

Structural analysis of the PATZ1 BTB domain homodimer

Sofia Piepoli,^a Aaron Oliver Alt,^b Canan Atilgan,^{a,c} Erika Jazmin Mancini^{b*} and Batu Erman^{a*}

^aFaculty of Engineering and Natural Sciences, Sabanci University, Orta Mahalle, Üniversite Caddesi No. 27, Orhanlı, Tuzla, 34956 Istanbul, Turkey, ^bSchool of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom, and ^cSabancı University Nanotechnology Research and Application Center, SUNUM, 34956 Istanbul, Turkey. *Correspondence e-mail: erika.mancini@sussex.ac.uk, batu@sabanciuniv.edu

Received 24 January 2020

Accepted 16 April 2020

Edited by G. Cingolani, Thomas Jefferson University, USA

Keywords: BTB; POZ; PATZ1; transcription factors; co-repressors; dimerization interface; structure dynamics.

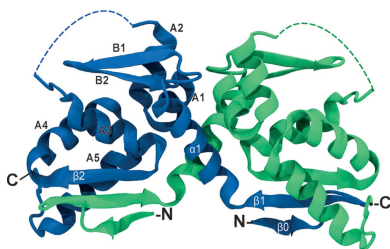
PDB references: mouse PATZ1 BTB domain, 6guv; zebrafish PATZ1 BTB domain, 6guw

Supporting information: this article has supporting information at journals.iucr.org/d

PATZ1 is a ubiquitously expressed transcriptional repressor belonging to the ZBTB family that is functionally expressed in T lymphocytes. PATZ1 targets the *CD8* gene in lymphocyte development and interacts with the p53 protein to control genes that are important in proliferation and in the DNA-damage response. PATZ1 exerts its activity through an N-terminal BTB domain that mediates dimerization and co-repressor interactions and a C-terminal zinc-finger motif-containing domain that mediates DNA binding. Here, the crystal structures of the murine and zebrafish PATZ1 BTB domains are reported at 2.3 and 1.8 Å resolution, respectively. The structures revealed that the PATZ1 BTB domain forms a stable homodimer with a lateral surface groove, as in other ZBTB structures. Analysis of the lateral groove revealed a large acidic patch in this region, which contrasts with the previously resolved basic co-repressor binding interface of BCL6. A large 30-amino-acid glycine- and alanine-rich central loop, which is unique to mammalian PATZ1 amongst all ZBTB proteins, could not be resolved, probably owing to its flexibility. Molecular-dynamics simulations suggest a contribution of this loop to modulation of the mammalian BTB dimerization interface.

1. Introduction

PATZ1 (POZ-, AT hook- and zinc-finger-containing protein 1), also known as ZBTB19, is a transcription factor that is present in all vertebrates (Fig. 1*a*). It was first discovered in a yeast two-hybrid (Y2H) experiment, in which it associated through its BTB (broad complex, tramtrack and bric-a-brac) domain with the BTB domain of the transcription factor BACH2 (BTB and CNC homology 2; Kobayashi *et al.*, 2000). PATZ1 is also referred to as MAZR (Myc-associated zinc finger-related) because of the close similarity between its zinc-finger (ZF) domain and that of MAZ (Myc-associated zinc finger). While its expression can be detected in many cell types and developmental stages, PATZ1/ZBTB19/MAZR is highly expressed specifically in the early stages of T-lymphocyte differentiation, where it negatively regulates *CD8* gene expression (Bilic & Ellmeier, 2007). 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 (Sakaguchi *et al.*, 2010, 2015; He *et al.*, 2010), and the expression of the FOXP3 transcription factor (Andersen *et al.*, 2019). The functions of PATZ1 are however not limited to lymphocytes as its expression is ubiquitous. An early embryonic role for PATZ1 has been suggested, as



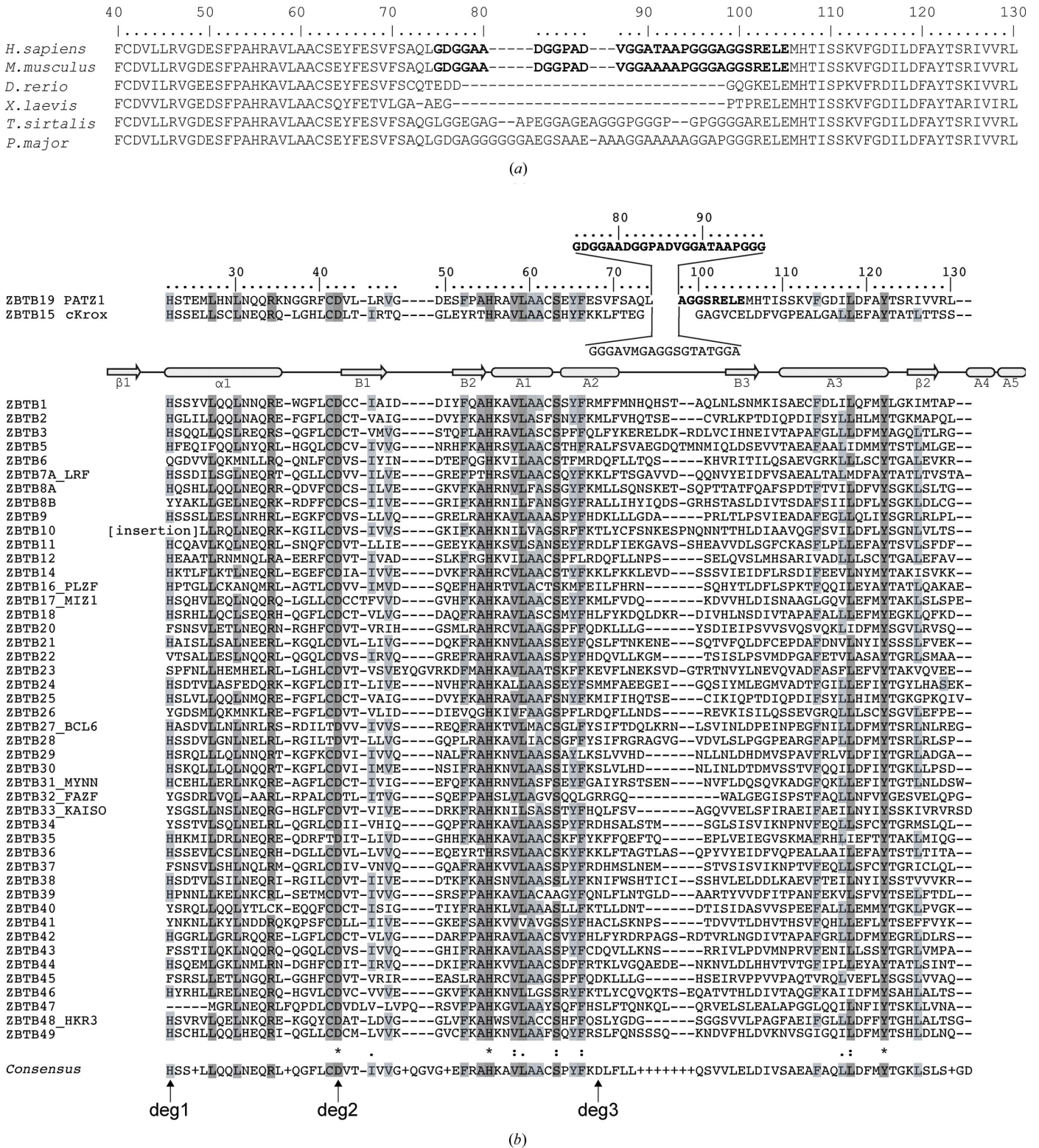


Figure 1 Sequence alignment of BTB domains of ZBTB transcription factors identifies a unique central region in PATZ1 that is conserved in mammals. (a) Sequence alignment of PATZ1 BTB domains from selected vertebrate species. The unique central sequence of the A2/B3 loop that is conserved in mammals and is missing in fish and amphibians is indicated in bold. (b) 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. The PATZ1 amino acids that correspond to this region without electron-density assignments from the crystal structure are shown in bold. 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 degenerate residues involved in BTB domain stability indicated by arrows.

PATZ1^{-/-} mice are embryonic lethal or born at non-Mendelian frequency and are small in size depending on the genetic background (Sakaguchi *et al.*, 2010). PATZ1 also negatively regulates induced pluripotent stem-cell (iPSC) generation (Ow *et al.*, 2014; Ma *et al.*, 2014). This function may be related to its interaction with the p53 tumour suppressor, as demonstrated by various studies (Valentino, Palmieri, Vitiello, Pierantoni *et al.*, 2013; Keskin *et al.*, 2015; Chiappetta *et al.*, 2015).

