Biochemistry and Biophysics Reports 42 (2025) 101972

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

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

Comprehensive in silico characterization of nonsynonymous SNPs in the human ezrin (EZR) gene and their role in disease pathogenesis

Sadia Akter [1], Mohtasim Fuad [1], Zimam Mahmud [*], Sonia Tamanna, Mohammad Sayem, Khalid Hasan Raj, Md. Zakir Hossain Howlader

Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, 1000, Bangladesh

ARTICLE INFO

Keywords:
EZR gene
Nonsynonymous mutations
FERM domain
ERM binding protein 50
Ezrin binding domain
Molecular docking

ABSTRACT

Ezrin (EZR) is a crucial linker between the actin cytoskeleton and the plasma membrane. It interacts with proteins involved in cancer-related signaling pathways. To assess the impact of nonsynonymous single nucleotide polymorphisms (nsSNPs) on EZR structure and function, we employed bioinformatics tools (SIFT, PolyPhen-2, PROVEAN, PhD-SNP, SNPs&GO, SuSPect, and FATHMM) and identified deleterious variants. Stability analyses using MUpro, mCSM, I-Mutant 2.0, and DynaMut2 revealed six destabilizing nsSNPs (F240S, H288D, I248T, L59Q, L125S, and L225P). Structural modeling using HOPE, MutPred2, AlphaFold, Swiss-Model, and protein-protein docking using HADDOCK 2.4 assessed the impact on the EZR-EBP50 complex. Binding free energy calculations, salt bridge analysis, and interface residue mapping further confirmed that the L225P, F240S, and I248T mutations significantly impaired EZR-EBP50 interaction, potentially disrupting key signaling pathways. Molecular dynamics simulations indicated that mutant EZR proteins exhibited reduced stability, flexibility, and hydrogen bonding. This first comprehensive in silico analysis of EZR highlights pathogenic nsSNPs that may contribute to disease progression. These findings provide a foundation for experimental validation and may inform targeted therapies for EZR-related pathologies.

1. Introduction

Ezrin (EZR) belongs to the family of ezrin/radixin/moesin (ERM) proteins, which function as linkers between the cytoskeleton and membrane and include actin-binding proteins [1]. EZR has been found exclusively in the apical canalicular membrane of parietal cells within the stomach epithelium [2]. There are three domains that make up ERM proteins: the C-terminal actin-binding domain (C-ERMAD), the helical domain and the N-terminal FERM domain (N-ERMAD), which again consists of three subdomains: F1, F2 and F3 [3]. When ezrin is in its dormant form, the head-to-tail joining of FERM and the C-ERMAD domains forms a stable complex, masking F-actin along with membrane protein binding sites and when ezrin is in its active form, it links the actin cytoskeleton to membranes [3,4]; disrupting this stable connection which is necessary for the adoption of the open active conformation that can bind F-actin [5]. EZR is involved in biochemical activities such as adhesion, migration, cytokinesis, and surface structure development [6]. According to recent in vivo investigations, the activity of ERM proteins depends on the binding of the FERM domain to phosphatidylinositol 4,5-bisphosphate [7]. EZR plays a role in forming dedicated domains of the plasma membrane by acting as linkers between actin and the plasma membrane; its function in tubulogenesis has also been studied [8]. The persistence of epithelial cells is thus affected by ezrin through stimulation of the PI3K/Akt pathway, where two distinct sites of ezrin bind with the regulatory subunit (p85) of phosphatidylinositol 3-kinase: the amino-terminal domain and the phosphorylated Tyr-353 residue [8]. Through an extra EB (ERM binding) protein-interacting domain, NHERF1/EBP50 (Na+/H+ exchanger 3 regulating factor 1; ezrin-radixin-moesin (ERM) binding phosphoprotein 50) binds EZR, allowing cell surface proteins to connect to the intracellular actin cytoskeleton beneath the plasma membrane [9].

Some studies have shown that the gene expression of EZR is increased in malignant solid carcinoma and is correlated with patient prognosis, for example, cervical cancer in women [10], gastric cancer [11], breast cancer [12] and pancreatic cancer [13]. Ezrin alters the infiltration ability of cancer cells in two ways: by forming a complex

* Corresponding author.
E-mail address: zimam@du.ac.bd (Z. Mahmud).
[1] Sadia Akter and Mohtasim Fuad contributed equally as first authors.

https://doi.org/10.1016/j.bbrep.2025.101972
Received 17 October 2024; Received in revised form 21 February 2025; Accepted 3 March 2025

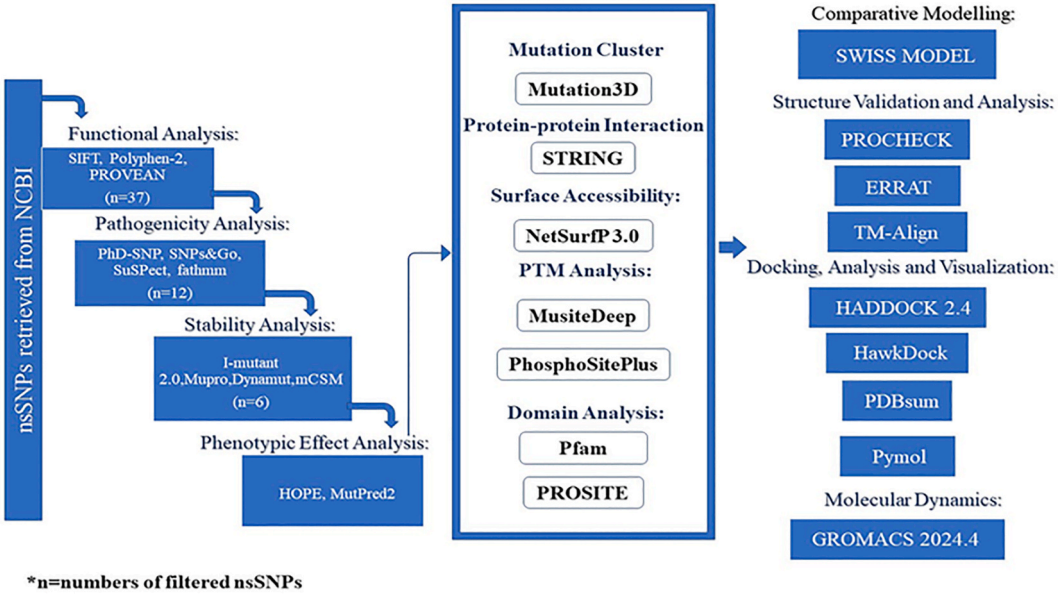


Fig. 1. Diagrammatic illustration of the study **workflow.** The total procedure can be summed up as a progression of filtration stages taken to identify the most harmful nsSNPs of EZR and subsequently through analysis of specific nsSNPs.

with calyx glycoprotein to reduce adhesion between cancer cells through an antiadhesion role and by activating glycoproteins to contribute to cytoskeleton transformation, pseudopodia formation, and improved cell migration [12]. Ezrin is expressed mostly on surface structures in normal cells, including microvilli (rich in actin), pseudopodia but in tumor cells, it is also upregulated [14]. According to an in vitro binding analysis, the cytoplasmic domain of the surface protein CD44, which is implicated in migration and metastasis, binds directly to the amino-terminal half of ERM proteins [15]. Certain CD44 isoforms are exclusive to cancerous cells and cell lines produced from tumors [16]. In in-vitro and in-vivo studies, ezrin has been shown to enhance breast cancer proliferation, migration, invasion, and angiogenesis. Mechanistic analysis of this effect revealed that ezrin interacted with AKT, promoting its kinase activity and activating its downstream signaling, which in turn led to increased angiogenesis and metastasis in BC cells [17]. Another study revealed that EZR activated the FAK/AKT signaling pathway to promote the development of pancreatic cancer [13]. The presence of microvilli and the localization of NHERF1/EBP50 depend on the interaction between ERM proteins and EBP50 [18]. In the sub-plasma membranous area of cells, NHERF1 is linked to the cortical actin cytoskeleton through the PDZ1 domain and the ezrin-binding (EB) domain [19]. This interaction of the EB domain with ERM proteins prevents NHERF1 from folding, which makes PDZ domains available to bind PDZ ligands such as PTEN or β-catenin [20].

As the most common genetic variation in the human genome, SNPs occur approximately every 100–300 bp and are related to variations in a single nucleotide at a particular genomic location among individuals [21]. SNPs in introns do not impose any change in the amino acid sequence of the proteins, and the majority of them are neutral; however, some of them alter protein function, which may increase the risk of developing certain diseases [22]. Currently, a common method for identifying precise point mutations of a certain protein that are connected to diseases or malignancies is in-silico research. These in-silico methods have made extensive use of a variety of technologies to filter out harmful functional nonsynonymous SNPs (nsSNPs).

Working with a large number of nsSNPs in wet laboratories is strenuous and can also be very costly. However, the use of a computational approach with many relevant bioinformatics tools makes it easy to develop selective research ideas while making the whole process cost effective and more precise. These computational tools use many algorithms to predict changes in structure as well as functions. They also assign scores that have specific meanings for understanding the result. Numerous algorithms are available and are being developed for the prediction of the functional, structural, phenotypic effects of nsSNPs [23]. Recently, many refined tools have been developed to study the effects of mutations on proteins [24].

To identify harmful nsSNPs in the EZR protein and assess their pathogenic or negative effects, an investigation was conducted in this study to identify harmful nsSNPs in the EZR protein and assess their pathogenic or negative effects. We classified high-risk nsSNPs via several prediction techniques and examined their structural and functional effects on the EZR protein. Moreover, molecular docking computations were employed to determine the impact of nsSNPs on the structure of the protein. Consequently, the findings of this in-silico study might offer relevant information for the investigation and creation of targeted medications and therapeutic approaches that address a range of diseases, including cancer.

2. Materials and methods

The complete process for selecting the most deleterious nsSNPs of EZR and analyzing their structures is presented in Fig. 1.

2.1. Collecting nsSNPs

A total of 656 nsSNPs of the human EZR gene were identified via the NCBI database (https://www.ncbi.nlm.nih.gov/) [25]. The human EZR protein sequence was acquired from the UniProt database (UniProt ID: P15311) [26].

2.2. Prediction of deleterious SNPs in the EZR gene

SIFT [27], PolyPhen-2 [28] and PROVEAN [29] were employed to predict deleterious nsSNPs that affect protein structure and function. These bioinformatic techniques were used to sort the harmful nsSNPs that were collected from dbSNP of NCBI. Two popular bioinformatics methods for predicting pathogenic nsSNPs are SIFT and Polyphen-2. SIFT makes predictions about whether changing the amino acid sequence will be beneficial or detrimental to proteins. A query position is predicted to be “Affected” if the SIFT score is less than 0.05. On the

other hand, if a query position's SIFT score is equal to or higher than 0.05, it signals that these are tolerated [27]. SNPs predicted as "Affected" in SIFT are given to PolyPhen-2 and PROVEAN. Polyphen-2 forecasts each substitution's potential impact on a protein's structural and functional characteristics. The structural and evolutionary conservation analysis yielded this prediction. Polyphen-2, in contrast to SIFT, typically provides three prediction categories: "Probably Damaging," "Possibly Damaging," and "Benign." The range of the Polyphen-2 score is 0–1, where 0 refers to benign and 1 refers to the most damaging [28]. PROVEAN anticipates the functional effectiveness of the variants. In PROVEAN, a score of < -2.5 indicates a harmful SNP; however, a lower score (e.g., −4.1) can also be utilized as a threshold to increase specificity and confidence [29].

