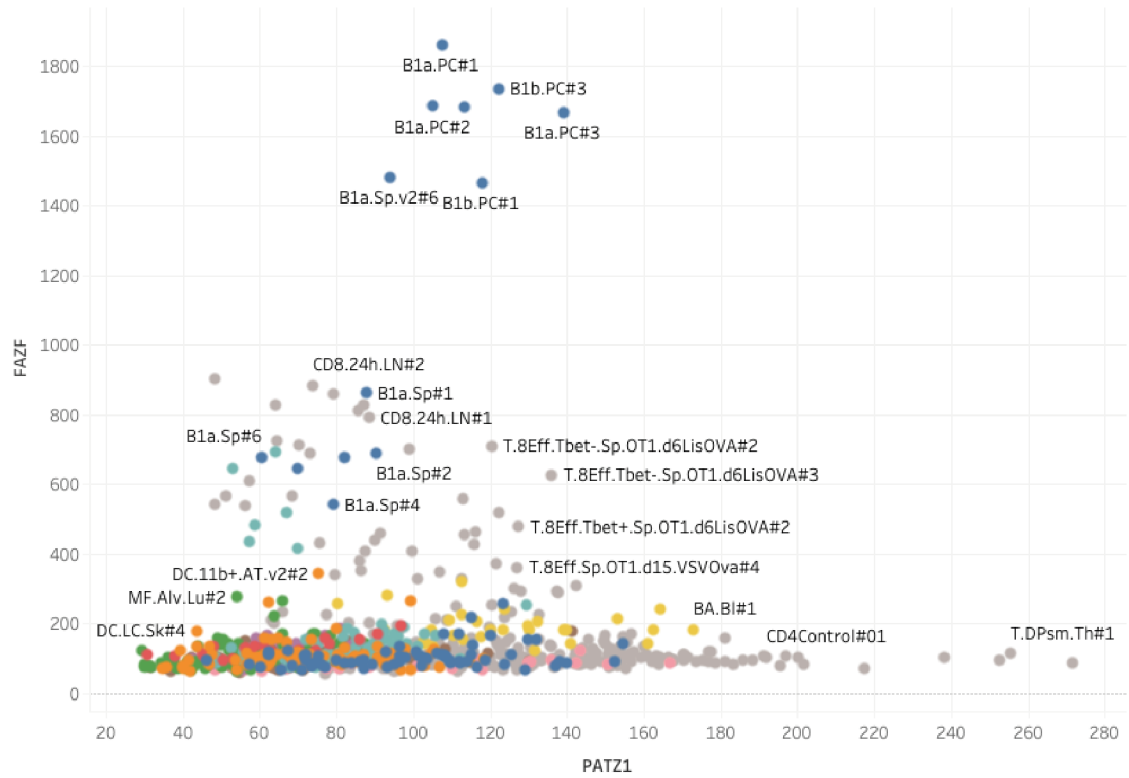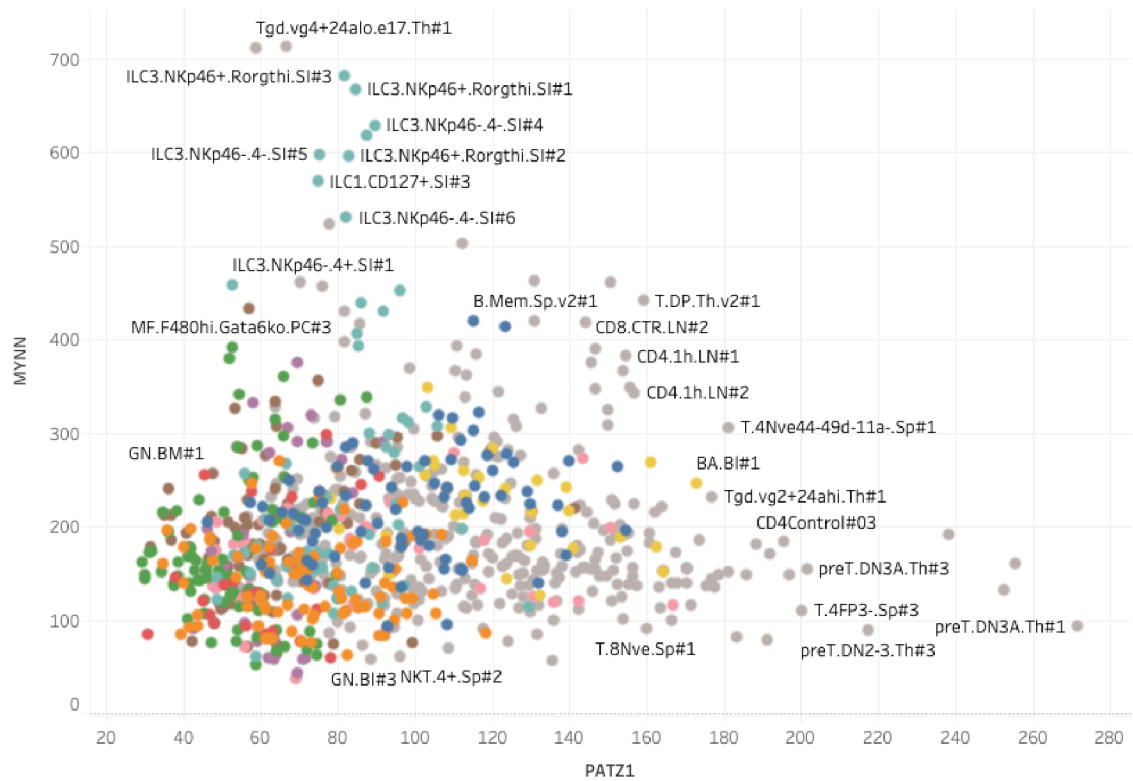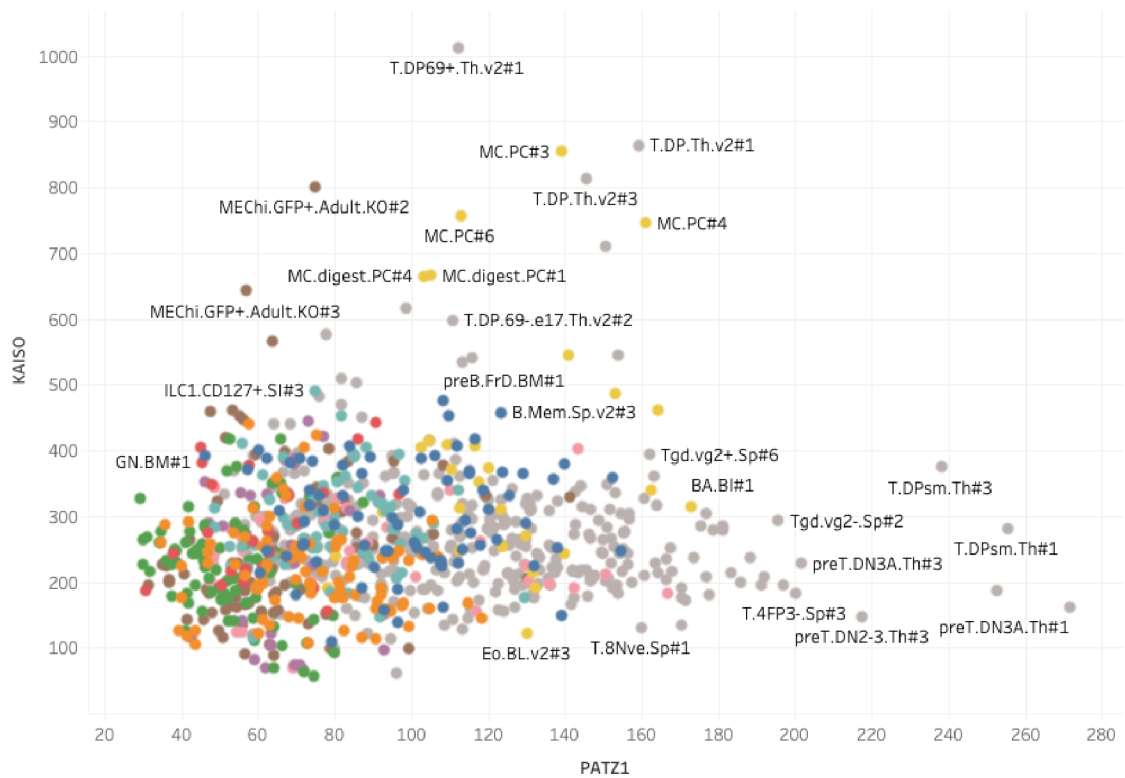
Figure D.3 Immune cells RNA expression data plotted for ZBTB genes pairs. The normalized data from the ZBTB genes were collected from ImmGen-Microarray Phase 1-2 datasets and compared to PATZ1 in nine pairs. Part 1/5: Lrf and cKrox vs Patz1; part 2/5: Plzf and Miz1 vs Patz1; part 3/5: Patz2 and Bcl6 vs Patz1; part 4/5: Mynn and Fazf vs Patz1; part 5/5: Kaiso vs Patz1 and legend.

