bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sibling rivalry among the ZBTB transcription factor family: homo vs. heterodimers

Sofia Piepoli[a,b], Liyne Nogay[a,c], Umit Akkose[a], Sarah Barakat[a], Hakan Taskiran[a,c], Nazife Tolay [a,d], Melike Gezen[a], Ogun Adebali[a], Canan Atilgan[a] and Batu Erman[b]

[a]Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Tuzla, Istanbul, Turkey
[b]Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Bogazici University, 34342 Bebek, Istanbul, Turkey
[c]Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, 79108 Freiburg im Breisgau, Germany.
[d]Department of Biochemistry, Biocenter, University of Würzburg, Würzburg, Germany

Keywords: ZBTB Transcription factor, BTB domain, homodimer, heterodimer, binding free energy, MM-GBSA

Abstract

The BTB domain is an oligomerization domain found in over 200 proteins encoded in the human genome. In the family of BTB domain and Zinc Finger-containing (ZBTB) transcription factors, 49 members share the same protein architecture. The N-terminal BTB domain is structurally conserved among the family members and serves as the dimerization site while the C-terminal zinc finger motifs mediate DNA binding. The available BTB domain structures from this family reveal a natural inclination for homodimerization. In this study we investigated the potential for heterodimer formation in the cellular environment. We selected five BTB homodimers and four heterodimer structures. We performed *in vitro* binding assays with fluorescent protein-BTB domain fusions to assess dimer formation. We tested the binding of several BTB pairs, and we were able to confirm the heterodimeric physical interaction between the BTB domains of PATZ1 and PATZ2, previously reported only in an interactome mapping experiment. We also found this pair to be co-expressed in several immune system cell types. Finally, we used the available structures of BTB domain dimers and newly constructed models in extended molecular dynamics simulations (500 ns) to understand the energetic determinants of homo and heterodimer formation. We conclude that heterodimer formation, although frequently described as less preferred than homodimers, is a possible mechanism to increase the combinatorial specificity of this transcription factor family.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

BTB (Broad complex, Tramtrack and Bric-à-brac) domains are protein-protein interaction domains that are found in about 200 human genome encoded proteins including the N-termini of 49 Zinc finger and BTB (ZBTB) proteins[1, 2]. The X-ray structures of the 9 members of this family that have been solved to date are from ZBTB7a (LRF)[3], ZBTB16 (PLZF)[4], ZBTB17 (MIZ1) and ZBTB32 (FAZF)[5], ZBTB19 (PATZ1)[6], ZBTB27 (BCL6)[7], ZBTB31 (MYNN)[8], ZBTB33 (KAISO)[9], ZBTB48 (HKR3)[10]. These structures indicate that the BTB domain forms obligate dimers. Dimerization likely facilitates target DNA binding through the C-terminal zinc finger motifs found in the DNA binding domains of these ZBTB transcription factors[11]. In addition to mediating homodimerization, the BTB dimer forms a scaffold for other ligands that modify the transcriptional regulation of target genes[12].

We recently solved the crystal structure of one ZBTB family member, PATZ1 (ZBTB19 or MAZR) from mouse and zebrafish[6]. This work highlighted the similarity of the structures of known BTB domains. The structural similarity among the family members led us to question whether heterodimerization was possible.

To date, 14 BTB domain pairs in the ZBTB family were reported to form heterodimeric structures. These studies employ techniques that range from mass spectrophotometry to yeast two-hybrid screening (BioGRID database[13]). PATZ1, the focus of our studies, was originally identified in a two-hybrid screen with the BACH2 BTB domain used as a bait[14]. It is not clear that this or any other reported heterodimer has any biological function. One “forced” heterodimer X-ray structure indicates that MIZ1 and BCL6 can form stable heterodimers when expressed as a fusion protein[15], but whether this interaction has a physiological significance is not clear. Moreover, for many reported interactions, it is not clear that the BTB domain is sufficient for heterodimer formation, leaving the possibility that additional C-terminal residues/domains might be necessary for mediating heterodimerization.

In the present study, we employ a fluorescent two-hybrid assay (F2H) in mammalian tissue culture cells to assess the homo vs. hetero dimerization of selected BTB domains, identifying only a single pair that can stably form heterodimers. Moreover, using ImmGen data[16], we analyze positive and negative correlations among gene expression profiles of all ZBTB proteins in cells of the immune system. Finally, we employ molecular dynamics (MD) on a set of BTB homo and heterodimer structures to identify if formation of homodimers or heterodimers is energetically more favorable, and to determine the driving forces that contribute to dimer stability. While one BTB domain containing transcription factor, BACH2 contains a disulfide bond holding the obligate homodimer together in the crystal structure as well as in cell extracts[17], our structural analysis indicates that mostly electrostatic interactions and hydrophobicity are responsible for dimer formation and stability. Among the BTB domains analyzed, only a single pair, PATZ1 and PATZ2 are co-expressed in many cell types, form

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

heterodimers and have favorable binding energies. Thus, a subset of target genes of these two transcription factors are likely regulated not by homodimers but rather heterodimers.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

BTB domain and GBP-LacI cloning. The coding sequence of the BTB domain of selected ZBTB family proteins was amplified from cDNA derived from the human HCT116 cell line using Q5 High-Fidelity DNA Polymerase (NEB). Specifically for the Patz1 construct, the BTB domain was amplified from a murine Patz1 cDNA. Amplified fragments (Table S1) were cloned into the pcDNA 3.1/Myc-His(-)B- expression vector that contained either a TagGFP cDNA with a nuclear localization signal (NLS) or a TagRFP cDNA with no signal. BTB cDNAs were cloned into the XhoI and NotI restriction sites for TagGFP and between SmaI and NotI for TagRFP vectors, such that they encoded NLS-tagGFP-BTB or tagRFP-BTB proteins. The recombinant plasmid DNA was sequenced and transfected into Baby Hamster Kidney fibroblasts (BHK-1 cells) that were modified to contain concatemeric *E.coli* Lactose operator (Lac O) sequences inserted into a single locus (Chromotek, Germany).

For targeting the GFP fusion protein to the Lac O locus, we constructed a plasmid containing the Lac repressor sequence (Lac I) fused to a nanobody specific to GFP (GFP binding protein-GBP). This fusion gene was amplified and cloned into the pcDNA 3.1/Myc-His(-)B- expression vector using NheI and BamHI digested amplicons generated from the F2H platform mixture as template with forward and reverse oligonucleotides (Table S1). This plasmid encodes a fusion protein that has a 107 amino acid N-terminal GBP fused to 355 amino acid C-terminal Lac I domain separated by a 7 amino acid linker.

Transfection, live-cell microscopy and F2H assay. The Fluorescent Two-Hybrid (F2H) assay[18-20] was used to study dimer formation between pairs of BTB domains. $1.5x10^5$ BHK-1 cells were seeded into 6-well plates with coverslip bottoms and transfected with polyethyleneimine (PEI) reagent at a ratio of 1:3 (DNA:PEI wt/wt). Equal mixtures of NLS-tagGFP-BTB or tagRFP-BTB and GBP-Lac I encoding plasmids were transiently co-transfected. 24 hours after transfection, adherent cells were visualized using an invert fluorescent microscope (ZEISS Axio Observer Z1) with 10-20X magnification. Excitation was performed using either an HXP 120V fluorescent light source or a Colibri7 light source with LED470 or LED-Neutralwhite (540-580nm) and Filterset 38 (Excitation 470/40 BP; dichroic 495LP; emission 525/50 BP) or Filterset 43 (Excitation 545/25 BP; dichroic 570LP; emission 605/70 BP) for tagGFP and for tagRFP visualization respectively. Emission was detected either using a Zeiss Axiocam 503 mono or MRc5 camera.

In the F2H assay, GFP foci were only evident when the GBP-Lac I and tagGFP-BTB encoding plasmids were included in the transfection mixture. GFP-RFP colocalization was only evident when GBP-Lac I, tagGFP-BTB and tagRFP-BTB encoding plasmids were included in the transfection mixture. No foci were observed if the GBP-Lac I encoding plasmid was omitted from the transfection mixture. Because the F2H-BHK cells were not synchronized in their cell cycle, some cells were in S-phase and

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contained two tagGFP foci, resulting from duplicated chromosomes. For these instances both foci were scored as independent events. Colocalization analysis was performed manually or by using the JACoP plugin of the Fiji software Version 2.1.0/1.53c[21-23].

ImmGen cell type analysis of RNA co-expression. The gene expression data of 46 of the 49 ZBTB family genes were obtained from the Immunological Genome project (ImmGen) Microarray Phase 1 and Phase 2 datasets[16]. Probes for ZBTB21 (ZNF295), ZBTB35 (ZNF131) and ZBTB47 (ZNF651) were missing in the dataset and were not analyzed. The dataset contained gene expression data from primary murine cells from multiple immune lineages including B Lymphocytes, Monocytes, Mast, Basophil and Eosinophil (MBE), Stromal Cells, Innate Lymphocytes, Granulocytes, Macrophages, Dendritic Cells, Stem Cells and T Lymphocytes. Correlation coefficients of all pairs were calculated using least-squares linear regression and two-sided p-value was used for hypothesis testing.

Conservation Analysis. To retrieve homologs for each of the 6 BTB-domain proteins, a Blast[24] search was conducted locally against a non-redundant database (downloaded from Uniprot[25] -August 2019 release) including a canonical isoform for each protein. MAFFT[26] was used to build a multiple sequence alignment (MSA). We reconstructed a phylogenetic tree for each protein separately with FastTree[27]. We selected orthologous protein sequences from each tree, by traversing the phylogenetic tree starting from the query sequence until the node having the next human protein sequence as an eventual child. The previous node was selected as the monophyletic clade including the orthologous sequences only. Then, a new multiple sequence alignment and a new phylogenetic tree were build using the orthologs (Fig. S6). The MSA was constructed with 101 orthologous sequences for PATZ1, 118 for BCL6, 88 for MIZ1, 76 for LRF, 152 for PATZ2 and 75 for ThPOK. ConSurf web server[28] was used with the final MSA and phylogenetic tree as inputs to calculate the conservation scores of the positions.

Structure of heterodimers: docking and modelling. The four BTB heterodimer structures presented in this work were obtained from available crystal structures or newly modelled structures built by homology and docked monomers of homodimer structures. Amongst the BTB heterodimers between ZBTB family members, the MIZ1-BCL6 construct is currently the only one for which the crystal structure has been deposited[15] (PDB entry 4U2M-chain B). 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. The PATZ1-BCL6 heterodimer structure was created using BCL6-BTB monomer (PDB entry 1R29) and PATZ1-BTB monomer (PDB entry 6GUV). The three BCL6 residues mutated to aid the crystallization process (C8Q; C67R; C84N)[7] were back mutated to

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WT using the Mutate Residue plugin of VMD[29]. Missing residues in the A2/B3 loop (75-105) of the PATZ1 structure were homology modelled as described previously[6]. The LRF-BTB structure (PDB entry 2NN2) was similarly modelled to fill the missing coordinates for A2/B3 residues 66-71 with ModLoop[30]. The PATZ2/ZBTB24-BTB domain (1-126) was homology modelled with the PRIMO suite[31] using BACH1, BACH2, MIZ1, BCL6 and PATZ1 structures as templates. Similarly, the ThPOK/ cKrox/ZBTB15/ZBTB7b -BTB domain (1-144) was homology modelled using SWISSMODEL[32] using LRF/ZBTB7a as a template. All modelled heterodimer structures were generated with the PRISM docking server[33] by selecting the pose with the highest energy score.

Molecular Dynamics simulations. Molecular dynamics (MD) simulations were performed in NAMD using the CHARMM36 force field parameters[34, 35]. The simulation environment was prepared in VMD[29]. BTB dimer structures were centered in a solvent box padded with a 10 Å layer of water in every direction. The solvent was modelled using TIP3W water molecules, ionized with 0.15 M KCl. Periodic boundary conditions were applied in which long-range electrostatic interactions were treated using the particle mesh Ewald method[36] with a cutoff distance of 12 Å. The structural analysis by molecular simulation includes an initial run of minimization at constant temperature and constant volume (NVT). In the case of PATZ1-BCL6 and LRF-ThPOK heterodimers, the protein dimers were minimized for 30000 steps. A series of short runs (2 ns) with ramping temperature at 10 K intervals (from 280 to 310 K) was performed to reach the final running temperature of 310 K. All simulations were then performed at a constant temperature of 310 K in isothermal and isobaric conditions (NPT) after minimization, for a total of 500 ns.