2.3. Prediction of disease-related SNPs

To predict disease-associated SNPs in the EZR gene, we used four websites: i) PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html), ii) SNPs&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html), iii) SuSPect (https://www.sbg.bio.ic.ac.uk/suspect/) and iv) fathmm (https://fathmm.biocompute.org.uk/). A gene's nsSNPs can be classified by PhD-SNP as neutral or disease-causing mutations [30]. Conversely, SNPs&GO uses GO keywords to predict variants linked to disease [31]. On the basis of their tendency to cause human diseases, the SNPs were divided into two categories: "Disease" and "Neutral." For further assessment, the SNPs with the label "Disease" were sorted. In PhD-SNP and SNPs&GO, the reliability index (RI) has a range of 0–10, with 10 denoting the maximum reliability. SuSPect generates a table with values ranging from 0 to 100. A score of 50 is advised to distinguish between neutral and disease-causing variations, with higher scores indicating more confidence predictions [32]. We applied the "Inherited Disease" option in the fathmm server to provide predictions that can distinguish between neutral polymorphisms and mutations that cause disease [33]. The harmful nsSNPs associated with human diseases were selected for further analysis.

2.4. Prediction of protein stability changes

We employed four distinct tools to identify any changes in protein stability: MUpro, mCSM, I-Mutant 2.0, and Dynamut2. We applied I-Mutant 2.0 (https://folding.biofold.org/i-mutant/i-mutant2.0.html) to predict alterations in protein stability due to mutation. They use an RI (reliability index) of 0–10, where 10 is the most dependable, to make predictions about the outcome. It also evaluates the free energy change value or delta del G (DDG), which is a scale that indicates an increase or decrease in stability and the degree of protein instability. Protein stability decreases when the DDG value is less than 0, whereas protein stability increases when the DDG value is greater than 0 [34]. Moreover, the MUpro (http://mupro.proteomics.ics.uci.edu/) server was also utilized to evaluate stability. This web server was constructed via support vector machines and neural networks. Using a positive or negative scoring system, these techniques calculate the effects of point modifications of amino acids on protein stability and increase or reduce stability in the same way as I-mutant 2.0 does [35]. Additionally, mCSM (https://biosig.lab.uq.edu.au/mcsm/stability) makes three predictions: stabilizing, destabilizing, and highly destabilizing [36]. The more negative the value of DDG is, the greater the degree of destabilization of the mutant protein. Dynamut2 (https://biosig.lab.uq.edu.au/dynamut2/) also predicts mutation as stabilizing or destabilizing in the same manner [37].

2.5. Analyzing the structural and functional effects of the nsSNPs

To examine the phenotypic effects of the selected deleterious nsSNPs, the HOPE (https://www3.cmbi.umcn.nl/hope/) server was used [38]. This automated mutant analyzer generates a report that shows how the mutation affects the charge, size, bonding pattern, and interaction of the protein with other proteins and molecules. We also used MutPred2 (http://mutpred.mutdb.org/) to analyze structural changes due to these mutations. MutPred2 uses a neural network technique for predicting how amino acid changes may affect function and structure [39]. The effects could include altered DNA binding, changes in the allosteric site, changes in solvent accessibility all of which could result in significantly changed protein phenotypic characteristics. For the FASTA sequence of the EZR protein, the relevant amino acid changes were given as input into this server, and the P value was 0.05, which was set as the default value.

2.6. Predicting mutation clusters

To determine if there is any mutation cluster that could be segments of DNA prone to genetic alteration, Mutation3D (http://mutation3d.org/) was used [40]. This tool is able to detect amino acid clusters on protein models and structures via a complete-linkage clustering approach.

2.7. ***Predicting*** *protein–protein interactions*

STRING (https://string-db.org/) was used to find out the proteins that interact with EZR the most. It provides a brief overview of the protein's function, regulation, molecular action [41]. The highest confidence score was used to find out the interacting proteins.

2.8. Surface accessibility and posttranslational modification (PTM) analysis

NetSurfP-3.0 (https://services.healthtech.dtu.dk/services/NetSurfP-3.0/) was used to estimate the surface accessibility of each amino acid residue of the EZR protein [42]. This web tool is sequence-based and makes predictions for each residue of the query sequences, such as relative solvent accessibility, secondary structure (e.g., alpha helix, beta sheet), and structural disorder. Post-translational changes in protein molecules are also very important. By modifying intra and inter-molecular interactions, they introduce novel functions and exercise control over protein structure and function. MusiteDeep (https://www.musite.net/) was used in this investigation to predict putative PTM sites and the associated likely alterations in the EZR protein sequence [43]. Additionally, we used PhosphoSitePlus (https://www.phosphosite.org/homeAction.action), a tool for researching the post-translational changes observed in experimental studies [44].

2.9. Domain analysis of the EZR

We utilized Pfam (https://www.ebi.ac.uk/interpro/entry/pfam/#table) [45] and PROSITE (https://prosite.expasy.org/) to determine the functional domain of the EZR protein. Pfam predicts a protein's domains via hidden Markov models and multiple sequence alignment [45], whereas PROSITE uses functional and structural data [46].

2.10. Modeling of the structures of mutant proteins

Swiss Model (https://swissmodel.expasy.org/), a homology modeling tool, was used to model mutated proteins containing all filtered harmful nsSNPs [47]. Owing to the absence of the complete crystal structure of EZR, the full structure predicted by Alphafold (https://alphafold.ebi.ac.uk/) was used as the template (structure ID: AF-P15311-F1), as the structure predicted by Alphafold is highly accurate [48]. To validate the use of the Alphafold-predicted structure as a template, alignment between this structure and the crystal structure was necessary. TM-Align [49] was used to obtain the TM score, and PyMOL [50] was used to obtain the root mean square deviation (RMSD) score. Two crystal structures were acquired from the RCSB PDB: 4RM9 (https:

//www.rcsb.org/structure/4RM9), in which the modeled residue count is 368 amino acids, and 4RMA (https://www.rcsb.org/structure/4RMA), which is the crystal structure of the FERM domain of human Ezrin and the modeled residue count is 295 amino acids at each chain. The SAVES v6.0 website (https://saves.mbi.ucla.edu/) and the SWISS-MODEL structure assessment program were used to evaluate the quality of the projected 3D structures [51]. SAVES v6.0 has five distinct validation tools. To validate our modeled mutant structures, we used ERRAT [51] and PROCHECK [52]. Following the prediction and verification of each mutant model and the wild-type protein model, we measured the TM score and the root mean square deviation (RMSD) value via TM align and to obtain Z scores, ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) was utilized. ProSA calculates the Z score for a specific input structure and indicates overall model quality [53] along with a plot of Z scores of all available experimentally solved protein structures in the Protein Data Bank; thus, we can check whether our obtained Z score is within the range of Z scores normally observed for previously solved similar-sized proteins [53].

2.11. Docking by HADDOCK 2.4 and analysis of complexes by HawkDock and PDBsum

HADDOCK 2.4 (https://rascar.science.uu.nl/haddock2.4/) was used which is a flexible docking approach for protein-protein docking purpose. The wild-type and mutant EZR were docked with the NHERF1/EBP50 (Na+/H+ exchange regulatory factor 1; ERM binding phosphoprotein 50) molecule [54]. EBP50 protein (PDB ID: 2KRG) molecules and modeled natural and mutant proteins were employed. From the literature, we gathered the interaction (binding) residues between EZR and EBP50 [20,55]. Neighboring binding residues were utilized as passive residues, and the binding residues in EZR that were gathered along with those in the partner EBP50 were utilized as active residues. In our investigation, we used the HADDOCK default parameters. They provide HADDOCK scores along with electrostatic energy, restraint energy, buried surface area, desolvation energy, van der Waals energy, Z scores [54].

The HawkDock (http://cadd.zju.edu.cn/hawkdock/#MMGBSA_submit) server was used to evaluate the binding affinity of complexes of wild-type and mutant EZR complexes with EBP50 downloaded from HADDOCK v2.4 via the MM/GBSA (mechanics/generalized born surface area) free energy decomposition approach [56]. We also used the PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) website to study these complexes of wild-type and mutant EZR with EBP50. We used the "generate" option to create a PDB entry, which provides important information such as secondary structures, the interface area, the number of interface residues, salt bridges, disulfide bonds and nonbonded contacts [57].

2.12. Molecular dynamics simulation

Molecular dynamics (MD) simulations of the EZR wild-type protein and its variants were performed utilizing GROMACS version 2024.4 and the CHARMM36 force field [58]. The initial structure for the wild-type (WT) EZR protein was derived from the AlphaFold-predicted model (structure ID: AF-P15311-F1). Mutant protein structures were modeled using the WT structure as a template and generated through the SWISS-MODEL workspace to produce six variant structures. The proteins were solubilized in a dodecahedral simulation box containing CHARMM-modified TIP3P water molecules, and the addition of counterions neutralized the systems. Energy reduction of the neutralized systems was conducted utilizing the steepest descent algorithm until the highest force decreased below 10 kJ mol^{-1} nm^{-1}, with an upper limit of 5000 steps. Two position-restrained molecular dynamics simulations were performed to equilibrate the solvent and ions. The system's temperature was regulated at 310 K utilizing a V-rescale thermostat [59], while the pressure was controlled at 1 bar with a Parrinello-Rahman barostat [60]. Following equilibration, unrestrained MD simulations were run for 200 ns under the same conditions as the equilibration steps. The LINCS method [61] was utilized to limit hydrogen-containing bonds, while the particle mesh Ewald approach [62] was used to calculate long-range electrostatic interactions. Lennard–Jones interactions, in conjunction with short- and long-range electrostatic interactions, utilized a cut-off distance of 12 Å. A time step of 2 fs was established for the incorporation of Newton's equations of motion. For the WT and each mutant variant, a single MD simulation trial was conducted.

The trajectory files generated from molecular dynamics simulations were analyzed using GROMACS version 2024.4. Root mean square deviation (RMSD) was calculated to evaluate the system's structural stability over the simulation time. The radius of gyration (Rg) was analyzed to assess protein compactness, folding, and overall stability. Root mean square fluctuation (RMSF) was used to quantify the flexibility of individual residues, highlighting the most dynamic regions of the protein. Hydrogen bond analysis was performed to examine the formation and stability of both intramolecular and intermolecular hydrogen bonds. Additionally, solvent-accessible surface area (SASA) was calculated to determine the solvent-exposed surface area, providing insights into hydrophobic and hydrophilic regions. The results of these analyses were visualized and compared between the wild-type protein and mutant variants using the R package ggplot2 [63]. The trajectory files were also analyzed to compare the secondary structure of the wild-type EZR and its variants throughout the simulation, with the results visualized using the Matplotlib Python package [64].

3. Results

3.1. Collection of SNPs from the database

There are a total of 26525 SNPs for the EZR gene according to the NCBI dbSNP database where 24401 SNPs are intronic and 2124 SNPs are located in the coding sequence. Among these, 656 SNPs were found to be missense mutations and were retrieved. The numbers of synonymous SNPs, 3 prime UTR variants and 5 prime UTR variants in the EZR gene were 331, 916 and 139, respectively. Only 656 nsSNPs were selected for this study.

3.2. Prediction of deleterious nsSNPs via SIFT

By analyzing sequence homology and the physical characteristics of amino acids, SIFT predicts that damaging nsSNPs are "affected" [27]. The SIFT results of all 656 nsSNPs are given in Supplementary Table 1. Out of a total of 656 nsSNPs, 221 nsSNPs were predicted to be "affected" and were then analyzed via PolyPhen-2 and PROVEAN.

3.3. Analysis of deleterious SNPs

The results of the scoring of 221 nsSNPs via all three bioinformatic tools are given in Supplementary Table 2. Among 221 nsSNPs, 135 were predicted to be 'probably damaging' by Polyphen-2 [28]. According to Polyphen-2, scores above 0.895 classified as "probably damaging" nsSNPs. A score close to 0 indicate that the mutation is benign. Among the 221 nsSNPs, 181 were found to be deleterious by PROVEAN (cutoff value = −2.5) [29]. Among these, 37 nsSNPs were selected as the most harmful, as these nsSNPs scored 0 in SIFT and 1 in PolyPhen-2 and less than −4.1 in PROVEAN. These were analyzed further in the downstream tools.