References:

1.   Zollman S, Godt D, Prive GG, et al. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proc Natl Acad Sci U S A* 1994; 91: 10717-10721.

2.   Stogios PJ, Downs GS, Jauhal JJ, et al. Sequence and structural analysis of BTB domain proteins. *Genome Biol* 2005; 6: R82. DOI: 10.1186/gb-2005-6-10-r82.

3.   Stogios PJ and Prive GG. The BACK domain in BTB-kelch proteins. *Trends Biochem Sci* 2004; 29: 634-637. 2004/11/17. DOI: 10.1016/j.tibs.2004.10.003.

4.   Hudson AM, Mannix KM, Gerdes JA, et al. Targeted substrate degradation by Kelch controls the actin cytoskeleton during ring canal expansion. *Development* 2019; 146 2018/12/19. DOI: 10.1242/dev.169219.

5.   Dementieva IS, Tereshko V, McCrossan ZA, et al. Pentameric assembly of potassium channel tetramerization domain-containing protein 5. *J Mol Biol* 2009; 387: 175-191. 2009/04/14. DOI: 10.1016/j.jmb.2009.01.030.

6.   Smaldone G, Pirone L, Pedone E, et al. The BTB domains of the potassium channel tetramerization domain proteins prevalently assume pentameric states. *FEBS Lett* 2016; 590: 1663-1671. 2016/05/07. DOI: 10.1002/1873-3468.12203.

7.   Xie J, Jin Y and Wang G. The role of SCF ubiquitin-ligase complex at the beginning of life. *Reprod Biol Endocrinol* 2019; 17: 101. 2019/11/30. DOI: 10.1186/s12958-019-0547-y.

8.   Willems AR, Schwab M and Tyers M. A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim Biophys Acta* 2004; 1695: 133-170. 2004/12/02. DOI: 10.1016/j.bbamcr.2004.09.027.

9.   Cai W and Yang H. The structure and regulation of Cullin 2 based E3 ubiquitin ligases and their biological functions. *Cell Div* 2016; 11: 7. 2016/05/26. DOI: 10.1186/s13008-016-0020-7.

10.   Fekairi S, Scaglione S, Chahwan C, et al. Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* 2009; 138: 78-89. 2009/07/15. DOI: 10.1016/j.cell.2009.06.029.

11.   Oyake T, Itoh K, Motohashi H, et al. Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol Cell Biol* 1996; 16: 6083-6095. 1996/11/01. DOI: 10.1128/mcb.16.11.6083.

12.   Perez-Torrado R, Yamada D and Defossez PA. Born to bind: the BTB protein-protein interaction domain. *Bioessays* 2006; 28: 1194-1202. DOI: 10.1002/bies.20500.

13.   Bonchuk A, Denisov S, Georgiev P, et al. Drosophila BTB/POZ Domains of "ttk Group" Can Form Multimers and Selectively Interact with Each Other. *Journal of Molecular Biology* 2011; 412: 423-436. DOI: 10.1016/j.jmb.2011.07.052.

14.   Bardwell VJ and Treisman R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* 1994; 8: 1664-1677.

15.   Siggs OM and Beutler B. The BTB-ZF transcription factors. *Cell Cycle* 2012; 11: 3358-3369. DOI: 10.4161/cc.21277.

16.   Ahmad KF, Engel CK and Prive GG. Crystal structure of the BTB domain from PLZF. *Proc Natl Acad Sci U S A* 1998; 95: 12123-12128.

17.   Li X, Peng H, Schultz DC, et al. Structure-function studies of the BTB/POZ transcriptional repression domain from the promyelocytic leukemia zinc finger oncoprotein. *Cancer Res* 1999; 59: 5275-5282. 1999/10/28.