Estimating free energy differences by MM-GBSA calculations. Based on root mean square deviation (RMSD) calculations, we determined a time interval with the most stable conformation of each structure by calculating RMSD values over 500 ns. For each stable conformation, a coordinate file (pdb) and a trajectory file (dcd) were saved separately for monomers and for the complex (dimer) without solvent. The MD log file results obtained with NAMD were used to retrieve the energy components used in the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) calculations[37]. The free energy of dimerization (ΔG), neglecting the entropic contribution, is estimated by the equation:

$$\Delta G = \Delta E_{int} + \Delta E_{\mathrm{ele}} + \Delta G_{\mathrm{sol}} + \Delta E_{vdW} \qquad (1)$$

$$\Delta G_{\mathrm{sol}} = \Delta G_{sol}^{PB} + \Delta G_{sol}^{SA}$$

where ΔE_{int} represents the changes in intermolecular interactions calculated using the combined change in bond, angle, dihedral and improper energies. ΔE_{ele} and ΔE_{vdW} represent the change in electrostatic and van der Waals energies, respectively. ΔG_{sol}

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is the sum of the electrostatic solvation energy (polar contribution) ΔG_{sol}^{PB} calculated via the Poisson-Boltzmann (PB) approximation, and the non-electrostatic solvation component (non-polar contribution), ΔG_{sol}^{SA} that is related to the solvent accessibility (SA) of the residues. The Generalized Born implicit solvent (GBIS), based on the Poisson Boltzmann model, calculates the polar contribution while the non-polar energy is estimated by the solvent accessible surface area (SASA). Each energy component term was first extracted separately for the single monomers and for the dimer complex from the MD log files with the pynamd script[38]. To calculate each term in the final equation, the sum of the values of the individual monomers was subtracted from the value of the complex. For each frame, the sum of all finalized components was used to calculate the ΔG of binding using equation (1). The average ΔG of binding over the number of frames was corrected by the standard error. For example, the ΔE_{vdW} term is: $\Delta E_{vdW} = \langle \Delta E_{vdW}^{complex} \rangle - [\langle \Delta E_{vdW}^{monomer1} \rangle + \langle \Delta E_{vdW}^{monomer2} \rangle]$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

PATZ1-PATZ2 is a unique BTB domain heterodimer. The BTB domain is found in about 1% (~200) of the proteins encoded in the human genome. The core secondary structures of the BTB domain are well conserved and their three-dimensional fold is strikingly similar (Fig. 1a). BTB domains are composed of around 120 amino acids, of which 35-40% make up the conserved dimer interface. The residues forming the interface are found in secondary structure elements forming β strands, α helices and loops (β1, α1, α1/B1 loop, A1, A2, A3, A3/β2 loop, β2 and A5 highlighted in Fig. 1c). The presence of β1, α1 and β2 is a specific feature of the BTB domain of ZBTB proteins defined by Stogios et al., 2005[11]. To quantify the structural similarity of BTB domains we calculated pairwise root mean square deviation (RMSD) values for eight select ZBTB proteins, whose structures were solved, or models were easily constructed (Fig. 1b). While primary sequence conservation is only evident in sub regions of the domain (Fig. 1c), structural similarity ranges between 1 and 2.5 Angstroms (Fig. 1b).

In order to study the potential dimer formation *in vitro*, we setup a system to screen dimer formation of the eight aforementioned BTB domains in a pairwise fashion. We repurposed the commercially available F2H assay (Chromotek) (Fig. 2a). For this assay, each minimal BTB domain was expressed as an N-terminal fusion to either tagGFP or tagRFP fluorescent protein in the BHK-1 cell line engineered with the insertion of a large number of LacO sequences into a genomic locus. The fusion proteins were co-expressed with a fusion protein composed of the DNA binding domain of the LacI (lac repressor) protein fused to a GFP binding nanobody (GBP). BHK-1 cells transiently expressing these three fusion proteins were visualized under fluorescent microscopy. A GFP focus was detected where the Lac I anchored the BTB-tagGFP fusion protein captured by the tagGFP specific nanobody onto the locus containing the LacO sites. Association between the tagGFP and tagRFP tagged BTB domains also formed a co-localized red fluorescent focus indicating dimer formation. Microscopic images of the F2H assay conducted with all sixty-four BTB pairs were used to generate a matrix of homo and heterodimers (Fig. 2b). Of the pairs of BTB domains analyzed, we found that all could form homodimers (shown on the diagonal of the matrix and in Fig. 2d-e), but only the PATZ1-PATZ2 pair formed a heterodimer in this assay (Fig. 2c).

To identify possible restrictions on heterodimer formation, we investigated the expression profiles of all ZBTB proteins in various cell types of the immune system using ImmGen data[16]. We particularly focused on four candidate pairs of ZBTB proteins (PATZ1-PATZ2, BCL6-PATZ1, MIZ1-BCL6 and LRF-ThPOK), which were previously reported to form heterodimers[15, 39-41].
While the expression of these ZBTB genes were positively correlated in many immune system cell types, Patz1-Bcl6 expression was negatively correlated in dendritic, mast,

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

basophil and eosinophil cells (Table 1 and Fig. S1). The negative correlation between Lrf-ThPOK expression in pooled T lymphocyte data is not evident when individual subpopulations are evaluated[16]. In this analysis, while positive correlation does not imply physical association between ZBTB proteins, it provides evidence that the physical association between PATZ1 and PATZ2 demonstrated in the F2H assay is not restricted by expression in most immune cell types (Fig. S2).

Structurally conserved BTB domains use diverse mechanisms to stabilize homodimers. To better understand the potential of BTB domain heterodimerization, we assessed structural features that contribute to dimer stability. The interaction surface for dimerization in the ZBTB family is mostly hydrophobic and involves the N and C termini of the two monomers and the central α-helices and loops. This dimerization interface contains a central charged pocket that consists of two charged residues (an absolutely conserved negatively charged aspartate (D), located at the beginning of B1, and a positively charged lysine (K) or arginine (R) at the beginning of A1, which form inter- or intra-chain ionic bonds[42]. We analyzed either crystal structures or models of five homodimers and four putative heterodimers by MD simulations to identify the relevance of these features.

For each dimer pair, we ran MD simulations of 500 ns. Our analysis of the interface interactions focused on the lifetime of salt bridges (plotted as barcode graphs), that have a strong contribution in the electrostatic component of the total ΔG of binding for homodimers (Fig. 3) and putative heterodimers (Fig. 4). Of the homodimers analyzed, we find that the PATZ1-PATZ1 pair has the highest number of interchain charged interactions (Fig. 3a). The salt bridge formed between R39 and D42 (which is in the BTB domain charged pocket) was originally observed in the crystal structure of PATZ1 (PDB entry 6GUV) but was replaced by the R39-D76 interaction upon the construction of the missing loop model. The extended MD simulation recovers the R39-D42 salt bridge. Unlike the PATZ1 homodimer, which contains dynamic salt bridges, the homodimers of BCL6, MIZ1, LRF, PATZ2 (Fig. 3b-e) have stable salt bridges forming their conserved charged pockets. Curiously, the residues of the charged pocket of BCL6 (Fig. 3b) form intrachain electrostatic interactions rather than interchain bonds in the crystal structure (PDB entry 1R29) and continue to do so over the course of the simulation.

As evolutionary conservation is correlated with structural or functional roles of amino acids, we assessed the conservation score for every residue of the BTB domain. These scores are color coded (scored from 1 to 9) in the tertiary structure of the respective BTB domains (Fig. S3) and are annotated in Fig. 3. We find that PATZ1 R39 and D42 are absolutely conserved oppositely charged residues. As the PATZ1 A2/B3 loop is a feature only observed in mammals, this region, including D76 shows low conservation (Fig. S3 and Piepoli et al., 2020[6]).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We surmised that the choice between homo and heterodimer formation may be driven by the relative stability of each alternative pair. To understand the thermodynamic basis of dimerization, we calculated an estimate of the total ΔG of binding by summing the free energy of ΔE_{int}, ΔE_{ele}, ΔG_{sol} and ΔE_{vdW}, based on MM-GBSA calculations derived from MD trajectories of homodimers (Table 2). Calculations were restricted to the equilibrated portions of the trajectory, as shown by the boxed portions of the RMSD plots in Fig. 3. As expected from the stable homodimeric structure of BTB domains, the energy features contributing to the dimerization interface for all dimers resulted in energetically favorable interactions with negative ΔG values. We find that although the stabilization energy per residue varied in the interval [-2.1, -1.6] kcal/mol, the factors contributing to this energy were from different sources for each pair of homodimers. For the intramolecular interactions in the molecules making up the dimer, for all systems analyzed the bond stretching/bending/torsions (ΔE_{int}) which make up the local terms were all negative indicating that local strains were relieved upon dimerization, more so in some systems (e.g., LRF homodimer) than in others (e.g., BCL6 or MIZ1 homodimer). In terms of nonbonded interactions, we found that the PATZ1 homodimer is overwhelmingly stabilized by the large favorable electrostatic interactions (ΔE_{ele}), especially those established at the interface as is also corroborated by the salt bridges formed (Fig. 3a).

MM-GBSA calculations show that the PATZ1 BTB domain is the most favorable homodimer with binding free energy (ΔG) equal to -529.1 kcal/mol for the equilibrated conformation, averaged between two duplicate MD runs (Table 2). BCL6 BTB homodimer is a less favorable construct than that of PATZ1 having 0.2 kcal/mol higher binding free energy per amino acid (-1.8 vs -1.6 kcal/mol). In this homodimer, the energy component deriving from local constraints in bonds, angles, and dihedrals (ΔE_{int}) is the least favorable. Due to low variation in the RMSD (Fig. 3b), for BCL6 all trajectory was considered for MM-GBSA calculations. In the case of the MIZ1 BTB homodimer, we considered the equilibrated conformation between 100 and 500 ns (Fig. 3c and Table 2). The binding free energy is favorable and equal to -384.1 kcal/mol, yet along with the previous BCL6 case, is the least favorable among the other dimers analyzed in this study (-1.6 kcal/mol/AA).

Despite the similar binding energies, the factors contributing to the overall energy are different. The energy components contributing favorably to a loss of electrostatics are the van der Waals energy (ΔE_{vdW}) and the solvation free energy deriving from the non-polar contribution (ΔG_{sol}^{SA}) making a weak dimer interface for MIZ1 BTB protein. In comparison, the binding free energy for LRF homodimer is the most favorable (-2.1 kcal/mol/AA). Unlike in PATZ1 homodimer, this strength draws not from an abundance of electrostatic interactions at the interface, but rather is due to the local release of strains in bond stretching, bending and torsional angles.

Apart from the formation of salt bridges and energetic contributions, another factor influencing the choice between homo and heterodimer may be the surface area of a

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

monomer buried by dimerization. Thus, we extracted the solvent accessible surface area (SASA) of the dimers and of the forming monomers and calculated the resultant buried surface area (BSA) from the trajectories of the five BTB homodimers (Fig. S4 and Table 2). We find that PATZ1 and PATZ2 have the largest BSA, correlating with the largest calculated free energy change of homodimerization (Table 2). The variability of the BSA values over the course of the simulation shows the stability of all the interchain contacts, including ionic, polar and non-polar interactions. We therefore conclude that while the overall folds of the BTB domains are well conserved as implicated by the low RMSD values (Fig. 1b), energetically, dimerization is not facilitated by a single mechanism. In fact, it is predominantly the extensive salt bridge formation in PATZ1, release of local strains in LRF, the relatively low energy cost of electrostatic solvation for MIZ1, and hydrophobicity for PATZ2. For BCL6, it is a combination and compensation of all these factors that achieve the final homodimer.

Heterodimerization results from additional interactions. To determine the relative stability of heterodimers, compared to homodimers, we constructed models of four putative BTB heterodimers originating from monomer structures and ran MD simulations, performed MM-GBSA calculations and identified the BSA values. As before, for each heterodimer pair, MD simulations (500 ns) yielded information about interface interactions based on the lifetime of salt bridges (Fig. 4).