3.4. Prediction of disease-causing nsSNPs

The 37 selected nsSNPs were all predicted to be disease caused by PhD-SNPs (Supplementary Table 3). However, SNPs&GO revealed 33 disease-related nsSNPs, SuSPect predicted 13 affected nsSNPs (cutoff

Table 1
Scores and predictions of the most damaging and disease-causing mutations revealed by the 7 computational tools.

Variant ID	AA Variation	SIFT Score	Polyphen2 Score	PROVEAN Score	PhD-SNP Prediction	SNPs&GO Prediction	SuSPect Score	fathmm Prediction
rs1430541322	R579C	0	1	−6.505	Disease	Disease	97	Damaging
rs763195862	T567 M	0	1	−4.787	Disease	Disease	98	Damaging
rs940756446	Y291C	0	1	−8.904	Disease	Disease	91	Damaging
rs774565006	H288D	0	1	−8.913	Disease	Disease	96	Damaging
rs201942373	I248T	0	1	−4.719	Disease	Disease	96	Damaging
rs1334969871	F240S	0	1	−7.87	Disease	Disease	98	Damaging
rs1408150607	L225P	0	1	−6.851	Disease	Disease	97	Damaging
rs1341436950	Y191C	0	1	−8.779	Disease	Disease	96	Damaging
rs765797550	L125S	0	1	−5.997	Disease	Disease	98	Damaging
rs775710251	Y85C	0	1	−7.05	Disease	Disease	95	Damaging
rs1372500947	L59Q	0	1	−5.471	Disease	Disease	92	Damaging
rs1194854415	Y44C	0	1	−6.002	Disease	Disease	97	Damaging

Table 2
Free energy change (DDG) and predictions of the stability of EZR due to nsSNPs.

AA Position	I-mutant 2.0		MUpro		mCSM		Dynamut2	
	DDG (kcal/mol)	**Prediction**	**DDG (kcal/mol)**	**Prediction**	**DDG (kcal/mol)**	**Prediction**	**DDG (kcal/mol)**	**Prediction**
H288D	−0.2	DECREASE	−0.4438	DECREASE	−1.851	DESTABILIZING	−1.75	DESTABILIZING
I248T	−2.25	DECREASE	−1.8253	DECREASE	−2.603	HIGHLY DESTABILIZING	−2.61	DESTABILIZING
F240S	−2.2	DECREASE	−1.65795	DECREASE	−3.828	HIGHLY DESTABILIZING	−3.49	DESTABILIZING
L225P	−1.77	DECREASE	−1.75308	DECREASE	−1.031	DESTABILIZING	−0.8	DESTABILIZING
L125S	−4.02	DECREASE	−1.5397	DECREASE	−2.594	HIGHLY DESTABILIZING	−2.68	DESTABILIZING
L59Q	−1.72	DECREASE	−1.82236	DECREASE	−2.311	HIGHLY DESTABILIZING	−2.01	DESTABILIZING

value = 50) [32], and fathmm predicted 26 nsSNPs as damaging (Supplementary Table 3).

Here, we modified the threshold for SuSPect to 90 as the most likely cause of disease, as alterations scoring 90–100 represent the most disease-causing [32,65]. As a result, a total of 12 deleterious nsSNPs in the EZR gene were predicted to be disease causing according to all four tools (Table 1).

3.5. Prediction of the effect on the stability of the EZR

The selected 12 nsSNPs were analyzed by four servers to determine their stability, and the results are given in Supplementary Table 4. While analyzed by MUpro servers, we found that the DDG scores of all 12 nsSNPs were less than 0, indicating that they had decreased stability (Supplementary Table 4). On the other hand, the stability of 8 nsSNPs decreased in response to I-mutant 2.0 (Supplementary Table 4). mCSM predicted that 6 nsSNPs and 5 nsSNPs were destabilizing and highly destabilizing respectively (Supplementary Table 4). Dynamut2 predicted 9 nsSNPs as destabilizing (Supplementary Table 4). As a result, out of the 12 disease-related SNPs in the EZR gene, 6 nsSNPs were found to be destabilizing the protein by all four servers (Table 2).

3.6. Structural changes caused by nsSNPs

Project HOPE was used to understand the structural effects of a mutation, including changes in amino acid charge, size and hydrophobicity and modifications in interactions with other molecules. The findings from HOPE are mentioned below:

For the L59Q mutation (Fig. 2- a), glutamine (mutant) is larger and more hydrophobic than leucine. Glutamine may not fit into the core of the EZR where the native residue was concealed and might cause a loss of connections. For the L125S mutation (Fig. 2- b), the wild-type residue is 100 % conserved, which means that only leucine was found at this position, thus damaging the protein. The mutant residue is more hydrophobic, and its size is smaller than that of the native residue, resulting in a bare space, which results in important interactions in the fundamental part of the protein. For the L225P mutation (Fig. 2- c), leucine is highly conserved, and the mutation results in little space in the core of the EZR. Moreover, the mutant residue is smaller than the leucine residue. In the case of the F240S mutation (Fig. 2- d), the size of the mutant residue is relatively small, which might cause a potential loss of outer interactions and cost hydrophobic connections on the surface of EZR with many molecules. Phenylalanine (F) is also 100 % conserved, indicating possible damaging mutations in the structure of EZR.

In the case of the I248T mutation (Fig. 2- e), isoleucine is smaller and more hydrophobic than threonine. The position near a highly conserved residue is highly conserved; thus, the mutation is likely deleterious. Owing to this nsSNP, the loss of external and hydrophobic contacts with other molecules on the surface of EZR can also occur. Finally, for the H288D mutation (Fig. 2- f), there are differences in charge and size. The charge of the native residue is neutral, but the charge of the mutant residue (aspartate) is negative, which might cause repulsion between aspartate and other nearby residues. The residue is also 100 % conserved, thus indicating possible damaging mutations. The size of the mutant residue is small, and the change might cause a loss of external interactions.

Moreover, all the mutations are located within the FERM domain, which is important for binding with many crucial molecules important for EZR functions. Thus, mutations in this domain can destroy its function and binding properties. Among these 6 mutations, the H288D and I248T mutations are located within a distinctive region where interaction with SCYL3 occurs; thus, these mutations can interrupt its function.

Next, the MutPred2 server was used for similar purposes, and all 6 nsSNPs were found to be pathogenic, as they scored greater than 0.8. Two of the mutations, L225P and L125S, caused altered ordered interfaces. F240S and I248T caused an altered disordered interface. L125S caused the loss of the allosteric site at Y128, whereas I248T caused the loss of acetylation at K253 and the loss of N-linked glycosylation at N247. Loss of strands is caused by H288D. Additionally, I248T, F240S, and L225P caused altered transmembrane properties. I248T and L225P increased relative solvent accessibility. The Mutpred2 results revealed that L225P (score = 0.975) had the most alterations, including altered metal binding and a gain of sulfation at Y228. A summary of the analysis by Mutpred2 is given in Supplementary Table 5.

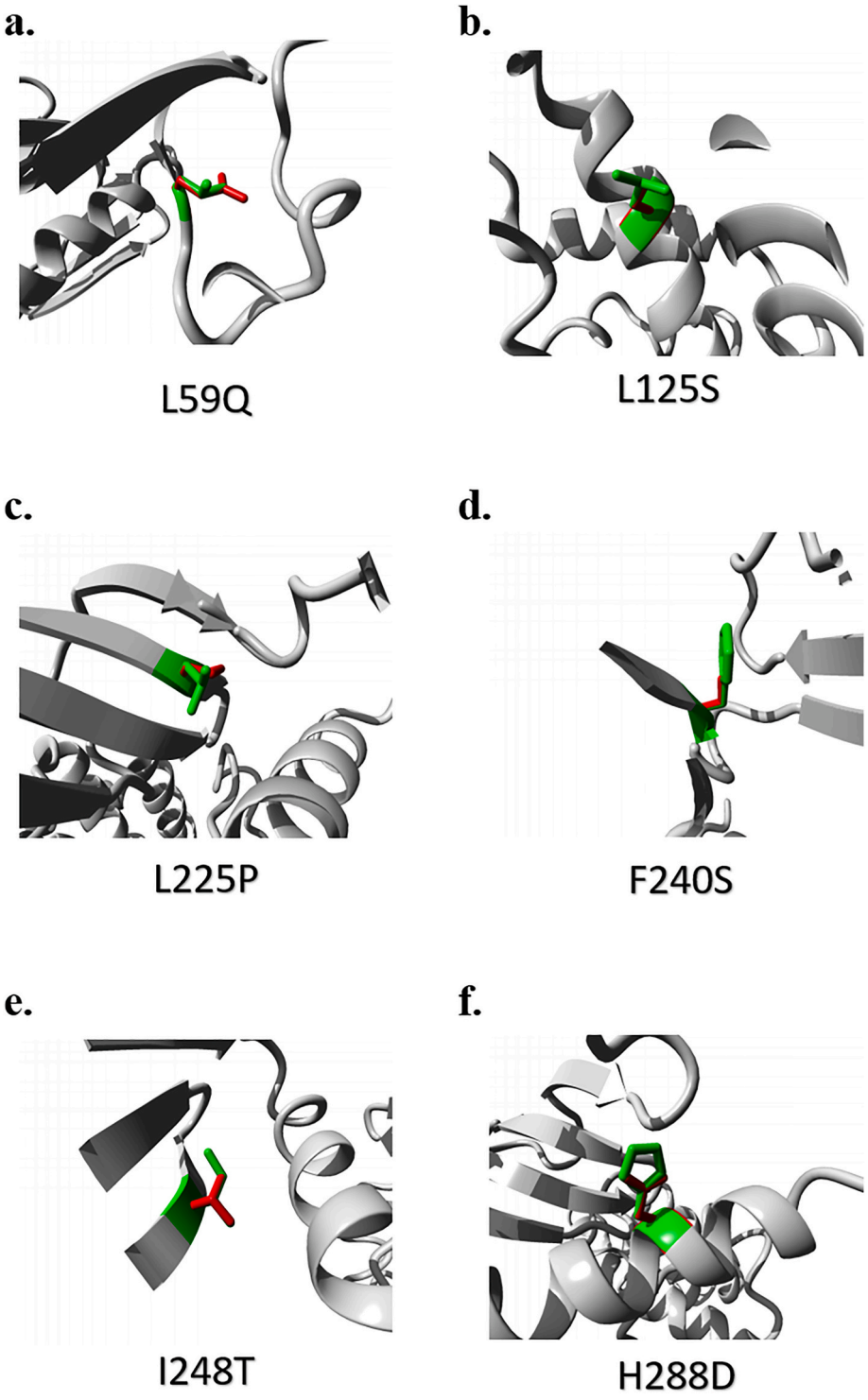


Fig. 2. Changes in amino acid structure caused by the 6 most damaging nsSNPs provided by Pro**ject-HOPE.** The other chains of EZR are shown in gray, and the side chains of the native and mutant residues are shown in green and red, respectively. Note: (a) L59Q (b) L125S (c) F240S (d) L225P (e) I248T (f) H288D.

3.7. Prediction of clusters of mutations

We analyzed the deleterious, disease-related and destabilizing 6 nsSNPs via the mutation3D server for predicting mutation clusters. The positions marked in red indicate clustered SNP locations, and the positions marked in blue indicate covered mutations (Fig. 3). The cluster contained four SNPs at positions 225, 240, 248 and 288, with a P value of 1.33e-4. These positions were observed to be part of the FERM_C domain (PF09380) in the UniProt database (Fig. 3).

3.8. Protein–protein interaction of EZR via STRING v12.0

From STRING we have found 10 proteins that interact with the EZR (Fig. 4). These proteins are principally involved in binding to membrane proteins, binding the actin cytoskeleton to the plasma membrane, signaling pathways, regulating diverse cellular processes. The minimum required interaction score was set to 0.9 for high confidence.