18.   Ahmad KF, Melnick A, Lax S, et al. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell* 2003; 12: 1551-1564.

19.   Schubot FD, Tropea JE and Waugh DS. Structure of the POZ domain of human LRF, a master regulator of oncogenesis. *Biochem Biophys Res Commun* 2006; 351: 1-6. 2006/10/21. DOI: 10.1016/j.bbrc.2006.09.167.

20.   Stogios PJ, Chen L and Prive GG. Crystal structure of the BTB domain from the LRF/ZBTB7 transcriptional regulator. *Protein Sci* 2007; 16: 336-342. DOI: 10.1110/ps.062660907.

21.   Ghetu AF, Corcoran CM, Cerchietti L, et al. Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer. *Mol Cell* 2008; 29: 384-391. DOI: 10.1016/j.molcel.2007.12.026.

22.   Stead MA, Rosbrook GO, Hadden JM, et al. Structure of the wild-type human BCL6 POZ domain. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2008; 64: 1101-1104. 2008/12/05. DOI: 10.1107/S1744309108036063.

23.   Cerchietti LC, Ghetu AF, Zhu X, et al. A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* 2010; 17: 400-411. 2010/04/14. DOI: 10.1016/j.ccr.2009.12.050.

24.   Stogios PJ, Cuesta-Seijo JA, Chen L, et al. Insights into strand exchange in BTB domain dimers from the crystal structures of FAZF and Miz1. *J Mol Biol* 2010; 400: 983-997. DOI: 10.1016/j.jmb.2010.05.028.

25.   Sakamoto K, Sogabe S, Kamada Y, et al. Discovery of high-affinity BCL6-binding peptide and its structure-activity relationship. *Biochem Biophys Res Commun* 2017; 482: 310-316. 2016/11/20. DOI: 10.1016/j.bbrc.2016.11.060.

26.   McCoull W, Abrams RD, Anderson E, et al. Discovery of Pyrazolo[1,5-a]pyrimidine B-Cell Lymphoma 6 (BCL6) Binders and Optimization to High Affinity Macrocyclic Inhibitors. *J Med Chem* 2017; 60: 4386-4402. 2017/05/10. DOI: 10.1021/acs.jmedchem.7b00359.

27.   Kamada Y, Sakai N, Sogabe S, et al. Discovery of a B-Cell Lymphoma 6 Protein-Protein Interaction Inhibitor by a Biophysics-Driven Fragment-Based Approach. *J Med Chem* 2017; 60: 4358-4368. 2017/05/05. DOI: 10.1021/acs.jmedchem.7b00313.

28.   Yasui T, Yamamoto T, Sakai N, et al. Discovery of a novel B-cell lymphoma 6 (BCL6)-corepressor interaction inhibitor by utilizing structure-based drug design. *Bioorg Med Chem* 2017; 25: 4876-4886. 2017/08/02. DOI: 10.1016/j.bmc.2017.07.037.

29.   Kerres N, Steurer S, Schlager S, et al. Chemically Induced Degradation of the Oncogenic Transcription Factor BCL6. *Cell Rep* 2017; 20: 2860-2875. 2017/09/21. DOI: 10.1016/j.celrep.2017.08.081.

30.   McCoull W, Cheung T, Anderson E, et al. Development of a Novel B-Cell Lymphoma 6 (BCL6) PROTAC To Provide Insight into Small Molecule Targeting of BCL6. *Acs Chemical Biology* 2018; 13: 3131-3141. DOI: 10.1021/acschembio.8b00698.

31.   Cheng HM, Linhares BM, Yu WB, et al. Identification of Thiourea-Based Inhibitors of the B-Cell Lymphoma 6 BTB Domain via NMR-Based Fragment Screening and Computer-Aided Drug Design. *Journal of Medicinal Chemistry* 2018; 61: 7573-7588. DOI: 10.1021/acs.jmedchem.8b00040.

32.   Stead MA, Trinh CH, Garnett JA, et al. A beta-sheet interaction interface directs the tetramerisation of the Miz-1 POZ domain. *J Mol Biol* 2007; 373: 820-826. DOI: 10.1016/j.jmb.2007.08.026.

33.		Kobayashi A, Yamagiwa H, Hoshino H, et al. A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol Cell Biol* 2000; 20: 1733-1746.

34.		Lee SU and Maeda T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol Rev* 2012; 247: 107-119. DOI: 10.1111/j.1600-065X.2012.01116.x.

35.		Beaulieu AM and Sant'Angelo DB. The BTB-ZF family of transcription factors: key regulators of lineage commitment and effector function development in the immune system. *J Immunol* 2011; 187: 2841-2847. DOI: 10.4049/jimmunol.1004006.

36.		Chevrier S and Corcoran LM. BTB-ZF transcription factors, a growing family of regulators of early and late B-cell development. *Immunol Cell Biol* 2014; 92: 481-488. DOI: 10.1038/icb.2014.20.

37.		Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015; 347: 1260419. 2015/01/24. DOI: 10.1126/science.1260419.

38.		Bilic I and Ellmeier W. The role of BTB domain-containing zinc finger proteins in T cell development and function. *Immunol Lett* 2007; 108: 1-9. DOI: 10.1016/j.imlet.2006.09.007.

39.		Sakaguchi S, Hombauer M, Bilic I, et al. The zinc-finger protein MAZR is part of the transcription factor network that controls the CD4 versus CD8 lineage fate of double-positive thymocytes. *Nat Immunol* 2010; 11: 442-448. DOI: 10.1038/ni.1860.

40.		Sakaguchi S, Hainberger D, Tizian C, et al. MAZR and Runx Factors Synergistically Repress ThPOK during CD8+ T Cell Lineage Development. *J Immunol* 2015; 195: 2879-2887. DOI: 10.4049/jimmunol.1500387.

41.		He X, Park K and Kappes DJ. The role of ThPOK in control of CD4/CD8 lineage commitment. *Annu Rev Immunol* 2010; 28: 295-320. DOI: 10.1146/annurev.immunol.25.022106.141715.

42.		Andersen L, Gulich AF, Alteneder M, et al. The Transcription Factor MAZR/PATZ1 Regulates the Development of FOXP3(+) Regulatory T Cells. *Cell Rep* 2019; 29: 4447-4459 e4446. 2019/12/26. DOI: 10.1016/j.celrep.2019.11.089.