Structurally, PATZ1 belongs to the POZ (pox virus and zinc finger) or ZBTB (zinc finger and BTB) family of transcription factors (Lee & Maeda, 2012). Proteins belonging to this family have been implicated in many biological processes, including transcriptional regulation and development (Chevrier & Corcoran, 2014), whilst their dysfunction in vertebrates has been linked to tumorigenesis. ZBTB proteins bind to DNA through their ZF domains, and use their BTB domains for oligomerization (Bonchuk *et al.*, 2011) and the recruitment of co-repressors and chromatin-remodelling factors (Bardwell & Treisman, 1994; Siggs & Beutler, 2012). The human genome encodes 49 members of the ZBTB family (Fig. 1*b*), 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 and PATZ2/ZBTB24).

The BTB domain is a structural feature that mediates functional interactions between proteins (Perez-Torrado *et al.*, 2006). Most of the available BTB domain structures from the ZBTB family are homodimers formed by the assembly of two identical monomers (Ahmad *et al.*, 1998, 2003; Li *et al.*, 1999; Schubot *et al.*, 2006; Stogios *et al.*, 2007, 2010; Ghetu *et al.*, 2008; Stead *et al.*, 2008; Cerchietti *et al.*, 2010; Sakamoto *et al.*, 2017; McCoull *et al.*, 2017, 2018; Kamada *et al.*, 2017; Yasui *et al.*, 2017; Kerres *et al.*, 2017; Cheng *et al.*, 2018). In the case of Myc-interacting zinc-finger protein 1 (MIZ1), the structures of both a homodimer (Stogios *et al.*, 2010) and a homotetramer (Stead *et al.*, 2007) have been reported.

The overall fold of the BTB domain is highly conserved, containing the following secondary-structure elements: β 1- α 1-B1-B2-A1-A2-B3-A3- β 2-A4-A5 (Stogios *et al.*, 2005). A unique feature of the PATZ1 BTB domain is a long glycine- and alanine-rich central loop between A2 and B3 (Kobayashi *et al.*, 2000; Fig. 1). This central loop is conserved in all vertebrate PATZ1 proteins, but it is absent in those from fish and amphibians. The two BTB monomers are known to homodimerize through a specific dimerization interface that includes β 1- α 1, A1-A2 and β 2-A5. Although homodimerization seems to be favoured, heterodimeric interactions between pairs of ZBTB proteins have also been documented. 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 (Stead & Wright, 2014). 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 centre B cells (Phan *et al.*, 2005), ZBTB4 (Weber *et al.*, 2008),

ZBTB36 (Lee *et al.*, 2012), ZBTB49 (Jeon *et al.*, 2014) and NAC1 (nucleus accumbens-associated 1; Stead & Wright, 2014). PATZ1 can also form heterodimers with other BTB-domain-containing proteins such as PATZ2 (Huttlin *et al.*, 2015), BACH1 and BACH2 (Kobayashi *et al.*, 2000).

The mechanism controlling the 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 (Kramer *et al.*, 2019). Recently, a dimerization quality-control mechanism for BTB proteins has been proposed in which the stability of a homodimer would exceed that of heterodimers because of the structural masking of destabilizing residues (Herhaus & Dikic, 2018; Mena *et al.*, 2018). 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.

Homodimerization of BCL6 and promyelocytic leukaemia zinc-finger (PLZF) proteins creates a charged groove that binds nuclear receptor co-repressors such as NCOR1, NCOR2 (SMRT) and BCOR (Huynh & Bardwell, 1998; Wong & Privalsky, 1998; Huynh *et al.*, 2000; Melnick *et al.*, 2000, 2002). NCOR1 and SMRT are structurally disordered proteins that share 45% identity (Granadino-Roldán *et al.*, 2014) and contain a conserved 17-amino-acid BCL6-binding domain (BBD). These co-repressors are components of large complexes containing histone deacetylases (Li *et al.*, 2000) that contribute to transcriptional silencing. It is not known whether co-repressor binding is a generalizable feature of BTB homodimers, as MIZ1, FAZF and LRF BTB homodimers do not interact with these co-repressors (Stogios *et al.*, 2007, 2010). The PATZ1 BTB domain has been shown to interact with NCOR1; however, it is not known whether the interaction is mediated by an interface similar to that of BCL6 and PLZF (Bilic *et al.*, 2006).

Prior to 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) and HKR3/TZAP (ZBTB48). In order to obtain biological insights into the binding of PATZ1 to co-repressors, we determined the atomic structure of the PATZ1 (ZBTB19) BTB domain. To probe the role of the unique A2/B3 central loop, we obtained crystal structures of both the mouse PATZ1 BTB domain and its zebrafish orthologue.

2. Materials and methods

2.1. Protein expression and purification

The human isoform 3 (Q9HBE1-3) and mouse (Q9JMG9) PATZ1 BTB sequences are 98.9% identical (as determined using *LALIGN*; Gasteiger *et al.*, 2003), diverging at only one residue (T91A). The mouse PATZ1 BTB protein-coding

sequence (12–166) was PCR-amplified from a CMV-HA plasmid construct (Keskin *et al.*, 2015) containing the full-length mouse PATZ1 cDNA and cloned into a pET-47b bacterial expression plasmid (Novagen) between SmaI and NotI restriction sites in frame with an N-terminal 6×His tag and an HRV 3C protease cleavage site for fusion-tag removal. The resulting plasmid was transformed into the *Escherichia coli* Rosetta 2(DE3) strain and grown at 310 K by shaking at 180 rev min⁻¹ in Terrific/Turbo Broth (TB) medium supplemented with 50 µg ml⁻¹ kanamycin and 33 µg ml⁻¹ chloramphenicol until the absorbance at 600 nm reached a value of 0.6. Expression of the fusion protein was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and growth was continued for 16 h at 291 K. The zebrafish (*Danio rerio*) PATZ1 BTB protein-coding sequence (1–135) was PCR-amplified from zebrafish genomic DNA (a kind gift from Dr S. H. Fuss), cloned and expressed in the same bacterial expression plasmid as described above.

The cells were harvested by centrifugation, resuspended in 25 ml lysis buffer (50 mM HEPES pH 7, 250 mM NaCl, 10 mM imidazole, 0.5 mM TCEP, DNase and protease-inhibitor cocktail) and disrupted by sonication on ice. The lysate was clarified by centrifugation at 26 700g for 45 min at 277 K. The supernatant was applied onto a HisPur Cobalt Resin column (Thermo Fisher) previously equilibrated with wash buffer (50 mM HEPES, 250 mM NaCl, 0.5 mM TCEP, 10 mM imidazole). Following a 10 min incubation at 227 K and the application of wash buffer, the protein was then eluted by the addition of elution buffer (50 mM HEPES, 250 mM NaCl, 0.5 mM TCEP, 300 mM imidazole). The collected eluate was concentrated to 2.6 mg ml⁻¹ using a Sartorius Vivaspinn 20 column (10K molecular-weight cutoff) for additional purification by size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 prep-grade column (GE Healthcare) in gel-filtration buffer (20 mM HEPES, 250 mM NaCl, 0.5 mM TCEP) at 277 K. Fractions were analyzed on a 14% SDS-PAGE gel by electrophoresis and those containing PATZ1 BTB were pooled and concentrated to 9 mg ml⁻¹. The zebrafish PATZ1 BTB domain (69% sequence identity to mouse BTB) was expressed and purified as described above for the mouse BTB domain. The final concentration of the protein was 8.5 mg ml⁻¹.

2.2. Crystallization

All crystallization experiments were performed at 291 K using the sitting-drop vapour-diffusion method. The initial screening of 768 conditions was performed at the University of Sussex crystallization facility using a Crystal Phoenix dispensing robot to pipette 0.1 µl protein solution and 0.1 µl precipitant solution into single drops in 96-well plates. Crystals of the mouse PATZ1 BTB domain appeared after 72 h in 0.1 M MMT (DL-malic acid, MES monohydrate, Tris), 25% (w/v) PEG 1500. Crystals of the zebrafish PATZ1 BTB domain appeared after 72 h 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.

Table 1

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Mouse PATZ1 BTB	Zebrafish PATZ1 BTB
Data-collection statistics		
Wavelength (Å)	0.97625	0.9795
Space group	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 3 ₁ 21
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.23, 43.23, 162.95	43.08, 43.08, 123.64
α , β , γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	43.23–2.29 (2.37–2.29)	41.8–1.8 (1.9–1.8)
No. of unique reflections	7573 (724)	12786 (1120)
<i>R</i> _{merge}	0.05 (0.67)	0.01 (0.49)
<i>I</i> / σ (<i>I</i>)	9.91 (1.39)	19.96 (1.27)
Completeness (%)	99.58 (99.18)	98.41 (87.99)
Multiplicity	2.0 (2.0)	2.0 (1.8)
Refinement statistics		
Resolution (Å)	41.79–2.29	41.2–1.8
No. of reflections (observed/ <i>R</i> _{free})	7549/417	12785/571
<i>R</i> _{work} / <i>R</i> _{free}	0.209/0.245	0.215/0.224
No. of atoms		
Total	1123	993
Protein	1057	940
Water	66	53
Average <i>B</i> , all atoms (Å ²)	54.3	50
R.m.s. deviations		
Bond lengths (Å)	0.002	0.007
Bond angles (°)	0.433	0.965
PDB code	6guv	6guw

2.3. Data collection and processing

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* (Winter, 2010) and *AIMLESS* (Evans & Murshudov, 2013). The space group was *P*4₁2₁2 (unit-cell parameters *a* = *b* = 43.23, *c* = 162.95 Å, α = β = γ = 90°), with one molecule in the asymmetric unit.