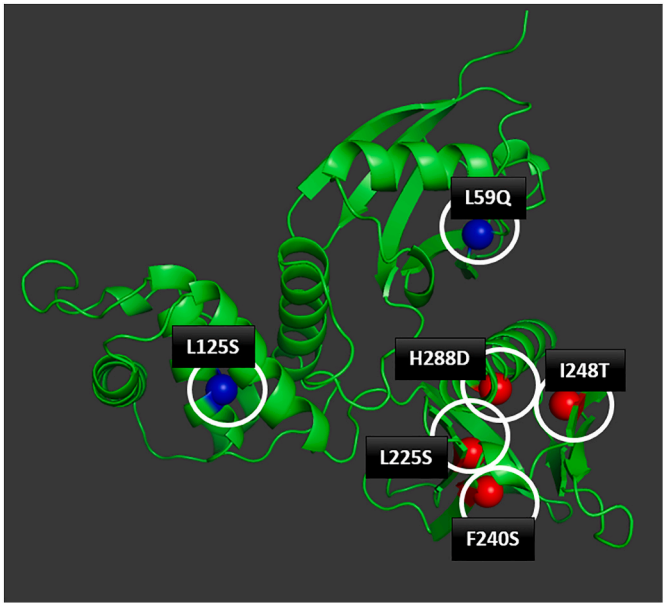


Fig. 3. Mutation cluster analysis via Mu**tation 3D.** The positions of clustered mutations are marked in red: 225, 240, 248, and 288 (somatic mutation hotspots), and the positions marked in blue indicate covered mutations: 59 and 125.

3.9. Surface accessibility analysis of the native and mutant EZR proteins

We analyzed the surface accessibility of native and mutant EZR via NetSurfP-3.0 with default parameters. The RSA (relative surface accessibility) threshold was set to 25 %. We found 3 SNPs among the 12 deleterious and disease-causing SNPs that scored greater than 25 %, indicating an exposed state (Supplementary Table 6). Among these, owing to the H288D mutation, the state was changed to the exposed state from that buried in the native residue. The other 9 nsSNPs scored less than 25 %, indicating a buried state. The prediction results of the most damaging 6 nsSNPs and T567M are given in Table 3, and the secondary structure revealed by NetSurfP-3.0 is given in Fig. 5. The comparison of surface accessibility and secondary structure between wild type and mutant EZR amino acids is also given in Supplementary Fig. 2.

3.10. Post-translational modification (PTM) and domain analysis

We used two different tools to analyze post-translational modifications of the EZR protein: MusiteDeep and PhosphoSitePlus. With a default cutoff of 0.5, MusiteDeep predicted that I248T is located between two PTM sites: N247, the site for glycosylation, and S249, the site for phosphorylation. Additionally, L59Q is located at the site just before N6-acetylation occurs at K60, where it forms N6-acetyllysine. All the PTM sites predicted by MusiteDeep with and without cutoff PTM scores are given in Supplementary Table 7. According to PhosphoSitePlus (Fig. 6), the FERM domain of the EZR is rich in PTM sites. PhosphoSitePlus also predicted a phosphorylation site at S249. They predicted K60 as a site for acetylation and succinylation. However, neither site predicted any PTM sites at our selected mutation sites.

PhosphoSitePlus also revealed that EZR consists of five domains: i) FERM_N, ii) FERM_M, iii) FERM_C, iv) ERM helical, and v) ERM_C. In our study, we also used the Pfam and PROSITE web servers to predict the functional domains of the EZR protein. Pfam predicted 5 domains of EZR: 1) the FERM central domain, 2) the ezrin/radixin/moesin family C-terminal domain, 3) the FERM N-terminal domain, 4) the FERM C-terminal PH-like domain, and 5) the alpha-helical domain. The positions of residues spanning each domain predicted by pfam are given in Table-4. PROSITE revealed 1 hit by the profile of the EZR protein sequence, which is a FERM domain spanning from 85 to 295 amino acids with a score of 85.295.

Table 3
Surface accessibility of amino acids in wild-type and mutated proteins with RSA (%).

Wild type residue	RSA (%)	State	Mutated residue	RSA (%)	State
H288	22 %	Buried	D	31 %	Exposed
I248	5 %	Buried	T	9 %	Buried
F240	7 %	Buried	S	6 %	Buried
L225	1 %	Buried	P	1 %	Buried
L125	1 %	Buried	S	1 %	Buried
L59	3 %	Buried	Q	7 %	Buried
T567	52 %	Exposed	M	46 %	Exposed

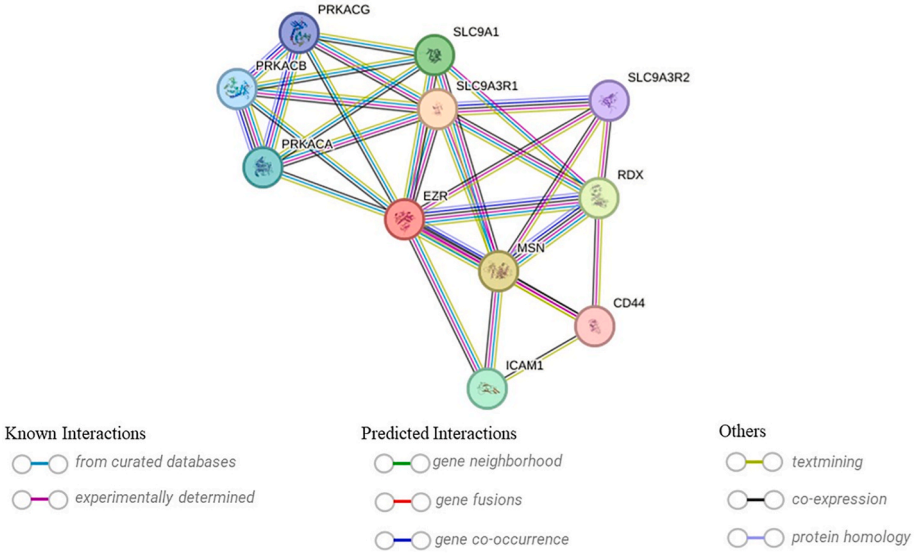


Fig. 4. Exploration of protein–protein interactions through the STRING v**12 server.** The node-colored red is the query protein (EZR). The edges represent protein–protein associations. The minimum required interaction score was set to 0.9 for high confidence.

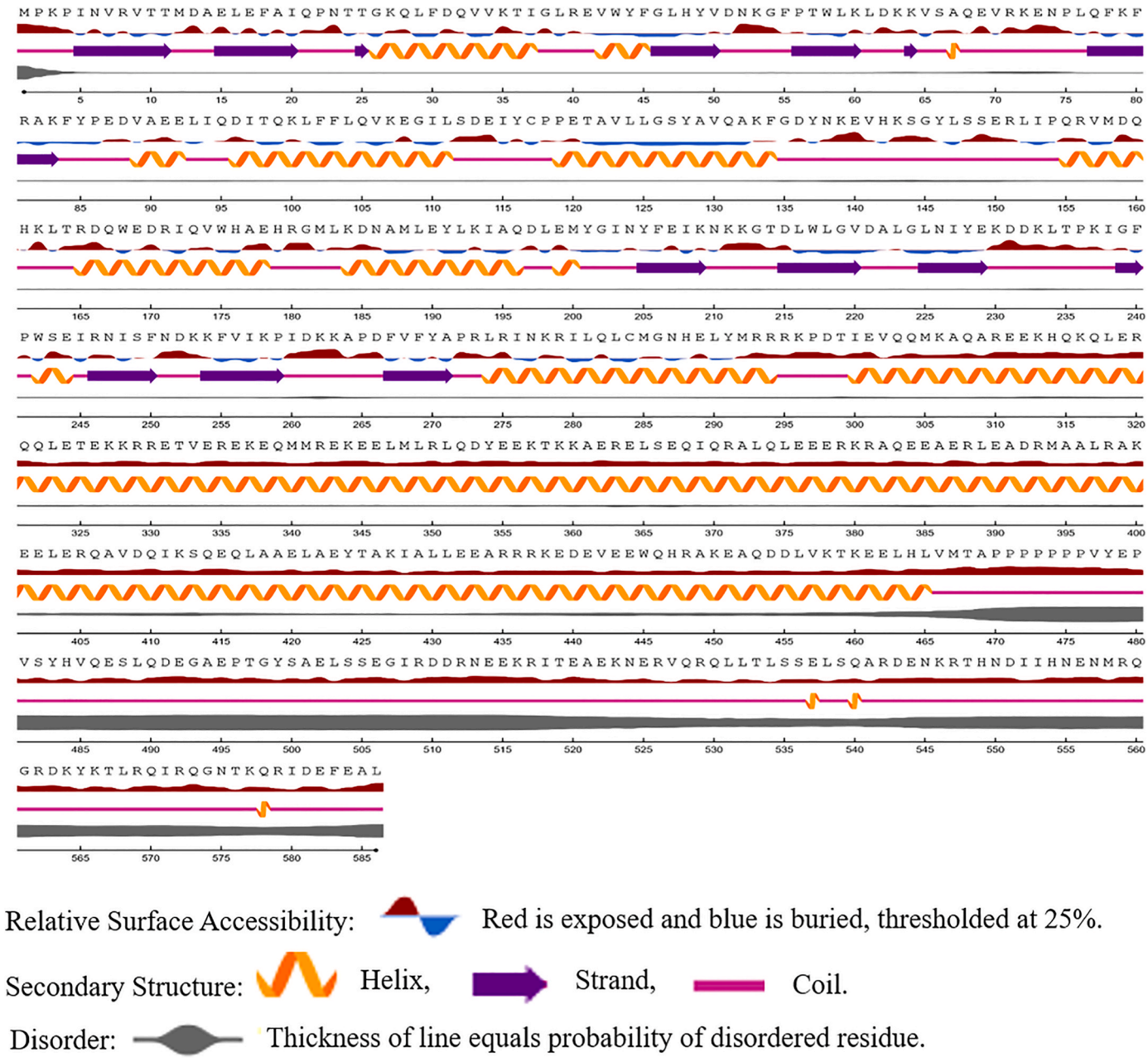


Fig. 5. Surface accessibility, secondary structure and disordered amino acids of the EZR protein predicted by NetSurfP-3.0.

3.11. Structural analyses of the protein mutations

Owing to the absence of a complete crystal structure of EZR, we used the Alphafold-predicted structure as our template (structure ID: AF-P15311-F1). The validity of this structure was checked by determining the RMSD via PYMOL and the TM score via TM-Align alignment of the Alphafold structure and experimental structures (PDB ID: 4RMA and 4RM9). The TM score ranges from 0 to 1, where a score >0.5 indicates similarity between native and mutant structures. However, smaller RMSD values reveal greater similarity between two structures. The scores and superimposition of the Alphafold and experimental structures are given in Table 5 and Figs. 7(a) and 1(b), respectively. Superimposition was performed by using PyMOL.

We used the SWISS-MODEL workspace with AF-P15311-F1 as a template to model all 6 mutant protein structures. After that, we assess the quality of the protein via a Ramachandran plot, the QMEAN Z score, and the MolProbity score from the "Structure Assessment" tool of SWISS MODEL; we also use ERRAT and PROCHECK from the SAVES server and determine Z scores via ProSA. All the predicted models in this study had more than 96 % residues in the Ramachandran favored region. QMEAN Z scores close to zero indicate similarity with native structures; all the models scored from 0.66 to 0.91 in this study. A low MolProbity score is accepted and we also found low scores in all our models. We used ProSA to determine whether the scores from our modeled structures were within the range of z scores normally found for "Protein Data Bank" deposited native proteins of comparable sizes [53]. The results are depicted in Table 6. The higher the score (>50) in ERRAT is, the higher the quality it indicates [51] and all our modeled structures scored more than 90. The Ramachandran plots generated by PROCHECK are given in Supplementary Fig. 1. Using TM-Align, we compared the structural parameters between the wild-type and the mutant EZR. The scores are given in Table 6. The table shows that all the mutated proteins had the same score (TM score = 0.99998 and RMSD = 0.04) and were not much different from those of the wild type.