While the PATZ1 interface has the largest number of salt bridges amongst the homodimers, the PATZ1-PATZ2 and BCL6-PATZ1 heterodimer interfaces established additional salt bridges (Fig. 4a-b). The PATZ1-PATZ2 dimer interface has a significant interaction between residues E60b and R39a, both well conserved residues, which is present for 60% of the trajectory. A second salt bridge formed between D38b and R56a, also a well conserved pair, is present for over 58% of the trajectory. Interestingly, while all other PATZ1 residues involved in the salt bridges between PATZ1-PATZ2 heterodimers also make similar interactions in the PATZ1 homodimer, PATZ1 residue R56 only makes salt bridges with PATZ2 (reconstituting an interchain charged pocket interaction). We find that in general the residues forming the charged pocket in the heterodimer models form interchain salt bridges, with the exception of the BCL6-PATZ1 pair, which retains stable intrachain salt bridges.

BTB domain N-terminal interactions have recently been proposed to mediate dimer stability[43]. N-terminal β-strand spontaneous complex dissociation could thus differentiate homodimers from heterodimers. While we observe the presence of stable N-terminal β-strand interactions in many of the homodimer structures, this feature is present only in the PATZ1-PATZ2 heterodimer model (Fig. 4a). Noticeably, the two symmetrical β-sheets formed at the dimerization interface between β1 and β2 of the two monomers are stable throughout the PATZ1-PATZ2 simulation, and do not show any sign of spontaneous unfolding. In contrast, in the BCL6-PATZ1 BTB heterodimer

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Fig. 4b) the N-terminal β1 strand of BCL6 (chain a) disengages from the β-sheet with β2 of PATZ1, leading to a partial unfolding of the dimer interface.

Also the MIZ1-BCL6 heterodimer has an unstable interface because, while the BCL6 homodimer interface relies on a sheet formed by the interaction of β1 and β2, MIZ1 lacks a complete β1-strand (Fig. 4c). MD trajectories reveal the accommodation of a new stable conformation for the short N-terminus of MIZ1 that swings from the initial docked position parallel to β2 in BCL6 to a new interaction with the N-terminal of BCL6. Explicitly, we can follow this conformational change by tracking the salt bridges formed by D2a initially interacting with R94b and then settling for R13b. Significantly, the spontaneous unfolding of one of the primary dimer interface β-sheets, may represent a target for dimer quality control mechanisms[43]. As for the MIZ1-BCL6 heterodimer, besides this local flexibility, the ionic interaction between the highly conserved charged pocket residues (D33b-K39a) is preserved and remains important in the trajectory of all dimers (Fig. 4c). Similarly, we can see the highly conserved charged pocket residues D35 and R49 of LRF-ThPOK BTB heterodimer (Fig. 4d) forming two strong symmetric salt bridges both present for at least 70% of the whole trajectory.

The A2/B3 loop of PATZ1 contributes to the large BSA of the PATZ1-PATZ2 and BCL6-PATZ1 heterodimers, which have an average area of 2348 and 2054 Å², respectively (Fig. S5). Unlike the first two cases, the MIZ1-BCL6 BTB heterodimer interface area is small, equal on average to 1548 Å² with the lowest percentage of the total residue count involved in the interface (Table 3). This is due to an asymmetric dimer interface between the two monomers. The fluctuations in the RMSD (Fig. 4c) reflect the adjustments related to the shorter β1 sequence of MIZ1-BTB.

As for the homodimers, all heterodimers show favorable interaction energy (Table 3). PATZ1-PATZ2 is the strongest heterodimer among the ones analyzed with binding free energy (ΔG) equal to -529.5 kcal/mol. BCL6-PATZ1 heterodimer is also a favorable construct with binding free energy equal to -470.9 kcal/mol. MIZ1-BCL6 heterodimer is the least favorable of the heterodimers considered in this study with binding free energy equal to -1.6 kcal/mol/AA. LRF-ThPOK heterodimer is a favorable construct with binding free energy of -1.8 kcal/mol/AA on the order similar to that of the PATZ1 homodimer. At the outset, a heterodimer is expected to form if the energy gain is lower than that expected from its homodimers. For example, for the PATZ1-PATZ2 heterodimer, an expected energy is the average from their homodimers, i.e., ca. -507 kcal/mol. Similarly, the BCL6-PATZ1 average energy is -465 kcal/mol and the MIZ1-BCL6 interaction is -393 kcal/mol. We find that, the ΔG for the PATZ1-PATZ2 heterodimer is -529.5 kcal/mol, i.e., it is ~22 kcal/mol lower than the average energy expected from the homodimers (Table 3). This is in contrast with the observations for BCL6-PATZ1 and MIZ1-BCL6 heterodimers whereby the expected and measured ΔG values are within ~5-6 kcal/mol of each other; i.e., there is no substantial need to prefer

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

heterodimers over homodimers for these pairs. A closer inspection of the various contributors to the final energies puts homo- vs. heterodimerization into perspective.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

This study documents that BTB domains can heterodimerize. We evaluated the dimerization potential of 64 pairs of BTB domains and find that while all pairs can generate homodimers, only one, PATZ1 (ZBTB19) and PATZ2 (ZBTB24) can form heterodimers *in vivo*. Energetic calculations confirmed that this heterodimer could form a favorable interaction interface, predominantly due to additional stable salt bridges. Despite the similar name, PATZ1 and PATZ2 only show 26.5% identity and 42.4% similarity in their BTB domain sequence. These two ZBTB family members are structurally related, being the only proteins in the ZBTB family that have an additional AT-hook motif (binding the minor groove of Adenine-Thymine rich DNA), that is thought to confer an alternative DNA binding specificity to these proteins. In our assays we used the minimal BTB domain consisting of 157 amino acids for PATZ1 and 133 for PATZ2, lacking the AT-hook motif. This demonstrates that the AT-hook is not necessary for heterodimer formation and that BTB domains are sufficient to form heterodimeric structures. These findings reveal that the PATZ1-PATZ2 heterodimer is as stable as the PATZ1 or PATZ2 homodimers in the cellular environment, a finding that is supported by the calculated binding free energy of these complexes. Electrostatic interactions in proteins are fine-tuned by the various niches in the cellular environment, with differences of pH or ionic strength[44]. The dominance of the electrostatic component in the PATZ1-PATZ2 heterodimer might confer its ubiquity in the different cell types where they are co-expressed (Table 1).

The demonstration of definitive heterodimer formation between PATZ1 and PATZ2 now will allow the questioning of the participation of each protein in the phenotypes observed in the mutation or knockout of the other factor. For example, mutations in the Zbtb24 gene result in the methylation defects observed in the immunodeficiency, centromeric instability and facial defect syndrome type 2 (ICF2)[45-47]. Does PATZ1 participate in this defect? How many of the previously identified 187 differentially expressed genes (DEG) in Patz1-/- cells[48] are controlled by PATZ1 in collaboration with PATZ2 is an open question.

We investigated the underlying structural factors behind BTB domain dimerization to understand the basis of homo vs. heterodimer choice. A functional consequence of homodimer formation in various ZBTB proteins is the formation of a lateral groove that is a docking site for co-repressor proteins[49]. While other BTB domains have been shown to interact with co-repressors, the only available co-crystal structure is that of BCL6 and its co-repressors[7, 50-52]. In these structures, the co-repressor peptides associate with the BTB homodimer as symmetrical pairs themselves. The interaction of BCL6 homodimers with co-repressor peptides has been studied in detail using molecular dynamics supported by MM-GBSA calculations, revealing potential sites that can be targeted by drugs[53]. With the definitive demonstration of the presence of heterodimers, we open the question of whether heterodimers can also form the landing pad structures for these co-repressors. If so, could the non-symmetrical lateral

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

grooves of BTB heterodimers provide a mechanism of altered specificity for co-repressors? Besides the BTB domain lateral groove interactions assisted by lower β-sheet extensions, exemplified by the BCOR/NCOR1/NCOR2 interactions with BCL6, a novel interaction site on BTB domains was recently revealed[54]. The interaction of a β-strand containing peptide from HUWE1 with the flexible B3 region of MIZ1 can result in an upper β-sheet extension. Whether these interactions can form in other BTB pairs is not known. An obvious candidate for such an interaction would be the flexible top region containing BTB domains such as PATZ1 and PATZ1 containing (heterodimeric) complexes.

Formation of BTB heterodimers would dramatically increase the combinatorial target specificity of this transcription factor family. Obviously, such heterodimer formation would be restricted by the tissue and stage specific expression of the individual proteins. Mechanistic constraints in the synthesis of these proteins, such as the recently reported co-translational dimerization pathways[55], may impart restrictions on the formation of heterodimers, possibly favoring the formation of homodimers co-translated on polysomes. However, the combinatorial specificity may not be regulated only at the level of the formation of homo- or heterodimers, but in the cellular half-life of these alternative protein structures. A recent study proposed the presence of evolutionarily conserved degron residues which preferentially targets BTB heterodimers for degradation[56]. Although this study examined the degradation properties of non-transcription factor BTB domain containing proteins, degron structures may likely be conserved in ZBTB proteins as well, making unwanted BTB heterodimers prone to degradation. Furthermore, according to the BTB quality control hypothesis[43], heterodimers can be targeted for degradation based on the identity of the N-terminal β1 sequence that forms a critical interface surface. In fact, we identified an N-terminal sequence in the PATZ1 crystal structure that preferentially stabilizes homodimeric structures[6]. The propensity of this region to result in aggregation that potentially targets BTB domains for degradation has also been observed in the BCL6 protein crystal structure, which can be used as a means for co-crystallization[50].

The F2H assay we introduce in this study is built on a previous iteration that tested the interaction between the minimal interaction domains of the p53 and MDM2/MDM4 proteins[19]. This system can be used as a high-throughput screening tool to test for drugs that block interaction[57]. In its current version, this assay can be used to not only discover new heterodimers and their third-party interactors, but also inhibitors of dimers. As BTB domains form obligate homodimers, it is surprising that heterodimers can in fact be observed in this assay. Because the system is set up with one monomer (tagGFP partner) with a nuclear localization signal (NLS) and a second monomer (tagRFP partner) without any such signal, we find that the interaction between BTB monomers is strong enough to recruit BTB domains with no NLS into the nucleus. Significantly, the PATZ1-PATZ2 interaction that scores positive with a GFP-RFP pair

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also does so with an RFP-GFP pair, indicating the robustness of the system to recapitulate *in vivo* interactions (Fig. 2).

In this study, we determined the driving forces that contribute to dimer stability. We find by MD simulations that all heterodimers are favorable. Different mechanisms contribute to homo and heterodimer stability. Significantly, homo and heterodimer interfaces are typically characterized by numerous and sometimes short-lived electrostatic interactions. Thus, evolution has favored conserving the fold which serves as a template for catering to the overall functions attributed to these systems while diverse mechanisms have been utilized to compensate for the variations observed in family members (siblings) introduced to enable those functions. The analysis of the energy components contributing to dimerization also paves the way to design stable BTB heterodimers particularly by engineering interface residues and limiting accessibility to degron positions. Our analysis confirms that heterodimerization among ZBTB family members is infrequent and that homodimers are preferred. Nevertheless, the absence of energetic restrictions for BTB domain-mediated heterodimers suggest that more pairs of heterodimers could possibly form, increasing transcription factor combinatorial specificity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Captions

Figure 1. Structural conservation in the BTB domain. A cartoon representation of the BTB domain **(a)**, with annotated secondary structural elements between N- and C-termini, is colored based on a metric for structural alignment (Q-score) ranging from blue to red to show the most and the least conserved regions, respectively. The nine overlapped structures belong to the BTB monomers of BCL6, KAISO, PLZF, ThPOK, LRF, PATZ1, PATZ2, MIZ1 and 90K human proteins. The structural alignment is measured in terms of RMSD (Å) of the C_α atoms for each pair of BTB domain structures **(b)**. The RMSD among this set of BTB structures is under 2 Å except for the two cases of PATZ1-PATZ2 and PATZ1-MIZ1. The secondary structure labeling follows the convention for the BTB fold as used in Stogios et al., 2005[11]. The structure and sequence of the human BTB-containing protein 90K (PDB entry 6GFB)[58] is only used here as a divergent example to underline the similarity of the BTB domain in ZBTB proteins. In the corresponding sequence alignment **(c)**, the residues forming the BTB homodimer interface are highlighted. The residues in the BTB characteristic charged pocket are indicated with a box. The three absolutely conserved positions are indicated with an asterisk (*). The secondary structures are annotated on the sequences for orientation with part **(a)**. The unlabeled β-strand between A2 and B3 indicates an additional secondary structure revealed in the model of PATZ1[6].