43.		Morii E, Oboki K, Kataoka TR, et al. Interaction and cooperation of mi transcription factor (MITF) and myc-associated zinc-finger protein-related factor (MAZR) for transcription of mouse mast cell protease 6 gene. *J Biol Chem* 2002; 277: 8566-8571. DOI: 10.1074/jbc.M110392200.

44.		Pejler G, Ronnberg E, Waern I, et al. Mast cell proteases: multifaceted regulators of inflammatory disease. *Blood* 2010; 115: 4981-4990. 2010/03/18. DOI: 10.1182/blood-2010-01-257287.

45.		Ow JR, Ma H, Jean A, et al. Patz1 regulates embryonic stem cell identity. *Stem Cells Dev* 2014; 23: 1062-1073. 2014/01/02. DOI: 10.1089/scd.2013.0430.

46.		Ma H, Ow JR, Tan BC, et al. The dosage of Patz1 modulates reprogramming process. *Sci Rep* 2014; 4: 7519. 2014/12/18. DOI: 10.1038/srep07519.

47.		Valentino T, Palmieri D, Vitiello M, et al. PATZ1 interacts with p53 and regulates expression of p53-target genes enhancing apoptosis or cell survival based on the cellular context. *Cell Death Dis* 2013; 4: e963. DOI: 10.1038/cddis.2013.500.

48.		Keskin N, Deniz E, Eryilmaz J, et al. PATZ1 Is a DNA Damage-Responsive Transcription Factor That Inhibits p53 Function. *Mol Cell Biol* 2015; 35: 1741-1753. DOI: 10.1128/MCB.01475-14.

49.		Chiappetta G, Valentino T, Vitiello M, et al. PATZ1 acts as a tumor suppressor in thyroid cancer via targeting p53-dependent genes involved in EMT and cell migration. *Oncotarget* 2015; 6: 5310-5323. DOI: 10.18632/oncotarget.2776.

33.		Kobayashi A, Yamagiwa H, Hoshino H, et al. A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol Cell Biol* 2000; 20: 1733-1746.

34.		Lee SU and Maeda T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol Rev* 2012; 247: 107-119. DOI: 10.1111/j.1600-065X.2012.01116.x.

35.		Beaulieu AM and Sant'Angelo DB. The BTB-ZF family of transcription factors: key regulators of lineage commitment and effector function development in the immune system. *J Immunol* 2011; 187: 2841-2847. DOI: 10.4049/jimmunol.1004006.

36.		Chevrier S and Corcoran LM. BTB-ZF transcription factors, a growing family of regulators of early and late B-cell development. *Immunol Cell Biol* 2014; 92: 481-488. DOI: 10.1038/icb.2014.20.

37.		Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015; 347: 1260419. 2015/01/24. DOI: 10.1126/science.1260419.

38.		Bilic I and Ellmeier W. The role of BTB domain-containing zinc finger proteins in T cell development and function. *Immunol Lett* 2007; 108: 1-9. DOI: 10.1016/j.imlet.2006.09.007.

39.		Sakaguchi S, Hombauer M, Bilic I, et al. The zinc-finger protein MAZR is part of the transcription factor network that controls the CD4 versus CD8 lineage fate of double-positive thymocytes. *Nat Immunol* 2010; 11: 442-448. DOI: 10.1038/ni.1860.

40.		Sakaguchi S, Hainberger D, Tizian C, et al. MAZR and Runx Factors Synergistically Repress ThPOK during CD8+ T Cell Lineage Development. *J Immunol* 2015; 195: 2879-2887. DOI: 10.4049/jimmunol.1500387.

41.		He X, Park K and Kappes DJ. The role of ThPOK in control of CD4/CD8 lineage commitment. *Annu Rev Immunol* 2010; 28: 295-320. DOI: 10.1146/annurev.immunol.25.022106.141715.

42.		Andersen L, Gulich AF, Alteneder M, et al. The Transcription Factor MAZR/PATZ1 Regulates the Development of FOXP3(+) Regulatory T Cells. *Cell Rep* 2019; 29: 4447-4459 e4446. 2019/12/26. DOI: 10.1016/j.celrep.2019.11.089.

43.		Morii E, Oboki K, Kataoka TR, et al. Interaction and cooperation of mi transcription factor (MITF) and myc-associated zinc-finger protein-related factor (MAZR) for transcription of mouse mast cell protease 6 gene. *J Biol Chem* 2002; 277: 8566-8571. DOI: 10.1074/jbc.M110392200.

44.		Pejler G, Ronnberg E, Waern I, et al. Mast cell proteases: multifaceted regulators of inflammatory disease. *Blood* 2010; 115: 4981-4990. 2010/03/18. DOI: 10.1182/blood-2010-01-257287.

45.		Ow JR, Ma H, Jean A, et al. Patz1 regulates embryonic stem cell identity. *Stem Cells Dev* 2014; 23: 1062-1073. 2014/01/02. DOI: 10.1089/scd.2013.0430.

46.		Ma H, Ow JR, Tan BC, et al. The dosage of Patz1 modulates reprogramming process. *Sci Rep* 2014; 4: 7519. 2014/12/18. DOI: 10.1038/srep07519.

47.		Valentino T, Palmieri D, Vitiello M, et al. PATZ1 interacts with p53 and regulates expression of p53-target genes enhancing apoptosis or cell survival based on the cellular context. *Cell Death Dis* 2013; 4: e963. DOI: 10.1038/cddis.2013.500.

48.		Keskin N, Deniz E, Eryilmaz J, et al. PATZ1 Is a DNA Damage-Responsive Transcription Factor That Inhibits p53 Function. *Mol Cell Biol* 2015; 35: 1741-1753. DOI: 10.1128/MCB.01475-14.

49.		Chiappetta G, Valentino T, Vitiello M, et al. PATZ1 acts as a tumor suppressor in thyroid cancer via targeting p53-dependent genes involved in EMT and cell migration. *Oncotarget* 2015; 6: 5310-5323. DOI: 10.18632/oncotarget.2776.

50. Guadagno E, Vitiello M, Francesca P, et al. PATZ1 is a new prognostic marker of glioblastoma associated with the stem-like phenotype and enriched in the proneural subtype. *Oncotarget* 2017; 8: 59282-59300. 2017/09/25. DOI: 10.18632/oncotarget.19546.

