Computers in Biology and Medicine 191 (2025) 110144

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

BCOR-rearranged sarcomas: *In silico* insights into altered domains and BCOR interactions

Kristóf Madarász [a], János András Mótyán [b], Yi-Che Chang Chien [a], Judit Bedekovics [a], Szilvia Lilla Csoma [a], Gábor Méhes [a], Attila Mokánszki [a,*]

[a] Department of Pathology, Faculty of Medicine, University of Debrecen, 4032, Debrecen, Hungary
[b] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, 4032, Debrecen, Hungary

ARTICLE INFO

Keywords:
BRS
BCOR
AlphaFold3
Sarcoma
Bioinformatics

ABSTRACT

BCOR (BCL-6 corepressor) rearranged small round cell sarcoma (BRS) represents an uncommon soft tissue malignancy, frequently characterized by the *BCOR*::*CCNB3* fusion. Other noteworthy fusions include *BCOR*::*MAML3*, *BCOR*::*CLGN*, *BCOR*::*MAML1*, *ZC3H7B*::*BCOR*, *KMT2D*::*BCOR*, *CIITA*::*BCOR*, *RTL9*::*BCOR*, and *AHR*::*BCOR*. The *BCOR* gene plays a pivotal role in the Polycomb Repressive Complex 1 (PRC1), essential for histone modification and gene silencing. It interfaces with the Polycomb group RING finger homolog (PCGF1). This study employed comprehensive *in silico* methodologies to investigate the structural and functional effects of *BCOR* fusion events in BRS. The analysis revealed significant alterations in the domain architecture of BCOR, which resulted in the loss of *BCL6*-regulated transcriptional repression. Furthermore, IUPred3 prediction indicated a significant increase in disorder in the C-terminal regions of the BCOR in the fusion proteins. A detailed analysis of the physicochemical properties by ProtParam revealed a decrease in isoelectric point, stability, and hydrophobicity. The analysis of protein structures predicted by AlphaFold3 using the PRODIGY algorithm revealed statistically significant deviations in binding affinities for BCOR-PCGF1 dimers and a non-canonical PRC1 variant tetramer compared to the wild-type BCOR. The findings provide a comprehensive summary and elucidation of the fusion proteome associated with BRS, suggesting a substantial impact on the stability and functionality of the fusion proteins, thereby contributing to the oncogenic mechanisms underlying BRS. In this study, we provide the first compilation and comparative analysis of the known BCOR fusions of BRS and introduce a new *in silico* approach to enhance a better understanding of the molecular basis of BRS.

1. Introduction

In silico methods have significant importance in modern cancer research, particularly in the context of the analysis of single nucleotide polymorphisms (SNPs), mapping of hotspots, virtual screening, molecular docking as well as investigation and prediction of intermolecular interactions [1–4]. Computational approaches facilitate the identification and annotation of SNPs through genome-wide association studies and next-generation sequencing, enabling the prediction of their functional impact on protein function and cancer risk [5,6]. The latest iteration of AlphaFold2, the AlphaFold3 (AF3) has further advanced the field of protein modeling, offering enhanced capabilities for predicting the three-dimensional structures of even larger fusion proteins and protein complexes [7,8]. Computational modeling techniques, including protein structure prediction and molecular dynamics simulations, permit researchers to visualize the impact of fusion events on the three-dimensional structure of proteins. This enables an assessment of how the relative positioning of distinct domains affects protein folding, stability, and function [7,9].

BCOR (BCL-6 corepressor) rearranged small round cell sarcoma (BRS) is a rare soft tissue tumor occurring in the bones of young patients. Based on the 5th edition of the WHO Classification of soft tissue and bone tumors, BRS belongs to the third distinct subset of "undifferentiated small round cell sarcomas of bone and soft tissue" [10] with histological features, immune profile, and expression signatures, that differ from round cell sarcomas with *EWSR1* gene fusion with non-ETS (non-E26 transformation-specific) family members and from Capicua

* Corresponding author.
E-mail addresses: madarasz.kristof@med.unideb.hu (K. Madarász), motyan.janos@med.unideb.hu (J.A. Mótyán), dr.changchien.yiche@med.unideb.hu (Y.-C. Chang Chien), bedekovics.judit@med.unideb.hu (J. Bedekovics), csoma.szilvia@med.unideb.hu (S.L. Csoma), gabor.mehes@med.unideb.hu (G. Méhes), mokanszki.attila@med.unideb.hu (A. Mokánszki).

https://doi.org/10.1016/j.compbiomed.2025.110144
Received 7 October 2024; Received in revised form 1 April 2025; Accepted 3 April 2025
Available online 13 April 2025

Abbreviations

AA - Amino Acids
ADD - Acetyltransferase Domain
AI - Aliphatic Index
AF - AlphaFold
ANK - Ankyrin Repeats
BCOR - BCL-6 Corepressor
BCOR-2 - BCOR Isoform 2
BRS - BCOR-Rearranged Small Round Cell Sarcoma
CCNB3 - G2/Mitotic-Specific Cyclin-B3
CDD - Conserved Domain Database
CIC - Capicua Transcriptional Repressor
CLGN - Calmegin
D-box - Destruction Box
EWSR1 - RNA-binding protein EWS
EZH2 - Enhancer of Zeste Homolog 2
FYRC - FY-Rich Domain C-Terminal Region
FYRN - FY-Rich Domain N-Terminal Region
GO - Gene Ontology
GRAVY - Grand Average Hydrophobicity
hESC - Human Embryonic Stem Cell
HOX - Homeobox
HMG-box - High Mobility Group-Box
IDRs - Intrinsically Disordered Regions
ITDs - Internal Tandem Duplications
KDM2B - Lysine-Specific Demethylase 2B
KMT2D - Histone-Lysine N-Methyltransferase 2D
LD motif - Leucine-Aspartate Repeat Motif
MAML1 - Mastermind-Like Protein 1
MAML3 - Mastermind-Like Protein 3
MD – Molecular dynamics
mESC - Mouse Embryonic Stem Cell
MM-PBSA - Molecular Mechanics Poisson-Boltzmann Surface Area
NGS - Next-Generation Sequencing
NLS1 - Nuclear Localization Signal 1
PARP - Poly ADP-Ribose Polymerase
PCGF1 - Polycomb Group RING Finger Homolog
PRC1 - Polycomb Repressive Complex 1
PUFD - PCGF Ubiquitin-Like Fold Discriminator
RAWUL - WD40-Associated Ubiquitin-Like
RGAG1 - Retrotransposon Gag Domain-Containing Protein 1
Rg - radius of gyration
RMSD – Root Mean Square Deviation
RMSF – Root Mean Square Fluctuation
RTL9 - Retrotransposon Gag-Like Protein 9
RT-PCR - Real-Time PCR
Sec/SPI - Sec/Signal Peptides
SKP1 - S-Phase Kinase-Associated Protein 1
SNPs - Single Nucleotide Polymorphisms
TCR-T - TCR-engineered T cells
TpI - Theoretical Isoelectric Point
UPR - Unfolded Protein Response
wBCOR - Wild-Type BCOR
WHO - World Health Organization
ZC3H7B - Zinc Finger CCCH Domain-Containing Protein 7B

transcriptional repressor (*CIC*)-rearranged sarcomas [11].

The *BCOR* gene is located in the Xp11.4 chromosomal region and mediates apoptotic and oncogenic activities of cells through *BCL6*-regulated transcriptional repression via epigenetic signaling mechanisms [12–14]. BCOR (Uniprot ID: Q6W2J9) binds to Polycomb group RING finger homolog (PCGF1) protein of Polycomb repressive complex 1 (PRC1) through its PCGF ubiquitin-like fold discriminator (PUFD) domain, which facilitates epigenetic modification of histones by the addition of a ubiquitin moiety to histone H2A at Lys119 (H2AK119) [15,16]. The BCOR PUFD and the PCGF1 WD40-associated ubiquitin-like (RAWUL) domain interact in a hierarchical mode of assembly. Upon binding of PCGF1, the termini of BCOR's PUFD domain become structurally ordered, enabling stable association with lysine-specific demethylase 2B (KDM2B). This interaction is critical for the assembly of the non-canonical PRC1 variant known as PRC1.1 [17,18]. The interaction between RAWUL and PUFD depends on the β-sheet and loop interaction surfaces for selective binding of their respective partners [15]. PRC1 silences several genes, including the homeobox (HOX) group of genes [16,19]. It inhibits the mesoderm and endoderm specification genes and assumes an integral part in regulating the primed pluripotent state. Notably, BCOR seems to be non-essential for mouse embryonic stem cell (mESC) pluripotency similar to other recently reported human ESC (hESC) regulators [20]. Recruitment of BCOR to the target sites is primarily mediated by its C terminus, which plays a necessary and sufficient role in this process. The PUFD domain is likely to play a key role in targeting BCOR to PRC1.1. In addition, the N terminus of BCOR contributes to transcriptional repression independently of the C terminus [21].

Internal tandem duplications (ITDs) and translocations leading to gene fusions are the most frequently occurring aberrations of *BCOR*. The G2/mitotic-specific cyclin-B3 (*CCNB3*) was identified as the most prevalent fusion partner in BRS [22,23]. Following the increasing use of molecular diagnostics methods, especially next-generation sequencing (NGS), Spacht et al. reported two novel fusions in 2016, namely *BCOR*::*MAML3* (mastermind-like protein 3) and *ZC3H7B*::*BCOR* (ZC3H7B: zinc finger CCCH domain-containing protein 7B) [24]. Subsequently, Kao et al. discovered the *KMT2D*::*BCOR* fusion (KMT2D: histone-lysine N-methyltransferase 2D) in 2018 [11], while Yoshida et al. detected *CIITA*::*BCOR* (CIITA: MHC class II transactivator) and *ZC3H7B*::*BCOR* fusions in 2020 [25]. More recently, in 2022, Vassella et al. identified an *RTL9*::*BCOR* gene fusion (*RTL9:* retrotransposon gag-like protein 9; the protein is also referred to as retrotransposon gag domain-containing protein 1, *RGAG1*). As compared to the other fusions described above, this fusion protein uniquely contains a non-canonical BCOR that lacks its 1168–1201 region and thus corresponds to isoform 2 of BCOR (BCOR-2) [26]. An additional fusion has been reported by our research group in 2023, we described the *BCOR*::*CLGN* (*CLGN*: calmegin) fusion in BRS [27]. In addition, Cocchi et al. found two novel BCOR-2 isoform fusion proteins in 2024, *BCOR*::*MAML1* (*MAML1*: mastermind-like protein 1) and *AHR*::*BCOR* (*AHR*: aryl hydrocarbon receptor) [28].

In this study, we aimed (i) to collect the currently available literature data on BRS, including cDNA sequences derived from NGS and Real-Time-PCR (RT-PCR) data, to assemble the amino acid sequences of the fusion proteins, (ii) to use *in silico* approaches for the analysis of the chimeric sequences as well as structures, and (iii) to build the 3D structures of the chimeric proteins, to reveal the oncogenic mechanism of *BCOR* fusion small round cell sarcomas.

2. Materials and methods

2.1. Protein information

The nucleotide sequences of the nine fusion genes – that were determined by Sanger sequencing – were obtained from the scientific literature [11,22,24–27]. Ensemble (Genome assembly: GRCh37.p13; GCA_000001405.14) was used to find the breakpoints of the fusions [29] and to manually assemble the sequences. The coding sequences were translated *in silico* by using the ExPASy: Translate online translation tool

(https://web.expasy.org/translate/) [30]. Protein information for the BCOR(-2), CCNB3, MAML3, ZC3H7B, KMT2D, CIITA, RTL9, and CLGN, MAML1, and AHR proteins or their assembled fusion proteins were obtained from the UniProt database [31]. The fusion of genes and proteins is indicated by a double colon () [32].

The domain information was obtained from UniProt, the Conserved Domain Database (CDD) [33], and InterPro (version 5.67–99.0) [34] databases. The Gene Ontology terms (GO) were determined based on the amino acid (AA) sequences we assembled on InterPro. The PANTHER GO terms were used [35–38].

2.2. Calculation of physicochemical properties

The physicochemical properties were calculated by using ExPASy's ProtParam tool (https://web.expasy.org/protparam/) [39], including the theoretical isoelectric point (TpI), molecular weight, the total number of positive and negatively charged residues, extinction coefficient (EC) [40–42], instability index (II) [43], aliphatic index (AI) [44], and grand average hydrophobicity (GRAVY) [45]. The extinction coefficient estimates the molar extinction coefficient based on the amino acid composition. This value is essential for accurate quantification of protein concentration in spectrophotometric analysis. The instability index provides an estimate of the *in vitro* stability of a protein. The aliphatic index of a protein is considered a factor increasing thermostability of globular proteins and is specifically defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). The GRAVY score is calculated as the sum of the hydropathy values of all amino acids (AA) divided by the number of residues in the sequence. The following formula was used to calculate the change in physicochemical values between the wild-type partner and the fusion proteins:

$$change\ (\%) = \frac{fusion\ protein\ value - wild\ type\ protein\ value}{|wild\ type\ protein\ value|}$$

2.3. Prediction of signal peptides and disorder propensities

The SignalP 6.0 predictor was utilized to identify signal peptides in the proteins. SignalP-6.0 employed a transformer protein language model combined with a conditional random field for structured prediction. Only Sec/Signal Peptides (Sec/SPI) found in eukaryotes were searched, as the selected input organism was Eukarya [46]. DeepLoc 2.0 is a multi-label predictor that was utilized to estimate the intracellular localization of the proteins, this tool is based on the use of a transformer-based protein language model, providing accurate predictions and interpretability through attention mechanisms. Only those probability scores were considered that were higher than the default thresholds of DeepLoc 2.0: cytoplasm (0.4761), nucleus (0.5014), and endoplasmic reticulum (0.6090) [47]. To validate the results of the predictions, the predicted data were compared to those of the UniProt database and The Human Protein Atlas [31,48]. Additionally, the NetGPI-1.1 GPI-anchored predictor [49] – a deep learning approach based on recurrent neural networks that incorporates an attention mechanism - was applied to determine the potential glycosylphosphatidylinositol (GPI) anchoring of the proteins. The three predictors were reached on the https://services.healthtech.dtu.dk/website.

To predict intrinsically disordered regions (IDRs) within the proteins, we utilized IUPred3 (https://iupred3.elte.hu/) which online tool that employs a biophysics-based model to identify regions lacking stable structure under native conditions. Using the IUPred3 web server, we performed a comparative analysis of the residues 1448–1633 (region$^{ANK\ +\ linker}$) and 1634–1748 (region^{PUFD}) between the wild-type BCOR (wBCOR) and its fusion proteins [50].

2.4. Analysis of protein structures

Three-dimensional (3D) models of the proteins involved in the BCOR-PCGF1 and PRC1.1 complexes were generated using AlphaFold3 (AF3) (https://alphafoldserver.com/) [7]. We constructed dimers of BCOR and PCGF1 and tetramers of BCOR, PCGF1, KDM2B, and S-phase kinase-associated protein 1 (SKP1) proteins. 50 models were prepared for each complex. The predicted models were compared with the experimentally determined structure (if available) of the BCOR-PCGF1 dimer (PDB ID: 4hpl) [15] and with the predicted models of the wild-type proteins. Due to the length limitation of the AlphaFold Server (beta), we could not predict the structure of KMT2D::BCOR consisting, of 5480 amino acid residues.

The PRODIGY (PROtein binDIng enerGY prediction) web server (https://rascar.science.uu.nl/prodigy/) [51] was employed to predict the binding affinity as described by the Gibbs free energy change (ΔG) in kcal mol^{-1}. PRODIGY is a web server that combines structural and energetic parameters to predict the Gibbs free energy change (ΔG) upon binding, based on the 3D structures of protein-protein complexes. The algorithm integrates intermolecular contacts and non-interacting surface properties to accurately estimate binding affinity. The alterations in the ΔG of BCOR-PCGF1 dimers were investigated across the entire protein sequence (full-length), specifically the BCOR PUFD region from 1634 to 1748 and the PCGF1 RAWUL region from 167 to 255 (RAWUL-PUFD domain length). The regions of the RAWUL domain encompassing 167–177 and 185–254, and the PUFD domain at 1636–1748 (4hpl length), were examined [15]. Additionally, varying interaction surfaces and key residues were investigated. This allowed us to compare the predicted models with the experimental structure of the BCOR-PCGF1 dimer and to limit the PRODIGY predictions to the relevant sequences. PyMOL was used to visualize the protein structures [52]. We omitted the 178–184 residues of the RAWUL domain from the analyses, as this loop is missing from the crystal structure (PDB ID: 4hpl). Although this region does not interact with the PUFD domain, we removed it from the modeled complexes using PyMOL [52] for consistency. We also generated BCOR-PUFD wild-type dimers based on the 4hpl crystal structure using AF3 [7] as a control. We predicted and analyzed the heterotetramers of BCOR, PCGF1, KDM2B, and SKP1 using their full-length sequences (BCOR 1–1755, PCGF1 1–259, KDM2B 1–1336, SKP1 1–163). In instances where the BCOR-2 isoform was present, the isoform sequence corresponding to the canonical protein was utilized.

The contact map was generated using the MAPIYA contact map webserver (https://mapiya.lcbio.pl/) with a distance cut-off of 5.5 Å. MAPIYA calculates residue-residue interactions by determining pairwise distances based on protein structure. Residues within the specified distance threshold were in contact, allowing for visualization of the protein's structural organization and interaction networks. Possible interaction forces were also described in brackets after the contacts [53].

PDB 4hpl was downloaded from the RCSB Protein Data Bank [54].

2.5. Molecular dynamics (MD) simulation

We performed 10 ns molecular dynamics (MD) simulations using the GROMACS software suite [55–62]. For each of the nine BCOR-PCGF1 dimers (wild-type and fusion variants), a single structure was selected from the 50 replicates generated by AlphaFold, based on the highest iPTM and pTM scores, ensuring the most stable and reliable conformations were analyzed. Binding affinities between BCOR proteins and PCGF1 were calculated using the gmx_MMPBSA tool, which applies the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method to estimate interaction energies with high accuracy [63,64].

2.6. Statistics

Friedman's tests followed by Dunn's multiple comparisons were

performed to compare the IUPred3 IDRs scores. Ordinary one-way ANOVA, Kruskal-Wallis, Brown-Forsythe ANOVA, and post-hoc tests were used to compare the binding affinity values of the complexes. Statistical tests were performed using GraphPad Prism version 9.5.1. [65–67]. *P* values were considered significant where $P < 0.05$.

For graph and Figure construction, GraphPad Prism [67] version 9.5.1 and IBS (illustrator of biological sequences) were used [68].

3. Results

3.1. Sequence analysis

There are considerable challenges in the diagnosis and therapy of BRS due to its rarity and overlapping features with other sarcomas. Nevertheless, the *BCOR* fusions serve as a diagnostic marker, helping the differentiation of BRS from similar malignancies [69]. Moreover, understanding the structural and functional changes caused by these fusions offers insights into the tumorigenic processes. To investigate this rare malignancy, we collected the known *BCOR* fusions of BRS. After collecting the nucleotide sequences (), we performed the translation of the sequences *in silico*, followed by the analysis of the fusion proteins' sequences of the fusion proteins and the comparison to those of the wild-types (by using Blastp tool).

The seven studied fusion proteins were classified into two groups based on the position of the wild-type BCOR (wBCOR) within the fusion protein. In four of the nine cases, the wBCOR protein was found to be located at the N terminus, while in the other five cases at the C terminus of the fusion proteins. These groups are referred to in this paper as $BCOR^{NT}$ and $BCOR^{CT}$, respectively.

The BCOR is present at the N terminus of the BCOR::CCNB3, BCOR::CLGN, BCOR::MAML3, and BCOR-2:MAML1 fusion proteins (Fig. 1). The BCOR::CCNB3 (UniProt ID: H9A532; 3038 AA) and BCOR::CLGN (2070 AA) encompasses the full-length wBCOR (1755 AA), while BCOR-2:MAML1 (2699AA) the whole BCOR-2 (UniProt ID: Q6W2J9-2). In these cases, the fusion causes the loss of a short N-terminal region of CCNB3 (<10 % of the full-length protein sequence), while the truncation is more remarkable (~50 %) in CLGN where 1–295 of the 610 residues is deleted upon the fusion, as reported previously [27]. The BCOR-2:AHR protein's C terminus AHR partner suffers a 39 AA minor loss. The BCOR::MAML3 fusion proteins contain the nearly full-length wBCOR, the 1–1751 residues are present and only the 1751-1755 C-terminal residues are missing. The 1752–2734 region of the fusion protein corresponds to the MAML3 protein having a truncated N terminus, and 1–156 of the 1138 residue long protein is missing. The amino acid residue at position 1752 was coded by guanine, alanine of *BCOR,* and guanine base of *MAML3*.