Figure 2. F2H assay is readapted for the screening of the BTB homo and heterodimer formation *in vitro*. Schematic description of the experimental setup **(a)**. The co-transfected plasmids of the recombinant sequences of BTB domains tagged with green or red fluorescent proteins (GFP or RFP) and the GFP-binding nanobody (GBP) fused to Lac I sequence are represented as white circles before the corresponding outline of the expressed fused proteins. Below, the nuclear interaction site on the DNA of BHK-1 cells, scaffolds the co-localization experiment. In matrix representation **(b)**, the summary of the different dimer combinations tested. For each experimental pair, the colocalization signal is either not detected (ND) or detected in the reported percentage of the total number of cells analyzed. The only heterodimer identified with this assay is between PATZ1 and PATZ2 BTB domains. **(c)** Two examples of the cells in which co-localization of the fluorescent signals from GFP-tagged PATZ1 and RFP-tagged PATZ2 BTB domains was detected. Different microscope filters detect the two different signals captured in separate images, later overlapped for the merged section. **(d)** Representative fluorescent microscopy images of colocalized tagGFP or tagRFP fusion BTB domains. Only the positive scored interactions from part **(a)** are shown. Three channel images displayed GFP (top row), RFP (middle row) fluorescence and brightfield (bottom row). **(e)** Quantification of the colocalization assay. The bar graph shows the percentage of GFP focus positive cells that also displayed an RFP focus (positive) or not (negative). Numbers inside the bar graphs indicate the total number of cells analyzed for each case. Colors refer to part **(a)** where each column displays data from cells transfected with GFP and RFP tagged versions of the indicated BTB

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

domains. The only heterodimers that interact were GFP tagged PATZ1-BTB (GP1) with RFP tagged PATZ2-BTB (RP2) and GFP tagged PATZ2-BTB (GP2) with RFP tagged PATZ1-BTB (RP1).

Figure 3. MD simulation analyses for the BTB domain homodimers. RMSD plots, salt bridge formation barcodes, and a cartoon representation of the BTB domain dimer structure are shown for, PATZ1 **(a)**, BCL6 **(b)**, MIZ1 **(c)**, LRF **(d)** and PATZ2 **(e)**. The RMSD plot shows the structural distance (Å) of the protein atoms coordinates (C_α) as a function of time (ns) and contains the snapshots of the significant conformational changes of the dimer structure. Every salt bridge between a pair of charged amino acids with a distance within the 3.0 Å cut-off, is represented with a bar in the barcode plot and reported if present over the 8% of the total simulation time. The amino acids belonging to one monomer (a) or the other (b) involved in the interchain interactions are labelled with one-letter-code. For each residue in these interchain salt bridges, the conservation score is displayed next to its label in the range [1,9] increasing from variable (1) to conserved (9) as calculated via the ConSurf webserver.

Figure 4. MD simulation analyses for the BTB domain heterodimers. RMSD plots, salt bridge formation barcodes, and a cartoon representation of the BTB domain dimer structure are shown for PATZ1-PATZ2 **(a)**, BCL6-PATZ1 **(b)**, MIZ1-BCL6 **(c)** and LRF-ThPOK **(d)**. See caption to Fig. 3 for details.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Expression correlation of four pairs of ZBTB genes*

	B Cells	T Cells	Monocytes	Stem Cells	Stromal Cells	Innate Lymphocytes	Macrophages	Dendritic Cells	MBE	Granulocytes
PATZ1-PATZ2	**0.433**	**0.229**	0.099	**0.354**	**0.627**	0.245	**0.446**	**0.331**	**0.507**	**0.581**
PATZ1-BCL6	0.077	**0.248**	-0.086	-0.024	**0.683**	0.157	0.097	**-0.361**	**-0.573**	**0.637**
BCL6-MIZ1	-0.194	**0.356**	-0.122	0.176	0.197	**0.603**	**0.336**	0.046	-0.026	**0.598**
LRF-ThPOK	**0.675**	**-0.185**	0.036	0.334	**0.370**	**0.455**	0.132	-0.080	**0.479**	**0.499**

*R-values with *p*-value ≤ 0.05 (significant) are shown in bold. The grouped Mast, Basophil and Eosinophil is shortened as "MBE". Additional information on these values is in Fig. S1.

Table 2. Binding energies for BTB homodimers and individual contributions to the total energy*

	MM-GBSA			ΔG_{sol}						
Homodimers	ΔG binding kcal/mol	ΔE_{int}	ΔE_{ele}	ΔG_{sol}^{PB}	ΔG_{sol}^{SA}	ΔE_{vdW}	#AA dimer	ΔG/AA	avg BSA (Å²)	AA in interface (%)
PATZ1 (300-500ns)	**-529.1 ±0.7**	-353.9 ± 0.4	**-418.0 ± 3.5**	437.1 ± 3.2	0 ± 0.1	-194.3 ± 0.5	290	-1.8	1864 ±14	40.7
BCL6 (1-500ns)	-401.6 ±0.2	-237.5 ± 0.1	-278.7 ± 0.9	330.6 ± 0.8	-25.4 ± 0.1	-190.5 ± 0.1	250	-1.6	1899 ± 6	34.8
MIZ1 (100-500ns)	-384.1 ±0.2	-282.8 ± 0.1	-170.0 ± 1.0	**235.9 ± 1.0**	-18.8 ± 0.1	-148.4 ± 0.2	234	-1.6	1471 ± 4	37.2
LRF (280-500ns)	-520.3 ±0.3	**-380.0 ± 0.2**	-322.8 ± 1.5	399.6 ± 1.4	-22.3 ± 0.1	-194.7 ± 0.2	244	**-2.1**	1752 ± 4	36.9
PATZ2 (100-500ns)	-485.4 ±0.7	-335.5 ± 0.3	-247.8 ± 3.9	254.5 ± 3.8	**-43.3 ± 0.1**	**-200.0 ± 0.6**	258	-1.9	2037 ± 6	38.0

*Calculations are carried out for the equilibrated portion of the trajectory indicated in parenthesis and shown in Fig. 3. Most favorable energy values indicated in bold.

Table 3. Binding energies for BTB heterodimers and individual contributions to the total energy*

	MM-GBSA				ΔG_{sol}						
Heterodimers	ΔG binding kcal/mol (Expected)	ΔG binding kcal/mol	ΔE_{int}	ΔE_{ele}	ΔG_{sol}^{PB}	ΔG_{sol}^{SA}	ΔE_{vdW}	#AA dimer	ΔG/AA	avg BSA (Å²)	AA in interface (%)
PATZ1-PATZ2 (350-500ns)	-507.3 ±1.4	**-529.5 ±0.4**	-340.0 ± 0.2	**-471.0 ± 2.8**	535.4 ± 2.7	**-30.4 ± 0.1**	**-223.4 ± 0.3**	274	**-1.9**	2348 ± 5	40.1
BCL6-PATZ1 (60-500ns)	-465.4 ±0.9	-470.9 ±0.2	-306.6 ± 0.1	-461.6 ± 1.8	531.0 ± 1.7	-27.3 ± 0.1	-206.4 ± 0.2	270	-1.7	2054 ± 6	37.8
MIZ1-BCL6 (180-500ns)	-392.9 ±0.4	-397.6 ±0.2	-272.3 ± 0.1	-152.8 ± 0.8	**202.7 ± 0.7**	-19.6 ± 0.1	-155.7 ± 0.2	242	-1.6	1548 ± 4	33.1
LRF-ThPOK (350-500ns)		-471.3 ±0.3	**-348.6 ± 0.2**	-249.3 ± 1.5	316.4 ± 1.4	-21.0 ± 0.1	-168.8 ± 0.2	259	-1.8	1706 ± 4	34.7

*Calculations are carried out for the equilibrated portion of the trajectory indicated in parenthesis and shown in Fig. 4. Most favorable energy values indicated in bold.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Material

Figure S1. Statistical significance for values in Table 1. The R-value indicates positive (blue) or negative (red) correlation in the expression profiles of the genes pair. The statistical significance of the R-value is determined by a p-value < 0.05 that is otherwise highlighted in grey. The -log10 of the p-value gives a measure in integer numbers of the p-value significance with a higher number indicating higher significance.

Figure S2. Co-expression patterns of ZBTB proteins based on linear least-squares regression correlation coefficients. Blue indicates a significant positive correlation, red indicates a significant negative correlation and grey indicates non-significant (p-value > 0.05) correlation. Proteins are clustered based on the correlation coefficients. Four pairs of ZBTB proteins selected for the heterodimers analysis in this study (PATZ1-PATZ2, BCL6-PATZ1, MIZ1-BCL6 and LRF-ThPOK) are highlighted in green.

Figure S3. Ribbon representation of selected BTB homodimer structures. The conservation was calculated by ConSurf and color coded as shown in the scale from blue (variable) to purple (conserved). Secondary structure features are labelled on PATZ1 BTB dimer. Exposed protein surfaces are less conserved than buried and dimer interface regions. PATZ2 and ThPOK dimers are based on novel modelled structured.

Figure S4. Dimerization interface in the BTB homodimers of PATZ1 (a; two independent simulations), BCL6 (b), MIZ1 (c), LRF (d) and PATZ2 (e). For each protein, SASA values ($Å^2$) are calculated separately for the dimer (red) and the single monomers (black and grey in the graph) through the simulation. In green, the BSA ($Å^2$) obtained by subtracting the sum of the SASA of the two monomers from the SASA of the complex dimer. Below each figure, the minimum, maximum and average values of BSA ($Å^2$) are reported.

Figure S5. Dimerization interface in the BTB heterodimers of PATZ1-PATZ2 (a), BCL6-PATZ1 (b), BCL6-MIZ1 (c) and LRF-ThPOK (d). See caption to Fig. S4 for details.

Figure S6. Phylogenetic Tree of the ZBTB family. The BTB domain sequences of all ZBTB family proteins were used in Blast search and the phylogenetic tree was constructed from the MSA of the top 10 Blast hits.

Table S1. List of cloning primers

OLIGONUCLEOTIDE NAME	SEQUENCE	PURPOSE
BCL6-BTB-XhoI-Forward	TGGACTCGAGGGATGGCCTCGCCGGCT GACAG	F2H cloning

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BCL6-BTB-NotI-Reverse	GGACGCGGCCGCTTATTCACTGGCCTTAATAAACTTCCGGCAAG	F2H cloning
KAISO-BTB-XhoI-Forward	TGGACTCGAGGGATGGAGAGTAGAAAACTGATTTCTGC	F2H cloning
KAISO-BTB-NotI-Reverse	GGACGCGGCCGCTTACTGTGACAATGGGACACCAA	F2H cloning
LRF-BTB-XhoI-Forward	TGGACTCGAGGGATGGCCGGCGGCGTGGA	F2H cloning
LRF-BTB-NotI-Reverse	GGACGCGGCCGCTTAGATCTGCCGGTCCAGGAGGTCG	F2H cloning
MIZ1-BTB- XhoI-Forward	TGGACTCGAGGGATGGACTTTCCCCAGCACAGCCAGC	F2H cloning
MIZ1-BTB-NotI-Reverse	GGACGCGGCCGCTTAAGCAAGTGACTTGAGGGCATGGCAG	F2H cloning
PATZ2-BTB-XhoI-Forward	TGGACTCGAGGGATGGCAGAAACATCGCCAGAG	F2H cloning
PATZ2-BTB-NotI-Reverse	GGACGCGGCCGCTTAGCTATGATTATTTTGGAAGTCTGTGTAAGC	F2H cloning
PLZF-BTB-XhoI-Forward	TGGACTCGAGGGATGGATCTGACAAAAATGGGCATGA	F2H cloning
PLZF-BTB-NotI-Reverse	GGACGCGGCCGCTTACTGGATGGTCTCCAGCATCTTCAG	F2H cloning
PATZ1-BTB-XhoI-Forward	TGGACTCGAGGGATGGAGCGGGTCAACGACGCTTC	F2H cloning
PATZ1-BTB-NotI-Reverse	GGACGCGGCCGCTTAGGACTGTTTGATTACTTCCTGGCAGATC	F2H cloning
ThPOK-BTB-SmaI-Forward	TGGACCCGGGATGGGGAGCCCCGAGGATGAC	F2H cloning
ThPOK-BTB-XhoI-Forward	TGGACTCGAGGGATGGGGAGCCCCGAGGATGACCTGATT	F2H cloning
ThPOK-BTB-NotI-Reverse	GGACGCGGCCGCTTATTAACTGCCCTGCAGAATCTCCATGCAAGCA	F2H cloning
CMV Forward	CGCAAATGGGCGGTAGGCGTG	F2H cloning
TagGFP-XhoI&SmaI-Reverse	GGACCTCGAGGGACCCCGGGAGAACCGCTGTACAGCTCGTCCATGCC	F2H cloning
LacI-GBP-Forward	TCAGCTAGCATGGCCGATGTGCAGCTGGT	F2H cloning
LacI-GBP-Reverse	ATTGGATCCTCATCGGGAAACCTGTCGTGC	F2H cloning

Acknowledgements

We thank Professor Dr Rémy Bosselut for helpful comments during the preparation of the manuscript.
This work was supported by TUBITAK grant number [118Z015 and 20AG007].