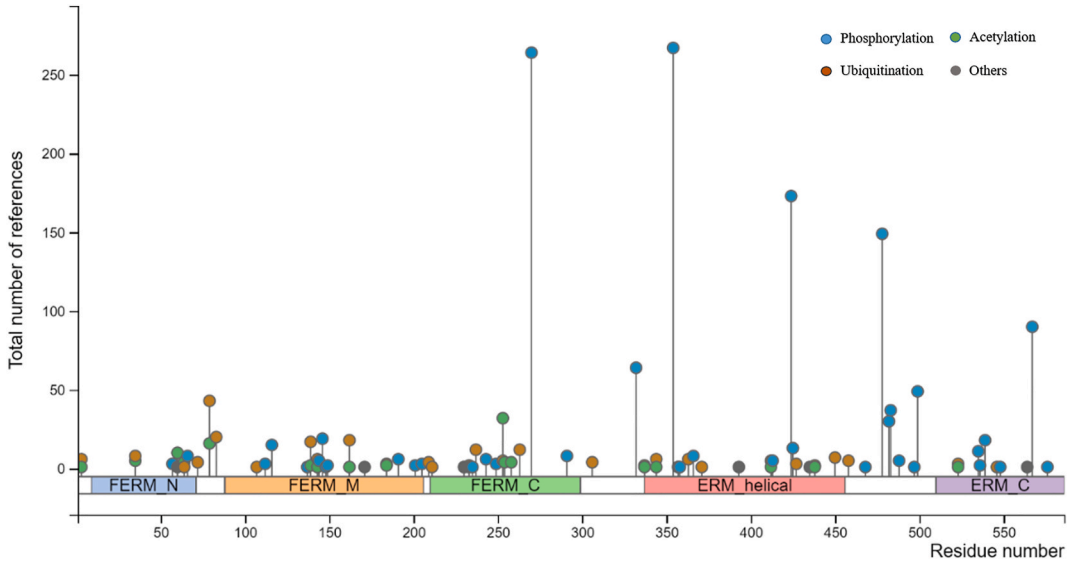


Fig. 6. Prediction of domains and posttranslational modification sites with the total number of references for PTM sites in the EZR protein via Phospho**SitePlus.** EZR consists of five domains: i) FERM_N (9–71), ii) FERM_M (88–206), iii) FERM_C (210–299), iv) ERM helical (337–456), and v) ERM_C (510–586), with hotspots of PTM sites in the FERM_N and FERM_C domains. High-confidence PTM sites are located downstream of the FERM_M domain.

Table 4
Prediction of EZR domains by Pfam.

Accession	Name	Source Database	Type	Sequence Matches
PF00373	FERM central domain	pfam	Domain	91–206
PF00769	Ezrin/radixin/moesin family C-terminal	pfam	Domain	511–586
PF09379	FERM N-terminal domain	pfam	Domain	9–68
PF09380	FERM C-terminal PH-like domain	pfam	Domain	210–298
PF20492	Alpha-helical domain	pfam	Coiled_coil	337–456

Table 5
RMSD (using PyMOL) and TM score (using TM-Align) alignment of the available crystal structures of EZR- and Alphafold-predicted structures.

	RMSD	TM-score	Aligned Residues
4RM9 and AF-P15311-F	0.501	0.98637	303 AA
4RMA and AF-P15311-F	0.471	0.97665	225 AA

3.12. Docking results of wild-type and selected mutated EZR with EBP-50 and binding free energy of docked complexes by hawkdock

We prepared all the proteins (wild type and mutants) for docking with EBP50. We used SwissPDB Viewer (for energy minimization) and Discovery Studio (to add polar hydrogens) then docking was carried out via the HADDOCK 2.4 tool. To comprehend how biological partners interact, the HADDOCK score needs to be calculated. This score enables the structures to be categorized [66]. As the first cluster given by HADDOCK is the best of the other given output clusters, the HADDOCK scores of the native EZR and destabilizing mutated EZR bound with the EBP-50 complex were taken and are shown in Table 7. The HADDOCK score was significantly lower for the L225P protein than for the WT protein, whereas L59Q and F240S scored almost the same as the WT protein. The buried surface area or BSA of the EZR is also determined. The protein complex with the H288D mutation had the lowest BSA value (1753.3±94.6), whereas the wild-type complex had a BSA value of

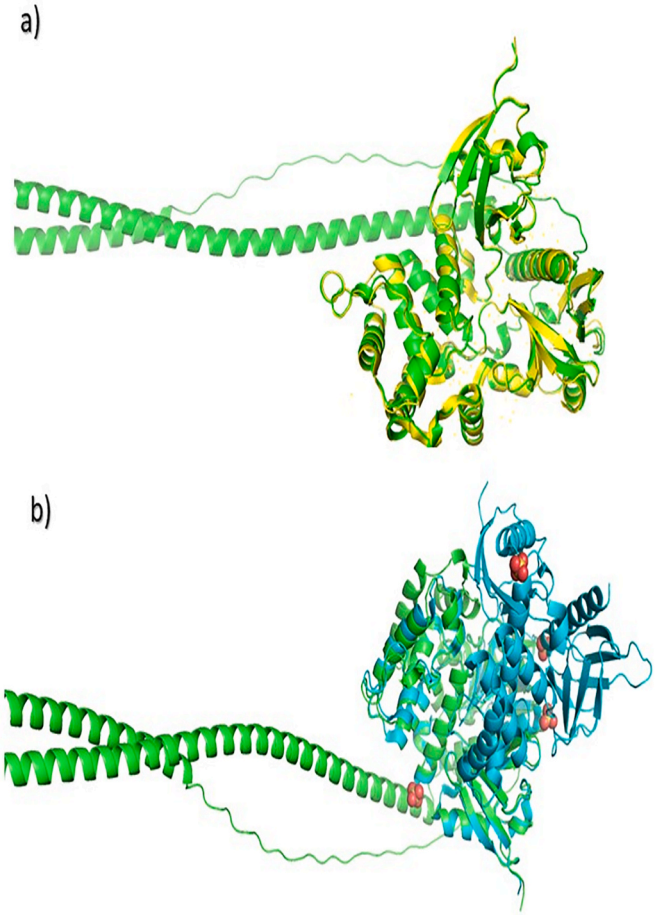


Fig. 7. Alignment of AF-P15311-F1 (labeled in green) with two incomplete crystal structures of EZR: a) 4RM9 (labeled in yellow), b) 4RMA (labeled in cyan).

Table 6
Validation and comparison of template and modeled mutant EZR structures.

Protein Model	SWISS MODEL			ERRAT	PROCHECK	ProSA	TM-Align	
	QMEAN	Ramachandran Favored Region Percentage	Molprobity Score	Score	Ramachandran Favored Region Percentage	Z Score	TM Score	RMSD Value
Template(AF-P15311-F1)	0.66	96.40 %	1.02	91.2477	94.80 %	−8.15	–	–
H288D	0.88	96.23 %	0.93	91.2477	95.00 %	−8.05	0.99998	0.04
I248T	0.76	96.40 %	0.95	91.4339	95.20 %	−8.06	0.99998	0.04
F240S	0.79	96.40 %	0.95	91.2477	95.20 %	−8.06	0.99998	0.04
L225P	0.91	96.40 %	0.95	91.4339	95.20 %	−8.09	0.99998	0.04
L125S	0.79	96.40 %	0.91	91.2477	95.20 %	−8.03	0.99998	0.04
L59Q	0.73	96.40 %	0.98	91.2477	95.20 %	−8.03	0.99998	0.04

Table 7
Docking analysis of wild-type and mutant EZR proteins with EBP-50 via HADDOCK 2.4 and Hawkdock.

Protein Complex	HADDOCK 2.4							HawkDock
	HADDOCK Score	Electrostatic energy (Kcal mol− 1)	Van der Waals energy (Kcal mol− 1)	Desolvation energy (Kcal mol− 1)	Buried Surface Area (A°2)	Restraints violation energy (Kcal mol− 1)	Z Score	Binding free energy of complex (Kcal/mol)
Wild type-EBP-50	−91.4±12.1	−419.7±52.0	−60.7±8.0	4.7±3.2	2175.2±88.7	486.0±47.9	−1.6	−58.61
L59Q-EBP-50	−92.5±10.1	−486.5±34.3	−41.2±5.6	9.1±1.9	1876.1±198.9	369.3±46.5	−1.7	−66.36
L125S-EBP-50	−89.3±8.1	−421.5±58.3	−67.7±10.2	8.8±5.5	2316.7±164.2	538.8±42.8	−1.8	−56.98
L225P-EBP-50	−80.5±5.8	−386.7±74.7	−58.9±3.4	5.3±2.0	2195.9±94.4	504.4±66.9	−1	−44.87
F240S-EBP-50	−92.7±5.0	−436.0±21.5	−48.5±3.5	8.4±0.9	1783.0±88.7	346.1±78.2	−1.6	−51.32
I248T-EBP-50	−88.1±3.1	−437.0±25.2	−59.9±6.3	7.5±1.4	2219.2±75.5	516.7±11.8	−1.3	−49.25
H288D-EBP-50	−84.1±3.8	−463.3±62.4	−43.1±4.8	12.3±8.0	1753.3±94.6	393.1±39.1	−1.3	−54.03

2175.2±88.7 (Table 7). The energy of the normal van Dar Walls complex is −60.7±8.0, whereas that of the mutant protein complex ranges from −67.7±10.2 to −41.2±5.6. To further confirm this, we calculated the binding free energy of these complexes via the MM/GBSA method. **HawkDock** was used to obtain the binding free energy (DDG) values of the complexes, which are given in Table 7. Wild-type EZR binds with EBP50 with a free energy of −58.61 kcal/mol, whereas in the case of the L59Q mutant protein, the binding free energy is the most negative (−66.36 kcal/mol), and the L225P EZR-EBP50 protein complex contains the least negative binding free energy (−44.87 kcal/mol) among the other complexes. All the docked structures were visualized by PyMOL (Fig. 8 and Supplementary Fig. 4).

3.13. Analysis of the structures of complexes from HADDOCK 2.4 by PDBsum

The results from PDBsum are given in Table 8. Here, a chain refers to EZR, and the B chain refers to EBP50. The wild-type EZR-EBP50 complex contains 19 H-bonds. In addition to L59Q, the other mutated complexes contained fewer hydrogen bonds than the wild type did. Moreover, the I248T complex lost 7 hydrogen bonds. However, the L225P complex has lost 4 hydrogen bonds and 1 salt bridge. Other important interaction information from PDBsum, including the number of interface residues, the interface area, and the number of nonbonded contacts, is also given in Table 8. The positions of the interacting residues collected from PDBsum are given in Supplementary Table 8 which suggest that the identified mutations are positioned within or near the interaction interface and may influence the structural integrity of the FERM_C domain, potentially affecting binding with NHERF1/EBP50.

3.14. Molecular dynamics simulation

The structural stability of the wild-type EZR protein was compared to its variants by calculating the RMSD of the alpha carbon (Cα) atoms for the structures obtained over a 200-ns MD simulation relative to the initial reference structure. The RMSD of the wild-type protein stabilized after approximately 20 ns, indicating structural equilibrium. However, the RMSD values of the EZR variants showed significant deviations compared to the wild type, suggesting lower structural stability compare to wild type (Fig. 9- a). Among the variants, L59Q demonstrated the highest deviation, with a marked increase in RMSD after 60 ns. The L125S variant displayed the least stability overall, characterized by consistently elevated RMSD values throughout the 200-ns simulation period. Variants F240S, H288D, and L225P showed moderate deviations, reflecting intermediate structural changes and a moderately reduced stability compared to the wild-type protein.

The radius of gyration (Rg) is a measure of a molecule's compactness, defined as the root mean square distance of all protein atoms from the center of mass. The wild-type protein exhibited an average Rg of 5.09 ± 0.12 nm, which was higher than that of all mutant variants, indicating that the mutations promote increased compactness of the protein structure. Among the variants, L125S showed a notable decrease in Rg (1.79 ± 0.43 nm), suggesting a significant reduction in structural expansion (Fig. 9- b).