In the case of the $BCOR^{CT}$ group, the fusion causes a more remarkable shortening of the BCOR protein sequence, except RTL9:BCOR-2. In the other three cases, there is a 50–62 % decrease in the length of the wBCOR sequence (Fig. 1, Table 1). The length of the ZC3H7B::BCOR fusion protein is 1108 AA, its 1–433 region corresponds to the truncated ZC3H7B protein while the 434–1108 region to the 1081–1755 residues of wBCOR. In the case of *KMT2D*::*BCOR*, the fusion results in a 5480 AA long fusion protein, representing the 1–4667 and 943–1755 residues of KMT2D and wBCOR, respectively. Accordingly, 16 % and 54 % of the respective full-length KMT2D and BCOR sequences are missing from the fusion proteins.

The most remarkable sequence truncation was observed for the fusion partners in the case of the CIITA::BCOR, which is the shortest of the studied fusion proteins (it contains 821 AA). The fusion protein contains only 13 % of the CIITA and 38 % of the wBCOR sequences, representing the 1-145 N-terminal and 1081-1755 C-terminal regions of the wild-type proteins, respectively. The triplet coding for the Arg146 residue was coded by a cytosine of the *CIITA* and guanine and cytosine of the *BCOR* gene.

The RTL9:BCOR-2 fusion protein – consisting of 2891 residues – contains the RTL9 protein truncated at its C terminus (the 1200–1388 region is missing) as well as the nearly full-length BCOR-2 which lacks only its 1-29 C-terminal residues in the fusion protein.

The AHR::BCOR-2 fusion protein was found to be 1676 AA long. The AHR was truncated at residue 801, and because of the fusion event, it continued with residues 802 to 806 from its intron 10. The C terminus of the fusion protein ended with the largely truncated 868–1721 part of the BCOR-2 protein (Fig. 1).

In summary, the sequence analysis revealed considerable variability both for BCOR and its fusion partners, thereby emphasizing the structural diversity caused by the fusion events. The degree of sequence retention observed for BCOR ranged from 38 % to 100 %, with an average value of approximately 79 %. It is worthy of note that the most extensive truncation was observed in the CIITA::BCOR fusion, where BCOR protein retains only 38 % of its sequence, while BCOR remains entirely intact (100 % retention) in fusions such as BCOR::CCNB3, BCOR::MAML3, and BCOR::CLGN. As compared to BCOR, the fusion partners exhibit more pronounced truncations, they retain 13–98 % of their original sequences (the average retention was approximately 65 %). The CIITA protein exhibits the most substantial loss, showing only 12.8 % sequence retention in the CIITA::BCOR fusion, while RTL9 retains almost its full sequence (98.3 %) in the RTL9:BCOR-2 fusion protein. These findings highlight the significant structural alterations induced by fusion events, which are likely to contribute to the functional consequences observed in *BCOR*-rearranged sarcomas.

3.2. Physicochemical characteristics of fusion proteins

The physicochemical properties of the fusion proteins were calculated by using the ProtParam web server (Fig. 2). In all instances, we observed alterations in the physicochemical properties of the fusion proteins as compared to the wild-types (Table 2).

The mean alteration across all cases was 29.54 %, and the mean alteration of absolute values was 53.43 % (|mean|). The + R(%) and -R (%) values - indicating the percentage of the charged amino acid residues to the total number of the residues - were 9.79 % and 8.68 % respectively, with standard deviations of 1.98 and 1.87. The mean change was 62.47 % and the absolute mean was 53.43 % for the $BCOR^{NT}$ group, while the respective 3.20 % and 42.61 % values were obtained for the $BCOR^{CT}$ group (Fig. 2, Table 2).

The MAML1 protein exhibited the most substantial mean change (132.20 %), which can be explained by its large truncation due to the fusion event. In contrast, the smallest mean difference (0.47 %) and the smallest |mean| change (3.56 %) were observed for ZC3H7B and KMT2D, respectively. The mean change of the GRAVY values was −26.13 % (Fig. 2C), with the smallest and largest differences observed for BCOR-2 in BCOR-2:MAML1 (−3.18 %) and CIITA (−305.10 %). The mean change in AI values was negligible (−0.19 %), although the CIITA showed a 24.29 % decrease, resulting in a decrease in thermostability. The mean alterations of the instability index were more noticeable (4.69 %), with the greatest change found in the case of BCOR protein in KMT2D::BCOR fusion, which increased by 39.63 %, indicating a decrease in protein thermostability.

The mean percentage change of +R and -R was 62.61 %–69.71 %, with the highest recorded values being 258.67 % and 301.41 %, both of which were observed in MAML1, due to its drastic truncation of the fusion event.

The mean change of EC of the fusion proteins was 74,18 % compared to their wild-type (Fig. 2C). The BCOR-2:MAML1 showed the largest increase in EC, rising by 352.6 % from MAML1. Conversely, CIITA::BCOR displayed the largest decrease in EC, dropping by 19.2 % from CIITA (Fig. 2A).

The mean pI of the fusion proteins was found to decrease by −6.31 % as compared to the pI values of the wild-type proteins (Fig. 2C). It is noteworthy that ZC3H7B::BCOR exhibited the most pronounced reduction of pI, with a 27.67 % decline as compared to ZC3H7B. In

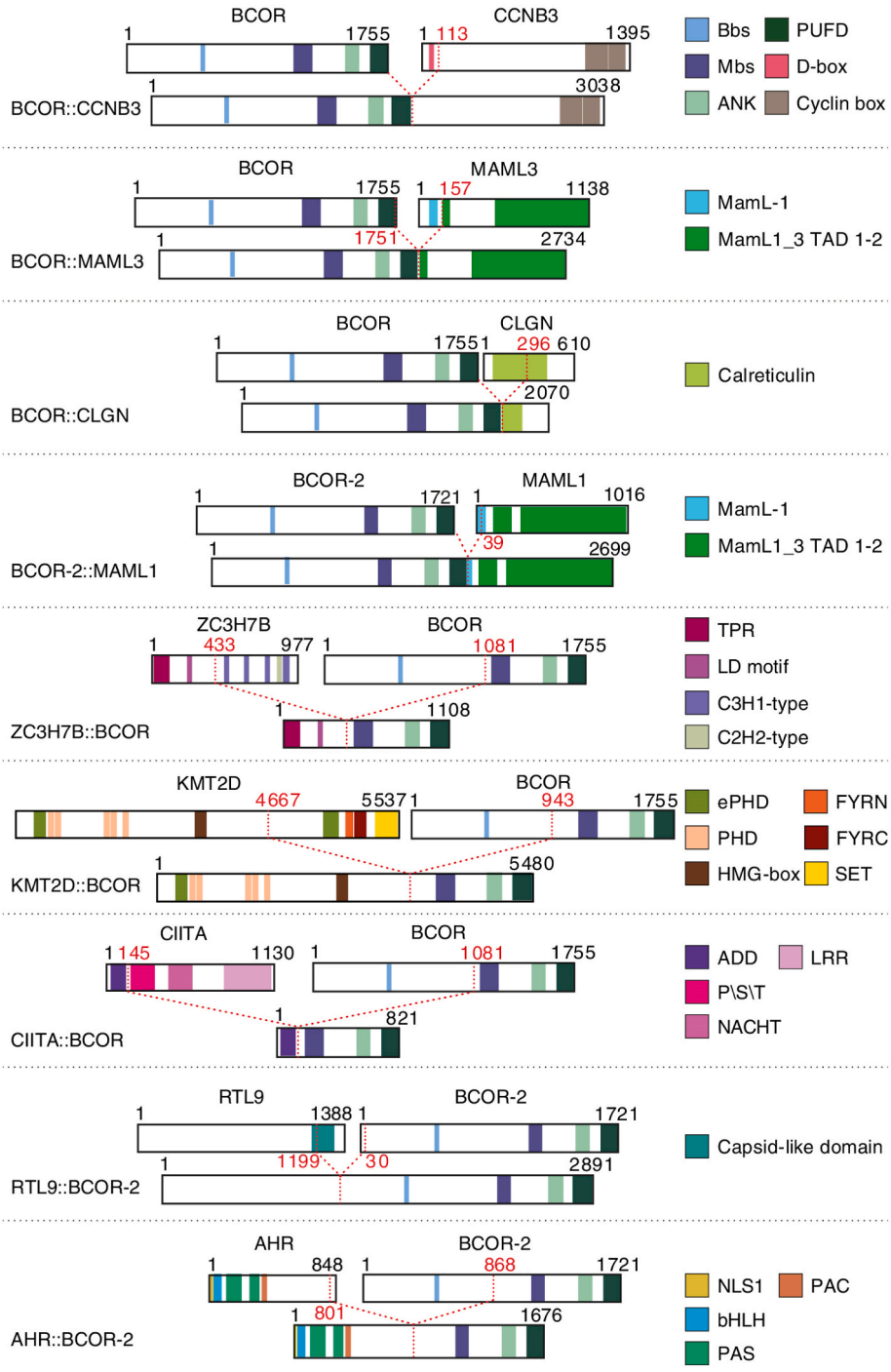

Fig. 1. Domains and functional sites of known fusion proteins from *BCOR*-rearranged sarcomas. As a result of the fusion events, the new fusion protein has lost several regions, both short and long, along with the domains and functional sites when compared to the wild-type fusion partners. Schematic representation is shown for each fusion partner and fusion protein, in the order corresponding to their N- or C-terminal positions in the fusion proteins. The red dotted lines and labels indicate the sites of the breakpoints at the protein level. The lengths of KMT2D and KMT2D::BCOR proteins, except for the domains, are not representative due to the large size of these two proteins, all other proteins are represented proportionally. Bbs: BCL6 biding site, Mbs: MLLT biding site, ANK: ankyrin repeats; PUFD: PCGF Ubiquitin-like fold discriminator; D-box: destruction box, MamL-1: Neurogenic mastermind-like, N-terminal domain, Maml1_3 TAD 1-2: Mastermind-like 1/3: transactivation domain; TPR: Tetratricopeptide repeats, LD motif: leucine-aspartate repeat motif, C3H1-type: CCCH type, C2H2-type: CCHH type; ePHD: extended plant homeodomain, PHD: plant homeodomain, HMG-box: high mobility group -box, FYRN: FY-rich domain N-terminal region, FYRC: FY-rich domain C-terminal region, SET: pre-SET and post-SET region; ADD: acetyltransferase domain, P\S\T: Proline-serine-threonine rich domain, LRR: Leucine-rich repeat ribonuclease inhibitor; NLS1: Nuclear localization signal 1; bHLH: basic helix-loop-helix; PAS: Per-ARNT-Sim domain; PAC: C terminal subunit of PAS; BCOR-2: BCOR isoform 2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Fusion events and breakpoints (BP) in *BCOR*-rearranged sarcomas. AA: amino acids, WT: wild-type. The fusion protein of RTL9, MAML1, and AHR with BCOR contained the second isoform of BCOR (BCOR-2). This isoform consists of 1721 residues. In the case of the other fusions, the BCOR corresponds to the canonical form containing 1755 residues. A schematic representation of the fusion proteins is shown in Fig. 1. The lengths are shown for the fusion partners if they are part of the fusion protein (partner 1 or 2) or not (WT). int.: Intron.

Index	Gene information					Protein information		
	Fusions	Gene 5′ BP	Gene 3′ BP	Chromosomal partners	Exonic BP 5' - 3′	Partner 1/WT length (AA)	Partner 2/WT length (AA)	Fusion protein length (AA)
1	***BCOR::CCNB3***	39 911 366	50 051 504	chr X - chr X	15. - 5.	1755/1755	1283/1395	3038
2	***BCOR::MAML3***	39 911 374	141 074 014	chr X - chr 4	15. - 1.	1751/1755	982/1138	2734
3	***BCOR::CLGN***	39 911 366	141 317 359	chr X - chr 4	15. - 9.	1755/1755	316/610	2070
4	***BCOR::MAML1***	40 051 795	179 733 227	chr X - chr 5.	10. - 4.	1721/1721	978/1016	2699
4	***ZC3H7B::BCOR***	41 738 632	39 923 852	chr X - chr X	12. - 7.	433/977	675/1755	1108
5	***KMT2D::BCOR***	49 424 063	39 931 774	chr 12 - chr X	42. - 4.	4667/5537	813/1755	5480
6	***CIITA::BCOR***	10 992 859	39 923 852	chr 16 - chr X	5. - 7.	145/1130	675/1755	821
7	***RTL9::BCOR***	109 697 442	39 937 097	chr X - chr X	2. - 3.	1199/1388	1692/1721	2891
9	***AHR::BCOR***	17 340 228	40 072 790	chr 7 - chr X	10. (+int 10.) 4.	801/848 + 6 (int 10)	853/1721	1676

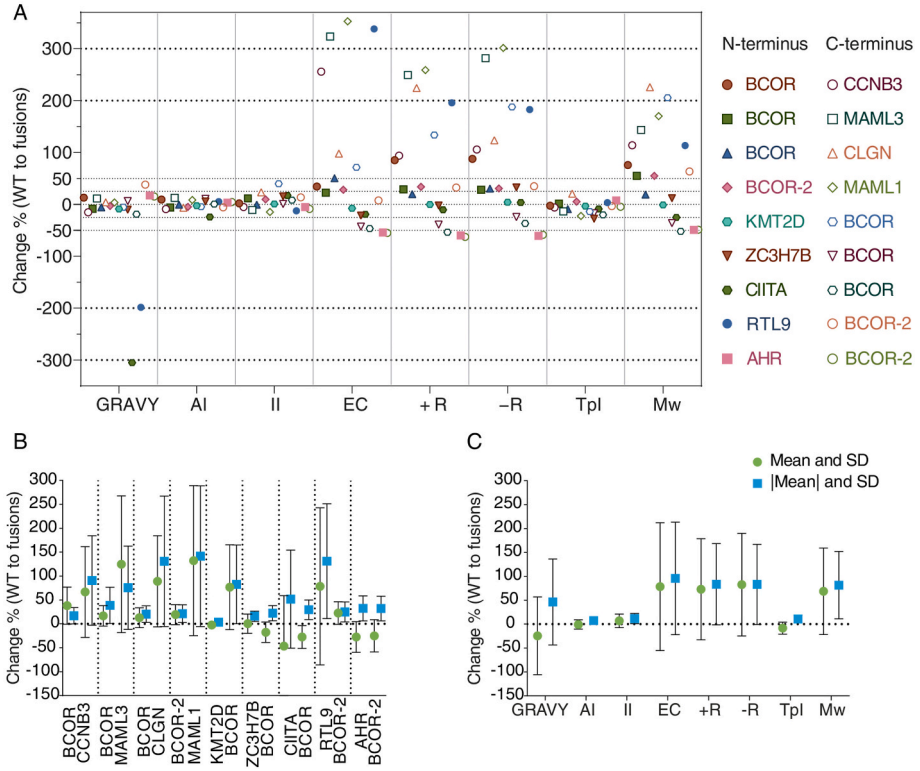

Fig. 2. Changes in the fusion proteins' physicochemical properties as compared to wild-type partner proteins. (A) Differences of eight investigated features (%) between wild-type and fusion proteins. GRAVY: grand average hydropathy, AI: aliphatic index, II: instability index, EC: extinction coefficient at 280 nm, +R: the number of positively charged residues, −R: the number of negatively charged residues, TpI: the theoretical isoelectric point, Mw: molecular weight, WT: wild-type. (B) The mean and standard deviation of the percentage changes in proteins. (C) The mean and standard deviation of the percentage changes in eight attributes. |Mean|: mean of absolute values, Error bars show standard deviation (SD).

contrast, the pI of BCOR::CLGN was found to be 20.6 % higher than that of CLGN (Fig. 2A). However, none of the fusion proteins exceeded the pH range of 4.5–10, indicating nuclear localization [70].

The CIITA protein's AI and GRAVY values decreased by −305.10 % and −24.29 %, respectively, indicating a decrease in protein thermostability and an increase in hydrophilicity. It is noteworthy that the BCOR protein in the KMT2D::BCOR fusion showed an increased II, despite already having a high II value and a longer protein length than BCOR. This resulted in a considerable decrease in the stability of the BCOR while a more moderate change in the stability of KMT2D (Fig. 2C).

Overall, considerably lower SD were observed for the AI, II, and TpI values than for the other parameters (Fig. 2C). The highly similar aliphatic index values (AI) imply similar thermal stability for the fusion proteins. The instability indexes are also similar, the highest value was calculated for KMT2D::BCOR, indicating the lowest relative stability for

Table 2
The physicochemical parameters were calculated for the wild-type and the fusion proteins. Partner's average: the mean of the values for each pair of proteins involved in the fusion. Partner's average represents the mean of the values for each pair of proteins involved in the fusion, while Fusion's average represents the mean of the fusion protein's value. GRAVY: grand average hydropathy, AI: aliphatic index, II: instability index, EC: extinction coefficient at 280 nm, +R: the number of positively charged residues, −R: the number of negatively charged residues, TpI: the theoretical isoelectric point, Mw: molecular weight, Sl: sequence length, AA: amino-acids, BCOR-2: Isoform 2 of BCOR.

	GREY	AI	II	EC	+ R	+ R (%)	−R	−R (%)	TpI	Mw (Da)	Sl (AA)
BCOR	−0.668	68.38	55.26	160115	203	8.65	226	7.77	6.06	192188.64	1755
CCNB3	−0.504	82.01	59.29	60500	194	7.19	206	6.77	6.28	157916.13	1395
BCOR::CCNB3	−0.581	74.72	56.34	215115	376	8.08	424	7.17	5.91	337723.51	3038
MAML3	−0.817	57.01	69.23	46340	75	15.17	76	14.97	7.12	122292.88	1138
BCOR::MAML3	−0.725	64.3	61.71	196235	262	10.44	290	9.43	6.15	297852.28	2734
CLGN	−0.734	72.74	44.52	121850	75	8.13	132	4.62	4.57	70038.6	610
BCOR::CLGN	−0.706	67.92	54.94	240930	243	8.52	295	7.02	5.51	228230.9	2070
BCOR-2	−0.661	68.43	55.38	153125	201	8.56	218	7.89	6.22	188202.33	1721
MAML1	−0.71	60.11	71.17	43360	75	13.55	71.00	14.31	8.45	108054.21	1016
BCOR-2:MAML1	−0.68	65.21	60.70	196235	269	10.03	285.00	9.47	6.56	291722.14	2699
KMT2D	−0.648	67.03	76.61	296935	475	11.66	623	8.89	5.4	593389.41	5537
KMT2D::BCOR	−0.704	65.27	77.16	274350	474	11.56	650	8.43	5.22	587199.74	5480
ZC3H7B	−0.567	71.72	48.26	116710	127	7.69	129	7.57	6.94	109858.00	977
ZC3H7B::BCOR	−0.624	75.73	55.68	91565	124	8.94	171	6.48	5.02	122865.76	1108
CIITA	−0.196	90.72	50.97	105780	105	10.76	138	8.19	5.30	123513.99	1130
CIITA::BCOR	−0.794	68.68	59.75	85480	94	8.73	143	5.74	4.84	92428.75	821
RTL9	−0.137	61.27	71.76	37650	90	15.42	104	13.35	5.81	144279.61	1388
RTL9:BCOR-2	−0.409	64.52	62.82	164960	266	10.87	294	9.83	6.01	307716.11	2891
AHR	−0.56	71.53	50.52	68855	75	11.31	90.00	9.42	5.94	96147.36	848
AHR::BCOR-2	−0.68	69.14	53.23	149865	186	9.01	229.00	7.32	5.52	188836.99	1676
Partner's average	−0.591	70.09	57.68	135684	183	9.75	213	8.49	6.00	190187.34	1739
Fusion's average	−0.649	68.73	61.20	181234	263	9.59	324	7.73	5.52	282002.44	2592

this protein. The comparable averages that were calculated for the wild-type partners and the fusion proteins indicate only moderate change for these parameters.

Based on the comparison of the mean values of the partner proteins and the fusion proteins, the Gravy, AI, -R(%) and +R(%) values of the partner proteins showed a decrease, while II was increased. This led to an increase in hydrophobicity, a decrease in thermostability, a lower percentage of the charged amino acid residues, and a higher probability of protein degradation or denaturation.

3.3. Protein domains and structural changes

The CDD, UniProt, and InterPro databases were utilized to map domains and interaction sites for both wild-type and fusion proteins. Notably, the ankyrin (ANK) repeats and PUFD domain of BCOR are retained in all fusion proteins (Fig. 1). The MLLT3 binding site (Mbs) has been identified in all fusion proteins. However, in the BCOR-2 isoforms only 89 residues are presented, in contrast to the full 123 residues observed in the canonical BCOR. The BCL6 binding site (Bbs) is present in all $BCOR^{NT}$ fusions, but only in RTL9:BCOR-2 of the $BCOR^{CT}$ group. The fusion proteins in other partners exhibit different degrees of changes in the functional domains.