Footnotes

Author contributions were as follows: SP, CA and BE designed the study; LN, SB, HT, NT and MG performed and analyzed F2H assay; SP and CA performed and analyzed MD simulations; UA and OA Performed Bioinformatic analysis; SP, CA, OA and BE wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The BTB dimer models generated in this study are available in ModelArchive (modelarchive.org) with the accession codes ma-olypj (PATZ1 homodimer), ma-1iskk (PATZ2 homodimer), ma-zhxm1 (ThPOK homodimer), ma-hf06e (PATZ1-PATZ2 heterodimer), ma-ql2m8 (BCL6-PATZ1 heterodimer), ma-wrsln (LRF-ThPOK heterodimer).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Siggs OM and Beutler B. The BTB-ZF transcription factors. *Cell Cycle* 2012; 11: 3358-3369. DOI: 10.4161/cc.21277.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Ahmad KF, Melnick A, Lax S, et al. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell* 2003; 12: 1551-1564.
8. Cooper CDOM, J.W.; Bullock, A.; Pike, A.C.W.; von Delft, F.;, Filippakopoulos PS, E.; Edwards, A.; Arrowsmith, C.H.; Bountra, C.; and Weigelt JK, S. Crystal structure of the BTB domain of human myoneurin. 2008. DOI: 10.2210/pdb2VPK/pdb.
9. Stogios PJC, L.; Prive, G.G. Crystal structure of the BTB domain from Kaiso/ZBTB33, form I. 2010. DOI: 10.2210/pdb3M4T/pdb.
10. Filippakopoulos PB, A.; Cooper, C.; Keates, T.; Salah, E.; Pilka, E.;, Pike ACWvD, F.; Arrowsmith, C.H.; Edwards, A.M.; Weigelt, J.; and Knapp SSGCS. Crystal structure of the human BTB domain of the Krueppel related Zinc Finger Protein 3 (HKR3). 2007. DOI: 10.2210/pdb3B84/pdb.
11. 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.
12. 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.
13. Oughtred R, Rust J, Chang C, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* 2021; 30: 187-200. 2020/10/19. DOI: 10.1002/pro.3978.
14. 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.
15. 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.
16. Heng TS and Painter MW. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 2008; 9: 1091-1094. 2008/09/19. DOI: 10.1038/ni1008-1091.
17. Rosbrook GO, Stead MA, Carr SB, et al. The structure of the Bach2 POZ-domain dimer reveals an intersubunit disulfide bond. *Acta Crystallogr D Biol Crystallogr* 2012; 68: 26-34. 2011/12/24. DOI: 10.1107/s0907444911048335.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Rothbauer U, Zolghadr K, Tillib S, et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* 2006; 3: 887-889. 2006/10/25. DOI: 10.1038/nmeth953.
19. 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.
20. Zolghadr K, Rothbauer U and Leonhardt H. The fluorescent two-hybrid (F2H) assay for direct analysis of protein-protein interactions in living cells. *Methods Mol Biol* 2012; 812: 275-282. 2012/01/06. DOI: 10.1007/978-1-61779-455-1_16.
21. Bolte S and Cordelières FP. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* 2006; 224: 213-232. 2007/01/11. DOI: 10.1111/j.1365-2818.2006.01706.x.
22. Schneider CA, Rasband WS and Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012; 9: 671-675. 2012/08/30. DOI: 10.1038/nmeth.2089.
23. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; 9: 676-682. 2012/06/30. DOI: 10.1038/nmeth.2019.
24. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25: 3389-3402. 1997/09/01. DOI: 10.1093/nar/25.17.3389.
25. 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.
26. Katoh K, Misawa K, Kuma K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; 30: 3059-3066. 2002/07/24. DOI: 10.1093/nar/gkf436.
27. Price MN, Dehal PS and Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; 5: e9490. 2010/03/13. DOI: 10.1371/journal.pone.0009490.
28. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 2016; 44: W344-350. 2016/05/12. DOI: 10.1093/nar/gkw408.
29. Humphrey W, Dalke A and Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996; 14: 33-38, 27-38.
30. Fiser A and Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 2003; 19: 2500-2501. 2003/12/12.
31. 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.
32. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018; 46: W296-W303. 2018/05/23. DOI: 10.1093/nar/gky427.
33. 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.
34. Best RB, Zhu X, Shim J, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

χ(1) and χ(2) dihedral angles. *J Chem Theory Comput* 2012; 8: 3257-3273. 2013/01/24. DOI: 10.1021/ct300400x.

35. 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.

36. 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.

37. Hou T, Wang J, Li Y, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 2011; 51: 69-82. 2010/12/02. DOI: 10.1021/ci100275a.

38. Radak B. pynamd. GitHub2021.

39. 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.

40. 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.

41. Vacchio MS, Wang L, Bouladoux N, et al. A ThPOK-LRF transcriptional node maintains the integrity and effector potential of post-thymic CD4+ T cells. *Nat Immunol* 2014; 15: 947-956. 2014/08/19. DOI: 10.1038/ni.2960.

42. 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.

43. 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.

44. Sensoy O, Atilgan AR and Atilgan C. FbpA iron storage and release are governed by periplasmic microenvironments. *Phys Chem Chem Phys* 2017; 19: 6064-6075. 2017/02/14. DOI: 10.1039/c6cp06961d.

45. 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.

46. Wu H, Thijssen PE, de Klerk E, et al. Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. *Hum Mol Genet* 2016; 25: 4041-4051. 2016/07/29. DOI: 10.1093/hmg/ddw243.

47. Thompson JJ, Kaur R, Sosa CP, et al. ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res* 2018; 46: 10034-10051. 2018/08/08. DOI: 10.1093/nar/gky682.

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. 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.

50. Zacharchenko T and Wright S. Functionalization of the BCL6 BTB domain into a noncovalent crystallization chaperone. *IUCrJ* 2021; 8: 154-160. 2021/03/13. DOI: 10.1107/s2052252520015754.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

51. 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.

52. 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.

53. 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.

54. Orth B, Sander B, Möglich A, et al. Identification of an atypical interaction site in the BTB domain of the MYC-interacting zinc-finger protein 1. *Structure* 2021 2021/06/30. DOI: 10.1016/j.str.2021.06.005.

55. Bertolini M, Fenzl K, Kats I, et al. Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science* 2021; 371: 57-64. 2021/01/02. DOI: 10.1126/science.abc7151.

56. 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.

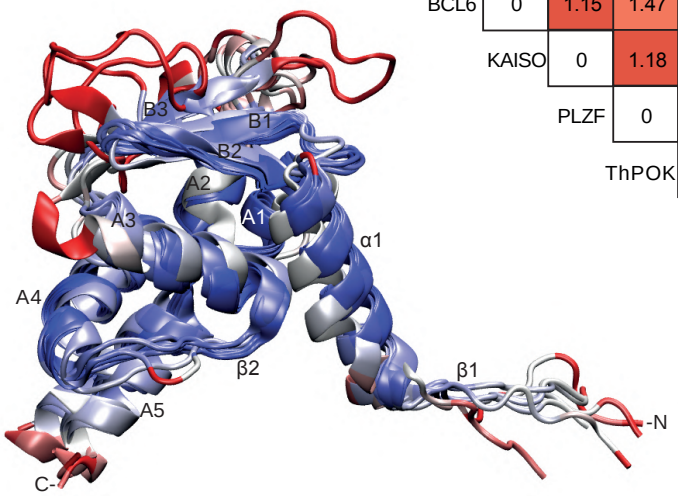
57. Yurlova L, Derks M, Buchfellner A, et al. The fluorescent two-hybrid assay to screen for protein-protein interaction inhibitors in live cells: targeting the interaction of p53 with Mdm2 and Mdm4. *J Biomol Screen* 2014; 19: 516-525. 2014/01/31. DOI: 10.1177/1087057113518067.

58. Lodermeyer V, Ssebyatika G, Passos V, et al. The Antiviral Activity of the Cellular Glycoprotein LGALS3BP/90K Is Species Specific. *J Virol* 2018; 92 2018/05/11. DOI: 10.1128/jvi.00226-18.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1

(a)



(b)

RMSD (Å)	BCL6	KAISO	PLZF	ThPOK	LRF	PATZ1	PATZ2	MIZ1	90K
BCL6	0	1.15	1.47	1.32	1.53	1.77	1.79	1.86	2.56
KAISO		0	1.18	1.15	1.25	1.59	1.64	1.88	2.59
PLZF			0	1.02	1.13	1.55	1.93	1.71	2.72
ThPOK				0	1.03	1.63	1.84	1.99	2.69
LRF					0	1.67	1.94	1.97	2.76
PATZ1						0	2.46	2.32	2.5
PATZ2							0	1.78	2.98
MIZ1								0	3.16
90K									0

(c)