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* (Kabsch, 2010). The space group was *P*3₁21 (unit-cell parameters *a* = *b* = 43.08, *c* = 123.64 Å, α = β = γ = 90°), with one molecule in the asymmetric unit.

Detailed X-ray data-collection and refinement statistics are given in Table 1.

2.4. Structure solution and refinement

The structure of mouse PATZ1 BTB was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the PLZF BTB structure (PDB entry 1buo; Ahmad *et al.*, 1998) as a search template. The identified solution was then subjected to rounds of manual rebuilding with *Coot* (Emsley *et al.*, 2010) and refinement with *Phenix* (Liebschner *et al.*, 2019) to give final *R*_{work} and *R*_{free} factors of 20.9% and 24.5%, respectively. The final structure was validated with *MolProbity* (Williams *et al.*, 2018) and deposited together with the structure factors in the Protein Data Bank as entry 6guv.

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.

Detailed X-ray data-refinement statistics are given in Table 1.

2.5. Sequence and structure analysis

The sequences of PATZ1 proteins from different organisms were retrieved from the NCBI Reference Sequence Database (RefSeq; Pruitt *et al.*, 2007). Except for *Homo sapiens* (NP_114440.1) and *Mus musculus* (NP_001240620.1) for mammals, one organism only was chosen for each group of different species of vertebrates: *Danio rerio* (XP_009300883.1) for fish, *Xenopus laevis* (XP_018117120.1) for amphibians, *Thamnophis sirtalis* (XP_013922905.1) for reptiles and *Parus major* (XP_015499085.1) for birds. Sequence limits were determined based on the annotations of the BTB domain in the UniProt database (The UniProt Consortium, 2019). The 49 members of the human ZBTB protein family were retrieved from Swiss-Prot and a multiple sequence alignment was obtained using the *PROMALS3D* online tool (Pei *et al.*, 2008). This alignment incorporates the structural information from the available PDB structures of these proteins. The secondary-structure nomenclature refers to that of Stogios *et al.* (2005) (Figs. 1*b*, 2*a*, 2*c* and 2*f*). ZBTB4 was excluded from this alignment because of its N-terminal serine-rich repetitive insertions, which are beyond the scope of this description. Shading was added according to the percentage of similarity for every position in the alignment visualized in *Jalview 2* (Waterhouse *et al.*, 2009) that generated a consensus sequence. A list of residues involved in homodimer interaction interfaces was obtained by *PDBePISA* (Krissinel & Henrick, 2007) and was graphically rendered on the protein structure with *VMD* (Humphrey *et al.*, 1996). Intra-chain and inter-chain interactions were retrieved by *PIC* (*Protein Interactions Calculator*; Tina *et al.*, 2007) and the *VMD Salt Bridges* and *Timeline* plugins. Structural alignments were calculated using the *MultiProt* server (Shatsky *et al.*, 2004).

2.6. Modelling

The central loop of the mammalian PATZ1 BTB structure (residues 83–113) was modelled as a monomer using the *PRIMO* suite (Hatherley *et al.*, 2016) based on the *MODELLER* program (Šali & Blundell, 1993). Structural information from *MEME* motif analysis of the loop sequence was added to the template (Bailey & Elkan, 1994). The obtained model was aligned with the deposited structure (PDB entry 6guv) 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* (Schneidman-Duhovny *et al.*, 2005) was used to reconstruct the dimer conformation.

The stability of the new structure was tested by molecular-dynamics (MD) simulations in *NAMD* (Phillips *et al.*, 2005). The protein structure model was centred in a solvent box built according to the protein size and padded with at least a 10 Å layer of water in every direction. The solvent was modelled explicitly using TIP3W water molecules. 0.15 M KCl was added to ionize the solvent. The MD simulation was performed using the CHARMM27 force field (Brooks *et al.*, 2009) in *NAMD*. Periodic boundary conditions were applied in which long-range electrostatic interactions were treated using the particle mesh Ewald method (Darden *et al.*, 1999) and the cutoff distance was set to 12 Å. All simulations were run as duplicates at a constant temperature of 310 K for at least 200 ns.

In one of the MD simulations of human PATZ1 BTB, the flexible loop region in one of the monomers formed an extra β -strand structure. The model was further refined by once more duplicating this monomer into the dimeric form. This refinement was processed by taking a frame from the run in which the root-mean-square deviation (r.m.s.d.) with the initial structure, excluding the loop, was minimal (3.04 Å) and by mirroring the information from the monomer with the formed secondary structures to the other monomer with *M-ZDOCK* (Pierce *et al.*, 2005), recreating the dimer by symmetry. The stability of the new structure was confirmed by additional MD simulations (twice, 200 ns each). *ModLoop* (Fiser & Sali, 2003) was used to model the coordinates of the missing residues (70–76) in the zebrafish PATZ1 BTB structure (PDB entry 6guw) and the stability of the modelled structure was assessed by MD as before.

3. Results and discussion

3.1. Structural features of the murine and zebrafish PATZ1 BTB domains

Here, we report the crystal structures of the BTB domains of murine and zebrafish PATZ1 (Fig. 2). In a similar way to other members of the ZBTB family, both PATZ1 BTB crystal structures reveal a strand-exchange homodimer, organized as a core fold BTB domain preceded by an N-terminal extension that interacts with the partner chain in the dimer. The characteristic secondary structures of the dimerization interfaces ($\beta 1$ – $\alpha 1$, A1–A2 and $\beta 2$ –A5) are conserved. Size-exclusion chromatography data for both the murine and zebrafish BTB domains suggest that the homodimeric complex is the most abundant oligomerization state found in solution (Supplementary Fig. S1). 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 (Figs. 2*a* and 2*b*). The last ten of these amino acids (ALEVLFQGP) are visible in the structure and fold into a β -strand ($\beta 0$) anti-parallel to the first N-terminal PATZ1 β -strand ($\beta 1$) (Supplementary Fig. S2). Superposition of this structure with