In the BCOR::CCNB3 fusion, a short N-terminal region of CCNB3 containing its destruction box (D-box) is lost, while both C-terminal cyclin boxes are retained. The BCOR::MAML3 fusion protein lacks the neurogenic mastermind-like, N-terminal (MamL-1) domain and part of the MamL1/3 transactivation domain 1 (MamL1/3 TAD1), but retains the MamL1/3 TAD2 domain. The BCOR::CLGN fusion results in the loss of 232 out of the 368 residues of the calreticulin domain, which is crucial for calcium ion binding [27].

The BCOR-2:MAML1 fusion protein retains the integrity of BCOR-2, while MAML1 loses its initial 39 residues, affecting 26 amino acid residues of the MamL-1 domain. In KMT2D, only the extended plant homeodomain (ePHD) zinc finger, five PHDs, and high mobility group (HMG) box domains are preserved, while the second ePHD, the F/Y-rich N terminus (FYRN), the F/Y-rich C terminus (FYRC) and the SET (including the post-SET) domains are absent. The ZC3H7B::BCOR fusion protein retains the LD motif and three tetratricopeptide repeats (TPR) but loses all zinc fingers. The CIITA::BCOR fusion protein contains only the acetyltransferase domain of CIITA, while its P\S\T rich and NACHT domains, as well as leucine-rich repeats (LRRs), are missing.

The RTL9:BCOR-2 fusion causes a breakpoint at the 1199th position of RTL9, resulting in the deletion of the entire capsid-like domain (1169–1316 region) near its C-terminus. For the AHR::BCOR-2 fusion, AHR retains all functional domains but loses 47 C-terminal residues, while BCOR-2 loses its Bbs domain (Fig. 1).

3.4. Gene ontology

We conducted an InterPro search to acquire PANTHER GO terms and define the functional attributes as well as compare the biological processes, molecular functions, and cellular components of the wild-type and fusion proteins. The results suggest that the GO terms associated with the BCOR protein, including negative regulation of transcription by RNA polymerase II (GO:0000122), transcription corepressor activity (GO:0003714), and nucleus (GO:0005634), were prevalent in the fusion proteins, as well. The GO terms of the wild-type proteins were not assigned to almost all fusion partners except BCOR. Interestingly, the novel KMT2D::BCOR protein retained the GO terms of only the wild-type KMT2D but not those of the BCOR (Table A.1). GO terms were not found for RTL9 and the RTL9:BCOR-2 proteins, as no domain information was available for these proteins in the database.

3.5. Predicted intracellular localization and signal peptides

The SignalP 6.0 tool was utilized to identify standard secretory signal peptides (Sec/SPI), while the DeepLoc 2.0 tool was used to estimate intracellular localization via calculating the probability values for localization in the cytoplasm, nucleus, extracellular cell membrane, mitochondrion, plastid, endoplasmic reticulum (ER), lysosome, Golgi

apparatus, and peroxisome. Additionally, the NetGPI-1.1 predictor was used to identify GPI-anchoring signals (Table 3).

According to the results of SignalP 6.0 analysis, only the CLGN protein contains a signal peptide, especially a secretory signal peptide with a probability of 99.92 %, although, the N-terminal region of CLGN encompassing this sequence motif is missing from the BCOR::CLGN fusion protein.

The DeepLoc 2.0 prediction indicated that seven of the ten wild-type proteins (BCOR, CCNB3, MAML3, MAML1, KMT2D, CIITA, and AHR) were most likely located in the nucleus, while CLGN was predicted to be located in the endoplasmic reticulum, which was consistent with the data available in the UniProt database. Based on the data available in the Human Protein Atlas database, KMT2D may have plasma membrane and cytosol localization. The subcellular localization of ZC3H7B seems to be uncertain, as the UniProt database information implies its nuclear localization but - in agreement with the data of Human Protein Atlas - we predicted that it was located in the cytoplasm (score: 0.6175). Data for the RTL9 protein was not available in the UniProt database. The predicted score of cytoplasmic localization was 0.6454, while that of the nuclear localization was also similar (score: 0.5456). Most recent findings revealed that mouse RTL9 protein is intracellular and is highly expressed in microglial lysosomes in the neonatal brain [71]. All fusion proteins were predicted to be located in the nucleus except BCOR::CLGN protein which may be localized in the endoplasmic reticulum (predicted score: 0.6739).

Based on DeepLoc 2.0 prediction all wild-type and fusion proteins excepting CLGN, BCOR::CLGN, CIITA::BCOR, and RTL9 proteins contain a nuclear export signal (NES). The CLGN protein was predicted to contain a signal peptide and a transmembrane domain, while BCOR::CLGN only contains a signal peptide. The presence of a nuclear localization signal (NLS) was predicted for the ZC3H7B::BCOR, KMT2D::BCOR, and RTL9:BCOR-2 next to the NES, while CIITA::BCOR and RTL9 protein contains NLS according to DeepLoc 2.0 prediction (Table 3). It should be mentioned that the prediction accuracy for NES is the lowest (<50 %), this is why NES was identified in most of the studied proteins.

Overall, the subcellular localization predicted by DeepLoc 2.0 was in agreement with the data available in UniProt and The Human Protein Atlas database in most cases (Table 3). No GPI-anchoring signals were predicted by using the NetGPI-1.1 predictor, which implied that neither the wild-type proteins nor the fusion proteins contain a lipid anchor and are not cell-surface proteins.

3.6. Prediction of intrinsically disordered regions

We used the IUPred3 online tool to identify the intrinsically disordered regions (IDRs) of the proteins. Subsequently, we compared the region$^{ANK\ +\ linker}$ and region^{PUFD} of the fusion proteins to the BCOR protein and in addition, the IUPred3 scores were also integrated into our statistical analyses by conducting Friedman's tests on the values of each region separately (Fig. 3).

Region$^{ANK\ +\ linker}$ of wBCOR (containing 1448–1633 residues) consists of ANK repeats and a linker region, region^{PUFD} includes the PUFD domain (consisting of the 1634–1748 residues).

We did not find significant differences between the IUPred3 IDR scores of the proteins within the region$^{ANK\ +\ linker}$ in at least two groups ($Q = 10.63$, $P = 0.3022$). However, significant differences were found within the region^{PUFD} in at least two groups (region^{PUFD}: $Q = 682$, $P < 0.0001$). Dunn's multiple comparison tests revealed significant rank means differences between the IUPred3 IDR scores of the BCOR and BCOR^{NT} proteins in both region$^{ANK\ +\ linker}$ and region^{PUFD}. There were significant differences in the rank means of BCOR vs. BCOR::CCNB3 ($Z = 8.178$, $P < 0.0001$), BCOR vs. BCOR::MAML3 ($Z = 11.63$, $P < 0.0001$), BCOR vs. BCOR::CLGN ($Z = 12.59$, $P < 0.0001$), and BCOR vs. BCOR-2:MAML1 ($Z = 9.256$, $P < 0.0001$). There was no statistically significant difference in the rank means between BCOR and ZC3H7B::BCOR, KMT2D::BCOR, CIITA::BCOR, RTL9:BCOR-2, and AHR::BCOR-2 fusion proteins (Fig. 3C and D).

The Z-score in Dunn's multiple comparisons serves as a statistic that indicates the magnitude of the difference between the two BCOR groups. In our analysis, the largest difference within the BCOR^{NT} group compared to the BCOR was observed in the BCOR::CLGN fusion ($Z = 12.59$), followed by the BCOR::MAML3 fusion ($Z = 11.63$), BCOR vs. BCOR-2:MAML1 ($Z = 9.256$), and then the BCOR::CCNB3 fusion protein ($Z = 8.13$). Despite these substantial differences, it is noteworthy that

Table 3
Intracellular localization and signals predicted for the wild-type and fusion proteins by using DeepLoc 2.0. NES: nuclear export signal,: Sec signal peptide (Sec/SPI), transmembrane domain, NLS: nuclear localization domain,: endoplasmic reticulum,: no data available, PML: Promyelocytic leukemia nuclear bodies, her: no signal peptides were found. For comparison, subcellular localization is indicated based on the UniProt database, if available. For DeepLoc predictions, only localizations reaching the threshold are shown.

Proteins	SignalP 6.0 (probability)	DeepLoc 2.0					UniProt data
		Predicted localization		Probability		Predicted signals	
BCOR	other: 1.0	Nucleus		0.8702		NES	Nucleus
CCNB3	other: 1.0	Cytoplasm	Nucleus	0.5879	0.7553	NES	Nucleus
BCOR::CCNB3	other: 1.0	Cytoplasm	Nucleus	0.5760	0.7130	NES	NA
MAML3	other: 1.0	Nucleus		0.8818		NES	Nucleus speckle
BCOR::MAML3	other: 1.0	Cytoplasm	Nucleus	0.4820	0.7875	NES	NA
CLGN	Signal Peptide (Sec/SPI): 0.9992	Endoplasmic reticulum		0.8276		SP, TD	Endoplasmic reticulum
BCOR::CLGN	other: 1.0	Endoplasmic reticulum		0.6739		SP, TD	NA
MAML1	other: 1.0	Nucleus		0.8781		NLS, NES	Nucleus
BCOR-2:MAML1	other: 1.0	Cytoplasm	Nucleus	0.4892	0.7800	NES	NA
ZC3H7B	other: 1.0	Cytoplasm		0.6175		NES	Nucleus
ZC3H7B::BCOR	other: 1.0	Cytoplasm	Nucleus	0.5808	0.7169	NLS, NES	NA
KMT2D	other: 1.0	Nucleus		0.7295		NES	Nucleus
KMT2D::BCOR	other: 1.0	Nucleus		0.8675		NLS, NES	NA
CIITA	other: 1.0	Cytoplasm	Nucleus	0.5189	0.6742	NLS, NES	Nucleus, PML body
CIITA::BCOR	other: 1.0	Nucleus		0.9452		NLS	NA
RTL9	other: 1.0	Cytoplasm	Nucleus	0.6454	0.5456	NLS	NA
RTL9:BCOR-2	other: 1.0	Nucleus		0.8576		NLS, NES	NA
AHR	other: 1.0	Cytoplasm	Nucleus	0.5320	0.802	NES	Cytoplasm, Nucleus
AHR::BCOR-2	other: 1.0	Cytoplasm	Nucleus	0.4890	0.8070	NES	NA

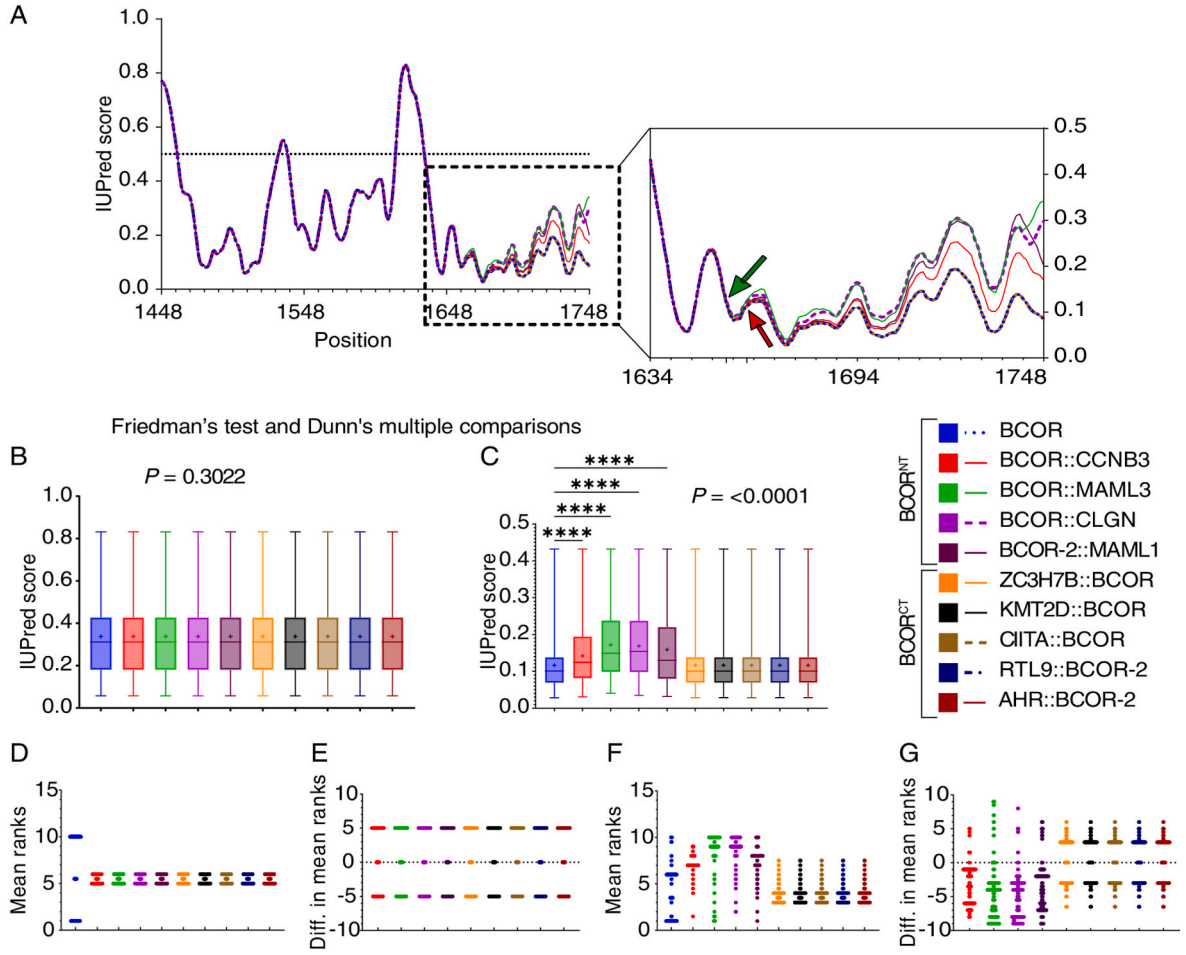

Fig. 3. Comparison of the disorder propensities of wild-type BCOR and fusion proteins. The disorder propensities were calculated by using the IUPred3 online tool. The region$^{ANK+linker}$ (encompassing 1448-1633 residues) and region^{PUFD} (encompassing 1634-1748 residues) of BCOR were investigated in the context of the BCOR and the fusion proteins. **(A)** IUPred3 scores for IDRs of BCOR and its fusion proteins in the coordinate system. The green arrow shows the first point of separation. The red arrow points to the groups' separation point. **(B,C)**, Friedman's test, and Dunn's multiple comparisons with *P* values for BCOR and the fusion proteins ($n = 8$). The horizontal line shows the median, plus symbol (+) the mean value, and error bars represent the min-max values. **(D,F)** The mean rank plot of Friedman's test. **(E,G),** Friedman's test's differences in mean ranks between pairs of groups (BCOR vs. each fusion). Region$^{ANK+linker}$ and region^{PUFD} are represented by (**A,B,D,E**) and (**A,C,F,G**) respectively. **BCOR^{NT}**: BCOR is located at the N-terminal end of the fusion protein. **BCOR^{CT}**: BCOR is located at the C-terminal end of the fusion protein. **Diff.**: Differences. ****P < 0.0001.

none of its IDR scores reached the 0.5 disordered cut-off threshold. Consequently, despite notable differences in IDR scores obtained for the region^{PUFD}, the IDR scores did not exceed the cut-off, indicating that this region retains its structured nature.

According to the IPred3 prediction, the BCOR^{NT} fusion proteins have increased disorder propensity in the region^{PUFD}, which might potentially affect the protein's interaction with the PCGF1 protein's RAWUL domain and with the KDM2B protein's leucine-rich repeat (LRR) region. Such changes were not predicted for the members of the BCOR^{CT} group.

3.7. Statistical binding affinity analyzes of the BCOR-PCGF1 dimer

As described above (Fig. 1), the PUFD domain retains the integrity of its sequence, but despite this, it is conceivable that the fusions may influence the interactions between BCOR and other proteins. This possibility has been investigated based on the predicted three-dimensional structures of the fusion proteins and the wBCOR complexed with PCGF1. The protein structures were built up by using AF3. Subsequent comparisons were made to assess variations in ΔG values in dimer complexes against wild-type counterparts (Fig. 4). Due to the complexity of the statistical analyses required, 50 predicted structures per protein were investigated, which was expected to provide sufficient reliability for the analyses (Fig. 4). Calculations of ΔG values for the dimers were conducted by using PRODIGY. A comprehensive study was conducted on the proteins in full-length (Fig. 4A), in RAWUL-PUFD domain length (Fig. 4B), and 4hpl length [15] (Fig. 4C). Additionally, six different interaction surfaces and prominent residues were analyzed, as described below (Fig. 4D–I), the RAWUL-PUFD binding sites were described by Junco et al. [15]. The nine analyses were carried out to determine the changes in different regions in binding and to identify how these were affected by the consequences of fusion events.

First, we analyzed the ΔG of the full-length proteins' dimers by using an ordinary one-way ANOVA test. We found a statistically significant difference in the mean ΔG values between at least two groups ($F(8, 441) = [111.9]$, $P < 0.0001$). Based on Dunnett's multiple comparisons test, the mean ΔG values were significantly different between the BCOR-PCGF1 and the fusion type groups. (Figs. 4A and 5). According to the mean ΔG values, the BCOR::CCNB3-, BCOR::MAML3-, BCOR-2:MAML1-, and RTL9:BCOR-2-PCGF1 had significantly lower, while - in contrast to this - the BCOR::CLGN-, ZC3H7B::BCOR-2-, CIITA::BCOR-, and AHR-BCOR-2-PCGF1 had higher ΔG values (Fig. 4A).

In the second comparison, the ΔG values of the nine groups within the RAWUL-PUFD domain length were subjected to analysis. The studied regions included 1634–1748 and 167–255 residues of BCOR and

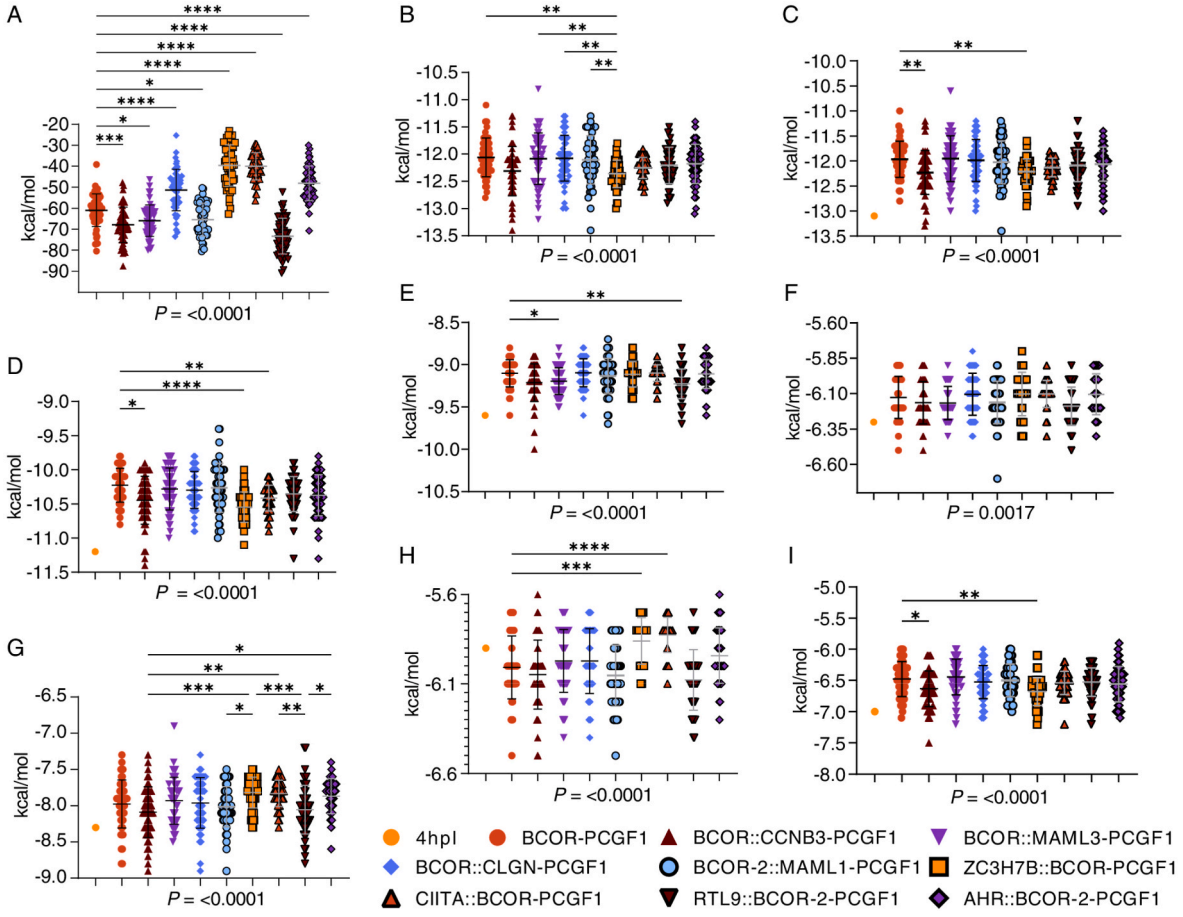

Fig. 4. Comparative binding affinities of BCOR-PCGF1 dimers with various BRS fusion proteins. We conducted 50 predictions for each of the nine protein dimers (one wild-type and eight fusion proteins), comprising BCOR or BRS fusion proteins and PCGF1. The variations in binding affinity (ΔG) were analyzed. Both the full-length proteins (<BA></B>) and the RAWUL-PUFD domains (B–I) were investigated. An experimentally determined RAWUL-PUFD dimer (PDB ID: 4hpl) served as an additional control in eight scopes (C–I). Statistical analyses included Ordinary one-way ANOVA (<BA></B>, <BI></B>) and Kruskal-Wallis tests (B–H). <A:B></B> Comparison of the full-length dimer BCOR-PCGF1 proteins revealed significant differences in ΔG values between the BCOR-PCGF1 complex and the eight fusion dimers. <BB></B>: Comparison in RAWUL-PUFD domain length, BCOR (1634–1748), and PCGF1 (167–255), demonstrated statistically significant differences in mean ΔG values. <BC></B>: Predicted structure binding affinities within the 4hpl range were compared against the 4hpl and wild-type predicted structure (BCOR-PCGF1). Two fusion dimers exhibited significantly lower ΔG values compared to BCOR-PCGF1. D–I: Binding affinities in various ranges (<BD></B>: ISNT/CT, <BE></B>: ISNT, <BF></B>: Phe1639/Phe1641, <BG></B>: ISLeu, <BH></B>: Leu1706, <BI></B>: ISCT) revealed significant differences between fusion protein-containing dimers and the BCOR-PCGF1 control. Interaction surfaces include Val1636-Asn1651, Ser1704-Asp1712, and Gly1738-Asp1748, with specific regions identified as Interaction-SurfaceNT/CT (ISNT/CT), Interaction-SurfaceLeu (ISLeu), and individual residues Leu1706 and Phe1639/Phe1641. Error bars represent mean values with standard deviations (*P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001).