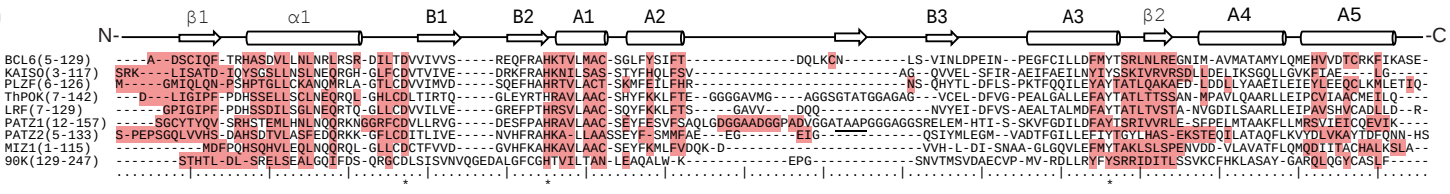


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

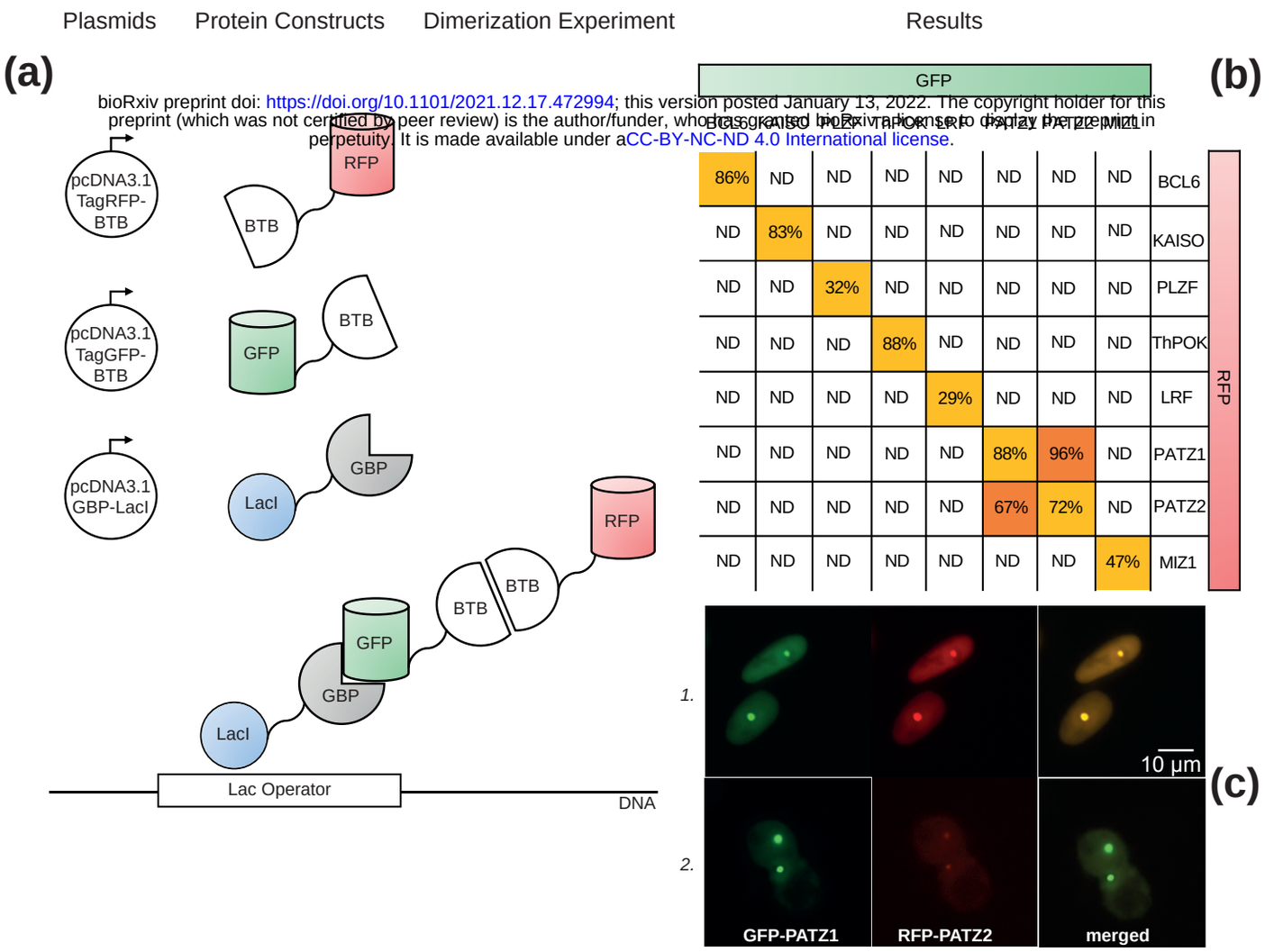
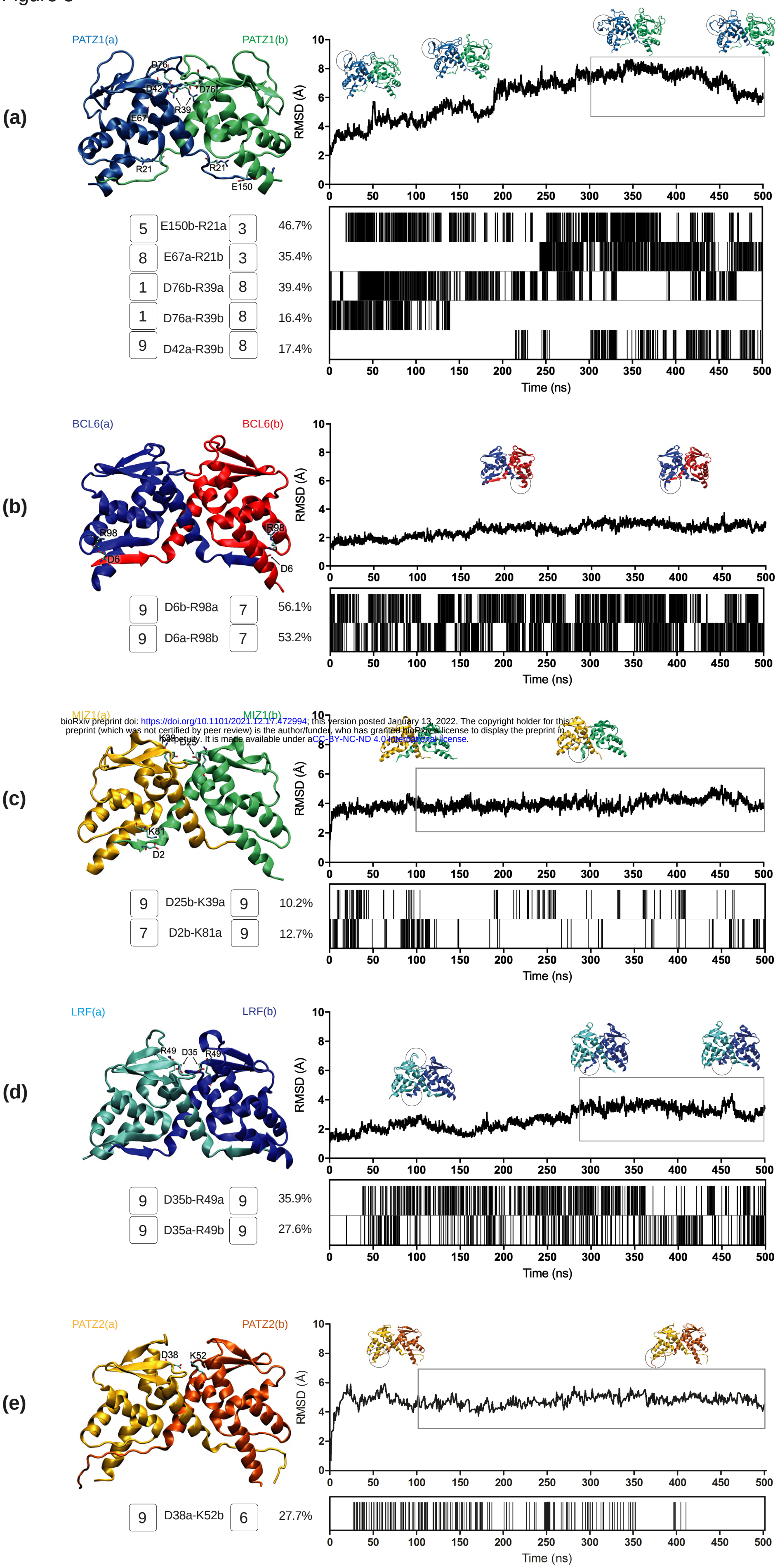
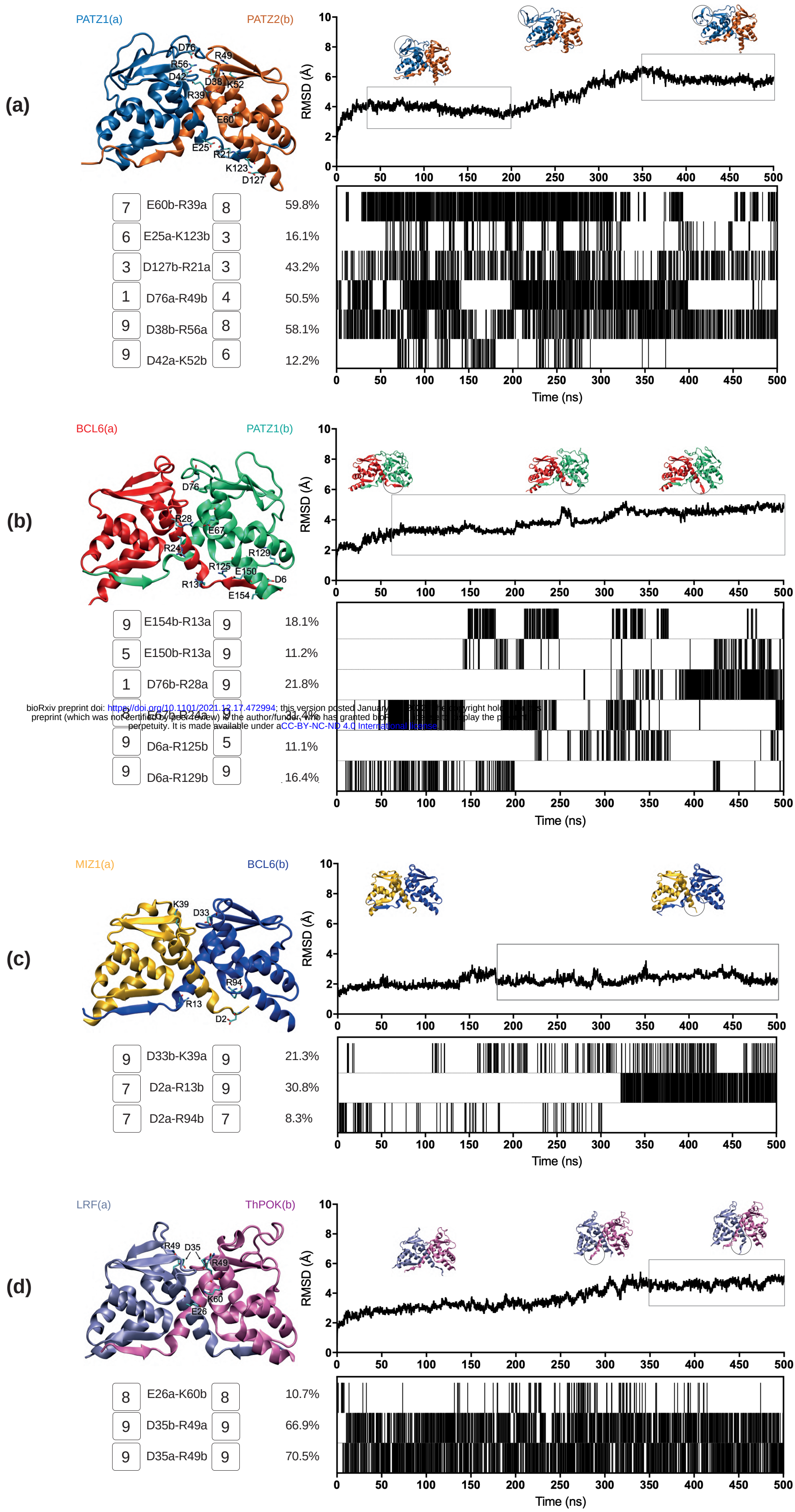


Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January [illegible] copyright hold[illegible] preprint (which was not certified by peer review) is the author/funder, who has granted bioR[illegible] display the p[illegible] perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S1

Patz1-Patz2			
Cell Type	R-value	P-value (0.05)	-log10(p)
B Cells	0.433	0.0000737	4.133
T Cells	0.229	0.0000231	4.636
Monocytes	0.099	0.5847	0.233
Stem Cells	0.354	0.0368	1.434
Stromal Cells	0.627	0.000000000802	9.096
Innate Lymphocytes	0.245	0.0569	1.245
Macrophages	0.446	0.0000268	4.572
Dendritic Cells	0.331	0.0016	2.783
Mast, Basophils and Eosinophil	0.507	0.0043	2.368
Granulocytes	0.581	0.0036	2.438

