

RESEARCH

Open Access



A computational approach for structural and functional analyses of disease-associated mutations in the human *CYLD* gene

Arpita Singha Roy^{1†}, Tasmiah Feroz^{1†}, Md. Kobirul Islam¹, Md. Adnan Munim¹, Dilara Akhter Supti², Nusrat Jahan Antora³, Hasan Al Reza⁴, Supriya Gosh², Newaz Mohammed Bahadur⁵, Mohammad Rahanur Alam^{2*} and Md Shahadat Hossain^{1*}

Abstract

Tumor suppressor cylindromatosis protein (CYLD) regulates NF- κ B and JNK signaling pathways by cleaving K63-linked poly-ubiquitin chain from its substrate molecules and thus preventing the progression of tumorigenesis and metastasis of the cancer cells. Mutations in *CYLD* can cause aberrant structure and abnormal functionality leading to tumor formation. In this study, we utilized several computational tools such as PANTHER, PROVEAN, PredictSNP, PolyPhen-2, PhD-SNP, PON-P2, and SIFT to find out deleterious nsSNPs. We also highlighted the damaging impact of those deleterious nsSNPs on the structure and function of the CYLD utilizing ConSurf, I-Mutant, SDM, Phyre2, HOPE, Swiss-PdbViewer, and Mutation 3D. We shortlisted 18 high-risk nsSNPs from a total of 446 nsSNPs recorded in the NCBI database. Based on the conservation profile, stability status, and structural impact analysis, we finalized 13 nsSNPs. Molecular docking analysis and molecular dynamic simulation concluded the study with the findings of two significant nsSNPs (R830K, H827R) which have a remarkable impact on binding affinity, RMSD, RMSF, radius of gyration, and hydrogen bond formation during CYLD-ubiquitin interaction. The principal component analysis compared native and two mutants R830K and H827R of CYLD that signify structural and energy profile fluctuations during molecular dynamic (MD) simulation. Finally, the protein–protein interaction network showed CYLD interacts with 20 proteins involved in several biological pathways that mutations can impair. Considering all these in silico analyses, our study recommended conducting large-scale association studies of nsSNPs of *CYLD* with cancer as well as designing precise medications against diseases associated with these polymorphisms.

Keywords Cylindromatosis, nsSNPs, USP domain, Molecular docking, Molecular dynamic simulations

[†]Arpita Singha Roy and Tasmiah Feroz contributed equally to this work.

*Correspondence:

Mohammad Rahanur Alam
rahanur.ftns@nstu.edu.bd
Md Shahadat Hossain
shahadat5.bge@nstu.edu.bd

¹ Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

² Department of Food Technology & Nutrition Sciences, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

³ Department of Genetic Engineering and Biotechnology, East West University, Dhaka 1212, Bangladesh

⁴ Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh

⁵ Department of Applied Chemistry and Chemical Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh



1 Introduction

CYLD, or cylindromatosis lysine 63 deubiquitinase, is a tumor suppressor protein that generally performs deubiquitinase activities essential for a variety of cellular and signaling processes [1]. CYLD is mainly a cytoplasmic protein that belongs to the ubiquitin-specific protease (USP) family [2] and is abundant in the brain [3], skeletal muscle [4], and immune cells [5]. CYLD processes larger substrate molecule by cleaving lysin63-linked ubiquitin chains from that molecule [6, 7] and thereby involved in corresponding cellular events, namely: cell cycle control [8], cellular differentiation [9], oncogenesis [10], cellular proliferation [11, 12], and apoptosis [13]. Mutation in *CYLD* can give rise to the constant activation and deregulation of cell survival proteins associated with tumorigenesis [10]. Several studies have demonstrated that mutated *CYLD* gene greatly contributes to the familial cylindromatosis, Brooke-Spiegler syndrome, and multiple familial trichoepitheliomas [14, 15].

CYLD gene encodes a 956 amino acids long protein with a weight of about 110 kD and is located on the chromosome 16q12-q13 [1] with 19 introns and 21 exons [16]. It contains a C-terminal conserved catalytic domain (USP) along with three N-terminal Cap-Gly domains [1]. Cap-Gly domain is crucial for the interaction with proteins involved in the NF- κ B pathway [17], while USP domain is important for hydrolyzing the ubiquitin chain. This carboxyl terminal (USP) catalytic domain changes its target proteins by deubiquitinating Lys63-linked ubiquitin chains of specific substrates that are vital in various cellular signaling events, especially in NF- κ B pathway [2, 18]. By deubiquitinating TRAF2/TRAF6, NF- κ B essential modifier (NEMO), CYLD, acts as a key regulator in the typical p65/NF- κ B pathway [19, 20]. CYLD also contributes greatly by preventing the Bcl3 from being localized in the nucleus and thus controls tumor development and proliferation [11]. Therefore, any mutation disrupting the deubiquitinating (DUB) activity of CYLD may lead to oncogenic function gain, as DUB activity is fundamental for CYLD as a tumor suppressor [21, 22].