PCGF1, respectively. A Kruskal-Wallis test indicated that there was a significant difference in the RAWUL-PUFD domain length across nine groups, (H)(8) = 35.37, (*P*) < 0.0001. The mean ranks of the Δ*G* values in the domain range were significantly different among the groups (Fig. 4B). Dunn's multiple comparisons test was also performed, and the analysis revealed significant differences between the BCOR-PCGF1 and CIITA::BCOR-PCGF1, where the fusion protein had lower Δ*G* values (Fig. 4B, Fig Fig. 5).

In the third analysis, the 4hpl crystal structure was used as a control, as it contains the established RAWUL-PUFD dimer. We used PyMOL to set the lengths of the AF3 structures to make them identical to the protein included in the crystal structure (4hpl). A comparison was then made between these structures and the BCOR-PCGF1 predicted structure with fusion type dimers. A Kruskal-Wallis test indicated a significant difference in the 4hpl length (PCGF1 RAWUL: from 167 to 177 and 185–254, BCOR PUFD: from 185 to 254) across 10 groups, (H)(9) = 33.96, (*P*) < 0.0001 (Fig. 4C). The Dunn's multiple comparisons test revealed that the BCOR::CCNB3-, and ZC3H7B::BCOR-PCGF1 had lower Δ*G* values. Nevertheless, no statistically significant difference was observed between 4hpl and the fusion dimers (Fig. 4C).

The PRODIGY analysis also revealed the intermolecular contacts (ICs) between the two proteins (BCOR-PCGF1), which agrees with the binding surfaces described previously by Junco et al. [15], and the MAPIYA tool identified the possible interaction forces. We also found that the interactions between RAWUL and PUFD domains were the densest in two main areas (Fig. 6A). The first involved the PUFD region connecting the known RAWUL β-sheet and loop binding site [15] (Fig. 6B). The regions of PUFD that interact with RAWUL's binding sites have been identified and include the Val1636 residue, the N-terminal β-sheet (Phe1637-Ser1642), the subsequent Glu1643-Asn1651 stretch, and the loop (Gly1738, Ser1739) that transitions into the C-terminal β-sheet (Ser1740-Leu1744) downstream residues His1745-Asp1748 (termed Interaction-Surface$^{NT/CT}$). The other interaction surface with frequent ICs in RAWUL-PUFD contact was centered around Leu1706 within Ser1704-Asp1712, named Interaction-Surface^{Leu} (Fig. 6). The Interaction-Surface$^{NT/CT}$ includes Phe1639/Phe1641 that engage in hydrophobic contact with RAWUL's Val206 (Fig. 6), while Interaction-Surface^{Leu}'s Leu1706 interacts with RAWUL's Leu Cage [15] (Fig. 6).

The prediction of binding affinities was made specifically for the

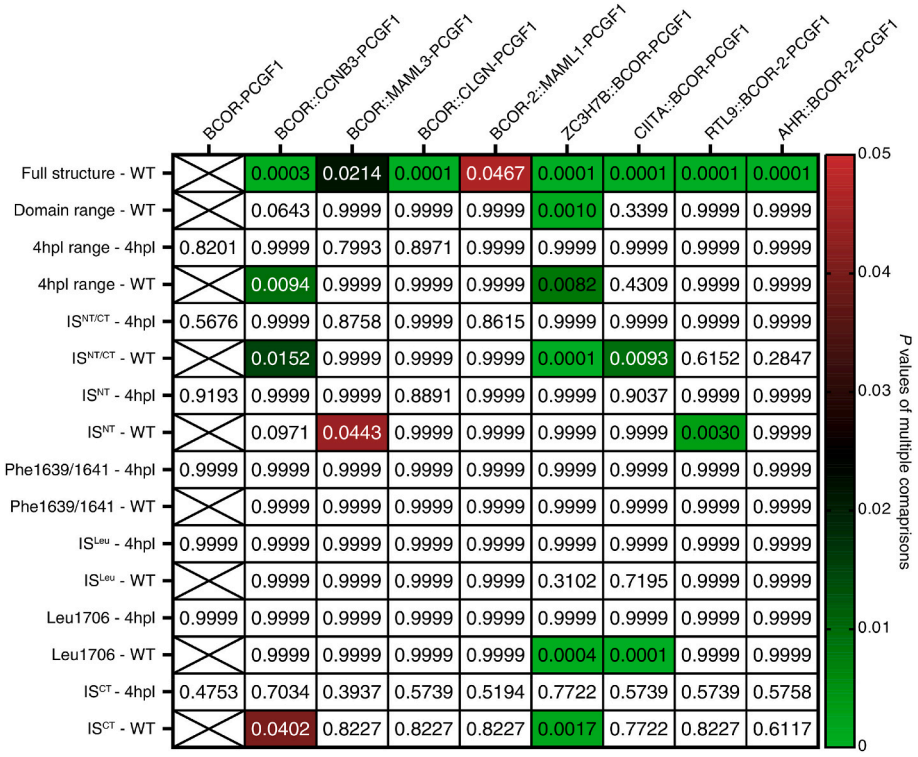

Fig. 5. Heatmap of post-hoc statistical comparisons of binding affinities Heatmap derived from post-hoc analysis of nine differential statistical multiple comparisons. Ordinary one-way ANOVA and Kruskal-Wallis tests revealed statistically significant differences between at least two groups within each scope. Multiple comparison tests were employed to discern differences between control groups—BCOR-PCGF1 (WT) and 4hpl—and the fusion proteins. Columns represent the scope and control group (Domain range – WT), while rows indicate the fusion protein in dimer with PCGF1. In the heatmap, numbers correspond to the *P* values of the comparisons. The most significant differences in binding affinity values were observed in the Full structure scope between the BCOR-PCGF1 and the fusion proteins. No significant differences were detected between the 4hpl and any other group in any scope, potentially due to its singular value.

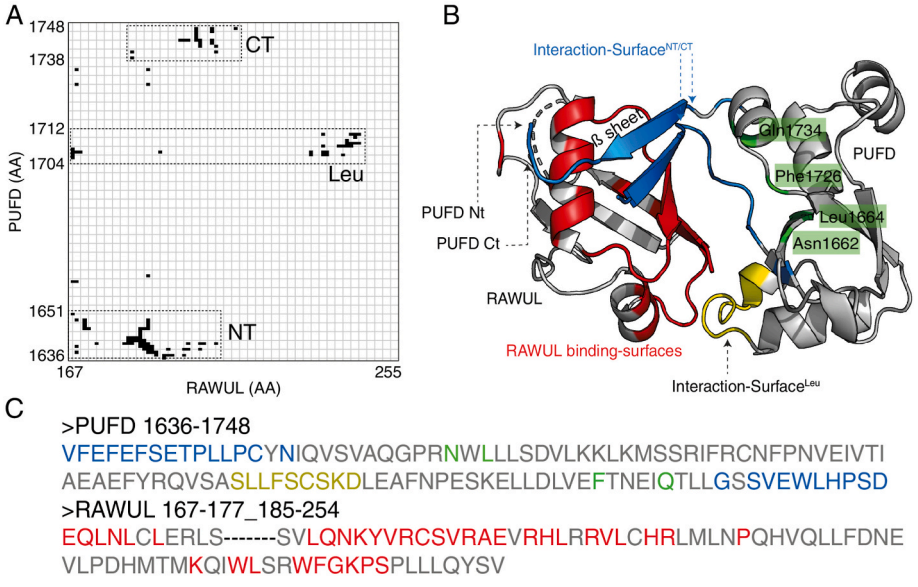

Fig. 6. Interaction surfaces of PUFD domain in the RAWUL-PUFD dimer. The analysis with PRODIGY revealed that the PUFD domain had three interaction surfaces. The Interaction-SurfaceNT (Val1636- Asn1651), Interaction-SurfaceCT (Gly1738-Asp1748), and the Interaction-SurfaceLeu (Ser1704-Asp1712). In these sections, the intermolecular contacts (ICs) with the RAWUL domain were concentrated. (<A)B></B> Interaction map of the RAWUL-PUFD dimer. (<B)B></B> The interaction surfaces and other ICs of the RAWUL-PUFD dimer (PDB ID:4hpl). (<C)B></B> The amino acid sequences (AA) of the protein dimer. Blue color marks the Interaction-SurfaceNT/CT, green the other ICs, yellow the Interaction-SurfaceLeu, grey the non-interacting residues, and red the interacting residues of the RAWUL domain. The coloring was consistent in (<BB></B>) and (<BC></B>). <Ntb>: N-terminus, <Ctb>: C-terminus.

interaction sites of the PUFD, focusing on crucial residues such as Phe1639/Phe1641, and Leu1706 [15], concerning their connection with the RAWUL domain.

Kruskal-Wallis test revealed a significant difference in the ΔG values of the $IS^{NT/CT}$ between the 10 groups, (H)(9) = 52.76, (P) < 0.0001 (Fig. 4D). Significantly lower values were obtained for the BCOR::CCCNB3-, ZC3H7B::BCOR-, and the CIITA::BCOR-PCGF1 groups as compared to the BCOR-PCGF1, based on the Dunn's multiple comparisons test (Fig. 4D, Fig Fig. 5).

A Kruskal-Wallis test indicated a significant difference in the ΔG values of the IS^{NT} range across 10 groups, (H)(9) = 38.91, (P) < 0.0001 (Fig. 4E). Dunn's multiple comparisons test found that the BCOR::MAML3,- and the RTL9:BCOR-2-PCGF1 groups had significantly lower ΔG values compared to the BCOR-PCGF1, but not to the 4hpl (Fig. 4E, Fig Fig. 5). However, the Kruskal-Wallis test indicated that there was also a significant difference in the ΔG values of the Phe1639/Phe1641 residues across 10 groups, (H)(9) = 26.51, (P) = 0.0017 (Fig. 4F), but the Dunn's multiple comparisons test could find the significant differences among the groups, not even between the fusion protein's (Fig. 4F, Fig Fig. 5).

A Kruskal-Wallis test indicated a significant difference in the ΔG values of IS^{Leu} across 10 groups, (H)(9) = 44.34, (P) < 0.0001 (Fig. 4G). Dunn's multiple comparisons test showed no significant differences between the control groups and the fusion type groups, only among the fusion type groups (Fig. 4G, Fig Fig. 5).

Kruskal-Wallis test indicated a significant difference in the ΔG values of Leu1706 residue across 10 groups, (H)(9) = 94.17, (P) < 0.0001 (Fig. 4H). There was also a significantly weaker binding affinity in the ZC3H7B::BCOR-, and the CIITA::BCOR-PCGF1 groups compared to the BCOR-PCGF1 (Fig. 4H, Fig Fig. 5).

Finally, we found a statistically significant difference in the mean ΔG values of the IS^{CT} range between at least two groups (F (8, 441) = [111.9], P < 0.0001). Holm-Šídák's multiple comparisons test found significantly lower ΔG values in the BCOR::CCNB3-, and ZC3H7B::BCOR-PCGF1 groups compared to the BCOR-PCGF1 (Figs. 4I and 5).

Our extensive analyses of the BCOR-PCGF1 dimers revealed significant differences in the binding affinities between various fusion proteins (Fig. 4, Fig Fig. 5). The fusion dimers exhibited altered ΔG values as compared to the wild-type, particularly within the full-length structure (Fig. 4A) and specific interaction surfaces (Fig. 4D -I, Fig. 5). Notably, four BCOR fusion proteins – BCOR::CCNB3, BCOR::MAML3, BCOR-2:MAML1 and RTL9:BCOR-2 – showed increased, while other four ones – BCOR::CLGN, ZC3H7B::BCOR, CIITA::BCOR, and AHR::BCOR-2 - decreased binding affinities in the dimers as compared to the wild-type. Of the nine comparisons, all had significant differences among the groups, and in seven the post hoc test found statistically significant differences between the BCOR-PCGF1 and at least one fusion dimer. As compared to the wild-type, the ZC3H7B::BCOR fusion protein had the most significant changes (significant difference in six cases) (Fig. 5), wherein in two comparisons, it decreased (Fig. 4A–H), and it indicated increased binding affinity in the dimers three times. In contrast, the BCOR::CCNB3 had the second most (four), and in all cases, it showed

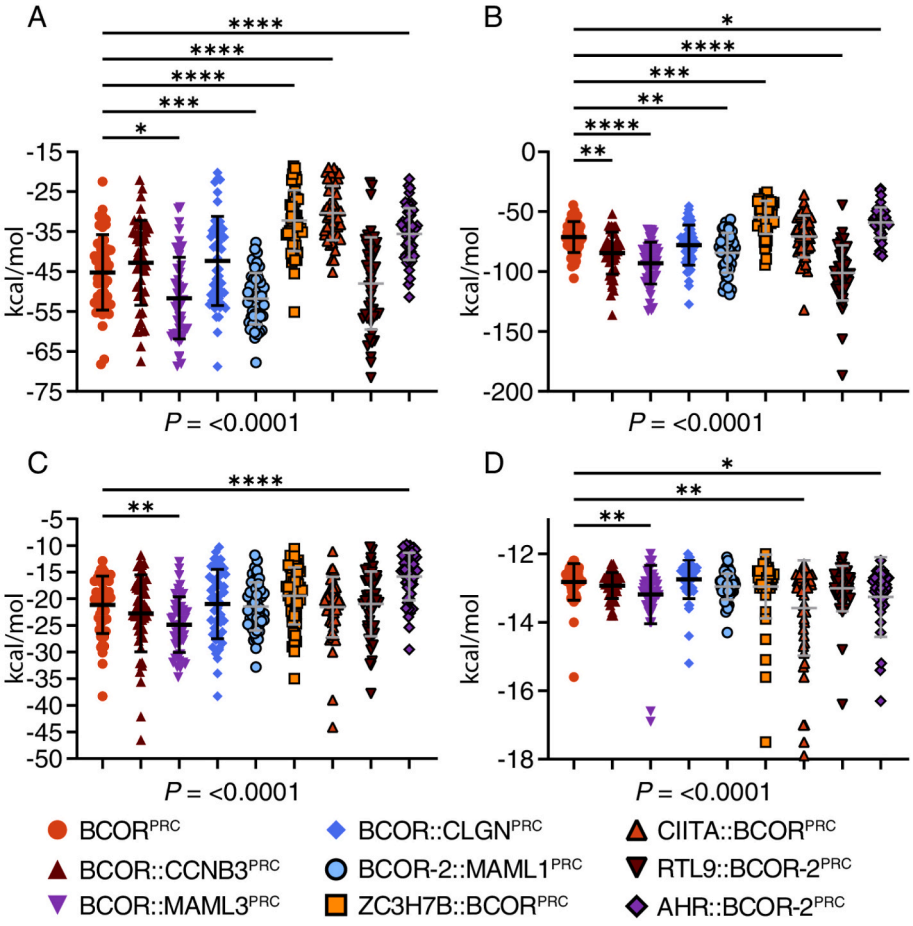

Fig. 7. Binding affinity comparisons of PRC1.1 partner proteins between wild and fusion type BCOR proteins. The variations in binding affinity among the subunits comprising the BRS fusion proteins and three additional PRC1.1 components (PCGF1, KDM2B, SKP1) were subjected to analysis. (<BA></B>) The investigation of the binding affinity between BCOR and PCGF1 using a Brown-Forsythe ANOVA test. (<BB></B>), (<BC></B>), and (<BD></B>) respectively analyze the binding affinities between BCOR and KDM2B, PCGF1 and KDM2B, and KDM2B and SKP1 using Kruskal-Wallis tests. Error bars show the mean values with standard deviation.*P < 0.05, **P < 0.01,***P < 0.001, ****P < 0.0001.

increased binding affinity (Fig. 4A–C,D,I) (Fig. 5). Out of the 72 BCOR-PCGF1 vs. fusion type multiple comparisons, we found 20 significant differences in binding affinity (28 %) (Fig. 5).

These findings indicate that the fusion proteins significantly impact the dimerization properties of BCOR-PCGF1, underscoring the potential functional implications of these alterations. No substantial differences were observed in the 4hpl control group, likely due to its limited dataset.

3.8. Influence of BRS Fusion Proteins on the Binding Affinities within the non-canonic Polycomb Repressive Complex1

To further expand our understanding on the role of BRS protein fusions in PRC1.1, we modeled the wild-type and fusion proteins with PCGF1, KDM2B, and SKP1 subunits. We predicted their ΔG values by using PRODIGY and applied statistical analyzes to determine the differences in the significant interactions, by comparing the binding affinities of nine fusion proteins (Fusion partners^{PRC}) to a wild-type complex (BCOR^{PRC}) (Fig. 7).

A Brown-Forsythe and Welch's ANOVA test was utilized to compare the ΔG values of the BCOR-PCGF1 pairs in the PCR1.1 complexes (Fig. 7A). The analyses found statistically significant differences between the groups (F*)(8.000, 375.2) = 38.01, (P) < 0.0001, (W)(8.000, 183.2) = 54.67, (P) < 0.0001. Dunnett's T3 multiple comparisons test found that values were significantly different between BCOR^{PRC} vs. BCOR::MAML3^{PRC} (P = 0.0114), BCOR-2:MAML1^{PRC} (P = 0.0008), ZC3H7B::BCOR^{PRC} (P = <0.0001), CIITA::BCOR^{PRC} (P = <0.0001), and vs. AHR::BCOR-2^{PRC}. There was no statistically significant difference between BCOR^{PRC} vs. BCOR::CCNB3^{PRC} (0.8719), BCOR::CLGN^{PRC} (0.7544), RTL9:BCOR-2^{PRC} (0.8188) (Fig. 7A).

To compare the ΔGs of the BCOR-KDM2B proteins among the nine groups we used the Kruskal-Wallis test (Fig. 7B). There were statistically significant differences between the groups (H)(8) = 201.4, (P) < 0.0001. Dunnett's multiple comparisons test found significant differences between BCOR^{PRC} vs. BCOR::CCNB3^{PRC} (P = 0.0080), BCOR::MAML3^{PRC} (P = <0.0001), BCOR-2:MAML1^{PRC} (P = 0.0076), ZC3H7B::BCOR^{PRC} (P = 0.0004), BCOR^{PRC} vs. RTL9-BCOR-2^{PRC} (P = <0.0001), and vs. AHR::BCOR-2^{PRC} (P = 0.148). There was no statistically significant difference between BCOR^{PRC} vs. BCOR::CLGN^{PRC} (0.7025), and CIITA-BCOR^{PRC} (>0.9999) (Fig. 7B).

We compared the PCGF1-KDM2B interactions' binding affinity as well (Fig. 7C). Kruskal-Wallis test found statistically significant differences between the groups (H)(8) = 70.17, (P) < 0.0001. Dunnett's multiple comparison tests found significant differences between the BCOR^{PRC} vs. BCOR::MAML3^{PRC} (P = 0.0082) and vs. RTL9:BCOR-2^{PRC} (P = <0.0001). There was no statistically significant difference between BCOR^{PRC} vs. the other six fusion proteins containing complex (P values were >0.9999 respectively).