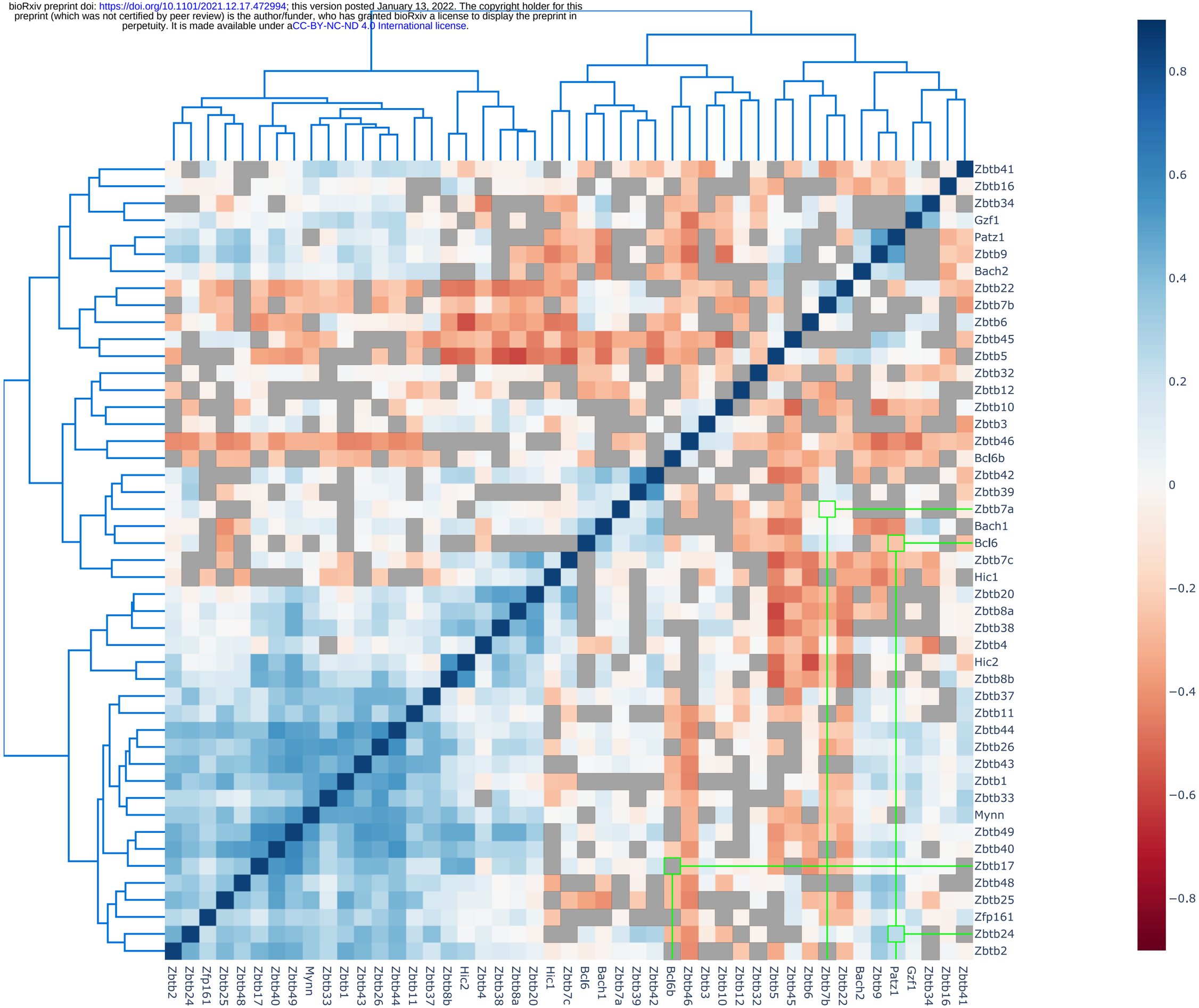
Patz1-Bcl6			
Cell Type			
B Cells	0.077	0.5042	0.297
T Cells	0.248	0.00000435	5.361
Monocytes	-0.086	0.6351	0.197
Stem Cells	-0.024	0.8902	0.051
Stromal Cells	0.683	0.00000000000552	11.258
Innate Lymphocytes	0.157	0.2259	0.646
Macrophages	0.097	0.3844	0.415
Dendritic Cells	-0.361	0.0005	3.263
Mast, Basophils and Eosinophil	-0.573	0.0009	3.027
Granulocytes	0.637	0.0011	2.964

Bcl6-Miz1			
Cell Type	R-value	P-value (0.05)	-log10(p)
B Cells	-0.194	0.0896	1.048
T Cells	0.356	0.0000000000201	10.698
Monocytes	-0.122	0.5002	0.301
Stem Cells	0.176	0.3106	0.508
Stromal Cells	0.197	0.0845	1.073
Innate Lymphocytes	0.603	0.000000270	6.569
Macrophages	0.336	0.0021	2.688
Dendritic Cells	0.046	0.6699	0.174
Mast, Basophils and Eosinophil	-0.026	0.8896	0.051
Granulocytes	0.598	0.0026	2.585

LRF-ThPOK			
Cell Type			
B Cells	0.675	0.0000000000121	10.916
T Cells	-0.185	0.0007	3.158
Monocytes	0.036	0.8425	0.074
Stem Cells	0.334	0.0501	1.301
Stromal Cells	0.370	0.0009	3.061
Innate Lymphocytes	0.455	0.0002	3.636
Macrophages	0.132	0.2360	0.627
Dendritic Cells	-0.080	0.4601	0.337
Mast, Basophils and Eosinophil	0.479	0.0074	2.131
Granulocytes	0.499	0.0152	1.817

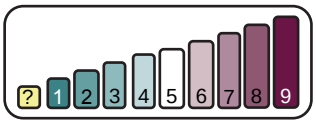
Figure S2

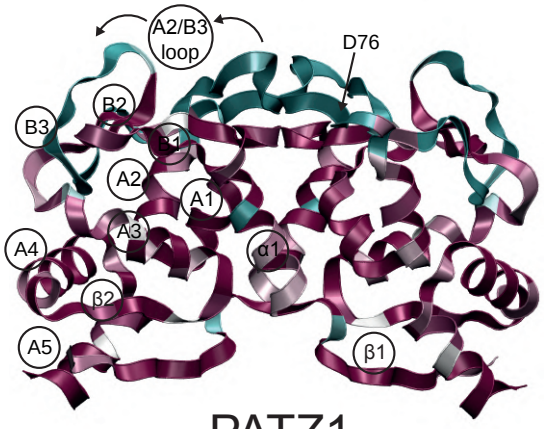
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



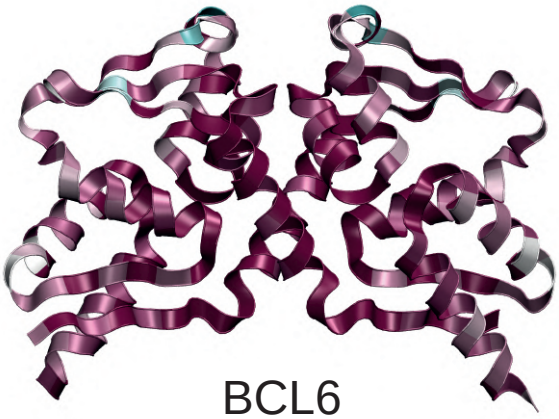
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S3

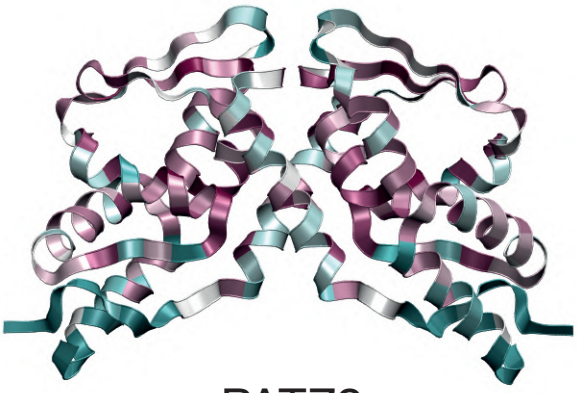




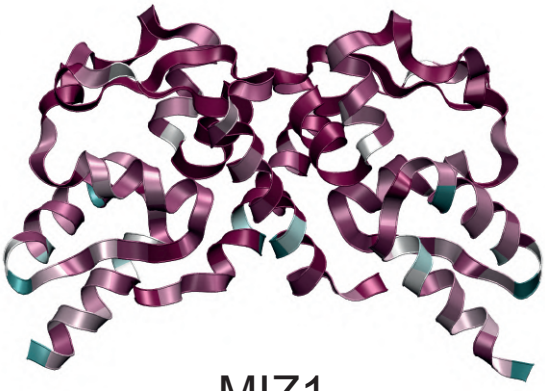
PATZ1



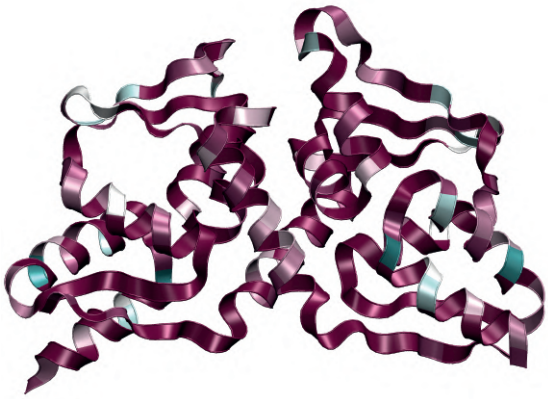
BCL6



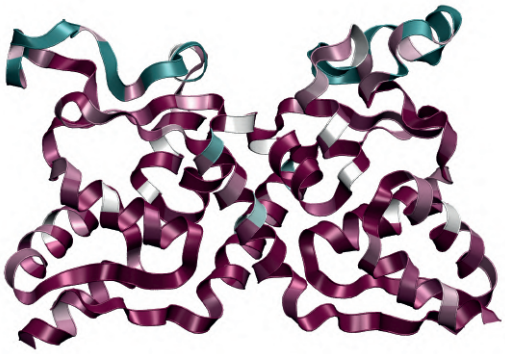
PATZ2



MIZ1



LRF

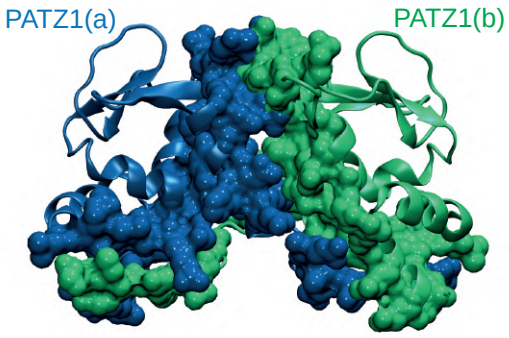


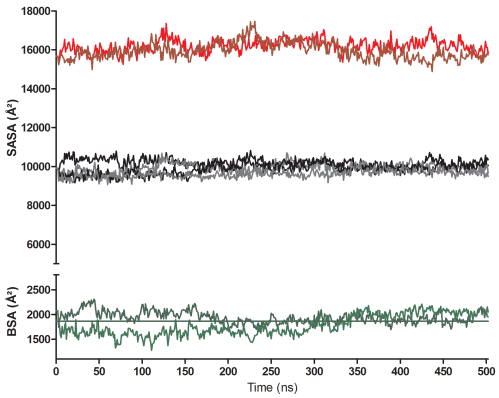
ThPOK

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S4

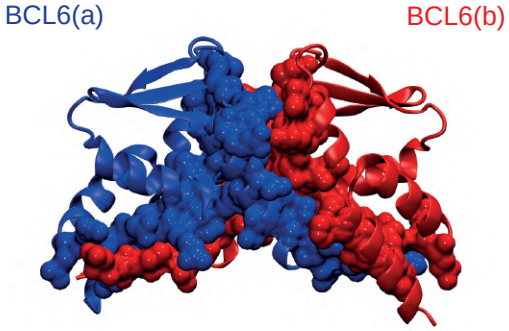
(a)

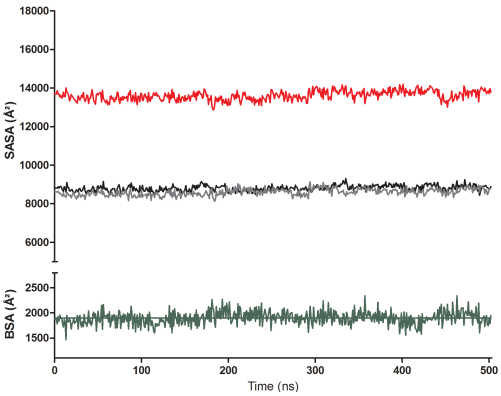




(run1) min: 1625; max: 2311; avg: 1945
(run2) min: 1279; max: 2210; avg: 1783

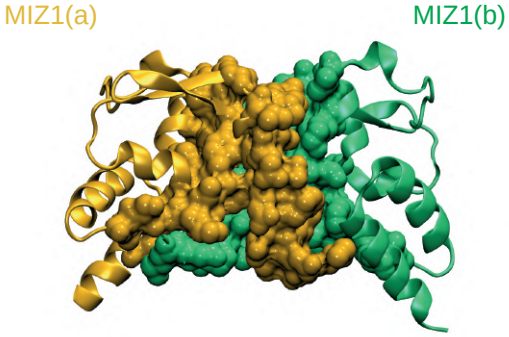
(b)