Several polymorphisms have been identified as being responsible for the impaired activity of the *CYLD* gene, which finally leads to the tumorigenesis [23, 24]. The consequence of missense mutation on *CYLD* and the manner by which it is associated with cancer formation are not fully explored yet using computational approaches. Therefore, in silico analysis on nsSNPs of *CYLD* will help to demonstrate the potential role of mutation contributing towards the molecular mechanisms of various cancer types.

By considering all these facts, we have conducted an extensive analysis and explored numerous bioinformatics tools to investigate the functional and structural effect of

various nsSNPs on the CYLD protein and narrow down the list of the high-risk nsSNPs for our present study. In addition, we performed structural stability analysis, conservation analysis, and protein–protein interactions analysis followed by molecular docking analysis with its interacting molecules. Cancer-associated nsSNP identification is further validated by molecular dynamic simulation analysis where root-mean-square deviations (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg) analysis, and H-bond analysis were taken into consideration. This study will help us to identify cancer-prone genotypes related to this CYLD protein as well as future research on *CYLD* mutations.

2 Methodology

2.1 Assortment of nsSNPs

The information about human CYLD protein along with its amino acid sequence was assembled from the National Center for Biotechnology Information (NCBI). Details of SNPs (reference ID, location, residual variations, global minor allele frequency) were retrieved from dbSNP [25] (<https://www.ncbi.nlm.nih.gov/projects/SNP/>), a publicly accessible database for genetic variations available in NCBI for further computational analyses.

2.2 Screening of most deleterious nsSNPs

We exploited seven different in silico nsSNP prediction tools (SIFT, PANTHER, PolyPhen-2, PROVEAN, PhD-SNP, PON-P2, and PredictSNP) for the assessment of most deleterious nsSNPs having significant effect on the structure and function of the CYLD protein [2].

SIFT [26] (Sorting Intolerant From Tolerant) (https://sift.jcvi.org/www/SIFT_seq_submit2.html), a sequence homology-based algorithm, determines the effect of amino acid substitution over the physical and functional properties of a protein. SIFT provides prediction score against our submitted rsID for query nsSNPs where prediction score < 0.05 is regarded as intolerant and > 0.05 regarded as tolerant. SIFT result was obtained from PredictSNP [27].

PANTHER [28] (<https://www.pantherdb.org/tools>) database integrates the evolutionary conservation history with hidden Markov models (HMMs) to analyze the probability of a damaging effect of nsSNPs on the functionality of a protein and their interacting ability with other proteins. PANTHER provides position-specific evolutionary conservation scores when protein sequences along with human missense SNPs are submitted as a query.

PolyPhen-2 [29] (Polymorphism Phenotyping v2) (<https://genetics.bwh.harvard.edu/pph2/>) is a tool that employs machine learning methods, considering multiple sequence alignment to classify the damaging impact

of allele change over the structure and function of a protein categorized as probably damaging with probabilistic score (0.85 to 1.0) and possibly damaging with probabilistic score (0.15 to 1.0). Information about amino acid substitution along with FASTA sequence of a protein is required for the query submission.

PROVEAN [30, 31] (Protein Variation Effect Analyzer) (<https://provean.jcvi.org/index.php>) predicts the deleterious consequences of single or multiple amino acid changes (insertion and deletion) on the biological function of a protein. PROVEAN considers -2.5 as a cut-off value where amino acid substitution score > -2.5 is regarded as deleterious mutation.

PhD-SNP [32] (Predictor of human Deleterious Single-Nucleotide Polymorphisms) (<https://snps.biofold.org/phd-snp/phd-snp.html>) applies support vector machines (SVMs) to distinguish a genetic disease-linked point mutation from the neutral polymorphisms. Protein sequences, mutation profile information such as position of mutation, and mutated residues are required as an input file.

PON-P2 [33] (<http://structure.bmc.lu.se/PON-P2/>), a machine learning-based algorithm, classifies the amino acid alteration into three categories: pathogenic, neutral, and unknown groups. This tool predicts probability score of variant tolerance with respect to sequence conservancy, biological and physical properties of amino acids, gene ontology features, and functional annotations of alteration sites.

PredictSNP [27] (<https://loschmidt.chemi.muni.cz/predictsnp>) is a consensus program based on protein mutant database and the UniProt database annotations. It confirms the accuracy of the results acquired from eight renowned prediction tools (MAPP, nsSNPAnalyzer, PANTHER, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP) which signifies the impact of amino acid alteration on functional activity of a protein.