Finally, we investigated and compared the KDM2B-SPK1 pairs in the PRC1.1 (Fig. 7D). The Kruskal-Wallis test revealed statistically significant differences between the groups (H)(8) = 42.91, (P) < 0.0001. The Dunnett multiple comparisons test revealed a statistically significant difference between the BCOR^{PRC} and BCOR::MAML3^{PRC} (P = 0.0029), CIITA-BCOR^{PRC} (P = 0.0047), and vs. AHR::BCOR-2^{PRC} (P = 0.0283). However, no statistically significant difference was observed in BCOR^{PRC} vs. the remaining five fusion protein-containing complexes (P values ranged from 0.1169 to >0.9999) (Fig. 7D).

The results demonstrate notable discrepancies in the ΔG values between the BCOR-PCGF1, BCOR-KDM2B, PCGF1-KDM2B, and KDM2B-SKP1 protein interactions within the PRC1.1 complex. The statistical analyses, including ANOVA and Kruskal-Wallis tests, and 32 multiple comparisons, demonstrate that the binding affinities of the fusion proteins were altered 16 times (50 %) as compared to the wild-type PRC1.1 complex (Fig. 7). Notably, of the eight fusion protein tetramer types, five (BCOR::CCNB3, BCOR::CLGN, ZC3H7B::BCOR, CIITA::BCOR, and AHR::BCOR-2) showed a decrease in binding affinity, while an increase was observed for three (BCOR::MAML3, BCOR-2:MAML1, and RTL9:BCOR-2) (Fig. 7A). Only the BCOR::MAML3 and the AHR::BCOR-2 fusion proteins exhibited statistically significant differences in all four comparisons (Fig. 7). The BCOR::MAML3 showed increase in four (Fig. 7), while the AHR::BCOR-2 showed decrease in three cases (Fig. 7A–C) and one instance of increased binding affinity within the tetramer complex (Fig. 7D). The PUFD domain was found to be intact, yet computational predictions indicated the potential for impacts on these protein-protein interactions, suggesting modifications in binding dynamics in the PRC1.1.

3.9. Molecular dynamic analysis of the dimer complexes

The MD simulations were conducted to complement structural and binding affinity analyses by providing dynamic insights into the stability, interaction strength, and thermodynamic properties of BCOR-PCGF1 dimers, including their fusion variants.

The MD simulations performed on BCOR-PCGF1 dimers, including various BRS fusion proteins, revealed notable differences in structural stability and interaction dynamics. Root mean square deviation (RMSD) analyses (Fig. 8A,) demonstrated significant variations in structural stability among the dimers, with the Kruskal-Wallis test confirming statistically significant differences (H)(9) = 23926, (P) < 0.0001. Dunnett's multiple comparison tests found significant differences between the BCOR-PCGF1 and all the dimers of the fusion proteins (Fig. 8B).

The RMSD analysis revealed that the wild-type BCOR-PCGF1, fusion type BCOR::CCNB3-PCGF1, BCOR-2:MAML1-PCGF1, and RTL9:BCOR-2-PCGF1 complexes reached stability more rapidly (within 3000 ps) and maintained lower average RMSD values (<0.4 nm) throughout the simulation, indicating a more stable conformation. The fusion variants BCOR::CCNB3-PCGF1, BCOR-2:MAML1-PCGF1, and RTL9:BCOR-2-PCGF1 also stabilized relatively quickly (within 3000 ps) with moderate RMSD values (<0.4 nm). In contrast, other fusion variants, particularly, BCOR::MAML3, BCOR::CLGN, and especially ZC3H7B::BCOR, CIITA::BCOR, and AHR::BCOR-2, showed elevated RMSD values and continued fluctuations even after 5000 ps, suggesting persistent structural instability and conformational flexibility in these fusion proteins.

In the Root mean square fluctuation (RMSF) analysis, the fusion type sequences were aligned to the wBCOR. The fusion proteins showed altered fluctuations compared to wBCOR, particularly within the functionally critical PUFD domain (1634–1748). The BCOR::MAML3 and CIITA::BCOR variants exhibited significantly higher residue-level fluctuations in this domain, with peak RMSF values reaching 0.3–0.4 nm, compared to the more stable wild-type complex (RMSF values < 0.3 nm). Notable the lowest fluctuation was observed in the case of the BCOR::CCNB3 (Fig. 8G), supporting its structural stability observed in the RMSD analysis.

Similarly, the gmx_MMPBSA binding affinity analyses (Fig. 8C,D) indicated distinct dimer thermodynamic profiles. Kruskal-Wallis test confirming statistically significant differences (H)(9) = 308.6, (P) < 0.0001. Dunnett's multiple comparison tests found significant differences between the BCOR-PCGF1 and the following dimers of the fusion proteins: BCOR::CLGN (P < 0.0001), ZC3H7B::BCOR (P < 0.0001), CIITA::BCOR (P < 0.0001), RTL9:BCOR-2 (P = 0.0007), AHR::BCOR-2 (P < 0.0001). All the fusion proteins showed reduced binding affinities compared to the wild-type BCOR-PCGF1 complex (mean ΔG = −142 kcal/mol). The weakest mean ΔG was found to be −90.24 kcal/mol in the case of CIITA::BCOR-PCGF1, while the strongest mean ΔG was calculated for the BCOR::CCNB3-PCGF1 (−141.4 kcal/mol) (Fig. 8D).

The determination of the number of hydrogen bonds-exclusively between the two protein chains as water-mediated hydrogen bonds were not included in the primary analysis - (Fig. 8E,F) further highlighted differences in the interactions' strength, with a significant variation in the number of the bonds between the dimers by Kruskal-Wallis test (H)(9) = 26191, (P) < 0.0001. The post-hoc Dunnett's multiple comparisons revealed statistically significant differences between

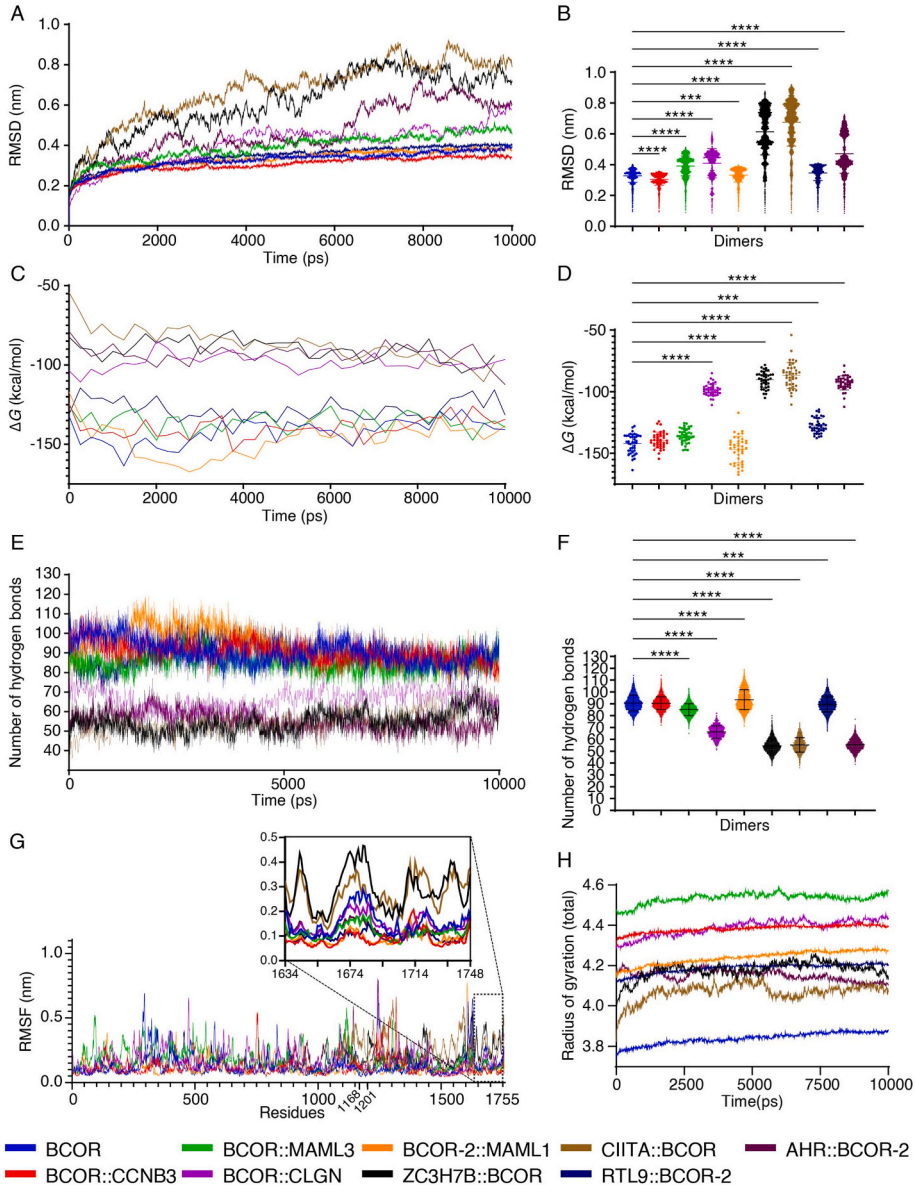

Fig. 8. Molecular dynamics simulations of BCOR-PCGF1 dimers with various BRS fusion proteins over 10 ns. (A, B) Root mean square deviation (RMSD), (C, D) binding affinity (ΔG), and (E, F) the number of hydrogen bonds were evaluated to assess the structural stability, interaction strength, and thermodynamic stability of the dimers. Statistical significance was determined using the Kruskal-Wallis test, revealing highly significant differences ($P < 0.0001$ in all tests) between the nine dimers in terms of RMSD (B), ΔG (D), and hydrogen bonds (F). Error bars represent mean values with standard deviation. (G) Root Mean Square Fluctuation (RMSF) analysis of BCOR sequences, highlighting changed flexibility in the PUFD domain for fusion variants compared to wild-type. The numbers 1168–1201 indicate the specific gap within the BCOR-2 protein sequences. (H) The radius of gyration analysis shows a notable increase in structural expansion for fusion variants relative to wild-type dimers. ***$P < 0.001$; ****$P < 0.0001$.

the BCOR-PCGF1 and all the fusion dimers (P values ranged from 0.0009 to <0.0001) excepting BCOR::CCNB3 ($P > 0.9999$). The lowest number of hydrogen bonds was determined for the ZC3H7B::BCOR-PCGF1 dimer (54.4 mean value), while the highest for the wild-type (90.41 mean value) and for the BCOR::CCNB3-PCGF1 dimer (90.27 mean value).

The gmx_MMPBSA binding affinities showed a correlation with the PRODGY results (Pearson test: $P = 0.00014$, $R^2 = 0.8885$). Also, there was a correlation between the number of hydrogen bonds and the RMSD (Spearman test: $P < 0.0001$, $R^2 = -0.530$) and the ΔG values ($P < 0.007$, $R^2 = -0.417$), but not between RMSD and ΔG values ($P = 0.862$). These correlations demonstrate that regions with high RMSD values correspond to fewer stable hydrogen bonds, creating a mechanistic link between residue flexibility and diminished interaction strength. The BCOR-PCGF1 dimers containing vBCOR and fusion partners, including CCNB3, MAML3, MAML1, and RTL9 formed a distinct group with higher hydrogen bond numbers and stronger binding affinities compared to those with CLGN, ZC3H7B, CIITA, and AHR fusion partners (Fig. 8C–D). Dimers with fewer hydrogen bonds exhibited weaker binding affinities, indicating fewer stable interactions. The reduced binding affinity of ZC3H7B, CIITA, and AHR fusion proteins, the reduced binding affinity can be attributed to the truncation of the BCOR protein in these fusions, which disrupts key interaction surfaces. However, for BCOR::CLGN, the entire BCOR protein is retained in the fusion protein. Despite this, the BCOR::CLGN dimer showed decreased binding affinity and reduced hydrogen bond numbers. This observation suggests that the structural alterations in the CLGN partner may impede its capacity to interact efficiently with PCGF1, potentially resulting in a change in the overall stability of the dimer.

The radius of gyration (Rg) analysis provided further insights into structural compactness across wild-type BCOR-PCGF1 dimers and their

fusion variants (Fig. 8H). The wild-type complex manifested the most compact conformation, exhibiting the lowest Rg values (3.8–3.9 nm), thereby indicating tight structural packing throughout the simulation. Conversely, all fusion proteins exhibited higher Rg values (4.0–4.6 nm), indicative of more extended conformations. Notably, although protein length variations influenced the height of Rg curves, many fusion proteins showed Rg values disproportionate to their sequence length, suggesting that structural rearrangements rather than mere size differences drive these conformational changes. The BCOR::MAML3 fusion exhibited the highest Rg values (4.4–4.6 nm), suggesting substantial structural expansion, while CIITA::BCOR exhibited the lowest Rg among fusion variants (4.0–4.1 nm). It was observed that all systems reached equilibrium within 2500 ps, after which Rg values stabilized with minor fluctuations, except in CIITA::BCOR and ZC3H7B::BCOR fusions. The findings indicate that fusion events notably alter the spatial arrangement and compactness of the BCOR complex, potentially affecting its interaction with PCGF1 and subsequent functional properties in the PRC1.1 complex assembly.

The combined analysis of RMSD, RMSF, hydrogen bond dynamics, and Rg values reveals a comprehensive picture of how fusion events alter BCOR-PCGF1 dimer stability. Two fusions, ZC3H7B::BCOR and CIITA::BCOR, exhibited the most pronounced structural instability, characterized by the lowest binding affinity, fewest hydrogen bonds, highest RMSD values, greatest PUFD domain flexibility, and most unstable Rg trajectories. This structure-dynamics-function relationship explains why fusion proteins with preserved domain architecture nevertheless display compromised binding properties, as local flexibility changes propagate to global structural destabilization. These findings are consistent with the trends in binding affinities observed in Fig. 4A for these dimers.

4. Discussion

BCOR was first identified as a novel interacting corepressor of *BCL*-6 that enhances BCL-6-mediated transcriptional repression [13]. Later, the specific BCL6 binding motif was found at the BCOR's 498–514 site [72]. Additionally, it was discovered that the BCOR directly interacts with the transcriptional regulator *AF9* (*MLLT3*) [73] and binds to chimeric MLL–AF9 proteins (MLL: mixed lineage leukemia), although, only some isoforms of BCOR bind AF9 [74]. BCOR is a known part of the PRC1.1 complex. During the assembly of the complex, an interaction is formed between the BCOR's PUFD domain and the PCGF1 protein's RAWUL domain. Additionally, the PUFD's ANK repeats and the linker region must also interact with the KDM2B's C terminus [17,18].

A new subset of gene fusion in bone sarcoma was described first in 2013 by Pierron et al. [22]. Since then six additional fusion genes have also been identified [11,22,24–27] and classified into the third distinct subset of "undifferentiated small round cell sarcomas of bone and soft tissue" [10,22]. The pathological features have been thoroughly investigated, but the exact fusion proteins have not been investigated at the structural level so far. In this study, we utilized *in silico* approaches to explore the properties of these uncommon fusion proteins. For this purpose, we aimed to analyze the sequences of the known *BCOR*-rearrangements in sarcomas and compare the fusion proteins' domain architectures (Fig. 1) and physicochemical characteristics (Table 1, Fig. 2), as well as to predict and analyze the PANTHER GO terms (Table A.1), signal peptides, intramolecular localizations (Table 3), and IDRs (Fig. 3). The 3D structures of BCOR-PCGF1 complexes were also investigated (Figs. 4, 5 and 6A, Fig.Fig. 7A), with a special emphasis on the binding affinity between the RAWUL-PUFD domains (Fig. 4B–I), of BCOR and partner proteins. In addition, new BCOR interaction surfaces of the RAWUL-PUFD dimer were also revealed (Fig. 6). Furthermore, the impact of protein fusions on the binding affinities was investigated at the level of tetramer PRC1.1 (Fig. 7).

The nine fusion proteins were classified into two groups based on the localization of BCOR in the fusion proteins: $BCOR^{NT}$ and $BCOR^{CT}$. The characteristics of the fusion proteins are described as follows. The BCOR::CCNB3 fusion protein has lost its destruction box motif. This motif acts as a recognition signal for the ubiquitin-proteasome pathway and is essential for the mitotic destruction of cyclin B3, which results in prolonged cell cycle progression and unchecked cell division, a hallmark of tumor formation [75,76]. Additionally, the ectopic expression of the fusion protein in NIH3T3 cells results in the aberrant expression of CCNB3 and a corresponding increase in the number of cells entering the S phase of the cell cycle [11,22,77]. The truncation in MAML3 and MAML1 affected its 1–157 and 1–39 region containing the N-terminal MamL-1 domain, which is responsible for the interaction with the ankyrin repeat region of the Notch proteins such as NOTCH1, NOTCH2, NOTCH3, and NOTCH4 [78–81]. The truncation of the MamL-1 domain in the *BCOR*::*MAML3* fusion disrupts Notch signaling, leading to aberrant cell fate determination and contributing to oncogenic potential [82, 83]. The BCOR::CLGN's calmegin lost almost half of its sequence, including most of its calreticulin sequence. This loss is likely to have an impact on the function of calmegin, impairing protein folding and disrupting ER chaperone functions. This could potentially lead to ER stress and the promotion of tumor progression through dysfunctional protein quality control [84], which in turn could affect male fertility [85].

In the case of ZC3H7B::BCOR, belonging to the $BCOR^{CT}$ group, the BCOR protein lost its Bbs site and, consequently, the role of the corepressor function with BCL6. Although ZC3H7B retained its TPR and LD motifs, it lost its C2H2- and C3H1-type zinc fingers. ZC3H7B recognizes the hairpins of miR-7-1, miR-16-2, and weaker miR-29a. Knockout of ZC3H7B in HEK293 cells reduced mature miR7 levels [86]. The truncation of the protein and loss of its C2H2 and C3H1 zinc fingers, which may interact with RNA [87–89], could potentially impact its interaction with miR-7, leading to compromised RNA-binding capabilities, disrupting post-transcriptional gene regulation and contributing to the dysregulation of crucial pathways such as miRNA biogenesis [90]. Due to the fusion of KMT2D and BCOR genes, the resulting chimera protein does not contain the Bbs of the BCOR. Consequently, the protein has lost its BCL6 corepressor function. The lost ePHD may have had a DNA binding function [91] while the FYRN/C and SET domains are crucial for protein-protein interactions within the COMPASS complex. This complex is crucial for the methylation of histone H3 at lysine 4 (H3K4) and plays an essential role in gene regulation and development [92]. The loss of the FYRN/C, ePHD, and SET domains destabilizes chromatin remodeling, leading to aberrant gene expression and genomic instability [93]. Regarding the CIITA::BCOR fusion, the BCOR lost its Bbs while the CIITA protein lost its main part (146–1130) excepting the N-terminal region; this change may potentially impact the protein's self-association [94] and the transcriptional activation of the major histocompatibility complex class II, possibly impairing MHC class II gene expression and potentially enabling immune evasion by tumor cells [95]. RTL9 is also known as sushi-ichi retrotransposon homolog 10 (SIRH10) or retrotransposon Gag domain-containing protein 1 (RGAG1). This protein belongs to the family of the so-called gag-like proteins whose domains are homologous to those of retroviral and retrotransposon polyproteins including the capsid structural protein [71,96]. The 1169–1316 region of RTL9, close to its C terminus, encompasses both the N- and C-terminal subdomains of the capsid-like domain. This domain shares a high structural similarity with the capsid domain of the human immunodeficiency virus and with the capsid-like domain of other mammalian gag-like proteins, such as activity-regulated cytoskeleton-associated proteins (Arc/Arg3.1). These proteins and other gag-like proteins including the members of the RTL family such as RTL1 (also known as PEG11) and RTL2 (also known as PEG10 or RGAG3) are known to have the ability to self-assemble into viral capsid-like particles, via oligomerization of the capsid-like domains [97]. The intact RTL9 protein may also be potentially able for self-assembly, but this has not been proved experimentally so far. The RTL9:BCOR-2 fusion protein contains only the 1–1199 residues of the RTL9, almost the entire capsid-like domain of RTL9 is deleted upon fusion. Consequently, the RTL9 cannot facilitate

the oligomerization of the fusion protein via its capsid-like domain. The AHR in the AHR::BCOR-2 chimeric protein suffers no loss of functional domain but, the addition of BCOR sequences likely alters its interaction dynamics, potentially disrupting xenobiotic metabolism and immune response pathways critical for tumor suppression [98] (Fig. 1).