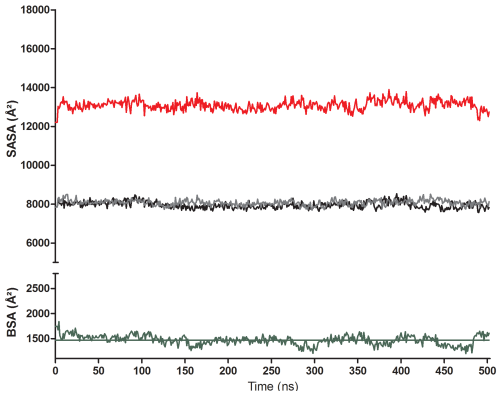




min: 1465; max: 2343; avg: 1899

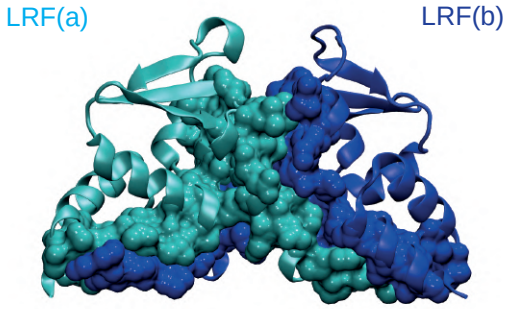
(c)

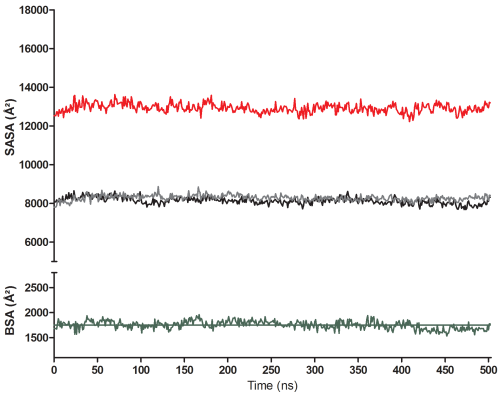




min: 1199; max: 1841; avg: 1471

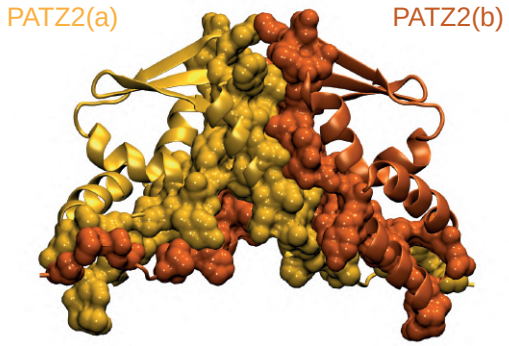
(d)

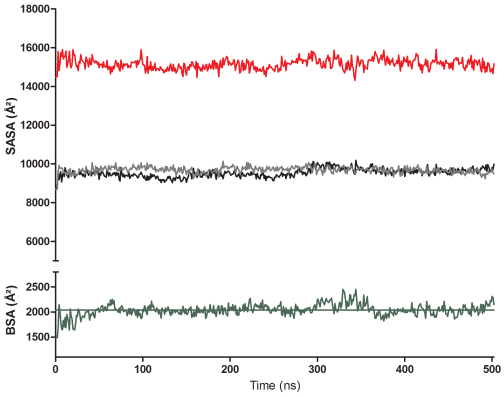




min: 1532; max: 1962; avg: 1752

(e)



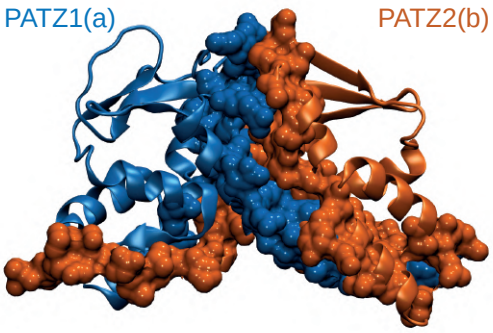


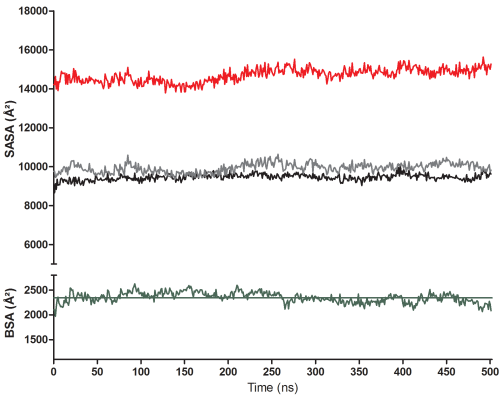
min: 1486; max: 2454; avg: 2037

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S5

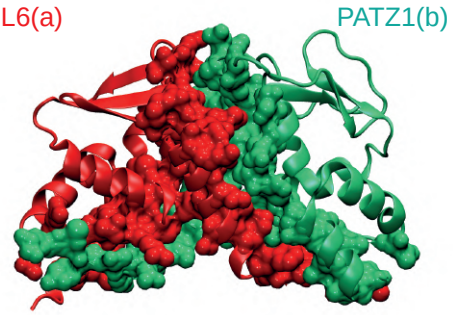
(a)

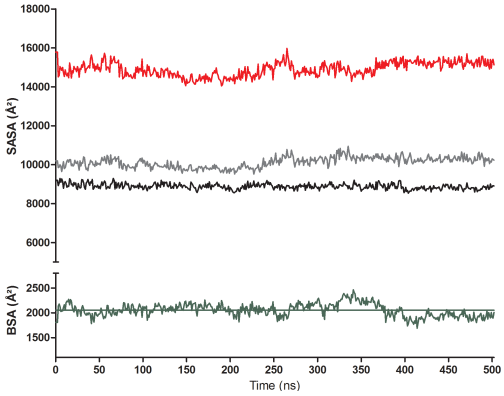




min: 1970; max: 2623; avg: 2348

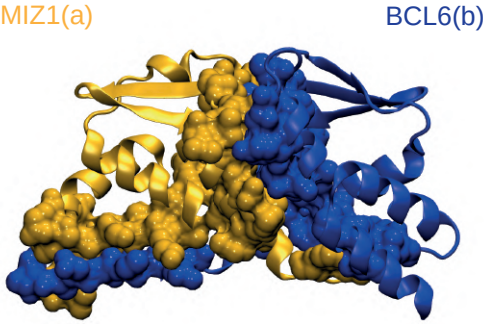
(b)

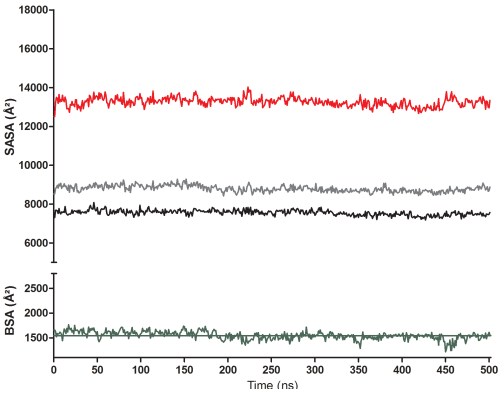




min: 1689; max: 2465; avg: 2054

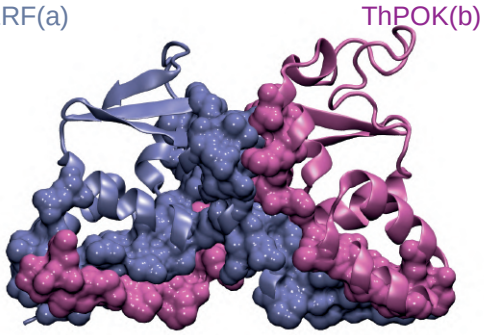
(c)

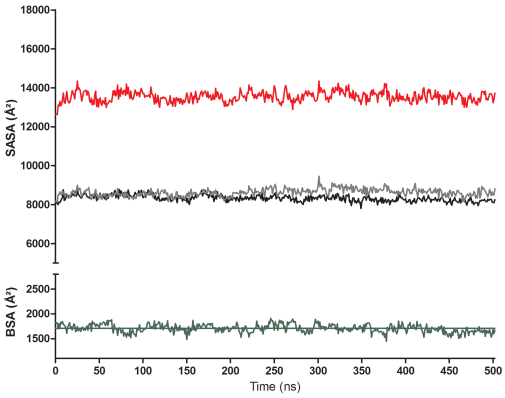




min: 1219; max: 1764; avg: 1548

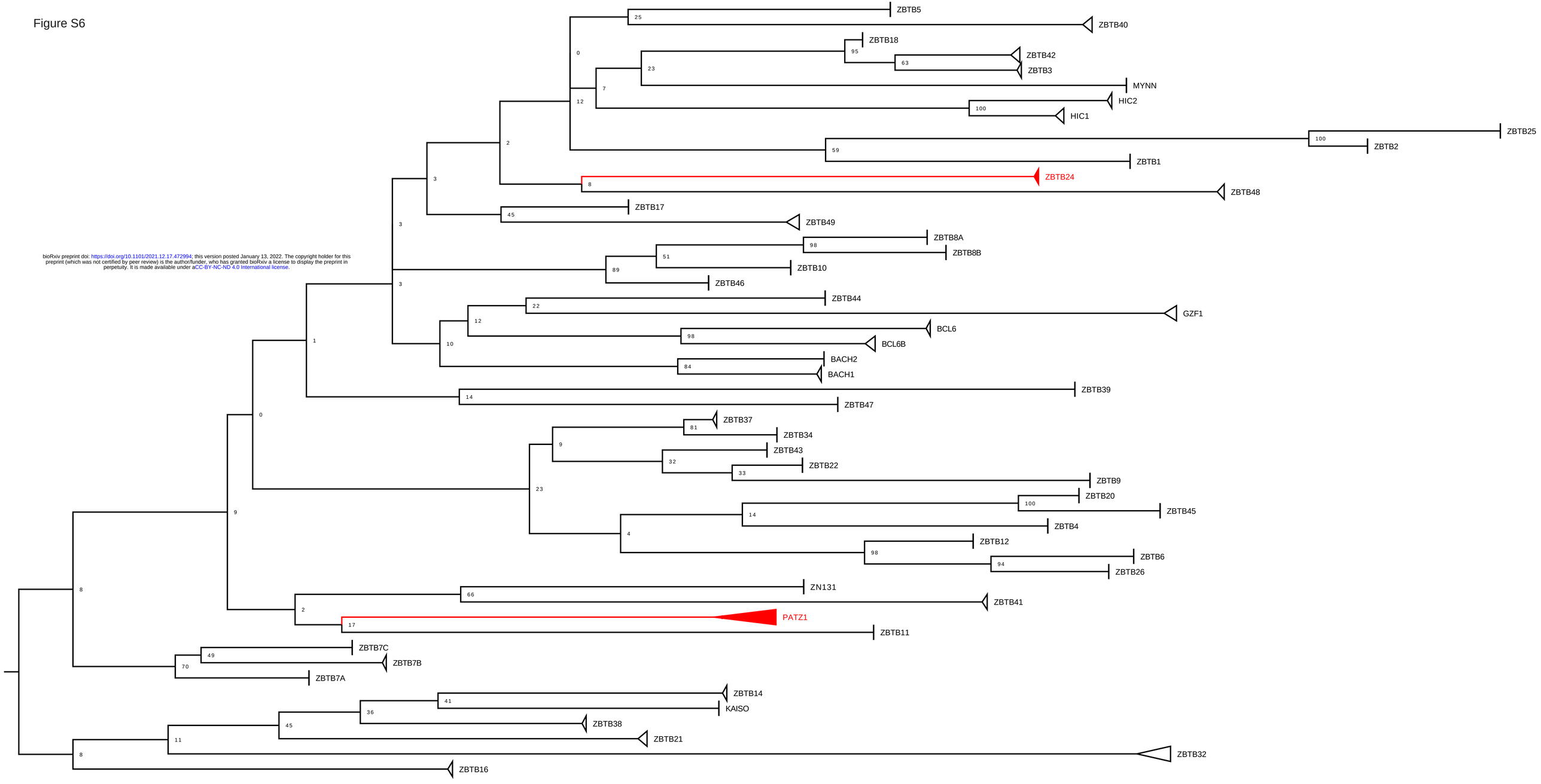
(d)





min: 1448; max: 1908; avg: 1706

Figure S6



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.17.472994; this version posted January 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.