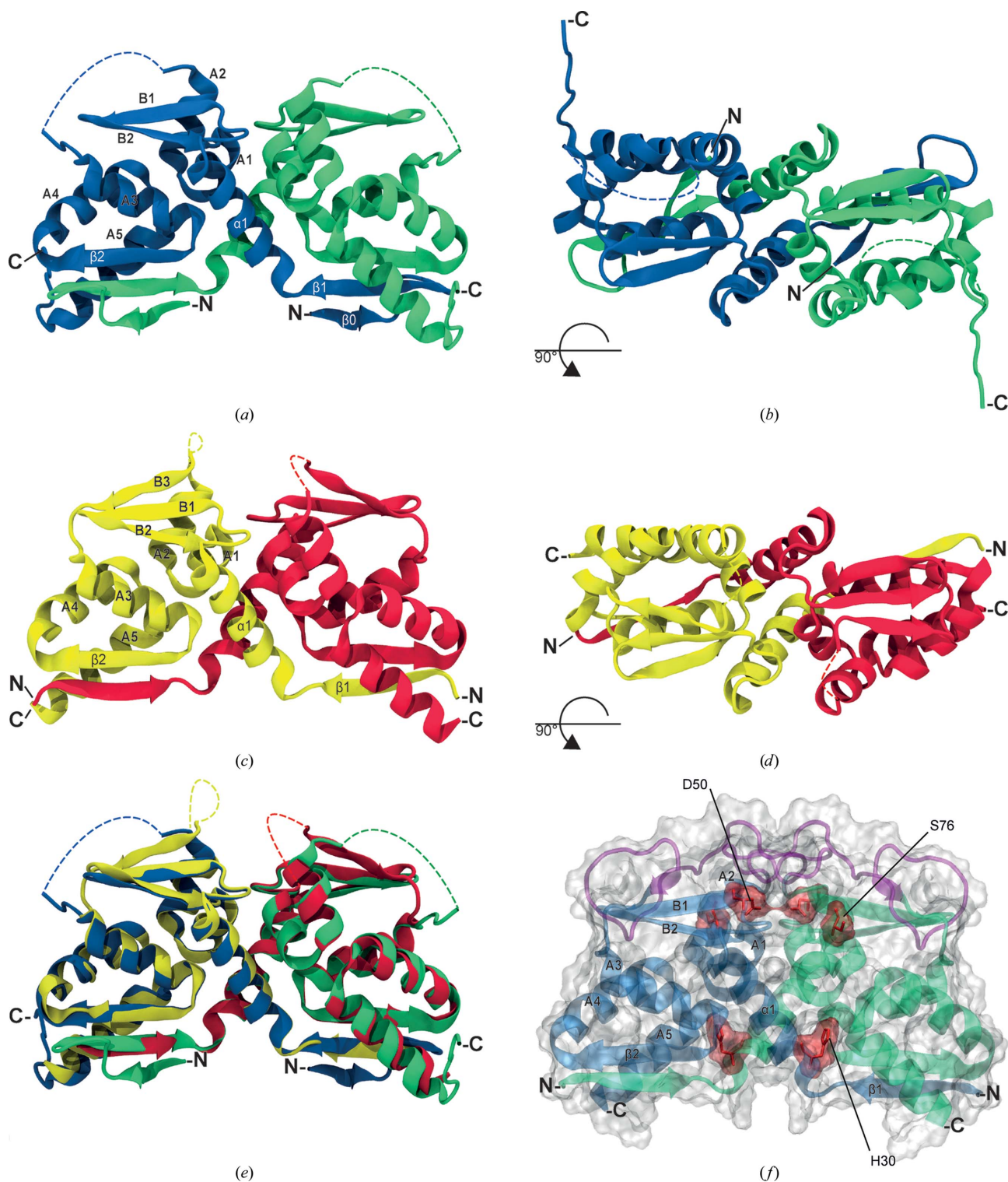


Figure 2
 Structure of the PATZ1 BTB domain. (a) Crystal structure of the BTB domain of the mouse PATZ1 protein (PDB entry 6guv) in cartoon representation (front view). The crystallographic asymmetric unit contains a PATZ1 BTB domain monomer (blue) with a second monomer (green) created by crystallographic symmetry. Secondary structures are indicated in capital and Greek letters. (b) Top view of the mouse PATZ1 BTB domain structure. N- and C-termini are indicated on one monomer. The coordinates of 31 residues in a central region, unique to mammalian PATZ1 BTB domains, could not be assigned (indicated by a dotted loop). (c) Crystal structure of the BTB domain of the zebrafish PATZ1 protein (PDB entry 6guw) with individual monomers coloured yellow and red (front view). (d) Top view of the zebrafish PATZ1 BTB structure. The coordinates of seven residues in the zebrafish PATZ1 BTB domain could not be assigned (indicated by a dotted loop). (e) Superimposition of the mouse (blue and green) and zebrafish (yellow and red) PATZ1 BTB domains (r.m.s.d. of 0.62 Å). (f) A space-filling representation of the mouse PATZ1 BTB domain structure. The predicted structure (in purple) of the central region was generated by homology modelling followed by conformation equilibration using MD simulations. Note that the modelled structure contains a predicted short β -strand. The three conserved degron residues, annotated at the bottom of Fig. 1(b) and predicted to play a role in BTB dimer degradation, are highlighted in red. Numbering refers to the residues in the crystal structure.

that of BCL6 BTB in complex with co-repressor peptides (PDB entries 1r2b and 3bim; Ahmad *et al.*, 2003; Ghetu *et al.*, 2008) suggests that these extra residues structurally mimic the β -strand-forming residues in both SMRT and BCOR (Supplementary Fig. S3), 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.

The mammalian PATZ1 (ZBTB19) 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 (Figs. 1a and 1b). 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). 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 structure no density could be assigned to the residues belonging to the A2/B3 loop, suggesting that these amino acids are partially disordered or flexible.

Interestingly, in the mouse PATZ1 BTB structure a seven-amino-acid stretch from the C-terminus of an adjacent molecule in the crystal unit cell extends into the region normally occupied by the B3 β -strand in other ZBTB proteins (Supplementary Fig. S4). 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, we crystallized a mouse PATZ1 BTB construct lacking the last seven C-terminal amino acids. 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.

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. 1a). The mammalian A2/B3 loop could be an evolutionarily acquired insertion sequence encoding an intrinsically disordered loop (IDL; Fukuchi *et al.*, 2006). To study the structure of this region in detail, we solved the structure of the zebrafish PATZ1 BTB domain to a resolution of 1.8 Å (Figs. 2c and 2d). 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. When excluding the central loop, the murine and zebrafish sequences share 83.7% identity, whilst the structures can be superimposed with an

r.m.s.d. of 0.62 Å (Fig. 2e). The zebrafish BTB domain also has the same quaternary structure as the mouse BTB domain: a strand-exchange homodimer. 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 zebrafish PATZ1 BTB domain also has a seven-amino-acid sequence between A2 and B3 for which no discernible electron density could be found. This loop was modelled using *ModLoop* (Fiser *et al.*, 2000; Fiser & Sali, 2003; Fig. 3d).

The human and mouse PATZ1 BTB domains have a single amino-acid difference (T91A). To detail the possible structure and dynamic behaviour of the A2/B3 loop in human PATZ1 BTB, we performed molecular modelling and 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 (Figs. 2f and 3c) and that at least ten amino acids within the modelled loop contribute to the homodimerization interface (representing 17.5% of the total interface of 57 amino acids). Interestingly, the β -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. Other BTB domains such as LRF and MIZ1 also contain flexible loops in this region (Stogios *et al.*, 2007; Stead *et al.*, 2007). In the case of MIZ1, the A2/B3 region mediates tetramerization of its BTB domain, whilst using size-exclusion chromatography (Supplementary Fig. S1) we found no evidence for such oligomerization in mouse or zebrafish PATZ1.

3.2. A highly charged and dynamic surface contributes to the homodimerization interface of the PATZ1 BTB domain

The dimerization interfaces of the mouse and zebrafish PATZ1 BTB domains are very similar. Using the *PIC* tool (Tina *et al.*, 2007), we determined that 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 (Figs. 3a and 3b). The interfaces retrieved from *PDBePISA* (Krissinel & Henrick, 2007) contain four basic and five acidic residues for the murine protein and four basic and three acidic residues for the zebrafish protein (Table 2).

In order to understand the dynamics of these interfaces, we assessed the number of contact-forming residues in the energy-minimized modelled BTB domains (Figs. 2f, 3c and 3d). 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 2). To understand the stability of these contacts, we assessed those that persist above a threshold value (15%) during the lifetime of the MD simulation (Figs. 3c and 3d). 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 show that it can contact a broader number of charged residues, including those from the flexible loop (Asp50, Asp84 and Asp89; Supplementary Fig. S6). We also find that two of the three degron residues (annotated in Figs. 1*b* and 2*f*) that are predicted to

play a role in BTB heterodimer degradation participate in the interaction interface of both BTB homodimer structures.

3.3. Co-repressor binding modalities are not conserved in different BTB domains

BTB domains have been shown to interact with the co-repressor proteins NCOR1, SMRT and BCOR (Huynh &