The physicochemical properties of the fusion proteins were investigated and compared to those of the wild-type proteins (Fig. 2 and Table 2). The study revealed the GRAVY values of fusion proteins decreased (35.95 %) compared to wild-type proteins (Fig. 2C). This reduction indicates a decrease in hydrophobicity, which can potentially impact protein folding and interactions significantly. This may result in improper assembly and function within cellular environments. Hydrophobic interactions play a pivotal role in stabilizing protein folding and assembling multi-protein complexes [99]. A reduction in hydrophobicity has the potential to disrupt the proper folding and stability of proteins, which may result in aggregation or loss of function [100]. Furthermore, alterations in hydrophobicity can impact protein-protein interactions and the overall stability of protein structures, particularly under conditions of stress [101]. However, the values did not reach a positive range. This indicates that the protein structures may still be predicted to adopt a globular configuration [45]. The AI values remained largely unaltered.

The increase in the II value points to a reduction in protein stability and a corresponding decrease in the *in vivo* half-life [43] (Fig. 2A–C). This instability may disrupt protein homeostasis (proteostasis), which is critical for maintaining cellular functions and signaling pathways. Disrupted proteostasis overwhelms the cell's quality control mechanisms, including molecular chaperones and proteasomes, leading to cellular dysfunction and disease [102]. The high expression of the BCOR protein is described in cases of BRS [103] and may serve as a marker for disrupted proteostasis. It has been demonstrated that overexpression of BCOR results in the forming of nuclear aggregates distinct from classical aggresomes [104]. In cancer cells, this deregulation promotes survival and proliferation under proteotoxic stress, aiding tumor progression [105]. Instability triggers the unfolded protein response (UPR), aiming to restore proteostasis by enhancing protein folding capacity, degrading misfolded proteins, and reducing protein synthesis. However, chronic UPR activation has been shown to support tumorigenesis by enabling cancer cells to survive and proliferate under stress [106].

The alterations in the EC values of the fusion proteins (Fig. 2) may have notable implications for their functional properties and detection methods. An elevated EC value is indicative of increased content of aromatic amino acids, particularly tryptophan, tyrosine, and cysteine residues, which exhibit ultraviolet light absorption at 280 nm [41]. This enhancement in absorbance facilitates more precise quantification of the proteins using spectrophotometric assays, which is a critical factor for accurate biochemical analyses and protein characterization [107]. However, alterations in the aromatic amino acid composition can impact the protein's folding and stability. Aromatic residues play a pivotal role in the hydrophobic interactions that are responsible for stabilizing protein tertiary structures [108]. Such alterations may affect the folding pathways, potentially leading to conformational changes that impact the protein's biological activity and its interactions with other macromolecules. Furthermore, alterations in the EC may indicate modifications in the protein's surface properties, which could influence cellular localization, protein-protein interactions, and susceptibility to post-translational modifications [40].

The observed alterations in the TpI of the fusion proteins in comparison to their partner proteins (Fig. 2 and Table 2) indicate notable implications for their biochemical behavior and potential role in disease pathogenesis. Most fusion proteins exhibited a reduction in pI values in comparison to their native counterparts (Fig. 2). A reduction in pI may result in decreased solubility at physiological pH, which could impact protein stability and enhance the likelihood of aggregation [109]. Proteins with lower pI values carry a higher negative charge at neutral pH, which can affect their cellular localization, interaction with other macromolecules, and participation in protein complexes [110]. Conversely, the observed increase in pI indicates a shift towards a more positive net charge under physiological conditions. This may also result in alterations to the protein's interaction dynamics and functional activity [111].

Our SignalP 6.0, and DeepLoc 2.0 findings were largely constant with the literature (Table 3). However, some already described signal peptides were not recognized by these tools. Only calmegin was found to have secreted signal peptides, while AHR also contains two NLS at 13–16 and 37–42 residues [112], which were not lost upon the fusion event. The DeepLoc 2.0 results indicate that almost all fusion proteins may be found in the nucleus, except BCOR::CLGN which may located in the endoplasmic reticulum. These results are consistent with the PANTHER GO term cellular component, which also indicates nuclear localization for most of the fusion proteins. Only for the BCOR::CLGN, we observed ER localization using DeepLoc 2.0, but nuclear localization using PATHER GO terms. The CLGN is known to contain an N-terminal hydrophobic signal peptide [113] at predicted position 1–19 [27], which is responsible for ER localization, and a hydrophilic C-terminus with a transmembrane domain and ER retention signal [85]. As a result of the fusion event, the N-terminal hydrophobic signal is lost with the first 295 AA of the protein, but the C-terminal sites remain intact (1931–1952 residues, according to BCOR::CLGN). The N-terminal hydrophobic signal peptide is crucial for directing calmegin protein to the ER. Without this signal peptide, calmegin may not be correctly targeted to the ER, which can affect its proper localization and function within the cell. The C-terminus of CLGN functions to anchor the protein to the ER membrane and ensure its retention within the ER for its chaperone functions. This signal is crucial for maintaining calmegin's localization within the ER, where it plays an important role in protein folding, quality control, and interactions related to spermatogenesis and sperm-egg interactions [85,114]. DeepLoc 2.0 identified the C-terminus of CLGN and located it in the ER, despite the absence of the N-terminus (Table 3). The KMT2D and KMT2D::BCOR's cellular component term is MLL3/4 complex (GO:0044666) which is a histone methyltransferase complex located in the nucleoplasm of the nucleus (Table A.1). No literature data were available for RTL9:BCOR-2.

BCOR-rearranged sarcoma is characterized by its small-sized tumors with fibrovascular stroma sometimes with myxoid change. The tumor nuclei can be round or ovoid, but the common feature is the scant cytoplasm, like Ewing sarcoma. Although fusion proteins could theoretically be detected in the cytoplasm or ER, their minimal amount makes the interpretation of immunohistochemical stains challenging.

The analysis of the disorder propensities by using IUPred3 revealed a significant increase in the disorder of $BCOR^{NTs}$ in $region^{PUFD}$ of the fusion proteins as compared to wBCOR (Fig. 3). This increase may affect the interactions of BCOR with the PCGF1's RAWUL domain and with the KDM2B's C terminus, potentially interfering with the formation of the PRC1.1 complex.

The fusion events were hypothesized to influence the interaction dynamics between BCOR and its binding partners. To investigate this hypothesis, we constructed the BCOR-PCGF1 and PRC1.1 complexes with AF3 and assessed their binding affinities (Figs. 4 and 5, Fig Fig. 7). Structural analyses of the 450-450 complexes revealed significant variations in binding affinity across both dimeric (Figs. 4 and 5) and tetrameric forms of BCOR (Fig. 7A and B) and its interaction partners (Fig. 7C and D). Of the 92 multiple comparisons conducted (Figs. 5 and 7), 32 (35 %) demonstrated altered ΔG values. Notably, two additional interaction surfaces within BCOR, designated $IS^{NT/CT}$ and IS^{Leu}, were identified (Fig. 6), both of which were impacted by the fusion events (Fig. 4D,E,G,I). Moreover, interactions involving BRS proteins indirectly affected binding affinities within the tetramer, particularly between PCGF1-KDM2B (Fig. 7C) and KDM2B-SKP1 (Fig. 7D).

Fusion proteins such as BCOR::CCNB3, BCOR::MAML3, BCOR-2: MAML1, and RTL9:BCOR-2 exhibited increased binding affinities with PRODGY, whereas BCOR::CLGN, ZC3H7B::BCOR, CIITA::BCOR, and

AHR::BCOR-2 displayed decreased affinities (Fig. 4). While reduced affinities indicate weakened binding, increased affinities suggest enhanced complex assembly. However, PRC1.1-mediated epigenetic regulation requires a balance between stability and flexibility to maintain dynamic assembly and disassembly essential for precise gene regulation [115]. Overstabilization can hinder these dynamics, impair cellular signal responsiveness, and disrupt gene expression [116].

Furthermore, molecular dynamics (MD) simulations (10 ns) revealed significant changes in RMSD, hydrogen bond pattern, and binding affinities among the nine most accurate dimer structures. The results of the gmx_MMPBSA binding affinity predictions demonstrated a strong correlation with the PRODIGY scores, thereby substantiating the reliability of the findings. The results obtained highlight the structural and energetic impact of BCOR fusion proteins on dimer stability and interaction dynamics with PCGF1.

The dynamic behavior of fusion proteins observed in the MD simulations provides mechanistic insights into how these rearrangements disrupt PRC1.1 assembly beyond what can be inferred from static structures alone. The consistent pattern of increased conformational flexibility suggests that even when essential binding domains are preserved in fusion proteins, their dynamic properties are fundamentally altered. Domain destabilization, in ZC3H7B::BCOR and CIITA::BCOR fusions, appears to propagate throughout the protein structure, as evidenced by their increased radius of gyration and reduced compactness.

These alterations in dynamics are likely to explain the paradoxical observation that fusion proteins retaining complete BCOR sequences still exhibit functional deficiencies in epigenetic regulation. The correlation between hydrogen bond instability and reduced binding affinity demonstrates how subtle changes in residue flexibility can dramatically impact interaction strength at critical interfaces. This paradigm provides a novel framework for comprehending the mechanisms by which cancer-associated fusion events disrupt biological processes, extending beyond the scope of simple domain truncation. A limitation of this study is the utilization of 10 ns MD simulations in favor of a longer 100 ns. Due to computational resource constraints and the necessity of analyzing nine BCOR-PCGF1 dimers, 10 ns simulations were employed, a duration that is sufficient for capturing key structural and interaction trends.

Excessive binding affinity may induce conformational changes that inhibit the recruitment of cofactors or enzymatic activities necessary for PRC1.1 function, such as RING1B's E3 ubiquitin ligase activity critical for histone H2A ubiquitination [117]. Fusion proteins may also exert dominant-negative effects by outcompeting wild-type BCOR in PRC1.1 complexes but failing to sustain normal function due to structural alterations [118]. The introduction of novel protein segments by fusions can disrupt BCOR's regulatory role, even if essential domains remain intact [119,120]. This phenomenon is observed in other cancers, where chimeric proteins alter localization and function despite the retention of regulatory elements [121].

Such disruptions may compromise BCOR's ability to recruit and assemble the PRC1.1 complex at target gene loci, which could result in ineffective repression of oncogenic genes, particularly those within the HOX gene cluster [122]. Impaired recruitment of PRC1.1 results in a reduction in the monoubiquitination of H2A-K119 (H2A-K119-ub1), an essential epigenetic mark for maintaining gene silencing [123]. Consequently, the aberrant upregulation of HOX genes due to disrupted gene silencing mechanisms can drive oncogenesis by promoting uncontrolled cell proliferation and interfering with normal differentiation pathways [122]. Additionally, these *BCOR* rearrangements may alter the epigenetic landscape, further contributing to gene expression dysregulation and tumorigenesis [124,125]. Therefore, the molecular changes resulting from *BCOR* rearrangements significantly impair its role in gene silencing, collectively contributing to the development and progression of oncological diseases.

Furthermore, physicochemical changes, such as decreased hydrophobicity and increased instability, may exacerbate dysfunction by affecting protein folding or promoting aggregation [100,101]. Protein aggregation can sequester functional components or form dysfunctional complexes, compromising PRC1.1's chromatin-modifying capacity and gene regulatory function. These findings underscore that the retention of essential domains is insufficient if fusion-derived sequences impair protein function or interactions.

In summary, our results indicate that BCOR fusion proteins disrupt the molecular interplay within PRC1.1, driving oncogenesis through dysregulated gene expression. Despite their theoretical ability to assemble into complexes, fusion proteins fail as effective epigenetic regulators, underscoring their role in the pathogenesis of *BCOR*-rearranged sarcomas.

Targeting these vulnerabilities requires multifaceted approaches, as no effective therapeutic options are specifically designed for BRS caused by fusion events [69]. However, it is crucial to investigate how existing or potential therapeutic agents can be applied to these fusions to improve treatment outcomes. CDK4/6 inhibitors, such as palbociclib, have shown promise in the treatment of different sarcoma subtypes via targeting cell cycle deregulation. Preclinical models and clinical trials also demonstrated disease control, particularly in tumors with CDK4 overexpression [126,127]. A phase II trial of palbociclib demonstrated activity in various sarcoma subtypes, selected by CDK4/CDKN2A expression, showing promising progression-free survival and overall survival in heavily pretreated patients with advanced sarcomas overexpressing CDK4 [127]. In parallel, conventional chemotherapy regimens, such as vincristine/doxorubicin/cyclophosphamide alternating with ifosfamide/etoposide, have demonstrated chemosensitivity and complete radiographic responses in infants with BRS [128]. While PARP inhibitors have been explored in various cancers, their specific role in BRS, particularly in combination with HMGA1-chromatin modifiers or ATR/CHK1 inhibitors to induce homologous recombination deficiency, requires further investigation [129,130]. Epigenetic modulation via EZH2 inhibitors and RING1B ubiquitin ligase blockers aims to restore PRC1.1 function, and PROTAC technologies might be able to selectively degrade dysfunctional BCOR fusion proteins, though direct evidence in BRS is still emerging [131,132]. Preclinical data suggest that *BCOR* fusion-induced destabilization of PRC1.1 leads to the deregulation of myeloid differentiation genes; however, more targeted research is needed [133]. Immunotherapeutic strategies, such as TCR-engineered T cells (TCR-T) targeting tumor-specific antigens, are also promising, particularly for tumors expressing intracellular antigens like *BCOR* fusions [134–137]. Furthermore, combination therapies also have considerable potential; meta-analyses suggest that non-Ewing treatment strategies (e.g. doxorubicin-based regimens) can be considered as a safe option, survival rates being comparable to Ewing protocols for BRS (20 % vs. 21.8 % mortality) can also be achieved [69]. Clinically, molecular diagnostics (e.g. *BCOR* FISH/RNA sequencing) are indispensable for identifying fusion-specific targets (e.g. *BCOR*::*CCNB3* vs. *BCOR*::*MAML3*), as therapeutic outcomes depend on genetic variants [103,138]. Given the rarity of these tumors and the lack of consensus guidelines, ongoing research (e.g., NCT04889924) and clinical trials evaluating PARP/checkpoint inhibitors are crucial to define optimal therapeutic strategies and advance personalized approaches [139–145].

5. Conclusions

This comparative structural study aimed to ascertain the impact of *BCOR*-rearrangements on the sequence and interaction characteristics of the BCOR protein and its fusion partners in BRS. Our findings demonstrate that the fusion events result in disruption of the fusion partners' domain composition and physicochemical properties, although, the majority of BCORs in the $BCOR^{NT}$ group remain intact. However, in the $BCOR^{NT}$ group, the absence of the Bbs site may result in oncogenic effects, as evidenced in murine models [146]. Additionally, the truncation of the N terminus may impede the transcriptional repression effect of BCOR [21] and might be associated with reduced levels of H2A-K119-ub1 at PRC1 target promoters, including Hoxa7, Hoxa9, and

Cebpa [146]. Interestingly, despite the presence of the Bbs site in BCOR in the case of fusion proteins belonging to the $BCOR^{CT}$ group, its interaction with BCL6 might be possibly attenuated, which may be attributed to physicochemical alterations. The fusion events alter domain architectures, binding affinities, and physicochemical properties of BCOR and its fusion partners, impairing the assembly and function of the PRC1.1 being a suppressor of critical differentiation programs [21]. These disruptions compromise histone H2A monoubiquitination, leading to dysregulation of oncogenic pathways, notably the upregulation of HOX genes [22,125]. Some possible limitations of this study are described as follows. The findings rely on the results of *in silico* analyses and predictions, therefore, the results might be confirmed by experimental studies, and the predicted changes of binding affinities and structural alterations are expected to be validated by biochemical and cellular assays. Additionally, the development of therapeutic strategies targeting the disrupted PRC1.1 complex or other pathways influenced by *BCOR* fusions represents a significant avenue for translational research. Furthermore, future investigations will be directed towards exploring the oncogenic effects of BCOR-ITDs.

Our findings highlight the critical role of BCOR in maintaining chromatin homeostasis and emphasize that structural integrity and physicochemical properties are essential for its function. Understanding the complex interplay between BCOR fusion proteins and PRC1.1 advances insights into the molecular mechanisms underlying BRS. These protein changes delineate the multifaceted nature of *BCOR* alterations and their consequential effects on protein's functionality, and interactions, through pluripotency, key regulator rule in various developmental processes, including embryogenesis, mesenchymal stem cell function, hematopoiesis, and lymphoid development [147].

These insights may contribute to the advancement of diagnostic, prognostic, and therapeutic strategies for these rare malignancies.

CRediT authorship contribution statement

Kristóf Madarász: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **János András Mótyán:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Yi-Che Chang Chien:** Project administration. **Judit Bedekovics:** Project administration. **Szilvia Lilla Csoma:** Project administration. **Gábor Méhes:** Writing – review & editing, Supervision, Resources. **Attila Mokánszki:** Writing – review & editing, Supervision, Conceptualization.

Data availability section

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The assembled cDNA and amino acid sequences can be accessed in Supplemental Data A.1. This study includes no data deposited in external repositories.

Ethics statement

Ethical approval was not required for this study because it exclusively involved computational predictions and analyses. The research was conducted entirely through bioinformatics methods and computational modeling, without any involvement of human subjects, animal experiments, or the handling of sensitive personal information. All data used in the study were obtained from publications and literature.

Funding

Kristóf Madarász was supported by PhD Excellence Scholarship from Count István Tisza Foundation for the University of Debrecen and by the ÚNKP-23-3-II-DE-330 New National Excellence Program of the Ministry of Innovation and Technology. This project was supported in part by the ÚNKP-23-5-DE-486 New National Excellence Program of the Ministry of Culture and Innovation from the source of the National Research, Development and Innovation Fund (NKFIH) (to J.AM.) and by the TKP2021-EGA-20 (Biotechnology) of NKFIH. János A. Mótyán is the receiver of the János Bolyai Research Scholarship provided by the Hungarian Academy of Sciences (BO/00110/23/5).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.110144.

References

[1] R. Rajasekaran, C.G.P. Doss, C. Sudandiradoss, K. Ramanathan, R. Purohit, R. Sethumadhavan, Effect of deleterious nsSNP on the HER2 receptor based on stability and binding affinity with herceptin: a computational approach, Comptes Rendus, Biologies 331 (2008) 409–417, https://doi.org/10.1016/j.crvi.2008.03.004.

[2] G. Tanwar, R. Purohit, Gain of native conformation of Aurora A S155R mutant by small molecules, J. Cell. Biochem. 120 (2019) 11104–11114, https://doi.org/10.1002/jcb.28387.

[3] R. Singh, V.K. Bhardwaj, J. Sharma, P. Das, R. Purohit, Identification of selective cyclin-dependent kinase 2 inhibitor from the library of pyrrolone-fused benzosuberene compounds: an in silico exploration, J. Biomol. Struct. Dyn. 40 (2022) 7693–7701, https://doi.org/10.1080/07391102.2021.1900918.

[4] R. Singh, S. Manna, H. Nandanwar, R. Purohit, Bioactives from medicinal herb against bedaquiline resistant tuberculosis: removing the dark clouds from the horizon, Microb. Infect. 26 (2024) 105279, https://doi.org/10.1016/j.micinf.2023.105279.

[5] P.C. Ng, S. Henikoff, SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Res. 31 (2003) 3812–3814, https://doi.org/10.1093/nar/gkg509.

[6] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat. Methods 7 (2010) 248–249, https://doi.org/10.1038/nmeth0410-248.

[7] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A.J. Ballard, J. Bambrick, S.W. Bodenstein, D.A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A.I. Cowen-Rivers, A. Cowie, M. Figurnov, F.B. Fuchs, H. Gladman, R. Jain, Y.A. Khan, C.M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E.D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J.M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature 630 (2024) 493–500, https://doi.org/10.1038/s41586-024-07487-w.

[8] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[9] L. Chen, Q. Li, K.F.A. Nasif, Y. Xie, B. Deng, S. Niu, S. Pouriyeh, Z. Dai, J. Chen, C. Y. Xie, AI-driven deep learning techniques in protein structure prediction, Int. J. Mol. Sci. 25 (2024) 8426, https://doi.org/10.3390/ijms25158426.

[10] M. Sbaraglia, E. Bellan, A.P. Dei Tos, The 2020 WHO classification of soft tissue tumours: news and perspectives, Pathologica 113 (2020) 70–84, https://doi.org/10.32074/1591-951X-213.

[11] Y.-C. Kao, A.A. Owosho, Y.-S. Sung, L. Zhang, Y. Fujisawa, J.-C. Lee, L. Wexler, P. Argani, D. Swanson, B.C. Dickson, C.D.M. Fletcher, C.R. Antonescu, BCOR-CCNB3 fusion positive sarcomas: a clinicopathologic and molecular analysis of 36 cases with comparison to morphologic spectrum and clinical behavior of other round cell sarcomas, Am. J. Surg. Pathol. 42 (2018) 604–615, https://doi.org/10.1097/PAS.0000000000000965.