51. Passariello A, Errico ME, Donofrio V, et al. PATZ1 Is Overexpressed in Pediatric Glial Tumors and Correlates with Worse Event-Free Survival in High-grade Gliomas. *Cancers (Basel)* 2019; 11 2019/10/17. DOI: 10.3390/cancers11101537.

52. Roy S, Dutta S, Khanna K, et al. Prediction of DNA-binding specificity in zinc finger proteins. *Journal of Biosciences* 2012; 37: 483-491. DOI: 10.1007/s12038-012-9213-7.

53. Huynh KD and Bardwell VJ. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* 1998; 17: 2473-2484. DOI: 10.1038/sj.onc.1202197.

54. Wong CW and Privalsky ML. Components of the SMRT corepressor complex exhibit distinctive interactions with the POZ domain oncoproteins PLZF, PLZF-RARalpha, and BCL-6. *J Biol Chem* 1998; 273: 27695-27702.

55. Huynh KD, Fischle W, Verdin E, et al. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev* 2000; 14: 1810-1823.

56. Melnick A, Ahmad KF, Arai S, et al. In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Molecular and Cellular Biology* 2000; 20: 6550-6567. DOI: Doi 10.1128/Mcb.20.17.6550-6567.2000.

57. Melnick A, Carlile G, Ahmad KF, et al. Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors. *Mol Cell Biol* 2002; 22: 1804-1818.

58. Granadino-Roldan JM, Obiol-Pardo C, Pinto M, et al. Molecular dynamics analysis of the interaction between the human BCL6 BTB domain and its SMRT, NcoR and BCOR corepressors: the quest for a consensus dynamic pharmacophore. *J Mol Graph Model* 2014; 50: 142-151. DOI: 10.1016/j.jmgm.2014.04.003.

59. Li J, Wang J, Wang J, et al. Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. *EMBO J* 2000; 19: 4342-4350. DOI: 10.1093/emboj/19.16.4342.

60. Bilic I, Koesters C, Unger B, et al. Negative regulation of CD8 expression via Cd8 enhancer-mediated recruitment of the zinc finger protein MAZR. *Nat Immunol* 2006; 7: 392-400. DOI: 10.1038/ni1311.

61. Schmitges FW, Radovani E, Najafabadi HS, et al. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res* 2016; 26: 1742-1752. 2016/11/18. DOI: 10.1101/gr.209643.116.

62. Persikov AV, Wetzel JL, Rowland EF, et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res* 2015; 43: 1965-1984. 2015/01/17. DOI: 10.1093/nar/gku1395.

63. Persikov AV, Rowland EF, Oakes BL, et al. Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. *Nucleic Acids Res* 2014; 42: 1497-1508. 2013/11/12. DOI: 10.1093/nar/gkt1034.

64. Flotho A and Melchior F. Sumoylation: a regulatory protein modification in health and disease. *Annu Rev Biochem* 2013; 82: 357-385. 2013/06/12. DOI: 10.1146/annurev-biochem-061909-093311.

65. Liang YC, Lee CC, Yao YL, et al. SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci Rep* 2016; 6: 26509. DOI: 10.1038/srep26509.

66. Su HL and Li SSL. Molecular features of human ubiquitin-like SUMO genes and their encoded proteins. *Gene* 2002; 296: 65-73. DOI: Pii S0378-1119(02)00843-0 Doi 10.1016/S0378-1119(02)00843-0.

67. Fedele M, Benvenuto G, Pero R, et al. A novel member of the BTB/POZ family, PATZ, associates with the RNF4 RING finger protein and acts as a transcriptional repressor. *J Biol Chem* 2000; 275: 7894-7901.

68. Tatham MH, Geoffroy MC, Shen L, et al. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* 2008; 10: 538-546. 2008/04/15. DOI: 10.1038/ncb1716.

69. Dhordain P, Albagli O, Honore N, et al. Colocalization and heteromerization between the two human oncogene POZ/zinc finger proteins, LAZ3 (BCL6) and PLZF. *Oncogene* 2000; 19: 6240-6250. 2001/02/15. DOI: 10.1038/sj.onc.1203976.

70. Franco R, Scognamiglio G, Valentino E, et al. PATZ1 expression correlates positively with BAX and negatively with BCL6 and survival in human diffuse large B cell lymphomas. *Oncotarget* 2016; 7: 59158-59172. DOI: 10.18632/oncotarget.10993.

71. Siegfried A, Rousseau A, Maurage C-A, et al. EWSR1-PATZ1 gene fusion may define a new glioneuronal tumor entity. *Brain Pathology* 2019; 29: 53-62. DOI: 10.1111/bpa.12619.

72. Bridge JA, Sumegi J, Druta M, et al. Clinical, pathological, and genomic features of EWSR1-PATZ1 fusion sarcoma. *Mod Pathol* 2019; 32: 1593-1604. 2019/06/14. DOI: 10.1038/s41379-019-0301-1.

73. Chougule A, Taylor MS, Nardi V, et al. Spindle and Round Cell Sarcoma With EWSR1-PATZ1 Gene Fusion: A Sarcoma With Polyphenotypic Differentiation. *Am J Surg Pathol* 2019; 43: 220-228. 2018/11/01. DOI: 10.1097/PAS.0000000000001183.

74. Sankar S and Lessnick SL. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genet* 2011; 204: 351-365. 2011/08/30. DOI: 10.1016/j.cancergen.2011.07.008.

75. Mastrangelo T, Modena P, Tornielli S, et al. A novel zinc finger gene is fused to EWS in small round cell tumor. *Oncogene* 2000; 19: 3799-3804. 2000/08/19. DOI: 10.1038/sj.onc.1203762.

76. Im YH, Kim HT, Lee C, et al. EWS-FLI1, EWS-ERG, and EWS-ETV1 oncoproteins of Ewing tumor family all suppress transcription of transforming growth factor beta type II receptor gene. *Cancer Res* 2000; 60: 1536-1540. 2000/04/05.

77. Burel-Vandenbos F, Pierron G, Thomas C, et al. A polyphenotypic malignant paediatric brain tumour presenting a MN1-PATZ1 fusion, no epigenetic similarities with CNS High-Grade Neuroepithelial Tumour with MN1 Alteration (CNS HGNET-MN1) and related to PATZ1-fused sarcomas. *Neuropathol Appl Neurobiol* 2020 2020/05/13. DOI: 10.1111/nan.12626.