To assess the local flexibility of wild type and its variants, RMSF values of the alpha carbon (Cα) atoms for each amino acid were calculated. Regions with higher RMSF values indicate greater flexibility. The wild type exhibits greater flexibility in certain regions compared to the mutants of EZR proteins. The L225P and I248T variants showed the lowest RMSF values between residues 10 and 80, indicating reduced

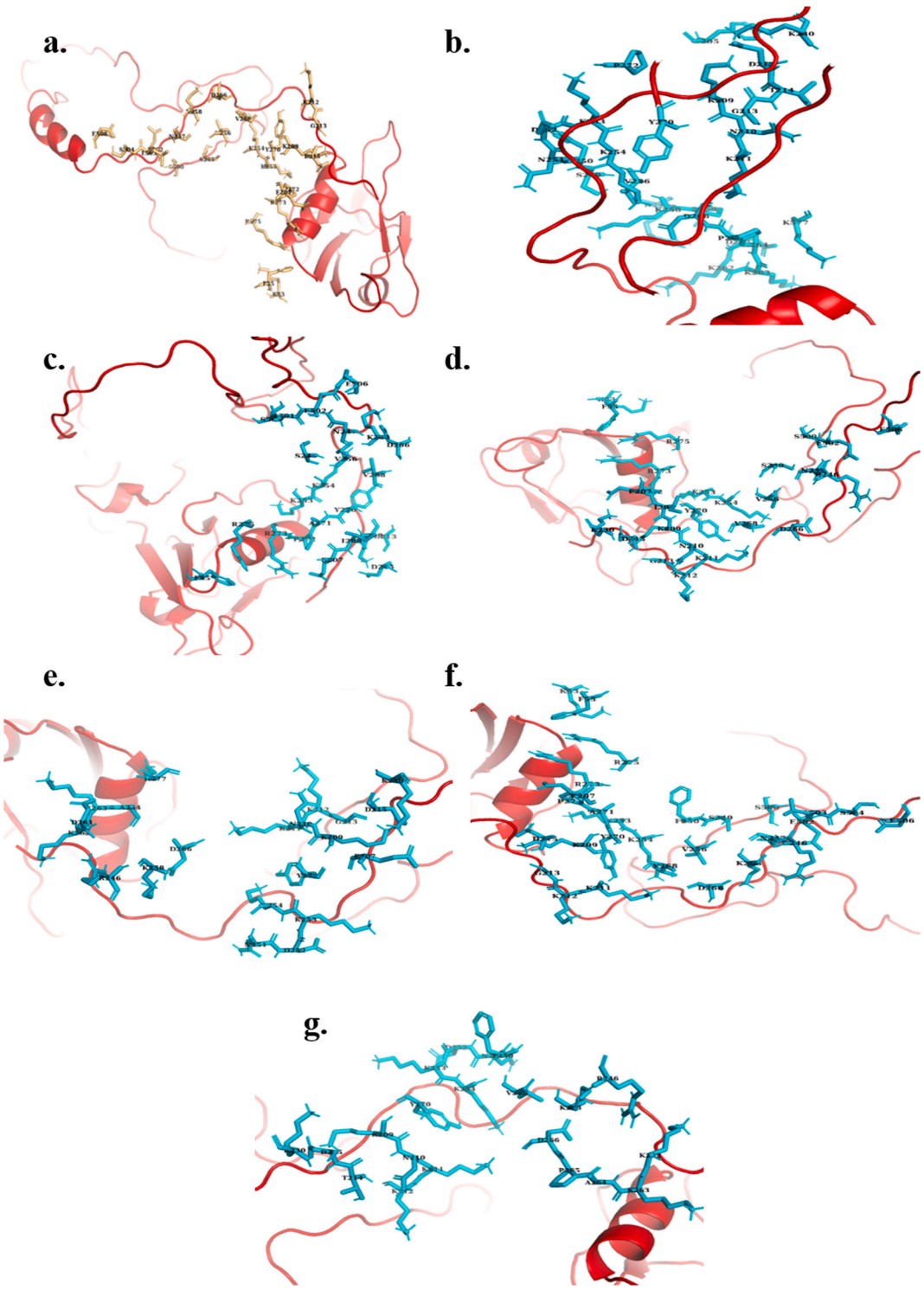


Fig. 8. Visualization of all docked complexes, wild-type (light orange) and mutated EZR proteins (cyan) with EBP50 (red) complexes **by PyMOL** (a) Wild-type EZR-EBP50, (b) L59Q-EBP50, (c) L125S-EBP50, (d) L225P-EBP50, (e) F240S-EBP50, (f) I248T-EBP50, (g) H288D-EBP50.

flexibility in this region. In contrast, residues 150 to 175 exhibited comparable RMSF values for the wild-type and L125S variant, both considerably higher than those of the other mutants. Notably, the L59Q variant displayed greater flexibility within the amino acid region spanning residues 360 to 470 compared to the wild type (Fig. 9- c). Beyond residue 475 and extending through most of the remaining protein, the wild-type structure demonstrated the highest RMSF values among all tested variants.

Hydrogen bond (H-bond) patterns offer critical insights into molecular interactions, atomic flexibility, and the structural organization of proteins. The average number of H-bonds throughout the simulation was higher for the native EZR protein (506.01 ± 11.61) compared to the analyzed variants, suggesting that the mutations may disrupt hydrogen bond formation and potentially affect the protein's stability (Fig. 9- d). Among the variants, L225P exhibited a slightly lower overall number of H-bonds, suggesting marginally reduced atomic flexibility compared to the other mutants. The Solvent Accessible Surface Area (SASA) was assessed to evaluate the exposure of amino acid residues to

Table 8
Docking analysis of the wild-type and mutant EZR proteins with EBP-50 via PDBsum.

Protein Type	Chain	Number of interface residues	Interface area (Å2)	No. of salt bridges	Number of hydrogen bonds	Number of nonbonded contacts
Wild type complex	A (EZR- WT)	24	1085	5	19	163
	B (EBP50)	23	1180			
L59Q complex	A (EZR-L59Q)	27	915	5	19	157
	B (EBP50)	18	1034			
L125S complex	A (EZR- L125S)	23	1177	5	17	172
	B (EBP50)	23	1228			
L225P complex	A (EZR- L225P)	25	1062	4	15	157
	B (EBP50)	22	1096			
F240S complex	A (EZR- F240S)	21	774	5	14	130
	B (EBP50)	15	868			
I248T complex	A (EZR- I248T)	28	1059	5	12	169
	B (EBP50)	23	1101			
H288D complex	A (EZR- H288D)	22	810	5	18	150
	B (EBP50)	16	896			

the solvent in both the native EZR protein and its variants. The average SASA value for the native protein (363.15 ± 6.08) was found to be similar to that of the majority of the variants, indicating minimal differences in overall solvent exposure (Fig. 9- e). However, the I225P and H288D variants exhibited slightly higher SASA values, suggesting a minor structural expansion compared to the wild type and other mutants.

Secondary structure analysis revealed that the native EZR protein exhibits a secondary structure profile similar to its variants. However, a notable increase in the proportion of κ-helix (polyproline helix) was observed around the 500th amino acid position in the F240S, I248T, and L59Q variants (Fig. 10, Supplementary Fig. 3).

4. Discussion

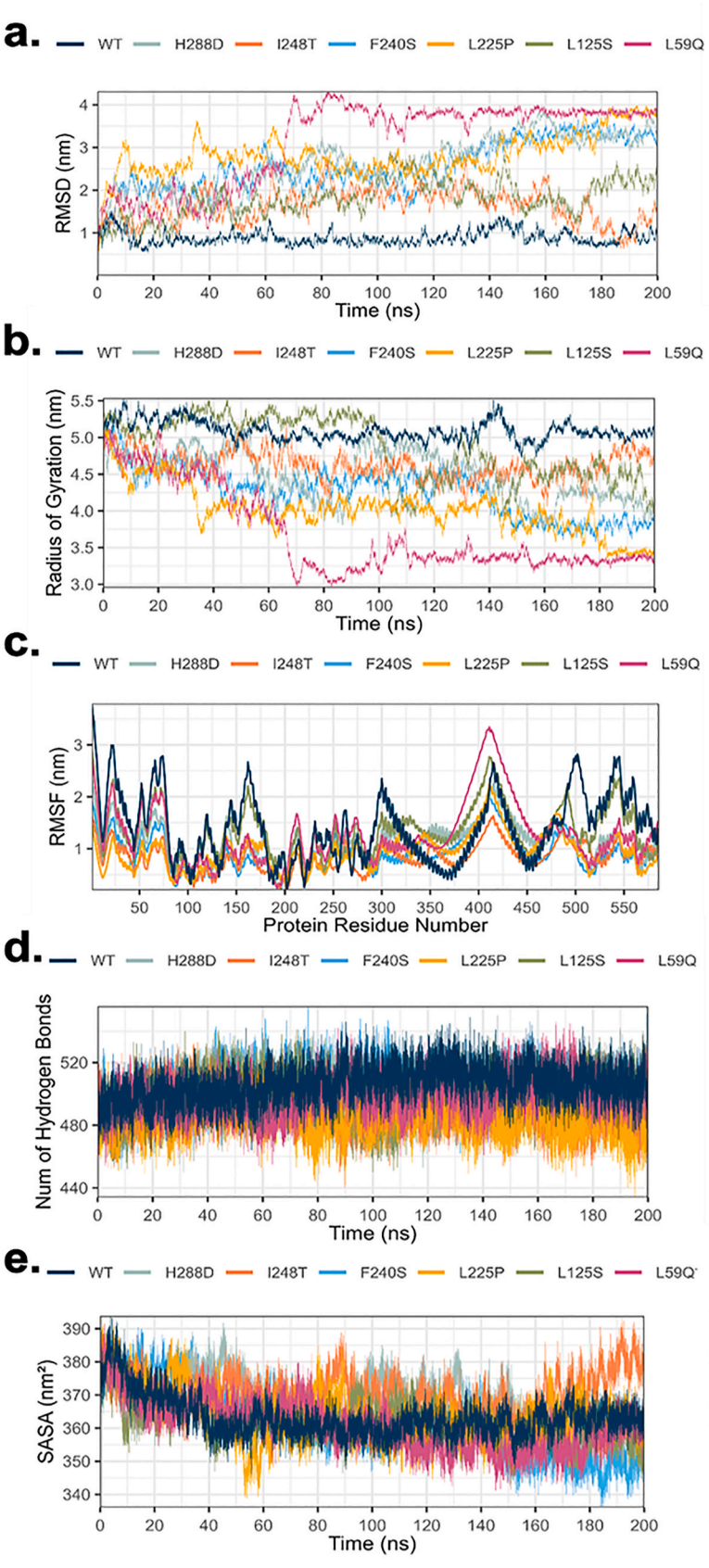
Using a wide range of computational methods, we conducted a thorough in silico review in this study to detect EZR gene pathogenic nsSNPs. To the best of our knowledge, no in silico investigations have been performed to assess nsSNPs in the human EZR gene. EZR acts as a linker between the actin cytoskeleton and the plasma membrane [8], thus helping to connect the intracellular actin cytoskeleton beneath the plasma membrane with many surface proteins, including proteins involved in signaling pathways [67]. High expression of EZR ultimately activates many signaling pathways involved in cancer [13,17] or promotes cancer cell migration and invasion [12]. Mutation of EZR can enhance or disrupt the binding of many proteins, which ultimately causes disease and cancer. To build a more precise and dependable estimation of mutational impact on the EZR, we employed a range of computational tools that take distinct approaches to differentiate between pathogenic and neutral variations. While these prediction techniques have been created to assess whether a particular alteration has a functional or pathogenic effect, the mechanisms by which the SNPs influence protein function or cause disease are not fully understood. Molecular docking and other experimental or computational methods can be used to investigate this issue.