[12] O. Albagli, D. Lantoine, S. Quief, F. Quignon, C. Englert, J.-P. Kerckaert, D. Montarras, C. Pinset, C. Lindon, Overexpressed BCL6 (LAZ3) oncoprotein triggers apoptosis, delays S phase progression and associates with replication foci, Oncogene 18 (1999) 5063–5075, https://doi.org/10.1038/sj.onc.1202892.

[13] K.D. Huynh, W. Fischle, E. Verdin, V.J. Bardwell, BCoR, a novel corepressor involved in BCL-6 repression, Genes Dev. 14 (2000) 1810–1823, https://doi.org/10.1101/gad.14.14.1810.

[14] J.K. Pagan, J. Arnold, K.J. Hanchard, R. Kumar, T. Bruno, M.J.K. Jones, D. J. Richard, A. Forrest, A. Spurdle, E. Verdin, M. Crossley, M. Fanciulli, G. Chenevix-Trench, D.B. Young, K.K. Khanna, A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP, J. Biol. Chem. 282 (2007) 15248–15257, https://doi.org/10.1074/jbc.M700246200.

[15] S.E. Junco, R. Wang, J.C. Gaipa, A.B. Taylor, V. Schirf, M.D. Gearhart, V. J. Bardwell, B. Demeler, P.J. Hart, C.A. Kim, Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs, Structure 21 (2013) 665–671, https://doi.org/10.1016/j.str.2013.02.013.

[16] N.P. Blackledge, N.R. Rose, R.J. Klose, Targeting Polycomb systems to regulate gene expression: modifications to a complex story, Nat. Rev. Mol. Cell Biol. 16 (2015) 643–649, https://doi.org/10.1038/nrm4067.

[17] N.P. Blackledge, A.M. Farcas, T. Kondo, H.W. King, J.F. McGouran, L.L. P. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H.K. Long, T. W. Sheahan, N. Brockdorff, B.M. Kessler, H. Koseki, R.J. Klose, Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation, Cell 157 (2014) 1445–1459, https://doi.org/10.1016/j.cell.2014.05.004.

[18] S.J. Wong, O. Senkovich, J.A. Artigas, M.D. Gearhart, U. Ilangovan, D. W. Graham, K.N. Abel, T. Yu, A.P. Hinck, V.J. Bardwell, C.A. Kim, Structure and role of BCOR PUFD in noncanonical PRC1 assembly and disease, Biochemistry 59 (2020) 2718–2728, https://doi.org/10.1021/acs.biochem.0c00285.

[19] E.C. Chittock, S. Latwiel, T.C.R. Miller, C.W. Müller, Molecular architecture of polycomb repressive complexes, Biochem. Soc. Trans. 45 (2017) 193–205, https://doi.org/10.1042/BST20160173.

[20] J.A. Wamstad, C.M. Corcoran, A.M. Keating, V.J. Bardwell, Role of the transcriptional corepressor bcor in embryonic stem cell differentiation and early embryonic development, PLoS One 3 (2008) e2814, https://doi.org/10.1371/journal.pone.0002814.

[21] Z. Wang, M.D. Gearhart, Y.-W. Lee, I. Kumar, B. Ramazanov, Y. Zhang, C. Hernandez, A.Y. Lu, N. Neuenkirchen, J. Deng, J. Jin, Y. Kluger, T.A. Neubert, V.J. Bardwell, N.B. Ivanova, A non-canonical BCOR-PRC1.1 complex represses differentiation programs in human ESCs, Cell Stem Cell 22 (2018) 235–251.e9, https://doi.org/10.1016/j.stem.2017.12.002.

[22] G. Pierron, F. Tirode, C. Lucchesi, S. Reynaud, S. Ballet, S. Cohen-Gogo, V. Perrin, J.-M. Coindre, O. Delattre, A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion, Nat. Genet. 44 (2012) 461–466, https://doi.org/10.1038/ng.1107.

[23] Y.-C. Kao, Y.-S. Sung, L. Zhang, S.-C. Huang, P. Argani, C.T. Chung, N.S. Graf, D. C. Wright, S.J. Kellie, N.P. Agaram, K. Ludwig, A. Zin, R. Alaggio, C.R. Antonescu, Recurrent BCOR internal tandem duplication and YWHAE-NUTM2B fusions in soft tissue undifferentiated round cell sarcoma of infancy: overlapping genetic features with clear cell sarcoma of kidney, Am. J. Surg. Pathol. 40 (2016) 1009–1020, https://doi.org/10.1097/PAS.0000000000000629.

[24] K. Specht, L. Zhang, Y.-S. Sung, M. Nucci, S. Dry, S. Vaiyapuri, G.H. Richter, C. D. Fletcher, C.R. Antonescu, Novel BCOR-MAML3 and ZC3H7B-BCOR gene fusions in undifferentiated small blue round cell sarcomas, Am. J. Surg. Pathol. 40 (2016) 433–442, https://doi.org/10.1097/PAS.0000000000000591.

[25] A. Yoshida, Y. Arai, N. Hama, H. Chikuta, Y. Bando, S. Nakano, E. Kobayashi, J. Shibahara, H. Fukuhara, M. Komiyama, S. Watanabe, K. Tamura, A. Kawai, T. Shibata, Expanding the clinicopathologic and molecular spectrum of BCOR-associated sarcomas in adults, Histopathology 76 (2020) 509–520, https://doi.org/10.1111/his.14023.

[26] M. Vasella, U. Wagner, C. Fritz, K. Seidl, L. Giudici, G.U. Exner, H. Moch, P. J. Wild, B. Bode-Lesniewska, Novel RGAG1-BCOR gene fusion revealed in a somatic soft tissue sarcoma with a long follow-up, Virchows Arch. 480 (2022) 1107–1114, https://doi.org/10.1007/s00428-021-03160-z.

[27] Y.-C. Chang Chien, K. Madarász, S.L. Csoma, J.A. Mótyán, H.-Y. Huang, G. Méhes, A. Mokánszki, Molecular identification and in silico protein analysis of a novel BCOR-CLGN gene fusion in intrathoracic BCOR-rearranged sarcoma, Cancers 15 (2023) 898, https://doi.org/10.3390/cancers15030898.

[28] S. Cocchi, M. Gambarotti, G. Gamberi, G. Magagnoli, M. Maioli, M. Stevanin, F. Samperi, A. Righi, S. Benini, The utility of FISH analysis in the diagnosis of BCOR-rearranged sarcomas, Pathol. Res. Pract. 255 (2024) 155209, https://doi.org/10.1016/j.prp.2024.155209.

[29] F.J. Martin, M.R. Amode, A. Aneja, O. Austine-Orimoloye, A.G. Azov, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S.K. Bhurji, A. Bignell, S. Boddu, P. R. Branco Lins, L. Brooks, S.B. Ramaraju, M. Charkhchi, A. Cockburn, L. Da Rin Fiorretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C.G. Giron, T. Genez, G.S. Ghattaoraya, J.G. Martinez, C. Guijarro, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Marques-Coelho, J.C. Marugán, G.A. Merino, L.P. Mirabueno, A. Mushtaq, S. N. Hossain, D.N. Ogeh, M.P. Sakthivel, A. Parker, M. Perry, I. Pilizota, I. Prosovetskaia, J.G. Pérez-Silva, A.I.A. Salam, N. Saraiva-Agostinho, H. Schuilenburg, D. Sheppard, S. Sinha, B. Sipos, W. Stark, E. Steed, R. Sukumaran, D. Sumathipala, M.-M. Suner, L. Surapaneni, K. Sutinen, M. Szpak, F.F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T.A. Walsh, B. Walts, E. Wass, N. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakiachvili, B. Flint, S. Giorgetti, L. Haggerty, G.R. Ilsley, J.E. Loveland, B. Moore, J.M. Mudge, J. Tate, D. Thybert, S.J. Trevanion, A. Winterbottom, A. Frankish, S.E. Hunt, M. Ruffier, F. Cunningham, S. Dyer, R.D. Finn, K.L. Howe, P.W. Harrison, A.D. Yates, P. Flicek, Ensembl 2023, Nucleic Acids Res. 51 (2023) D933–D941, https://doi.org/10.1093/nar/gkac958.

[30] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (2003) 3784–3788.

[31] The UniProt Consortium, UniProt: the universal protein knowledgebase in 2023, Nucleic Acids Res. 51 (2023) D523–D531, https://doi.org/10.1093/nar/gkac1052.

[32] E.A. Bruford, C.R. Antonescu, A.J. Carroll, A. Chinnaiyan, I.A. Cree, N.C.P. Cross, R. Dalgleish, R.P. Gale, C.J. Harrison, R.J. Hastings, J.-L. Huret, B. Johansson, M. Le Beau, C. Mecucci, F. Mertens, R. Verhaak, F. Mitelman, HUGO Gene Nomenclature Committee (HGNC) recommendations for the designation of gene fusions, Leukemia 35 (2021) 3040–3043, https://doi.org/10.1038/s41375-021-01436-6.

[33] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (2020) D265–D268, https://doi.org/10.1093/nar/gkz991.

[34] T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B.L. Pinto, G.A. Salazar, M.L. Bileschi, P. Bork, A. Bridge, L. Colwell, J. Gough, D.H. Haft, I. Letunić, A. Marchler-Bauer, H. Mi, D.A. Natale, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J.A. Sigrist, I. Sillitoe, N. Thanki, P.D. Thomas, S.C.E. Tosatto, C. H. Wu, A. Bateman, InterPro in 2022, Nucleic Acids Res. 51 (2023) D418–D427, https://doi.org/10.1093/nar/gkac993.

[35] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A. P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology, Nat. Genet. 25 (2000) 25–29, https://doi.org/10.1038/75556.

[36] H. Mi, A. Muruganujan, D. Ebert, X. Huang, P.D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, Nucleic Acids Res. 47 (2019) D419–D426, https://doi.org/10.1093/nar/gky1038.

[37] H. Mi, A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, P.D. Thomas, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0), Nat. Protoc. 14 (2019) 703–721, https://doi.org/10.1038/s41596-019-0128-8.

[38] The Gene Ontology Consortium, S.A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H.J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N.L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C.J. Mungall, A. Muruganugan, T. Mushayahama, P.W. Sternberg, P.D. Thomas, K. Van Auken, J. Ramsey, D.A. Siegele, R. L. Chisholm, P. Fey, M.C. Aspromonte, M.V. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. dos Santos, S. Marygold, V. Strelets, C.J. Tabone, J. Thurmond, P. Zhou, S.H. Ahmed, P. Asanitthong, D. Luna Buitrago, M.N. Erdol, M.C. Gage, M. Ali Kadhum, K.Y.C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S.C.C. Saverimuttu, R. Su, K.E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M.E. Dolan, H.J. Drabkin, D.P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J.L. De Pons, M.R. Dwinell, G.T. Hayman, M. L. Kaldunski, A.E. Kwitek, S.J.F. Laulederkind, M.A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A. Morgat, S. A. Aleksander, J.M. Cherry, S.R. Engel, K. Karra, S.R. Miyasato, R.S. Nash, M. S. Skrzypek, S. Weng, E.D. Wong, E. Bakker, T.Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M.-C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, H. Bye-A-Jee, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M.J. Martin, S. Orchard, P. Raposo, E. Speretta, N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka, M. Westerfield, The gene Ontology knowledgebase in 2023, Genetics 224 (2023), https://doi.org/10.1093/genetics/iyad031 iyad031.

[39] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J. M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, Totowa, NJ, 2005, pp. 571–607, https://doi.org/10.1385/1-59259-890-0:571.

[40] H. Edelhoch, Spectroscopic determination of tryptophan and tyrosine in proteins, Biochemistry 6 (1967) 1948–1954, https://doi.org/10.1021/bi00859a010.

[41] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326, https://doi.org/10.1016/0003-2697(89)90602-7.

[42] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423.

[43] K. Guruprasad, B.V. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. 4 (1990) 155–161, https://doi.org/10.1093/protein/4.2.155.

[44] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.

[45] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132, https://doi.org/10.1016/0022-2836(82)90515-0.

[46] F. Teufel, J.J. Almagro Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 6.0 predicts

all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (2022) 1023–1025, https://doi.org/10.1038/s41587-021-01156-3.
[47] V. Thumuluri, J.J. Almagro Armenteros, A.R. Johansen, H. Nielsen, O. Winther, DeepLoc 2.0: multi-label subcellular localization prediction using protein language models, Nucleic Acids Res. 50 (2022) W228–W234, https://doi.org/10.1093/nar/gkac278.
[48] P.J. Thul, L. Åkesson, M. Wiking, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L.M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C.M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J.M. Schwenk, Å. Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K.S. Lilley, M. Uhlén, E. Lundberg, A subcellular map of the human proteome, Science 356 (2017) eaal3321, https://doi.org/10.1126/science.aal3321.
[49] M.H. Gíslason, H. Nielsen, J.J. Almagro Armenteros, A.R. Johansen, Prediction of GPI-anchored proteins with pointer neural networks, Curr. Res. Biotechnol. 3 (2021) 6–13, https://doi.org/10.1016/j.crbiot.2021.01.001.
[50] G. Erdős, M. Pajkos, Z. Dosztányi, IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation, Nucleic Acids Res. 49 (2021) W297–W303, https://doi.org/10.1093/nar/gkab408.
[51] L.C. Xue, J.P. Rodrigues, P.L. Kastritis, A.M. Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, Bioinformatics 32 (2016) 3676–3678, https://doi.org/10.1093/bioinformatics/btw514.
[52] The PyMOL Molecular Graphics System, Version 2.5, Schrödinger, LLC, (n.d.).
[53] A.E. Badaczewska-Dawid, C. Nithin, K. Wroblewski, M. Kurcinski, S. Kmiecik, MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes, Nucleic Acids Res. 50 (2022), https://doi.org/10.1093/nar/gkac307. W474–W482.
[54] S.K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, H. Chao, L. Chen, P.A. Craig, G. V. Crichlow, K. Dalenberg, J.M. Duarte, S. Dutta, M. Fayazi, Z. Feng, J.W. Flatt, S. Ganesan, S. Ghosh, D.S. Goodsell, R.K. Green, V. Guranovic, J. Henry, B. P. Hudson, I. Khokhriakov, C.L. Lawson, Y. Liang, R. Lowe, E. Peisach, I. Persikova, D.W. Piehl, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, B. Vallat, M. Voigt, B. Webb, J.D. Westbrook, S. Whetstone, J.Y. Young, A. Zalevsky, C. Zardecki, RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning, Nucleic Acids Res. 51 (2023) D488–D508, https://doi.org/10.1093/nar/gkac1077.
[55] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25, https://doi.org/10.1016/j.softx.2015.06.001.
[56] S. Páll, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling exascale software challenges in molecular dynamics simulations with GROMACS, in: S. Markidis, E. Laure (Eds.), Solving Software Challenges for Exascale, Springer International Publishing, Cham, 2015, pp. 3–27, https://doi.org/10.1007/978-3-319-15976-8_1.
[57] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J. C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854, https://doi.org/10.1093/bioinformatics/btt055.
[58] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theor. Comput. 4 (2008) 435–447, https://doi.org/10.1021/ct700301q.
[59] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718, https://doi.org/10.1002/jcc.20291.
[60] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Model. 7 (2001) 306–317, https://doi.org/10.1007/s008940100045.
[61] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56, https://doi.org/10.1016/0010-4655(95)00042-E.
[62] M. Abraham, A. Alekseenko, C. Bergh, C. Blau, E. Briand, M. Doijade, S. Fleischmann, V. Gapsys, G. Garg, S. Gorelov, G. Gouaillardet, A. Gray, M. E. Irrgang, F. Jalalypour, J. Jordan, C. Junghans, P. Kanduri, S. Keller, C. Kutzner, J.A. Lemkul, M. Lundborg, P. Merz, V. Miletić, D. Morozov, S. Páll, R. Schulz, M. Shirts, A. Shvetsov, B. Soproni, D. van der Spoel, P. Turner, C. Uphoff, A. Villa, S. Wingbermühle, A. Zhmurov, P. Bauer, B. Hess, E. Lindahl, GROMACS 2023.3 Source code. https://doi.org/10.5281/zenodo.10017686, 2023.
[63] B.R.I. Miller, T.D.Jr. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, J. Chem. Theor. Comput. 8 (2012) 3314–3321, https://doi.org/10.1021/ct300418h.
[64] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, J. Chem. Theor. Comput. 17 (2021) 6281–6291, https://doi.org/10.1021/acs.jctc.1c00645.
[65] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, J. Am. Stat. Assoc. 32 (1937) 675–701, https://doi.org/10.1080/01621459.1937.10503522.
[66] O.J. Dunn, Multiple comparisons using rank sums, Technometrics 6 (1964) 241–252, https://doi.org/10.2307/1266041.
[67] GraphPad Prism version 9.5.1 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com, (n.d.).
[68] W. Liu, Y. Xie, J. Ma, X. Luo, P. Nie, Z. Zuo, U. Lahrmann, Q. Zhao, Y. Zheng, Y. Zhao, Y. Xue, J. Ren, IBS: an illustrator for the presentation and visualization of biological sequences, Bioinformatics 31 (2015) 3359–3361, https://doi.org/10.1093/bioinformatics/btv362.
[69] A. Kyriazoglou, P. Bagos, Meta-analysis of BCOR rearranged sarcomas: challenging the therapeutic approach, Acta Oncol. 60 (2021) 721–726, https://doi.org/10.1080/0284186X.2021.1890818.
[70] R. Schwartz, C.S. Ting, J. King, Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life, Genome Res. 11 (2001) 703–709, https://doi.org/10.1101/gr.158701.
[71] F. Ishino, J. Itoh, M. Irie, A. Matsuzawa, M. Naruse, T. Suzuki, Y. Hiraoka, T. Kaneko-Ishino, Retrovirus-derived RTL9 plays an important role in innate antifungal immunity in the eutherian brain, Int. J. Mol. Sci. 24 (2023) 14884, https://doi.org/10.3390/ijms241914884.
[72] A.F. Ghetu, C.M. Corcoran, L. Cerchietti, V.J. Bardwell, A. Melnick, G.G. Privé, Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer, Mol. Cell 29 (2008) 384–391, https://doi.org/10.1016/j.molcel.2007.12.026.
[73] J.H. Bushweller, C. Schmidt, N. Achille, A. Kuntimaddi, A. Boulton, B. Leach, S. Zhang, N.J. Zeleznik-Le, Direct binding of BCOR, but not CBX8, to MLL-AF9 is essential for MLL-AF9 leukemia via regulation of the EYA1/SIX1 gene network, Blood 132 (2018) 1316, https://doi.org/10.1182/blood-2018-99-111388.
[74] R.S. Srinivasan, A.C. de Erkenez, C.S. Hemenway, The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor, Oncogene 22 (2003) 3395–3406, https://doi.org/10.1038/sj.onc.1206361.
[75] H. Yamano, C. Tsurumi, J. Gannon, T. Hunt, The role of the destruction box and its neighbouring lysine residues in cyclin B for anaphase ubiquitin-dependent proteolysis in fission yeast: defining the D-box receptor, EMBO J. 17 (1998) 5670–5678, https://doi.org/10.1093/emboj/17.19.5670.
[76] T.B. Nguyen, K. Manova, P. Capodieci, C. Lindon, S. Bottega, X.-Y. Wang, J. Refik-Rogers, J. Pines, D.J. Wolgemuth, A. Koff, Characterization and expression of mammalian cyclin B3, a prepachytene meiotic cyclin, J. Biol. Chem. 277 (2002) 41960–41969, https://doi.org/10.1074/jbc.M203951200.
[77] T.L. Peters, V. Kumar, S. Polikepahad, F.Y. Lin, S.F. Sarabia, Y. Liang, W.-L. Wang, A.J. Lazar, H. Doddapaneni, H. Chao, D.M. Muzny, D.A. Wheeler, M.F. Okcu, S. E. Plon, M.J. Hicks, D. López-Terrada, D.W. Parsons, A. Roy, BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children, Mod. Pathol. 28 (2015) 575–586, https://doi.org/10.1038/modpathol.2014.139.
[78] M.Y. Chiang, M.L. Xu, G. Histen, O. Shestova, M. Roy, Y. Nam, S.C. Blacklow, D. B. Sacks, W.S. Pear, J.C. Aster, Identification of a conserved negative regulatory sequence that influences the leukemogenic activity of NOTCH1, Mol. Cell Biol. 26 (2006) 6261–6271, https://doi.org/10.1128/mcb.02478-05.
[79] H. Liu, S. Kennard, B. Lilly, NOTCH3 expression is induced in mural cells through an autoregulatory loop that requires endothelial-expressed JAGGED1, Circ. Res. 104 (2009) 466–475, https://doi.org/10.1161/circresaha.108.184846.
[80] Y. Nam, P. Sliz, L. Song, J.C. Aster, S.C. Blacklow, Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes, Cell 124 (2006) 973–983, https://doi.org/10.1016/j.cell.2005.12.037.
[81] L. Wu, I. Maillard, M. Nakamura, W.S. Pear, J.D. Griffin, The transcriptional coactivator Maml1 is required for Notch2-mediated marginal zone B-cell development, Blood 110 (2007) 3618–3623, https://doi.org/10.1182/blood-2007-06-097030.
[82] T. Oyama, K. Harigaya, N. Sasaki, Y. Okamura, H. Kokubo, Y. Saga, K. Hozumi, A. Suganami, Y. Tamura, T. Nagase, H. Koga, M. Nishimura, R. Sakamoto, M. Sato, N. Yoshida, M. Kitagawa, Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo, Development 138 (2011) 5235–5246, https://doi.org/10.1242/dev.062802.
[83] M. Kitagawa, Notch signalling in the nucleus: roles of Mastermind-like (MAML) transcriptional coactivators, J. Biochem. 159 (2016) 287–294, https://doi.org/10.1093/jb/mvv123.
[84] A. Venkatesan, L.S. Satin, M. Raghavan, Roles of calreticulin in protein folding, immunity, calcium signaling and cell transformation, in: L.B. Agellon, M. Michalak (Eds.), Cellular Biology of the Endoplasmic Reticulum, Springer International Publishing, Cham, 2021, pp. 145–162, https://doi.org/10.1007/978-3-030-67696-4_7.
[85] M. Ikawa, I. Wada, K. Kominami, D. Watanabe, K. Toshimori, Y. Nishimune, M. Okabe, The putative chaperone calmegin is required for sperm fertility, Nature 387 (1997) 607–611, https://doi.org/10.1038/42484.
[86] T. Treiber, N. Treiber, U. Plessmann, S. Harlander, J.-L. Daiß, N. Eichner, G. Lehmann, K. Schall, H. Urlaub, G. Meister, A compendium of RNA-binding proteins that regulate MicroRNA biogenesis, Mol. Cell 66 (2017) 270–284.e13, https://doi.org/10.1016/j.molcel.2017.03.014.
[87] E. Carballo, W.S. Lai, P.J. Blackshear, Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin, Science 281 (1998) 1001–1005, https://doi.org/10.1126/science.281.5379.1001.
[88] W.S. Lai, E. Carballo, J.R. Strum, E.A. Kennington, R.S. Phillips, P.J. Blackshear, Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA, Mol. Cell Biol. 19 (1999) 4311–4323, https://doi.org/10.1128/mcb.19.6.4311.
[89] R.S. Brown, Zinc finger proteins: getting a grip on RNA, Curr. Opin. Struct. Biol. 15 (2005) 94–98, https://doi.org/10.1016/j.sbi.2005.01.006.