78. Fukuchi S, Homma K, Minezaki Y, et al. Intrinsically disordered loops inserted into the structural domains of human proteins. *Journal of Molecular Biology* 2006; 355: 845-857. DOI: 10.1016/j.jmb.2005.10.037.

79. Dyson HJ and Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005; 6: 197-208. 2005/03/02. DOI: 10.1038/nrm1589.

80. Koike R, Amano M, Kaibuchi K, et al. Protein kinases phosphorylate long disordered regions in intrinsically disordered proteins. *Protein Sci* 2020; 29: 564-571. 2019/11/15. DOI: 10.1002/pro.3789.

81. Nei M, Xu P and Glazko G. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci U S A* 2001; 98: 2497-2502. 2001/02/28. DOI: 10.1073/pnas.051611498.

82.   Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420: 520-562. 2002/12/06. DOI: 10.1038/nature01262.

83.   Mestas J and Hughes CC. Of mice and not men: differences between mouse and human immunology. *J Immunol* 2004; 172: 2731-2738. 2004/02/24. DOI: 10.4049/jimmunol.172.5.2731.

84.   Rangarajan A and Weinberg RA. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer* 2003; 3: 952-959. 2004/01/23. DOI: 10.1038/nrc1235.

85.   Shay T, Jojic V, Zuk O, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci U S A* 2013; 110: 2946-2951. 2013/02/06. DOI: 10.1073/pnas.1222738110.

86.   Stead MA and Wright SC. Structures of heterodimeric POZ domains of Miz1/BCL6 and Miz1/NAC1. *Acta Crystallogr F Struct Biol Commun* 2014; 70: 1591-1596. DOI: 10.1107/S2053230X14023449.

87.   Phan RT, Saito M, Basso K, et al. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nat Immunol* 2005; 6: 1054-1060. 2005/09/06. DOI: 10.1038/ni1245.

88.   Weber A, Marquardt J, Elzi D, et al. Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. *EMBO J* 2008; 27: 1563-1574. 2008/05/03. DOI: 10.1038/emboj.2008.85.

89.   Lee KM, Choi WI, Koh DI, et al. The proto-oncoprotein KR-POK represses transcriptional activation of CDKN1A by MIZ-1 through competitive binding. *Oncogene* 2012; 31: 1442-1458. 2011/08/02. DOI: 10.1038/onc.2011.331.

90.   Jeon BN, Kim MK, Yoon JH, et al. Two ZNF509 (ZBTB49) isoforms induce cell-cycle arrest by activating transcription of p21/CDKN1A and RB upon exposure to genotoxic stress. *Nucleic Acids Res* 2014; 42: 11447-11461. 2014/09/24. DOI: 10.1093/nar/gku857.

91.   Huttlin EL, Ting L, Bruckner RJ, et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 2015; 162: 425-440. 2015/07/18. DOI: 10.1016/j.cell.2015.06.043.

92.   Kramer G, Shiber A and Bukau B. Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. *Annu Rev Biochem* 2019; 88: 337-364. 2018/12/07. DOI: 10.1146/annurev-biochem-013118-111717.

93.   Herhaus L and Dikic I. Dimerization quality control via ubiquitylation A specialized ubiquitin ligase enzyme approves only functional dimers. *Science* 2018; 362: 151-152. DOI: 10.1126/science.aav1391.

94.   Mena EL, Kjolby RAS, Saxton RA, et al. Dimerization quality control ensures neuronal development and survival. *Science* 2018; 362: 198-+. DOI: ARTN eaap8236 10.1126/science.aap8236.

95.   Koh DI, Choi WI, Jeon BN, et al. A novel POK family transcription factor, ZBTB5, represses transcription of p21CIP1 gene. *J Biol Chem* 2009; 284: 19856-19866. 2009/06/06. DOI: 10.1074/jbc.M109.025817.

96.   Morrison DJ, Pendergrast PS, Stavropoulos P, et al. FBI-1, a factor that binds to the HIV-1 inducer of short transcripts (IST), is a POZ domain protein. *Nucleic Acids Res* 1999; 27: 1251-1262. 1999/02/12. DOI: 10.1093/nar/27.5.1251.

97.   Davies JM, Hawe N, Kabarowski J, et al. Novel BTB/POZ domain zinc-finger protein, LRF, is a potential target of the LAZ-3/BCL-6 oncogene. *Oncogene* 1999; 18: 365-375. 1999/02/02. DOI: 10.1038/sj.onc.1202332.

98. Yoon JH, Choi WI, Jeon BN, et al. Human Kruppel-related 3 (HKR3) is a novel transcription activator of alternate reading frame (ARF) gene. *J Biol Chem* 2014; 289: 4018-4031. 2014/01/03. DOI: 10.1074/jbc.M113.526855.

99. Widom RL, Lee JY, Joseph C, et al. The hcKrox gene family regulates multiple extracellular matrix genes. *Matrix Biol* 2001; 20: 451-462. 2001/11/03. DOI: 10.1016/s0945-053x(01)00167-6.

100. Koken MH, Reid A, Quignon F, et al. Leukemia-associated retinoic acid receptor alpha fusion partners, PML and PLZF, heterodimerize and colocalize to nuclear bodies. *Proc Natl Acad Sci U S A* 1997; 94: 10255-10260.

101. Hoatlin ME, Zhi Y, Ball H, et al. A novel BTB/POZ transcriptional repressor protein interacts with the Fanconi anemia group C protein and PLZF. *Blood* 1999; 94: 3737-3747. 1999/11/26.

102. Wang J, Kudoh J, Takayanagi A, et al. Novel human BTB/POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161. *Biochem Biophys Res Commun* 2005; 327: 615-627. 2005/01/05. DOI: 10.1016/j.bbrc.2004.12.048.

103. Takenaga M, Hatano M, Takamori M, et al. Bcl6-dependent transcriptional repression by BAZF. *Biochem Biophys Res Commun* 2003; 303: 600-608. 2003/03/28. DOI: 10.1016/s0006-291x(03)00396-6.

104. Deltour S, Guerardel C and Leprince D. Recruitment of SMRT/N-CoR-mSin3A-HDAC-repressing complexes is not a general mechanism for BTB/POZ transcriptional repressors: the case of HIC-1 and gammaFBP-B. *Proc Natl Acad Sci U S A* 1999; 96: 14831-14836.