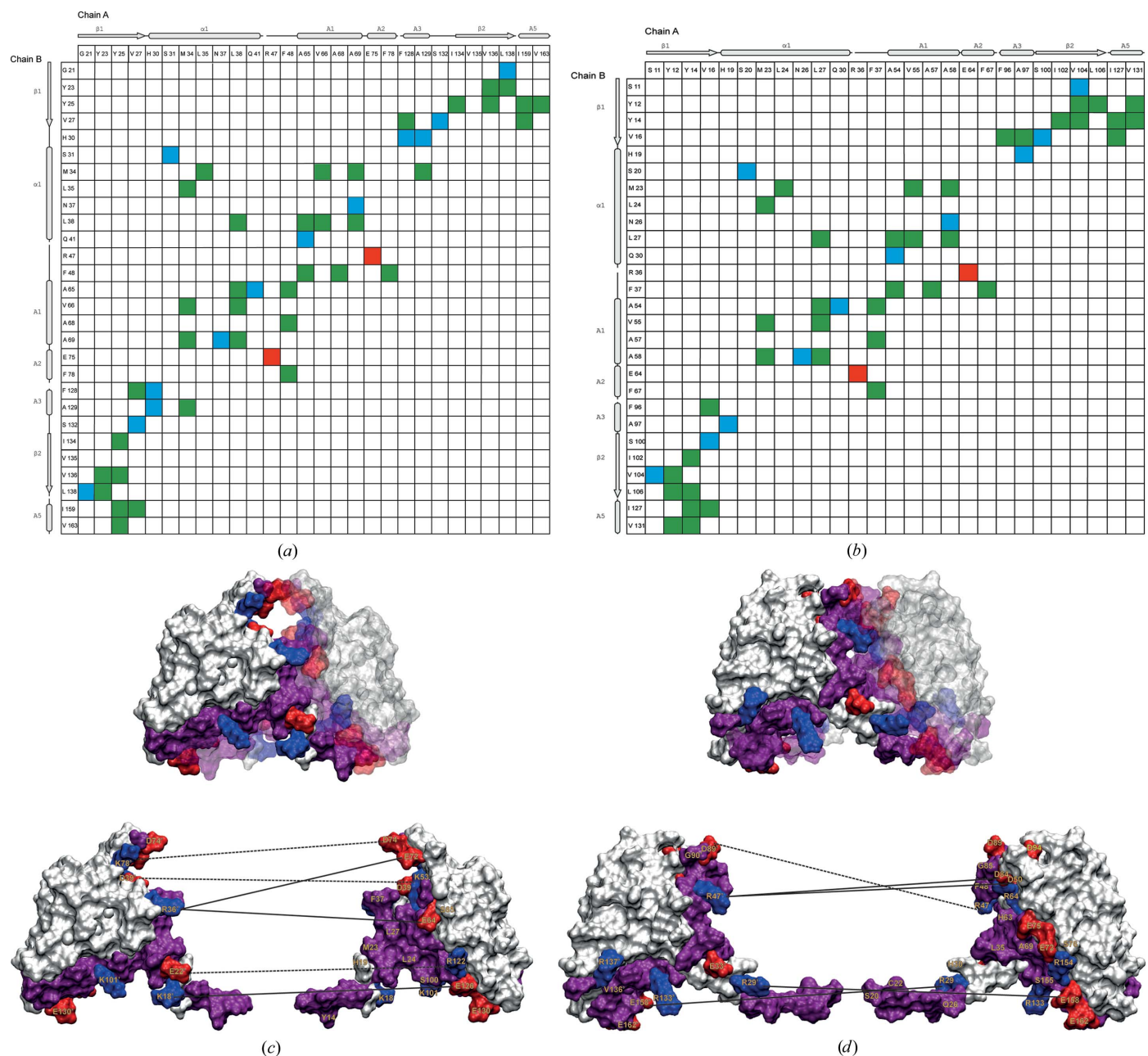


Figure 3 Identification of residues involved in homodimer interaction interfaces. Contact maps of inter-chain residue interactions of the mouse (a) and zebrafish (b) PATZ1 BTB domain crystal structures. Hydrophobic interactions (green), hydrogen bonds (blue) and ionic bonds (red) are indicated. The relevant elements of the secondary structures are shown for orientation. A highly charged dimerization interface mediates PATZ1 BTB homodimerization. (c) A split homodimer view in surface representation and completed with the modelled loop highlights the residues involved in the interaction interface; 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. (d) The zebrafish PATZ1 dimerization interface is also shown as a split homodimer view for comparison. Numbering refers to the crystal structure files.

Table 2
Inter-chain salt bridges in the dimerization interface of the BTB domain of PATZ1.

Residue numbers refer to the crystal structures.

BTB protein	PDB code	Inter-chain salt bridges	Frequency in MD (%)	Intra-monomer salt bridges	
				Intra-interface	All protein
Mouse PATZ1	6guv	Glu75–Arg47 (crystal only)	0	Asp50–Arg64	Asp127–Arg42
		Asp50–Arg47 (<i>B–A</i>)†	64.84	Asp124–Arg137	Glu111–Arg54†
		Asp84–Arg47†	7.54 (<i>BA</i>)–32.77 (<i>AB</i>)	Asp84–Arg64†	Glu111–Arg110 (<i>B–B</i>)†
		Asp89–Arg47 (<i>A–B</i>)†	11.14	Asp94–Arg110†	Asp127–Lys43†
		Glu158–Arg29†	32.32 (<i>AB</i>)–76.82 (<i>BA</i>)	Glu72–Arg154†	Glu143–Lys120†
				Asp50–Arg110 (<i>A–A</i>)†	Asp57–Arg54†
				Asp89–Arg64 (<i>A–A</i>)†	Asp57–Lys120†
				Asp89–Arg110†	Asp124–Lys120†
				Asp89–Arg47 (<i>B–B</i>)†	Glu58–Arg110 (<i>A–A</i>)†
				Asp127–Arg133†	Glu58–Lys120†
				Glu33–Arg29†	Glu58–Lys43 (<i>B–B</i>)†
				Glu75–Arg64 (<i>A–A</i>)†	Glu139–Lys120†
				Glu139–Arg137†	Asp94–Arg64 (<i>A–A</i>)†
				Glu143–Arg137†	
		Glu158–Arg154†			
		Glu162–Lys165†			
Zebrafish PATZ1	6guw	Glu64–Arg36	77.42 (<i>BA</i>)–85.7 (<i>AB</i>)	Asp92–Arg105	Asp95–Arg31
		Asp39–Lys53 (<i>A–B</i>)†	5.44	Asp33–Arg36†	Glu61–Arg122
		Asp74–Lys78†	5.89 (<i>AB</i>)	Asp39–Lys53†	Glu107–Lys88
		Glu22–Arg122 (<i>A–B</i>)†	12.84	Asp39–Lys78 (<i>A–A</i>)†	Asp33–Lys32†
		Glu72–Arg36 (<i>A–B</i>)†	93.06	Asp73–Lys53 (<i>A–A</i>)†	Asp92–Arg91†
		Glu126–Lys18†	19.08 (<i>AB</i>)–28.42 (<i>BA</i>)	Asp92–Lys101 (<i>A–A</i>)†	Asp92–Lys88†
				Asp95–Lys101†	Asp95–Arg91†
				Glu22–Lys18†	Asp95–Lys32†
				Glu22–Arg122 (<i>A–A</i>)†	Glu46–Arg43†
				Glu72–Lys53†	Glu46–Arg91 (<i>A–A</i>)†
				Glu107–Arg105†	Glu46–Lys88†
			Glu47–Arg91†		
			Glu47–Lys88†		
			Glu79–Arg43†		
			Glu79–Lys78†		
			Glu81–Arg43†		
			Glu111–Lys88†		
			Glu126–Arg122 (<i>A–A</i>)†		

† Added with MD simulation of the modelled structures.

Bardwell, 1998; Wong & Privalsky, 1998; Melnick *et al.*, 2002; Huynh *et al.*, 2000). 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 (Ahmad *et al.*, 2003; Fig. 4). Additionally, four residues, when mutated (L19S, N23H and L25S/R26L), interfere with co-repressor binding by preventing homodimerization of the BTB domain (Huynh & Bardwell, 1998; Ghetu *et al.*, 2008; Granadino-Roldán *et al.*, 2014). The PATZ1 BTB domain has also been shown to bind to NCOR1, suggesting that a similar lateral groove may be mediating this interaction (Bilic *et al.*, 2006). 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 (Bilic *et al.*, 2006).

Even though BCL6 and PATZ1 are structurally very similar, their corresponding co-repressor binding interface sequences are not conserved (Fig. 4d). To examine the structural similarity between the PATZ1 and BCL6 BTB domains, we calculated the r.m.s.d. (1.56 Å) between individual monomers. The structural similarity between BCL6 and PATZ1 was more evident when the flexible PATZ1 loop was excluded (r.m.s.d.

of 1.23 Å). Comparison of the surface charge distributions of the two proteins indicates major differences (Supplementary Fig. S5). 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. 4a). Surprisingly, the surface of PATZ1 corresponding to the BCL6 lateral groove did not contain as many basic residues (Fig. 4d and Supplementary Fig. S5). 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 co-repressors 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 (Stogios *et al.*, 2007). We also find that mouse PATZ1 residue Asp50 (Supplementary Fig. S5) is part of the charged pocket that is conserved between BCL6, LRF and PLZF (Stogios *et al.*, 2007). 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). The charged pocket that is formed by the participation of Asp50