[90] X. Li, M. Han, H. Zhang, F. Liu, Y. Pan, J. Zhu, Z. Liao, X. Chen, B. Zhang, Structures and biological functions of zinc finger proteins and their roles in hepatocellular carcinoma, Biomark. Res. 10 (2022) 2, https://doi.org/10.1186/s40364-021-00345-1.
[91] H.S. Lee, I. Bang, J. You, T.-K. Jeong, C.R. Kim, M. Hwang, J.-S. Kim, S.H. Baek, J.-J. Song, H.-J. Choi, Molecular basis for PHF7-mediated ubiquitination of histone H3, Genes Dev. 37 (2023) 984–997, https://doi.org/10.1101/gad.350989.123.
[92] Y.-W. Cho, T. Hong, S. Hong, H. Guo, H. Yu, D. Kim, T. Guszczynski, G. R. Dressler, T.D. Copeland, M. Kalkum, K. Ge, PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex, J. Biol. Chem. 282 (2007) 20395–20406, https://doi.org/10.1074/jbc.M701574200.
[93] Z.-L. Zhang, P.-F. Yu, Z.-Q. Ling, The role of KMT2 gene in human tumors, Histol. Histopathol. 37 (2022) 323–334, https://doi.org/10.14670/HH-18-447.
[94] M.W. Linhoff, J.A. Harton, D.E. Cressman, B.K. Martin, J.P. Ting, Two distinct domains within CIITA mediate self-association: involvement of the GTP-binding and leucine-rich repeat domains, Mol. Cell Biol. 21 (2001) 3001–3011, https://doi.org/10.1128/MCB.21.9.3001-3011.2001.
[95] J.A. León Machado, V. Steimle, The MHC class II transactivator CIITA: not (quite) the odd-one-out anymore among NLR proteins, Int. J. Mol. Sci. 22 (2021) 1074, https://doi.org/10.3390/ijms22031074.
[96] M. Campillos, T. Doerks, P.K. Shah, P. Bork, Computational characterization of multiple Gag-like human proteins, Trends Genet. 22 (2006) 585–589, https://doi.org/10.1016/j.tig.2006.09.006.
[97] M. Segel, B. Lash, J. Song, A. Ladha, C.C. Liu, X. Jin, S.L. Mekhedov, R.K. Macrae, E.V. Koonin, F. Zhang, Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery, Science 373 (2021) 882–889, https://doi.org/10.1126/science.abg6155.
[98] F. Bahman, K. Choudhry, F. Al-Rashed, F. Al-Mulla, S. Sindhu, R. Ahmad, Aryl hydrocarbon receptor: current perspectives on key signaling partners and immunoregulatory role in inflammatory diseases, Front. Immunol. 15 (2024) 1421346, https://doi.org/10.3389/fimmu.2024.1421346.
[99] B.W. Matthews, Hydrophobic interactions in proteins, in: Encyclopedia of Life Sciences, first ed., Wiley, 2001 https://doi.org/10.1038/npg.els.0002975.
[100] C.N. Pace, J.M. Scholtz, G.R. Grimsley, Forces stabilizing proteins, FEBS Lett. 588 (2014) 2177–2184, https://doi.org/10.1016/j.febslet.2014.05.006.
[101] G.G. Ferenczy, M. Kellermayer, Contribution of hydrophobic interactions to protein mechanical stability, Comput. Struct. Biotechnol. J. 20 (2022) 1946–1956, https://doi.org/10.1016/j.csbj.2022.04.025.
[102] E.T. Powers, R.I. Morimoto, A. Dillin, J.W. Kelly, W.E. Balch, Biological and chemical approaches to diseases of proteostasis deficiency, Annu. Rev. Biochem. 78 (2009) 959–991, https://doi.org/10.1146/annurev.biochem.052308.114844.
[103] L. Li, M. Zhang, S. Chen, X. Sun, H. Xu, L. Li, T. Zhang, X. Huang, H. Ye, Y. Ding, Detection of BCOR gene rearrangement in Ewing-like sarcoma: an important diagnostic tool, Diagn. Pathol. 16 (2021) 50, https://doi.org/10.1186/s13000-021-01114-2.
[104] E. Buchberger, M.E. Harchi, D. Payrhuber, A. Zommer, D. Schauer, I. Simonitsch-Klupp, M. Bilban, C. Brostjan, Overexpression of the transcriptional repressor complex BCL-6/BCoR leads to nuclear aggregates distinct from classical aggresomes, PLoS One 8 (2013) e76845, https://doi.org/10.1371/journal.pone.0076845.
[105] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013.
[106] S. Wang, R.J. Kaufman, The impact of the unfolded protein response on human disease, J. Cell Biol. 197 (2012) 857–867, https://doi.org/10.1083/jcb.201110131.
[107] R.K. Scopes, Protein Purification: Principles and Practice, Springer Science & Business Media, 1993.
[108] S.K. Burley, G.A. Petsko, Aromatic-aromatic interaction: a mechanism of protein structure stabilization, Science 229 (1985) 23–28, https://doi.org/10.1126/science.3892686.
[109] B. Bjellqvist, B. Basse, E. Olsen, J.E. Celis, Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions, Electrophoresis 15 (1994) 529–539, https://doi.org/10.1002/elps.1150150171.
[110] P.G. Righetti, T. Caravaggio, Isoelectric points and molecular weights of proteins: a table, J. Chromatogr. A 127 (1976) 1–28, https://doi.org/10.1016/S0021-9673(00)98537-6.
[111] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155, https://doi.org/10.1021/bi00483a001.
[112] R. Haidar, F. Henkler, J. Kugler, A. Rosin, D. Genkinger, P. Laux, A. Luch, The role of DNA-binding and ARNT dimerization on the nucleo-cytoplasmic translocation of the aryl hydrocarbon receptor, Sci. Rep. 11 (2021) 18194, https://doi.org/10.1038/s41598-021-97507-w.
[113] D. Watanabe, K. Yamada, Y. Nishina, Y. Tajima, U. Koshimizu, A. Nagata, Y. Nishimune, Molecular cloning of a novel Ca(2+)-binding protein (calmegin) specifically expressed during male meiotic germ cell development, J. Biol. Chem. 269 (1994) 7744–7749, https://doi.org/10.1016/S0021-9258(17)37349-0.
[114] M. Ikawa, T. Nakanishi, S. Yamada, I. Wada, K. Kominami, H. Tanaka, M. Nozaki, Y. Nishimune, M. Okabe, Calmegin is required for fertilin α/β heterodimerization and sperm fertility, Dev. Biol. 240 (2001) 254–261, https://doi.org/10.1006/dbio.2001.0462.
[115] N.J. Francis, R.E. Kingston, C.L. Woodcock, Chromatin compaction by a polycomb group protein complex, Science 306 (2004) 1574–1577, https://doi.org/10.1126/science.1100576.
[116] T. Cheutin, A.J. McNairn, T. Jenuwein, D.M. Gilbert, P.B. Singh, T. Misteli, Maintenance of stable heterochromatin domains by dynamic HP1 binding, Science 299 (2003) 721–725, https://doi.org/10.1126/science.1078572.
[117] Z. Gao, J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, D. Reinberg, PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes, Mol. Cell 45 (2012) 344–356, https://doi.org/10.1016/j.molcel.2012.01.002.
[118] C. Kadoch, G.R. Crabtree, Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma, Cell 153 (2013) 71–85, https://doi.org/10.1016/j.cell.2013.02.036.
[119] F. Mitelman, B. Johansson, F. Mertens, The impact of translocations and gene fusions on cancer causation, Nat. Rev. Cancer 7 (2007) 233–245, https://doi.org/10.1038/nrc2091.
[120] F. Mertens, B. Johansson, T. Fioretos, F. Mitelman, The emerging complexity of gene fusions in cancer, Nat. Rev. Cancer 15 (2015) 371–381, https://doi.org/10.1038/nrc3947.
[121] N. Stransky, E. Cerami, S. Schalm, J.L. Kim, C. Lengauer, The landscape of kinase fusions in cancer, Nat. Commun. 5 (2014) 4846, https://doi.org/10.1038/ncomms5846.
[122] Q. Cao, M.D. Gearhart, S. Gery, S. Shojaee, H. Yang, H. Sun, D.-C. Lin, J.-W. Bai, M. Mead, Z. Zhao, Q. Chen, W.-W. Chien, S. Alkan, T. Alpermann, T. Haferlach, M. Müschen, V.J. Bardwell, H.P. Koeffler, BCOR regulates myeloid cell proliferation and differentiation, Leukemia 30 (2016) 1155–1165, https://doi.org/10.1038/leu.2016.2.
[123] M.D. Gearhart, C.M. Corcoran, J.A. Wamstad, V.J. Bardwell, Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets, Mol. Cell Biol. 26 (2006) 6880–6889, https://doi.org/10.1128/MCB.00630-06.
[124] A. Roy, V. Kumar, B. Zorman, E. Fang, K.M. Haines, H. Doddapaneni, O. A. Hampton, S. White, A.A. Bavle, N.R. Patel, K.W. Eldin, M. John Hicks, D. Rakheja, P.J. Leavey, S.X. Skapek, J.F. Amatruda, J.G. Nuchtern, M. M. Chintagumpala, D.A. Wheeler, S.E. Plon, P. Sumazin, D.W. Parsons, Recurrent internal tandem duplications of BCOR in clear cell sarcoma of the kidney, Nat. Commun. 6 (2015) 8891, https://doi.org/10.1038/ncomms9891.
[125] H. Ueno-Yokohata, H. Okita, K. Nakasato, S. Akimoto, J. Hata, T. Koshinaga, M. Fukuzawa, N. Kiyokawa, Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney, Nat. Genet. 47 (2015) 861–863, https://doi.org/10.1038/ng.3338.
[126] M. Perez, S. Muñoz-Galván, M.P. Jiménez-García, J.J. Marín, A. Carnero, Efficacy of CDK4 inhibition against sarcomas depends on their levels of CDK4 and p16ink4 mRNA, Oncotarget 6 (2015) 40557–40574.
[127] J. Martin-Broto, J. Martinez-Garcia, D.S. Moura, A. Redondo, A. Gutierrez, A. Lopez-Pousa, J. Martinez-Trufero, I. Sevilla, R. Diaz-Beveridge, M.P. Solis-Hernandez, A. Carnero, M. Perez, D. Marcilla, J. Garcia-Foncillas, P. Romero, J. Fernandez-Jara, D. Lopez-Lopez, I. Arribas, N. Hindi, Phase II trial of CDK4/6 inhibitor palbociclib in advanced sarcoma based on mRNA expression of CDK4/CDKN2A, Signal Transduct. Targeted Ther. 8 (2023) 1–10, https://doi.org/10.1038/s41392-023-01661-8.
[128] N. Merjaneh, H. Kim, H. Escoto, J. Metts, A. Ray, A. Bukowinski, Z. LeBlanc, D. Fair, M. Watanbe, E. Alva, K. Todd, J. Daley, D. Hartt, S.L. Cramer, S. Szabo, J. G. Pressey, Strategies for the treatment of infantile soft tissue sarcomas with BCOR alterations, J. Pediatr. Hematol. Oncol. 45 (2023) 315–321, https://doi.org/10.1097/MPH.0000000000002620.
[129] H. Kim, E. George, R. Ragland, S. Rafial, R. Zhang, C. Krepler, M. Morgan, M. Herlyn, E. Brown, F. Simpkins, Targeting the ATR/CHK1 axis with PARP inhibition results in tumor regression in BRCA mutant models, Clin. Cancer Res. 23 (2017) 3097–3108, https://doi.org/10.1158/1078-0432.CCR-16-2273.
[130] L. Planas-Paz, A. Pliego-Mendieta, C. Hagedorn, D. Aguilera-Garcia, M. Haberecker, F. Arnold, M. Herzog, L. Bankel, R. Guggenberger, S. Steiner, Y. Chen, A. Kahraman, M. Zoche, M.A. Rubin, H. Moch, C. Britschgi, C. Pauli, Unravelling homologous recombination repair deficiency and therapeutic opportunities in soft tissue and bone sarcoma, EMBO Mol. Med. 15 (2023) e16863, https://doi.org/10.15252/emmm.202216863.
[131] N.A. Fursova, N.P. Blackledge, M. Nakayama, S. Ito, Y. Koseki, A.M. Farcas, H. W. King, H. Koseki, R.J. Klose, Synergy between variant PRC1 complexes defines polycomb-mediated gene repression, Mol. Cell 74 (2019) 1020–1036.e8, https://doi.org/10.1016/j.molcel.2019.03.024.
[132] M. Békés, D.R. Langley, C.M. Crews, PROTAC targeted protein degraders: the past is prologue, Nat. Rev. Drug Discov. 21 (2022) 181–200, https://doi.org/10.1038/s41573-021-00371-6.
[133] R.L. Lloyd, P.W.G. Wijnhoven, A. Ramos-Montoya, Z. Wilson, G. Illuzzi, K. Falenta, G.N. Jones, N. James, C.D. Chabbert, J. Stott, E. Dean, A. Lau, L. A. Young, Combined PARP and ATR inhibition potentiates genome instability and cell death in ATM-deficient cancer cells, Oncogene 39 (2020) 4869–4883, https://doi.org/10.1038/s41388-020-1328-y.
[134] F. Wei, X.-X. Cheng, J.Z. Xue, S.-A. Xue, Emerging strategies in TCR-engineered T cells, Front. Immunol. 13 (2022), https://doi.org/10.3389/fimmu.2022.850358.
[135] P. Shafer, L.M. Kelly, V. Hoyos, Cancer therapy with TCR-engineered T cells: current strategies, challenges, and prospects, Front. Immunol. 13 (2022) 835762, https://doi.org/10.3389/fimmu.2022.835762.
[136] Y. Zhang, Z. Liu, W. Wei, Y. Li, TCR engineered T cells for solid tumor immunotherapy, Exp. Hematol. Oncol. 11 (2022) 38, https://doi.org/10.1186/s40164-022-00291-0.
[137] E. Baulu, C. Gardet, N. Chuvin, S. Depil, TCR-engineered T cell therapy in solid tumors: State of the art and perspectives, Sci. Adv. 9 (n.d.) eadf3700. https://doi.org/10.1126/sciadv.adf3700.

[138] R.K. Gupta, B. Rekhi, M.C. Sharma, R. Ganga, M. Ravina, A. Kumar, D. Sahu, The BCOR-rearranged sarcoma involving the lung: diagnosis with clinical outcome and literature review, Indian J. Pathol. Microbiol. 67 (2024) 865, https://doi.org/10.4103/ijpm.ijpm_569_24.

[139] S. Sleijfer, I. Ray-Coquard, Z. Papai, A. Le Cesne, M. Scurr, P. Schöffski, F. Collin, L. Pandite, S. Marreaud, A. De Brauwer, M. van Glabbeke, J. Verweij, J.-Y. Blay, Pazopanib, a multikinase angiogenesis inhibitor, in patients with relapsed or refractory advanced soft tissue sarcoma: a phase II study from the European organisation for research and treatment of cancer-soft tissue and bone sarcoma group (EORTC study 62043), J. Clin. Oncol. 27 (2009) 3126–3132, https://doi.org/10.1200/JCO.2008.21.3223.

[140] E. Choy, J.E. Butrynski, D.C. Harmon, J.A. Morgan, S. George, A.J. Wagner, D. D'Adamo, G.M. Cote, Y. Flamand, C.H. Benes, D.A. Haber, J.M. Baselga, G. D. Demetri, Phase II study of olaparib in patients with refractory Ewing sarcoma following failure of standard chemotherapy, BMC Cancer 14 (2014) 813, https://doi.org/10.1186/1471-2407-14-813.

[141] S.P. D'Angelo, M.R. Mahoney, B.A. Van Tine, J. Atkins, M.M. Milhem, B. N. Jahagirdar, C.R. Antonescu, E. Horvath, W.D. Tap, G.K. Schwartz, H. Streicher, A non-comparative multi-center randomized phase II study of nivolumab +/− ipilimumab for patients with metastatic sarcoma (Alliance A091401), Lancet Oncol. 19 (2018) 416–426, https://doi.org/10.1016/S1470-2045(18)30006-8.

[142] J. Mateo, C.J. Lord, V. Serra, A. Tutt, J. Balmaña, M. Castroviejo-Bermejo, C. Cruz, A. Oaknin, S.B. Kaye, J.S. de Bono, A decade of clinical development of PARP inhibitors in perspective, Ann. Oncol. 30 (2019) 1437–1447, https://doi.org/10.1093/annonc/mdz192.

[143] T.A. Yap, M.G. Krebs, S. Postel-Vinay, A. El-Khouiery, J.-C. Soria, J. Lopez, A. Berges, S.Y.A. Cheung, I. Irurzun-Arana, A. Goldwin, B. Felicetti, G.N. Jones, A. Lau, P. Frewer, A.J. Pierce, G. Clack, C. Stephens, S.A. Smith, E. Dean, S. J. Hollingsworth, Ceralasertib (AZD6738), an oral ATR kinase inhibitor, in combination with carboplatin in patients with advanced solid tumors: a phase I study, Clin. Cancer Res. 27 (2021) 5213–5224, https://doi.org/10.1158/1078-0432.CCR-21-1032.

[144] S.B. Whittle, S. Fetzko, A. Roy, R. Venkatramani, Soft tissue and visceral organ sarcomas with BCOR alterations, J. Pediatr. Hematol. Oncol. 44 (2022) 195–200, https://doi.org/10.1097/MPH.0000000000002480.

[145] H. Mahadevia, B. Ponvilawan, A. Al-Obaidi, J. Buckley, J. Subramanian, D. Bansal, Exceptional synergistic response of PARP inhibitor and immune checkpoint inhibitor in esophageal adenocarcinoma with a germline BRCA2 mutation: a case report, Ther Adv Med Oncol 16 (2024) 17588359241242406, https://doi.org/10.1177/17588359241242406.

[146] S. Tara, Y. Isshiki, Y. Nakajima-Takagi, M. Oshima, K. Aoyama, T. Tanaka, D. Shinoda, S. Koide, A. Saraya, S. Miyagi, I. Manabe, H. Matsui, H. Koseki, V. J. Bardwell, A. Iwama, Bcor insufficiency promotes initiation and progression of myelodysplastic syndrome, Blood 132 (2018) 2470–2483, https://doi.org/10.1182/blood-2018-01-827964.

[147] P. Sportoletti, D. Sorcini, B. Falini, BCOR gene alterations in hematologic diseases, Blood 138 (2021) 2455–2468, https://doi.org/10.1182/blood.2021010958.