A SIFT score of 0, a PolyPhen-2 score of 1 and a value of −4.1 were used in PROVEAN to filter the most damaging nsSNPs, and 37 nsSNPs that met these threshold values were identified. These damaging 37 nsSNPs were subjected to filtration by PhD-SNP, SNPs&GO, SuSPect and fathmm to find the most pathogenic mutations, and 12 nsSNPs were found after the pathogenicity of functional nsSNPs was evaluated via all four bioinformatic tools. T567 is a significant location among them. It has been proposed that subsequent phosphorylation events via the ezrin protein take place after ezrin T567 is phosphorylated [68]. The essential step in activating ezrin is the phosphorylation of the C-terminal threonine residue (Thr567), which breaks the dormant state of head-to-tail connections and enables the actin filament binding domains to interact with other proteins [69]. Liver metastasis is associated with increased ezrin phosphorylation at the T567 location [70], indicating that T567 M mutation has a significant effect on ezrin function. This mutation is also very harmful, according to our findings. Since phosphorylation would not be possible at this location, EZR activation would be decreased and could have anticancer effects [70]. This hypothesis can be correlated with our result of surface accessibility by NetSurfP-3.0, where relative surface accessibility decreased in the mutant protein (52%–46%) to some extent, and stability analysis by mCSM and Dynamut2 revealed that the mutation was stabilizing. Navapour et al. set a SuSPect score of 90 as the threshold to identify the most deleterious SNPs of the CYP1A2 protein, which is why we also used this score to have high confidence in our results to find the most damaging nsSNPs of the EZR protein [65].

Through structural analysis by HOPE, we assessed the changes in amino acid structure and charge and their interactions due to mutation. We found that L125S, L225P and F240S caused smaller residues and a charge shift from neutral to negative occurred due to H288D. L225P has a Mutpred2 score of 0.975 and has more alterations than the other mutations do, which were subsequently proven to be damaging by docking analysis. Additionally, the mutation 3D server revealed cluster mutations (L225P, F240S, I248T, and H288D), which are located in the FERM_C domain and this domain is also the binding site for its docking partner NHERF1/EBP50. We found that the conserved FERM domain consists of three subdomains, where FERM_M and FERM_C contain most of the PTM sites and two other domains—the ERM alpha helical and C-terminal domains. Additionally, I248 is located between two PTM sites, indicating that the wild-type residue is important for its function. These findings further validated the importance to inspect docking between mutant proteins and EBP50.

STRING v12.0 has provided the most interacting proteins with EZR, using high-confidence interactions (minimum interaction score ≥0.9), ensuring that the depicted relationships are biologically significant and supported by strong evidence. We found SLC9A3R1 as one of the top 10 interacting proteins with EZR. Here, SLC9A3R1 is also known as the Na(+)/H(+) exchange regulatory cofactor (NHERF1) or ERM Binding Protein 50(EBP50) which is our docking partner. While modeling the mutated protein via SWISS MODEL, we used the Alphafold-predicted structure as the wild type and template to model other mutated EZR proteins, as the complete crystal structure of EZR is not available. To validate the structure of Alphafold (AF-P15311-F1), we aligned two available crystal structures and found significant similarity in the use of AF-P15311-F1 as a wild-type structure and as a template to model the mutated structure.

Ali et al. performed a study to elucidate short linear motif-based interactions of the FERM domains of ERM (Ezrin, Radixin, Moesin) proteins and Merlin and defined binding sites for multiple ligands, including the peptide sequence of NHERF1/EBP50, from which we can identify the C-terminal domain of EBP50 that binds to the F3 subdomain



Type	Average RMSD (nm)
Wild Type	0.888 ± 0.16
H288D	2.610 ± 0.63
I248T	1.639 ± 0.38
F240S	2.562 ± 0.59
L225P	2.832 ± 0.48
L125S	4.855 ± 0.42
L59Q	3.152 ± 0.96

Type	Average Rg (nm)
Wild Type	5.092 ± 0.12
H288D	4.482 ± 0.32
I248T	4.626 ± 0.18
F240S	4.266 ± 0.33
L225P	4.019 ± 0.32
L125S	1.792 ± 0.43
L59Q	3.694 ± 0.57

Type	Average RMSF (nm)
Wild Type	1.413 ± 0.65
H288D	1.111 ± 0.36
I248T	0.832 ± 0.27
F240S	0.915 ± 0.35
L225P	0.946 ± 0.34
L125S	1.351 ± 0.48
L59Q	1.273 ± 0.587

Type	Average H-bond num
Wild Type	506.010 ± 11.61
H288D	498.792 ± 11.83
I248T	490.137 ± 11.51
F240S	504.570 ± 12.43
L225P	485.83 ± 11.82
L125S	498.08 ± 12.95
L59Q	499. 57 ± 11.83

Type	Average SASA (nm²)
Wild Type	363.154 ± 6.08
H288D	367.711 ± 7.13
I248T	370.712 ± 6.05
F240S	362.088 ± 10.24
L225P	363.61 ± 7.95
L125S	361.82 ± 6.26
L59Q	360.956 ± 7.56

Fig. 9. Analysis of trajectory data for the native EZR protein and its **variants.** (a) Root Mean Square Deviation (RMSD) over time. (b) Radius of Gyration (Rg) over time. (c) Root Mean Square Fluctuation (RMSF) across protein residues. (d) Number of hydrogen bonds (H-bonds) over time. (e) Solvent Accessible Surface Area (SASA) over time.

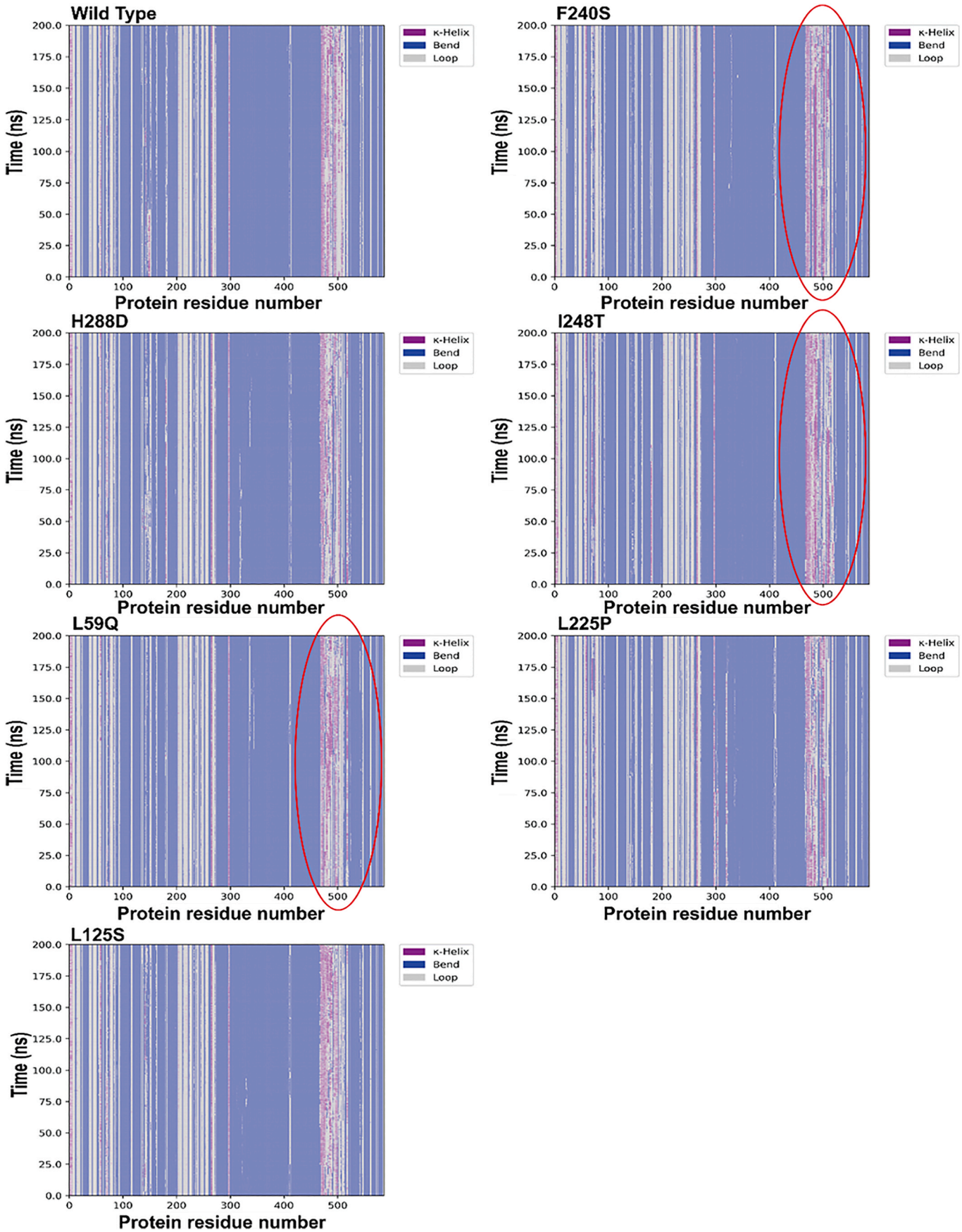


Fig. 10. Secondary structure changes over the course of the molecular dynamics simulation for the native EZR protein and its **variants.** Red circles highlight regions where the secondary structure in the variants differs from that of the wild-type protein.

(PF09380: 210–298 AA) of the FERM domain [55]. In another study, the ezrin binding (EB) domain of EBP50 ranged from 338 to 358 amino acids [20]. As the FERM domain spans amino acids 1–297 of EZR, we chose amino acids 210–297 of EZR and 338–358 of NHERF1/EBP50 (PDB ID: 2KRG) as direct interacting residues for docking via HADDOCK 2.4.

Among the most deleterious and disease-causing nsSNPs, H288D caused the secondary structure of EZR to shift from the buried state (RSA = 22%) to the exposed state (RSA = 31%), as reported by NetSurfP-3.0, making the mutation even more damaging than others. Similar findings were observed in the SASA analysis, where H288D exhibited slight structural expansion, suggesting altered surface accessibility. The output of the complex structures from docking by HADDOCK 2.4 revealed that the L225, I248 and F240 positions are important for its ability to bind with EBP50. Here, owing to the L225P mutation, the complex lost 4 hydrogen bonds and 1 salt bridge according to PDBsum, with the lowest free binding energy with EBP50 (Hawkdock MM/GBSA free binding energy = −44.87 kcal/mol) among the other complexes. Additionally, the I248T mutation causes the loss of 7 hydrogen bonds, affecting binding with EBP50. Compared with other mutated complexes, F240S causes the loss of 5 hydrogen bonds and the loss of more nonbonded contacts (from 163 to 130).

Analyses of RMSD, RMSF, secondary structure, Rg, SASA, and hydrogen bonding via molecular dynamics simulations revealed that these mutations exert varying effects on the stability of the EZR protein. Structural stability analysis revealed that the wild-type protein had a stable structure throughout the simulation period while the variants F240S, H288D, and L225P showed altered RMSD, indicating a loss of equilibrium and lower stability. Notably, the L125S mutation exhibited persistently elevated RMSD values, making it the least stable amongst the mutants. A notable decrease in Rg for L125S reflects a significant loss of structural expansion, likely restricting the protein's capacity to undergo essential conformational changes. Such compaction indicates reduced flexibility, which was also observed in the RMSF analysis. Mutants such as L225P and I248T showed diminished flexibility in the FERM domain, particularly between residues 10 and 80, suggesting that their rigid structures may hinder interactions with critical binding partners like NHERF1/EBP50. Mutations also disrupted interatomic interactions, with variants such as L225P displaying fewer H-bonds, indicating a weakened structural framework. This reduction in bonding strength is reflected in the docking results, where mutations like L225P and I248T caused significant losses of hydrogen bonds and nonbonded contacts in complexes with EBP50, resulting in decreased binding affinity. An increase in κ-helix content was observed near the 500th amino acid position in the F240S, I248T, and L59Q variants. This structural shift may have important functional implications, as the κ-helix, which has altered geometry and reduced flexibility compared to an alpha helix, could introduce localized rigidity and disrupt conformational changes required for protein-protein interactions [71]. Decreased binding of EZR with EBP50 causes head−to−tail interactions masking the PDZ domains of EBP50, thus sequestering the binding sites for many proteins, including β-catenin [20]. At the plasma membrane, disturbance of contacts between EBP50 and β-catenin causes a cascade of events, including decreased proliferation, enhanced migration capability, and stabilization of nuclear β-catenin, which leads to the development of cancer [72].