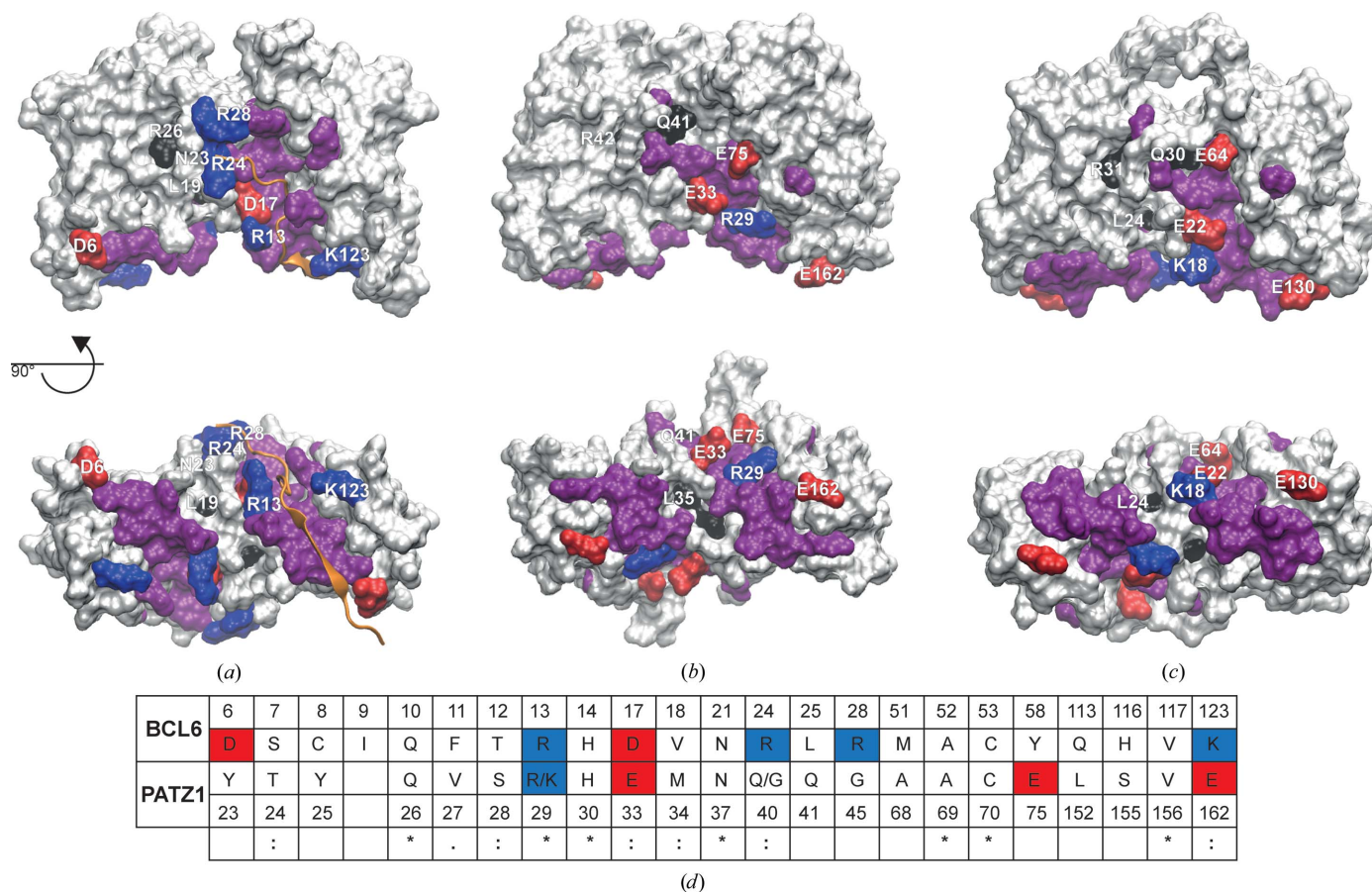


Figure 4
 Comparison of the lateral grooves of PATZ1 and BCL6 BTB domains. The structures of the BCL6 (PDB entry 1r2b) (a) and the energy-minimized modelled mouse (b) and zebrafish (c) PATZ1 BTB domains are shown in surface representation viewed from the front and bottom. The surface area of the residues in the lateral groove of the BCL6 BTB domain is buried upon formation of the BCL6–SMRT complex (Ahmad *et al.*, 2003). 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. Colours indicate residue type: positively charged, blue; negatively charged, red. All other residues in this region are coloured 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). 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 coloured 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.

residues from both monomers has been suggested as an alternative region for ligand binding (Supplementary Fig. S5; Melnick *et al.*, 2002).

4. Discussion

A role for PATZ1 has been demonstrated in various malignancies such as thyroid and testicular cancer (Fedele *et al.*, 2008, 2017; Valentino, Palmieri, Vitiello, Pierantoni *et al.*, 2013; Chiappetta *et al.*, 2015; Vitiello *et al.*, 2016; Monaco *et al.*, 2018). The interaction between PATZ1 and the tumour suppressor p53 (Valentino, Palmieri, Vitiello, Pierantoni *et al.*, 2013; Valentino, Palmieri, Vitiello, Simeone *et al.*, 2013; Chiappetta *et al.*, 2015; Keskin *et al.*, 2015) mediated by a motif in the zinc-finger DNA-binding domain rather than the BTB domain also links this protein to cancer. 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 (Siegfried *et al.*, 2019; Bridge *et al.*, 2019; Chougule *et al.*, 2019; Sankar & Lessnick, 2011; Mastrangelo *et al.*, 2000; Im *et al.*, 2000). The dimerization and co-repressor interaction properties of PATZ1 identified in this study may shed light on the mechanism of these sarcomas.

The gene targets of PATZ1 have not been extensively identified. ChIP-Seq and RNASeq experiments have identified 187 putative targets (Encode Project Consortium, 2012; Keskin *et al.*, 2015). 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 current study only highlights structural motifs that are likely to play a role in the interaction of the PATZ1 BTB domain with co-repressor proteins. Yet other interactions may be involved in the potential role of PATZ1 in gene upregulation.

Our present study identifies important structural features of the PATZ1 BTB domain. One unique feature of BTB domains is their ability to form homodimers as well as heterodimers as a result of their close structural homology. Dimer formation is necessary for interaction with co-repressor proteins for both PATZ1 and BCL6 (Bilic *et al.*, 2006). A lateral groove that BCL6 uses to bind co-repressors is structurally conserved in PATZ1 (Fig. 4). However, the discrepancy in charged amino acids in this groove (Supplementary Fig. S5) may indicate altered binding modalities and/or affinities for co-repressors. In fact, a structurally conserved charged pocket has previously been hypothesized to be involved in ligand interaction (Stogios *et al.*, 2007). While this charged pocket is surface-exposed in BCL6, LRF and PLZF structures, the PATZ1 A2/B3 loop could dynamically gate this site, potentially regulating ligand interaction.

Another feature that is common to PATZ1 and BCL6 is their BTB domain-mediated localization to nuclear speckles (Huynh *et al.*, 2000; Fedele *et al.*, 2000; Franco *et al.*, 2016). 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 (Pero *et al.*, 2002). A recent study identified three BTB domain degron residues that are surface-exposed preferentially in heterodimers (Mena *et al.*, 2018). The targeting of these degrons by ubiquitin ligases mediates the proteasome-dependent degradation of heterodimers over homodimers. It is not known whether this is a generalizable feature of BTB domains. Consistent with the stability of homodimers, in the current study we find that PATZ1 buries two of these three degron residues in the protein globular structure (Fig. 2*f*).

The MD simulation shows that alternative contacts are possible for several charged residues at the homodimer interface (Table 2). Because a single charged residue can contact more than one oppositely charged residue (Fig. 3), the dimer interface may be more resilient to the disruption of single contacts. 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. Nevertheless, there are an exceptional number of unpaired charged surface residues in the murine PATZ1 BTB domain. Some proteins, for example calmodulin, with large net surface charges are known to modulate their environment by redistributing nonspecific ions in the surrounding medium (Aykut *et al.*, 2013). Such modes of action are utilized to shift the population of the available conformational states, leading to fine-tuned functions.

BTB domains are attractive targets for anticancer compounds. Compounds that prevent homodimerization or result in the degradation of BCL6 (Kerres *et al.*, 2017) 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 (Fig. 1). 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.

Acknowledgements

We thank Professor Dr Stefan H. Fuss, Professor Dr Zehra Sayers and Professor Dr Wilfried Ellmeier for helpful comments for the preparation of the manuscript. Author contributions were as follows. SP and AA performed experiments and solved the crystal structures, SP and CA performed the MD, and SP, BE and EM supervised the overall study and wrote the manuscript.

Funding information

This work was funded by TUBITAK 1001 grant 118Z015 and The Royal Society Newton International Exchanges Grant NI140172. SP was supported by a TUBITAK BIDEB scholarship.