105. Deltour S, Pinte S, Guerardel C, et al. Characterization of HRG22, a human homologue of the putative tumor suppressor gene HIC1. *Biochem Biophys Res Commun* 2001; 287: 427-434. 2001/09/14. DOI: 10.1006/bbrc.2001.5624.

106. Donaldson NS, Nordgaard CL, Pierre CC, et al. Kaiso regulates Znf131-mediated transcriptional activation. *Exp Cell Res* 2010; 316: 1692-1705. 2010/03/23. DOI: 10.1016/j.yexcr.2010.03.011.

107. Humphrey W, Dalke A and Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996; 14: 33-38, 27-38.

108. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem* 2005; 26: 1781-1802. DOI: 10.1002/jcc.20289.

109. Hatherley R, Brown DK, Glenister M, et al. PRIMO: An Interactive Homology Modeling Pipeline. *PLoS One* 2016; 11: e0166698. 2016/11/18. DOI: 10.1371/journal.pone.0166698.

110. Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005; 33: W363-367. 2005/06/28. DOI: 10.1093/nar/gki481.

111. Pierce B, Tong WW and Weng ZP. M-ZDOCK: a grid-based approach for C-n symmetric multimer docking. *Bioinformatics* 2005; 21: 1472-1478. DOI: 10.1093/bioinformatics/bti229.

112. Sali A and Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993; 234: 779-815. 1993/12/05. DOI: 10.1006/jmbi.1993.1626.

113. Shatsky M, Nussinov R and Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins* 2004; 56: 143-156. 2004/05/27. DOI: 10.1002/prot.10628.

114.   Pei J, Kim BH and Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 2008; 36: 2295-2300. 2008/02/22. DOI: 10.1093/nar/gkn072.

115.   Soding J, Biegert A and Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005; 33: W244-248. 2005/06/28. DOI: 10.1093/nar/gki408.

116.   Bateman A, Martin MJ, Orchard S, et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 2019; 47: D506-D515. DOI: 10.1093/nar/gky1049.

117.   Baspinar A, Cukuroglu E, Nussinov R, et al. PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res* 2014; 42: W285-289. 2014/05/16. DOI: 10.1093/nar/gku397.

118.   Tina KG, Bhadra R and Srinivasan N. PIC: Protein Interactions Calculator. *Nucleic Acids Res* 2007; 35: W473-476. 2007/06/23. DOI: 10.1093/nar/gkm423.

119.   Zimmermann L, Stephens A, Nam SZ, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* 2018; 430: 2237-2243. 2017/12/21. DOI: 10.1016/j.jmb.2017.12.007.

120.   Woldeyes RA, Sivak DA and Fraser JS. E pluribus unum, no more: from one crystal, many conformations. *Curr Opin Struct Biol* 2014; 28: 56-62. 2014/08/13. DOI: 10.1016/j.sbi.2014.07.005.

121.   Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., et al. CHARMM: the biomolecular simulation program. *J Comput Chem* 2009; 30: 1545-1614. 2009/05/16. DOI: 10.1002/jcc.21287.

122.   Darden T, Perera L, Li LP, et al. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure with Folding & Design* 1999; 7: R55-R60. DOI: Doi 10.1016/S0969-2126(99)80033-1.

123.   Zhao Q, Xie Y, Zheng Y, et al. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res* 2014; 42: W325-330. DOI: 10.1093/nar/gku383.

124.   Ren J, Gao X, Jin C, et al. Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *PROTEOMICS* 2009; 9: 3409-3412. DOI: 10.1002/pmic.200800646.

125.   Hecker CM, Rabiller M, Haglund K, et al. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* 2006; 281: 16117-16127. DOI: 10.1074/jbc.M512757200.

126.   Kerscher O. SUMO junction-what's your function? New insights through SUMO-interacting motifs. *EMBO Rep* 2007; 8: 550-555. 2007/06/05. DOI: 10.1038/sj.embor.7400980.

127.   Winter G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 2010; 43: 186-190. DOI: 10.1107/S0021889809045701.

128.   Evans PR and Murshudov GN. How good are my data and what is the resolution? *Acta Crystallographica Section D-Biological Crystallography* 2013; 69: 1204-1214. DOI: 10.1107/S0907444913000061.

129.   Mccoy AJ, Grosse-Kunstleve RW, Adams PD, et al. Phaser crystallographic software. *Journal of Applied Crystallography* 2007; 40: 658-674. DOI: 10.1107/S0021889807021206.

130.   Emsley P, Lohkamp B, Scott WG, et al. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 2010; 66: 486-501. DOI: 10.1107/S0907444910007493.

131.   Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 2019; 75: 861-877. 2019/10/08. DOI: 10.1107/S2059798319011471.

132.   Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 2018; 27: 293-315. 2017/10/27. DOI: 10.1002/pro.3330.

133.   Bailey TL and Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994; 2: 28-36. 1994/01/01.

134.   Kabsch W. Xds. *Acta Crystallogr D Biol Crystallogr* 2010; 66: 125-132. 2010/02/04. DOI: 10.1107/S0907444909047337.

135.   Fiser A and Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 2003; 19: 2500-2501. 2003/12/12.

136.   Fiser A, Do RK and Sali A. Modeling of loops in protein structures. *Protein Sci* 2000; 9: 1753-1773. 2000/10/25. DOI: 10.1110/ps.9.9.1753.

137.   Krissinel E and Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007; 372: 774-797. 2007/08/08. DOI: 10.1016/j.jmb.2007.05.022.

138.   Janson J-C. Protein Purification: Principles, High Resolution Methods, and Applications. *Methods of Biochemica Analysis*. 3 ed.: Wiley, 2011, p. 532.

139.   Piepoli S, Alt AO, Atilgan C, et al. Structural analysis of the PATZ1 BTB domain homodimer. *Acta Crystallographica Section D* 2020; 76: 581-593. DOI: doi:10.1107/S2059798320005355.

140.   Yoon HG, Chan DW, Reynolds AB, et al. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell* 2003; 12: 723-734. 2003/10/07.

141.   Cifuentes-Diaz C, Bitoun M, Goudou D, et al. Neuromuscular expression of the BTB/POZ and zinc finger protein myoneurin. *Muscle Nerve* 2004; 29: 59-65. 2003/12/25. DOI: 10.1002/mus.10526.

142.   Li JS, Miralles Fuste J, Simavorian T, et al. TZAP: A telomere-associated protein involved in telomere length control. *Science* 2017; 355: 638-641. 2017/01/14. DOI: 10.1126/science.aah6752.