The combined results of in silico pathogenicity predictions, molecular docking, and molecular dynamics simulations pinpointed L225P, F240S, and H288D as the most deleterious SNPs. These mutations caused significant structural and functional disruptions in EZR, particularly affecting its binding interactions with EBP50. The collection of these important data can serve as a foundation for additional medical studies aimed at examining the connection between these genetic nonsynonymous variations and the possibility of a disease breaking out, as well as at creating precise treatment plans for numerous diseases associated with the aberrant behavior of this protein. However, appropriate laboratory-oriented research and studies on animals and humans are necessary to determine the exact roles of these missense SNPs and other clinically important mutations in EZR proteins.

CRediT authorship contribution statement

Sadia Akter: Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation. **Mohtasim Fuad:** Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation. **Zimam Mahmud:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Investigation, Formal analysis, Data curation, Conceptualization. **Sonia Tamanna:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Mohammad Sayem:** Writing – review & editing, Validation, Software, Methodology, Formal analysis, Data curation. **Khalid Hasan Raj:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **Md. Zakir Hossain Howlader:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis.

Ethical approval

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank the University of Dhaka, Bangladesh, for providing an International Publication Grant for APC payment and the Department of Biochemistry and Molecular Biology, Faculty of Biological Sciences, University of Dhaka, for providing necessary support to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2025.101972.

Data availability

All data generated or analyzed during this study are included in this manuscript [and its supplementary material files].

References

[1] S. Yonemura, et al., Concentration of an integral membrane protein, CD43 (leukosialin, sialophorin), in the cleavage furrow through the interaction of its cytoplasmic domain with actin-based cytoskeletons, The Journal of cell biology 120 (2) (1993) 437–449.
[2] D. Hanzel, et al., The secretion-stimulated 80K phosphoprotein of parietal cells is ezrin, and has properties of a membrane cytoskeletal linker in the induced apical microvilli, The EMBO journal 10 (9) (1991) 2363–2373.
[3] J.M. Phang, et al., Structural characterization suggests models for monomeric and dimeric forms of full-length ezrin, Biochem. J. 473 (18) (2016) 2763–2782.
[4] R. Gary, A. Bretscher, Ezrin self-association involves binding of an N-terminal domain to a normally masked C-terminal domain that includes the F-actin binding site, Mol. Biol. Cell 6 (8) (1995) 1061–1075.
[5] S. Bosk, et al., Activation of F-actin binding capacity of ezrin: synergism of PIP2 interaction and phosphorylation, Biophys. J. 100 (7) (2011) 1708–1717.
[6] K.W. Hunter, Ezrin, a key component in tumor metastasis, Trends Mol. Med. 10 (5) (2004) 201–204.

[7] J.J. Ramalho, et al., C-terminal phosphorylation modulates ERM-1 localization and dynamics to control cortical actin organization and support lumen formation during Caenorhabditis elegans development, Development 147 (14) (2020) dev188011.
[8] A. Gautreau, et al., Ezrin, a plasma membrane–microfilament linker, signals cell survival through the phosphatidylinositol 3-kinase/Akt pathway, Proc. Natl. Acad. Sci. USA 96 (13) (1999) 7300–7305.
[9] D. Reczek, M. Berryman, A. Bretscher, Identification of EBP50: a PDZ-containing phosphoprotein that associates with members of the ezrin-radixin-moesin family, J. Cell Biol. 139 (1) (1997) 169–179.
[10] J. Kong, et al., High expression of ezrin predicts poor prognosis in uterine cervical cancer, BMC Cancer 13 (1) (2013) 1–8.
[11] L. Li, et al., Ezrin is associated with gastric cancer progression and prognosis, Pathol. Oncol. Res. 17 (2011) 909–915.
[12] R. Zhang, et al., High expression of EZR (ezrin) gene is correlated with the poor overall survival of breast cancer patients, Thoracic Cancer 10 (10) (2019) 1953–1961.
[13] J. Xu, W. Zhang, EZR promotes pancreatic cancer proliferation and metastasis by activating FAK/AKT signaling pathway, Cancer Cell Int. 21 (1) (2021) 521.
[14] D. Sarrió, et al., Abnormal ezrin localization is associated with clinicopathological features in invasive breast carcinomas, Breast Cancer Res. Treat. 98 (1) (2006) 71–79.
[15] S. Yonemura, et al., Ezrin/radixin/moesin (ERM) proteins bind to a positively charged amino acid cluster in the juxta-membrane cytoplasmic domain of CD44, CD43, and ICAM-2, J. Cell Biol. 140 (4) (1998) 885–895.
[16] T.A. Martin, et al., The role of the CD44/ezrin complex in cancer metastasis, Crit. Rev. Oncol. Hematol. 46 (2) (2003) 165–186.
[17] N. Li, et al., Ezrin promotes breast cancer progression by modulating AKT signals, Br. J. Cancer 120 (7) (2019) 703–713.
[18] D. Garbett, D.P. LaLonde, A. Bretscher, The scaffolding protein EBP50 regulates microvillar assembly in a phosphorylation-dependent manner, J. Cell Biol. 191 (2) (2010) 397–413.
[19] R. Sheng, et al., Cholesterol modulates cell signaling and protein networking by specifically interacting with PDZ domain-containing scaffold proteins, Nat. Commun. 3 (1) (2012) 1249.
[20] F.C. Morales, et al., NHERF1/EBP50 head-to-tail intramolecular interaction masks association with PDZ domain ligands, Molecular and cellular biology 27 (7) (2007) 2527–2537.
[21] J.-E. Lee, et al., Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 573 (1) (2005) 195–204.
[22] X. Wang, et al., Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes, Toxicol. Appl. Pharmacol. 207 (2 Suppl) (2005) 84–90.
[23] A. Kumar, et al., Computational SNP analysis: current approaches and future prospects, Cell Biochem. Biophys. 68 (2) (2014) 233–239.
[24] A. Kumar, R. Purohit, Use of long term molecular dynamics simulation in predicting cancer associated SNPs, PLoS Comput. Biol. 10 (4) (2014) e1003318.
[25] S.T. Sherry, et al., dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (1) (2001) 308–311.
[26] T.U. Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (D1) (2020) D480–D489.
[27] N.-L. Sim, et al., SIFT web server: predicting effects of amino acid substitutions on proteins, Nucleic Acids Res. 40 (W1) (2012) W452–W457.
[28] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, Curr Protoc Hum Genet 76 (1) (2013) 7.20.1–7.20.41.
[29] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, Bioinformatics 31 (16) (2015) 2745–2747.
[30] E. Capriotti, R. Calabrese, R. Casadio, Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information, Bioinformatics 22 (22) (2006) 2729–2734.
[31] E. Capriotti, et al., WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation, BMC Genom. 14 (Suppl 3) (2013) S6. Suppl 3.
[32] C.M. Yates, et al., SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features, J. Mol. Biol. 426 (14) (2014) 2692–2701.
[33] H.A. Shihab, et al., Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models, Hum. Mutat. 34 (1) (2013) 57–65.
[34] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (suppl_2) (2005) W306–W310.
[35] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins 62 (4) (2006) 1125–1132.
[36] D.E.V. Pires, D.B. Ascher, T.L. Blundell, mCSM: predicting the effects of mutations in proteins using graph-based signatures, Bioinformatics 30 (3) (2013) 335–342.
[37] C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations, Protein Sci. 30 (1) (2021) 60–69.
[38] H. Venselaar, et al., Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinf. 11 (1) (2010) 548.
[39] V. Pejaver, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat. Commun. 11 (1) (2020) 5918.
[40] M.J. Meyer, et al., mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome, Hum. Mutat. 37 (5) (2016) 447–456.
[41] B. Snel, et al., STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene, Nucleic Acids Res. 28 (18) (2000) 3442–3444.
[42] M.H. Høie, et al., NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning, Nucleic Acids Res. 50 (W1) (2022) W510–W515.
[43] D. Wang, et al., MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization, Nucleic Acids Res. 48 (W1) (2020) W140–W146.
[44] P.V. Hornbeck, et al., PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, Nucleic Acids Res. 40 (Database issue) (2012) D261–D270.
[45] R.D. Finn, et al., Pfam: the protein families database, Nucleic Acids Res. 42 (Database issue) (2014) D222–D230.
[46] N. Hulo, et al., The PROSITE database, Nucleic Acids Res. 34 (Database issue) (2006) D227–D230.
[47] A. Waterhouse, et al., SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.
[48] J. Jumper, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[49] Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score, Nucleic acids research 33 (7) (2005) 2302–2309.
[50] L.L.C. Schrödinger, The PyMOL Molecular Graphics System, Schrödinger, LLC, New York, NY, 2020. Version 2.5.
[51] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (9) (1993) 1511–1519.
[52] R.A. Laskowski, et al., PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (2) (1993) 283–291.
[53] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (Web Server issue) (2007) W407–W410.
[54] G.C.P. van Zundert, et al., The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (4) (2016) 720–725.
[55] M. Ali, et al., Elucidation of short linear motif-based interactions of the FERM domains of ezrin, radixin, moesin, and merlin, Biochemistry 62 (11) (2023) 1594–1607.
[56] G. Weng, et al., HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA, Nucleic Acids Res. 47 (W1) (2019) W322–w330.
[57] R.A. Laskowski, et al., PDBsum: structural summaries of PDB entries, Protein Sci. 27 (1) (2018) 129–134.
[58] J. Huang, A.D. MacKerell Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J. Comput. Chem. 34 (25) (2013) 2135–2145.
[59] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, The Journal of chemical physics 126 (1) (2007).
[60] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (12) (1981) 7182–7190.
[61] B. Hess, et al., LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (12) (1997) 1463–1472.
[62] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N· log (N) method for Ewald sums in large systems, The Journal of chemical physics 98 (12) (1993) 10089–10092.
[63] H. Wickham, H. Wickham, Data Analysis, Springer, 2016.
[64] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (3) (2007) 90–95.
[65] L. Navapour, N. Mogharrab, In silico screening and analysis of nonsynonymous SNPs in human CYP1A2 to assess possible associations with pathogenicity and cancer susceptibility, Sci. Rep. 11 (1) (2021) 4977.
[66] I. Halperin, et al., Principles of docking: an overview of search algorithms and a guide to scoring functions, Proteins: Struct., Funct., Bioinf. 47 (4) (2002) 409–443.
[67] A. Bretscher, K. Edwards, R.G. Fehon, ERM proteins and merlin: integrators at the cell cortex, Nat. Rev. Mol. Cell Biol. 3 (8) (2002) 586–599.
[68] B. Bruce, et al., Expression of the cytoskeleton linker protein ezrin in human cancers, Clin. Exp. Metastasis 24 (2) (2007) 69–78.
[69] T. Matsui, et al., Rho-kinase phosphorylates COOH-terminal threonines of ezrin/radixin/moesin (ERM) proteins and regulates their head-to-tail association, J. Cell Biol. 140 (3) (1998) 647–657.
[70] P.D. Leiphrakpam, et al., Ezrin expression and cell survival regulation in colorectal cancer, Cell. Signal. 26 (5) (2014) 868–879.
[71] T. Meirson, et al., κ-helix and the helical lock and key model: a pivotal way of looking at polyproline II, Bioinformatics 36 (12) (2020) 3726–3732.
[72] T. Itou, et al., EBP50 depletion and nuclear β-catenin accumulation engender aggressive behavior of colorectal carcinoma through induction of tumor budding, Cancers 16 (1) (2023).