References

- Ahmad, K. F., Engel, C. K. & Privé, G. G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12123–12128.
- Ahmad, K. F., Melnick, A., Lax, S., Bouchard, D., Liu, J., Kiang, C.-L., Mayer, S., Takahashi, S., Licht, J. D. & Privé, G. G. (2003). *Mol. Cell*, **12**, 1551–1564.
- Andersen, L., Gülich, A. F., Alteneder, M., Preglej, T., Orola, M. J., Dhele, N., Stolz, V., Schebesta, A., Hamminger, P., Hladik, A., Floess, S., Krausgruber, T., Faux, T., Andrabi, S. B. A., Huehn, J., Knapp, S., Sparwasser, T., Bock, C., Laiho, A., Elo, L. L., Rasool, O., Lahesmaa, R., Sakaguchi, S. & Ellmeier, W. (2019). *Cell. Rep.* **29**, 4447–4459.
- Aykut, A. O., Atilgan, A. R. & Atilgan, C. (2013). *PLoS Comput. Biol.* **9**, e1003366.
- Bailey, T. L. & Elkan, C. (1994). *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36.
- Bardwell, V. J. & Treisman, R. (1994). *Genes Dev.* **8**, 1664–1677.
- Bilic, I. & Ellmeier, W. (2007). *Immunol. Lett.* **108**, 1–9.
- Bilic, I., Koesters, C., Unger, B., Sekimata, M., Hertweck, A., Maschek, R., Wilson, C. B. & Ellmeier, W. (2006). *Nat. Immunol.* **7**, 392–400.
- Bonchuk, A., Denisov, S., Georgiev, P. & Maksimenko, O. (2011). *J. Mol. Biol.* **412**, 423–436.
- Bridge, J. A., Sumegi, J., Druta, M., Bui, M. M., Henderson-Jackson, E., Linos, K., Baker, M., Walko, C. M., Millis, S. & Brohl, A. S. (2019). *Mod. Pathol.* **32**, 1593–1604.
- Brooks, B. R., Brooks, C. L., MacKerell, A. D. Jr, Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodosscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. (2009). *J. Comput. Chem.* **30**, 1545–1614.
- Cerchiatti, L. C., Ghetu, A. F., Zhu, X., Da Silva, G. F., Zhong, S., Matthews, M., Bunting, K. L., Polo, J. M., Farès, C., Arrowsmith, C. H., Yang, S. N., Garcia, M., Coop, A., MacKerell, A. D. Jr, Privé, G. G. & Melnick, A. (2010). *Cancer Cell*, **17**, 400–411.
- Cheng, H. M., Linhares, B. M., Yu, W. B., Cardenas, M. G., Ai, Y., Jiang, W. J., Winkler, A., Cohen, S., Melnick, A., MacKerell, A., Cierpicki, T. & Xue, F. T. (2018). *J. Med. Chem.* **61**, 7573–7588.
- Chevrier, S. & Corcoran, L. M. (2014). *Immunol. Cell Biol.* **92**, 481–488.

- Chiappetta, G., Valentino, T., Vitiello, M., Pasquinelli, R., Monaco, M., Palma, G., Sepe, R., Luciano, A., Pallante, P., Palmieri, D., Aiello, C., Rea, D., Losito, S. N., Arra, C., Fusco, A. & Fedele, M. (2015). *Oncotarget*, **6**, 5310–5323.
- Chougule, A., Taylor, M. S., Nardi, V., Chebib, I., Cote, G. M., Choy, E., Nielsen, G. P. & Deshpande, V. (2019). *Am. J. Surg. Pathol.* **43**, 220–228.
- Darden, T., Perera, L., Li, L. P. & Pedersen, L. (1999). *Structure*, **7**, R55–R60.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Fedele, M., Benvenuto, G., Pero, R., Majello, B., Battista, S., Lembo, F., Vollono, E., Day, P. M., Santoro, M., Lania, L., Bruni, C. B., Fusco, A. & Chiariotti, L. (2000). *J. Biol. Chem.* **275**, 7894–7901.
- Fedele, M., Crescenzi, E. & Cerchia, L. (2017). *Int. J. Mol. Sci.* **18**, 2524.
- Fedele, M., Franco, R., Salvatore, G., Paronetto, M. P., Barbagallo, F., Pero, R., Chiariotti, L., Sette, C., Tramontano, D., Chieffi, G., Fusco, A. & Chieffi, P. (2008). *J. Pathol.* **215**, 39–47.
- Fiser, A. & Sali, A. (2003). *Bioinformatics*, **19**, 2500–2501.
- Fiser, A., Do, R. K. G. & Šali, A. (2000). *Protein Sci.* **9**, 1753–1773.
- Franco, R., Scognamiglio, G., Valentino, E., Vitiello, M., Luciano, A., Palma, G., Arra, C., La Mantia, E., Panico, L., Tenneriello, V., Pinto, A., Frigeri, F., Capobianco, G., Botti, G., Cerchia, L., De Chiara, A. & Fedele, M. (2016). *Oncotarget*, **7**, 59158–59172.
- Fukuchi, S., Homma, K., Minezaki, Y. & Nishikawa, K. (2006). *J. Mol. Biol.* **355**, 845–857.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). *Nucleic Acids Res.* **31**, 3784–3788.
- Ghetu, A. F., Corcoran, C. M., Cerchietti, L., Bardwell, V. J., Melnick, A. & Privé, G. G. (2008). *Mol. Cell*, **29**, 384–391.
- Granadino-Roldán, J. M., Obiol-Pardo, C., Pinto, M., Garzón, A. & Rubio-Martínez, J. (2014). *J. Mol. Graph. Model.* **50**, 142–151.
- Hatherley, R., Brown, D. K., Glenister, M. & Tastan Bishop, O. (2016). *PLoS One*, **11**, e0166698.
- He, X., Park, K. & Kappes, D. J. (2010). *Annu. Rev. Immunol.* **28**, 295–320.
- Herhaus, L. & Dikic, I. (2018). *Science*, **362**, 151–152.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
- Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., Kolippakkam, D., Mintseris, J., Obar, R. A., Harris, T., Artavanis-Tsakonas, S., Sowa, M. E., De Camilli, P., Paulo, J. A., Harper, J. W. & Gygi, S. P. (2015). *Cell*, **162**, 425–440.
- Huynh, K. D. & Bardwell, V. J. (1998). *Oncogene*, **17**, 2473–2484.
- Huynh, K. D., Fischle, W., Verdin, E. & Bardwell, V. J. (2000). *Genes Dev.* **14**, 1810–1823.
- Im, Y.-H., Kim, H. T., Lee, C., Poulin, D., Welford, S., Sorensen, P. H. B., Denny, C. T. & Kim, S.-J. (2000). *Cancer Res.* **60**, 1536–1540.
- Jeon, B.-N., Kim, M.-K., Yoon, J.-H., Kim, M.-Y., An, H., Noh, H.-J., Choi, W.-I., Koh, D.-I. & Hur, M.-W. (2014). *Nucleic Acids Res.* **42**, 11447–11461.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kamada, Y., Sakai, N., Sogabe, S., Ida, K., Oki, H., Sakamoto, K., Lane, W., Snell, G., Iida, M., Imaeda, Y., Sakamoto, J. & Matsui, J. (2017). *J. Med. Chem.* **60**, 4358–4368.
- Kerres, N., Steurer, S., Schlager, S., Bader, G., Berger, H., Caligiuri, M., Dank, C., Engen, J. R., Etmayer, P., Fischerauer, B., Flotzinger, G., Gerlach, D., Gerstberger, T., Gmaschitz, T., Greb, P., Han, B., Heyes, E., Jacob, R. E., Kessler, D., Kölle, H., Lamarre, L., Lancia, D. R., Lucas, S., Mayer, M., Mayr, K., Mischerikow, N., Mück, K., Peinsipp, C., Petermann, O., Reiser, U., Rudolph, D., Rumpel, K., Salomon, C., Scharn, D., Schnitzer, R., Schrenk, A., Schweifer, N., Thompson, D., Traxler, E., Vареcka, R., Voss, T., Weiss-Puxbaum, A., Winkler, S., Zheng, X., Zoepfel, A., Kraut, N., McConnell, D., Pearson, M. & Koegl, M. (2017). *Cell. Rep.* **20**, 2860–2875.
- Keskin, N., Deniz, E., Eryilmaz, J., Un, M., Batur, T., Ersahin, T., Cetin Atalay, R., Sakaguchi, S., Ellmeier, W. & Erman, B. (2015). *Mol. Cell. Biol.* **35**, 1741–1753.
- Kobayashi, A., Yamagiwa, H., Hoshino, H., Muto, A., Sato, K., Morita, M., Hayashi, N., Yamamoto, M. & Igarashi, K. (2000). *Mol. Cell. Biol.* **20**, 1733–1746.
- Kramer, G., Shiber, A. & Bukau, B. (2019). *Annu. Rev. Biochem.* **88**, 337–364.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lee, K. M., Choi, W. I., Koh, D. I., Kim, Y. J., Jeon, B. N., Yoon, J. H., Lee, C. E., Kim, S. H., Oh, J. & Hur, M. W. (2012). *Oncogene*, **31**, 1442–1458.
- Lee, S.-U. & Maeda, T. (2012). *Immunol. Rev.* **247**, 107–119.
- Li, J., Wang, J., Wang, J., Nawaz, Z., Liu, J. M., Qin, J. & Wong, J. (2000). *EMBO J.* **19**, 4342–4350.
- Li, X., Peng, H., Schultz, D. C., Lopez-Guisa, J. M., Rauscher, F. J. & Marmorstein, R. (1999). *Cancer Res.* **59**, 5275–5282.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Ma, H., Ow, J. R., Tan, B. C., Goh, Z., Feng, B., Loh, Y. H., Fedele, M., Li, H. & Wu, Q. (2014). *Sci. Rep.* **4**, 7519.
- Mastrangelo, T., Modena, P., Torrielli, S., Bullrich, F., Testi, M. A., Mezzelani, A., Radice, P., Azzarelli, A., Pilotti, S., Croce, C. M., Pierotti, M. A. & Sozzi, G. (2000). *Oncogene*, **19**, 3799–3804.
- McCoull, W., Abrams, R. D., Anderson, E., Blades, K., Barton, P., Box, M., Burgess, J., Byth, K., Cao, Q., Chuaqui, C., Carbajo, R. J., Cheung, T., Code, E., Ferguson, A. D., Fillery, S., Fuller, N. O., Gangl, E., Gao, N., Grist, M., Hargreaves, D., Howard, M. R., Hu, J., Kemmitt, P. D., Nelson, J. E., O’Connell, N., Prince, D. B., Raubo, P., Rawlins, P. B., Robb, G. R., Shi, J., Waring, M. J., Whittaker, D., Wylot, M. & Zhu, X. (2017). *J. Med. Chem.* **60**, 4386–4402.
- McCoull, W., Cheung, T., Anderson, E., Barton, P., Burgess, J., Byth, K., Cao, Q., Castald, M. P., Chen, H., Chiarparin, E., Carbajo, R. J., Code, E., Cowan, S., Davey, P. R., Ferguson, A. D., Fillery, S., Fuller, N. O., Gao, N., Hargreaves, D., Howard, M. R., Hu, J., Kawatkar, A., Kemmitt, P. D., Leo, E., Molina, D. M., O’Connell, N., Petheruti, P., Rasmusson, T., Raubo, P., Rawlins, P. B., Ricchiuto, P., Robb, G. R., Schenone, M., Waring, M. J., Zinda, M., Fawell, S. & Wilson, D. M. (2018). *ACS Chem. Biol.* **13**, 3131–3141.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Melnick, A., Ahmad, K. F., Arai, S., Polinger, A., Ball, H., Borden, K. L., Carlile, G. W., Prive, G. G. & Licht, J. D. (2000). *Mol. Cell. Biol.* **20**, 6550–6567.
- Melnick, A., Carlile, G., Ahmad, K. F., Kiang, C.-L., Corcoran, C., Bardwell, V., Prive, G. G. & Licht, J. D. (2002). *Mol. Cell. Biol.* **22**, 1804–1818.
- Mena, E. L., Kjolby, R. A. S., Saxton, R. A., Werner, A., Lew, B. G., Boyle, J. M., Harland, R. & Rape, M. (2018). *Science*, **362**, eaap8236.
- Monaco, M., Palma, G., Vitiello, M., Capiluongo, A., D’Andrea, B., Vuttariello, E., Luciano, A., Cerchia, L., Chiappetta, G., Arra, C., Fusco, A. & Fedele, M. (2018). *Cancers*, **10**, 92.
- Ow, J. R., Ma, H., Jean, A., Goh, Z., Lee, Y. H., Chong, Y. M., Soong, R., Fu, X.-Y., Yang, H. & Wu, Q. (2014). *Stem Cells Dev.* **23**, 1062–1073.
- Pei, J., Kim, B.-H. & Grishin, N. V. (2008). *Nucleic Acids Res.* **36**, 2295–2300.
- Perez-Torrado, R., Yamada, D. & Defossez, P. A. (2006). *Bioessays*, **28**, 1194–1202.
- Pero, R., Lembo, F., Palmieri, E. A., Vitiello, C., Fedele, M., Fusco, A., Bruni, C. B. & Chiariotti, L. (2002). *J. Biol. Chem.* **277**, 3280–3285.