143.   Adhikary S, Peukert K, Karsunky H, et al. Miz1 is required for early embryonic development during gastrulation. *Mol Cell Biol* 2003; 23: 7648-7657. 2003/10/16. DOI: 10.1128/mcb.23.21.7648-7657.2003.

144.   Kosan C, Saba I, Godmann M, et al. Transcription factor miz-1 is required to regulate interleukin-7 receptor signaling at early commitment stages of B cell differentiation. *Immunity* 2010; 33: 917-928. 2010/12/21. DOI: 10.1016/j.immuni.2010.11.028.

145.   Maeda T, Merghoub T, Hobbs RM, et al. Regulation of B versus T lymphoid lineage fate decision by the proto-oncogene LRF. *Science* 2007; 316: 860-866. 2007/05/15. DOI: 10.1126/science.1140881.

146.   Dai MS, Chevallier N, Stone S, et al. The effects of the Fanconi anemia zinc finger (FAZF) on cell cycle, apoptosis, and proliferation are differentiation stage-specific. *J Biol Chem* 2002; 277: 26327-26334. 2002/05/03. DOI: 10.1074/jbc.M201834200.

147. Zolghadr K, Mortusewicz O, Rothbauer U, et al. A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. *Mol Cell Proteomics* 2008; 7: 2279-2287. 2008/07/16. DOI: 10.1074/mcp.M700548-MCP200.

148. de Greef JC, Wang J, Balog J, et al. Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am J Hum Genet* 2011; 88: 796-804. 2011/05/21. DOI: 10.1016/j.ajhg.2011.04.018.

149. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; 37: 1530-1534. 2020/02/06. DOI: 10.1093/molbev/msaa015.

150. Mena EL, Jevtic P, Greber BJ, et al. Structural basis for dimerization quality control. *Nature* 2020 2020/08/21. DOI: 10.1038/s41586-020-2636-7.

151. Immunological Genome P. ImmGen at 15. *Nat Immunol* 2020; 21: 700-703. 2020/06/25. DOI: 10.1038/s41590-020-0687-4.

152. Costoya JA. Functional analysis of the role of POK transcriptional repressors. *Brief Funct Genomic Proteomic* 2007; 6: 8-18. DOI: 10.1093/bfgp/elm002.

153. David G, Alland L, Hong SH, et al. Histone deacetylase associated with mSin3A mediates repression by the acute promyelocytic leukemia-associated PLZF protein. *Oncogene* 1998; 16: 2549-2556. 1998/06/17. DOI: 10.1038/sj.onc.1202043.

154. Dhordain P, Lin RJ, Quief S, et al. The LAZ3(BCL-6) oncoprotein recruits a SMRT/mSIN3A/histone deacetylase containing complex to mediate transcriptional repression. *Nucleic Acids Res* 1998; 26: 4645-4651. 1998/10/01. DOI: 10.1093/nar/26.20.4645.

155. Watson PJ, Fairall L and Schwabe JW. Nuclear hormone receptor co-repressors: structure and function. *Mol Cell Endocrinol* 2012; 348: 440-449. 2011/09/20. DOI: 10.1016/j.mce.2011.08.033.

156. Uversky VN, Gillespie JR and Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 2000; 41: 415-427. 2000/10/12. DOI: 10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7.

157. Xue B, Dunker AK and Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* 2012; 30: 137-149. 2012/06/19. DOI: 10.1080/07391102.2012.675145.

158. Deiana A, Forcelloni S, Porrello A, et al. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* 2019; 14: e0217889. 2019/08/20. DOI: 10.1371/journal.pone.0217889.

159. Fedele M, Franco R, Salvatore G, et al. PATZ1 gene has a critical role in the spermatogenesis and testicular tumours. *J Pathol* 2008; 215: 39-47. DOI: 10.1002/path.2323.

160. Monaco M, Palma G, Vitiello M, et al. Loss of One or Two PATZ1 Alleles Has a Critical Role in the Progression of Thyroid Carcinomas Induced by the RET/PTC1 Oncogene. *Cancers* 2018; 10. DOI: ARTN 92 10.3390/cancers10040092.

161. Fedele M, Crescenzi E and Cerchia L. The POZ/BTB and AT-Hook Containing Zinc Finger 1 (PATZ1) Transcription Regulator: Physiological Functions and Disease Involvement. *Int J Mol Sci* 2017; 18 2017/12/01. DOI: 10.3390/ijms18122524.

162. Vitiello M, Valentino T, De Menna M, et al. PATZ1 is a target of miR-29b that is induced by Ha-Ras oncogene in rat thyroid cells. *Sci Rep* 2016; 6: 25268. DOI: 10.1038/srep25268.

163.   Valentino T, Palmieri D, Vitiello M, et al. Embryonic defects and growth alteration in mice with homozygous disruption of the Patz1 gene. *J Cell Physiol* 2013; 228: 646-653. 2012/08/14. DOI: 10.1002/jcp.24174.

164.   Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57-74. 2012/09/08. DOI: 10.1038/nature11247.

165.   Suter DM. Transcription Factors and DNA Play Hide and Seek. *Trends Cell Biol* 2020; 30: 491-500. 2020/05/16. DOI: 10.1016/j.tcb.2020.03.003.

166.   Sicoli G, Vezin H, Ledolter K, et al. Conformational tuning of a DNA-bound transcription factor. *Nucleic Acids Res* 2019; 47: 5429-5435. 2019/04/26. DOI: 10.1093/nar/gkz291.

167.   Hilser VJ and Thompson EB. Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors. *J Biol Chem* 2011; 286: 39675-39682. 2011/09/23. DOI: 10.1074/jbc.R111.278929.

168.   Pero R, Lembo F, Palmieri EA, et al. PATZ attenuates the RNF4-mediated enhancement of androgen receptor-dependent transcription. *Journal of Biological Chemistry* 2002; 277: 3280-3285. DOI: 10.1074/jbc.M109491200.

169.   Schwanhausser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. *Nature* 2011; 473: 337-342. 2011/05/20. DOI: 10.1038/nature10098.

170.   Buccitelli C and Selbach M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 2020 2020/07/28. DOI: 10.1038/s41576-020-0258-4.

171.   Csardi G, Franks A, Choi DS, et al. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet* 2015; 11: e1005206. 2015/05/08. DOI: 10.1371/journal.pgen.1005206.