- Phan, R. T., Saito, M., Basso, K., Niu, H. & Dalla-Favera, R. (2005). *Nat. Immunol.* **6**, 1054–1060.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005). *J. Comput. Chem.* **26**, 1781–1802.
- Pierce, B., Tong, W. W. & Weng, Z. P. (2005). *Bioinformatics*, **21**, 1472–1478.
- Pruitt, K. D., Tatusova, T. & Maglott, D. R. (2007). *Nucleic Acids Res.* **35**, D61–D65.
- Sakaguchi, S., Hainberger, D., Tizian, C., Tanaka, H., Okuda, T., Taniuchi, I. & Ellmeier, W. (2015). *J. Immunol.* **195**, 2879–2887.
- Sakaguchi, S., Hombauer, M., Bilic, I., Naoe, Y., Schebesta, A., Taniuchi, I. & Ellmeier, W. (2010). *Nat. Immunol.* **11**, 442–448.
- Sakamoto, K., Sogabe, S., Kamada, Y., Sakai, N., Asano, K., Yoshimatsu, M., Ida, K., Imaeda, Y. & Sakamoto, J. I. (2017). *Biochem. Biophys. Res. Commun.* **482**, 310–316.
- Šali, A. & Blundell, T. L. (1993). *J. Mol. Biol.* **234**, 779–815.
- Sankar, S. & Lessnick, S. L. (2011). *Cancer Genet.* **204**, 351–365.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. (2005). *Nucleic Acids Res.* **33**, W363–W367.
- Schubot, F. D., Tropea, J. E. & Waugh, D. S. (2006). *Biochem. Biophys. Res. Commun.* **351**, 1–6.
- Shatsky, M., Nussinov, R. & Wolfson, H. J. (2004). *Proteins*, **56**, 143–156.
- Siegfried, A., Rousseau, A., Maurage, C.-A., Pericart, S., Nicaise, Y., Escudie, F., Grand, D., Delrieu, A., Gomez-Brouchet, A., Le Guellec, S., Franchet, C., Boetto, S., Vinchon, M., Sol, J.-C., Roux, F.-E., Rigau, V., Bertozzi, A.-I., Jones, D. T. W., Figarella-Branger, D. & Uro-Coste, E. (2019). *Brain Pathol.* **29**, 53–62.
- Siggs, O. M. & Beutler, B. (2012). *Cell Cycle*, **11**, 3358–3369.
- Stead, M. A., Rosbrook, G. O., Hadden, J. M., Trinh, C. H., Carr, S. B. & Wright, S. C. (2008). *Acta Cryst.* **F64**, 1101–1104.
- Stead, M. A., Trinh, C. H., Garnett, J. A., Carr, S. B., Baron, A. J., Edwards, T. A. & Wright, S. C. (2007). *J. Mol. Biol.* **373**, 820–826.
- Stead, M. A. & Wright, S. C. (2014). *Acta Cryst.* **F70**, 1591–1596.
- Stogios, P. J., Chen, L. & Privé, G. G. (2007). *Protein Sci.* **16**, 336–342.
- Stogios, P. J., Cuesta-Seijo, J. A., Chen, L., Pomroy, N. C. & Privé, G. G. (2010). *J. Mol. Biol.* **400**, 983–997.
- Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K. & Privé, G. G. (2005). *Genome Biol.* **6**, R82.
- The ENCODE Project Consortium (2012). *Nature*, **489**, 57–74.
- The UniProt Consortium (2019). *Nucleic Acids Res.* **47**, D506–D515.
- Tina, K. G., Bhadra, R. & Srinivasan, N. (2007). *Nucleic Acids Res.* **35**, W473–W476.
- Valentino, T., Palmieri, D., Vitiello, M., Pierantoni, G. M., Fusco, A. & Fedele, M. (2013). *Cell Death Dis.* **4**, e963.
- Valentino, T., Palmieri, D., Vitiello, M., Simeone, A., Palma, G., Arra, C., Chieffi, P., Chiariotti, L., Fusco, A. & Fedele, M. (2013). *J. Cell. Physiol.* **228**, 646–653.
- Vitiello, M., Valentino, T., De Menna, M., Crescenzi, E., Francesca, P., Rea, D., Arra, C., Fusco, A., De Vita, G., Cerchia, L. & Fedele, M. (2016). *Sci. Rep.* **6**, 25268.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
- Weber, A., Marquardt, J., Elzi, D., Forster, N., Starke, S., Glaum, A., Yamada, D., Defosse, P. A., Delrow, J., Eisenman, R. N., Christiansen, H. & Eilers, M. (2008). *EMBO J.* **27**, 1563–1574.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci.* **27**, 293–315.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Wong, C.-W. & Privalsky, M. L. (1998). *J. Biol. Chem.* **273**, 27695–27702.
- Yasui, T., Yamamoto, T., Sakai, N., Asano, K., Takai, T., Yoshitomi, Y., Davis, M., Takagi, T., Sakamoto, K., Sogabe, S., Kamada, Y., Lane, W., Snell, G., Iwata, M., Goto, M., Inooka, H., Sakamoto, J. I., Nakada, Y. & Imaeda, Y. (2017). *Bioorg. Med. Chem.* **25**, 4876–4886.