2.2.1 Identification of deleterious nsSNPs on different CYLD domains

InterPro [34] (<https://www.ebi.ac.uk/interpro/>) program ascertained the position of mutant nsSNPs on the CYLD protein according to the protein families, domains, and functional regions by integrating information from various protein-signature databases such as Pfam, PROSITE, PRINTS, ProDom, SMART, PIRSE, TIGRFAMs, PANTHER, the structure-based SUPERFAMILY, and Gene3D. Protein ID or FASTA sequences are used for the query searching.

2.2.2 Structural stability determination

I-Mutant and SDM tools specify the structural stability alteration of a protein as a result of deleterious point

mutation. I-Mutant [35] (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>) is a web server established on a support vector machine utilizing the thermodynamic database ProTherm to offer the changing free energy (DDG value) and the reliability index value (RI) of a protein and thus evaluate the level of changing structural stability in mutant proteins.

Site Director Mutator [36, 37] (SDM) is a web-based application that predicts the significant effect of mutation on protein stability. This tool computes stability score considering amino acid substitution patterns among homologous proteins from the same family and estimates free energy variation by comparing wild-type and mutant-type proteins.

2.2.3 Evolutionary conservation analysis

ConSurf [38] (<https://consurf.tau.ac.il>) applies either an empirical Bayesian method or a maximum likelihood (ML) method for the interpretation of the evolutionary conservancy of a particular amino acid at a specific position of a protein that signifies its structural and functional importance. To assess conservation score (ranges from 1 to 9) of an amino acid in a protein, ConSurf analyzes the phylogenetic relationship, multiple sequence alignment, and sequence homology of the protein. Conserved nsSNPs were considered for the further investigations.

2.2.4 Structural effect of nsSNPs on CYLD

Project HOPE [39] (<https://www3.cmbi.umcn.nl/hope>) is a web browser that uses Distributed Annotation System along with UniProt database to analyze the impact of a point mutation on the structure of a protein. Significant findings regarding structural variations between mutant and native residues are produced through 3D homology modelling using YASARA program. FASTA sequence or UniProt ID is submitted as query file.

Swiss-PdbViewer [40, 41] (v4.1.0) (<https://spdbv.vital-it.ch/>) computes the energy minimization of a protein for different amino acid substitutions. This tool utilizes its mutation tool and thereby selects best rotamer of the mutated protein and calculates the energy minimization state of a native and mutant 3D protein model using NOMAD-Ref server. This server performed the energy minimization of a 3D structure of a protein using GROMACS program as a default force field which is built on the methods of steepest descent, conjugate gradient, and L-BFGS (limited-memory Broyden–Fletcher–Goldfarb–Shanno) algorithm.

2.2.5 3D structure modelling to visualize the effect of nsSNPs

Three-homology modeling tools, namely: SWISS-MODEL, Phyre2, and I-TASSER, were used to create 3D structure of native and mutant proteins.

I-TASSER [42–44] (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) is a web server that operates replica-exchanged Monte Carlo simulations and thereby builds 3D structure of a full-length protein through splicing continuous threading alignments. This tool offers the comparative analysis of the I-TASSER models using confidence score, TM-score (template modelling score), and RMSD (root-mean-square deviation) value which is conducted by benchmark scoring system.

Phyre2 [45] (<https://www.sbg.bio.ic.ac.uk/phyre2>) is a web server based on advanced distant homology detection algorithms that generates 3D protein model and therefore provides analysis on the influence of amino acid variants on structure and function of a protein. Intensive mode was selected for developing 3D structure of CYLD protein. This mode constructs a full-length sequence model of a query protein combining different template models with high confidence score and sequence similarity. Then, TM-align tool [46] (<https://zhanglab.ccmb.med.umich.edu/TM-align/>) is incorporated to compare the mutant protein structure against the native one. TM-align calculates template modelling score (TM-score) and root-mean-square deviation (RMSD) based on structural similarities between two proteins. It generates TM-score ranges between 0 and 1, where TM-score 1 signifies perfect similarity between two protein structures. Significant deviation between native and mutant structure is estimated considering higher RMSD value.

SWISS-MODEL [47] (<https://swissmodel.expasy.org/>) server combines sequence alignment and template structure to develop three-dimensional structure of a protein. QMEAN scoring function applies for the model quality assessment to validate the reliability of the resultant models of both wild-type and mutant-type proteins.

2.3 Molecular docking analysis

HADDOCK [48, 49] (High Ambiguity Driven Protein–Protein Docking) (<https://wenmr.science.uu.nl/>), a web tool, was used to perform molecular docking analysis to understand the effect of deleterious point mutation over the binding affinity of CYLD with its interacting proteins. Protein–protein docking was carried out by the HADDOCK, with default settings for all parameters. The PDB structure of wild-type CYLD protein (PDB id-2VHF) was taken from SWISS-MODEL [50], and Ramachandran plot was used to validate the structure. Refinement was done before performing docking analysis using refinement interface in HADDOCK. CPORT [51] server (<http://haddock.chem.uu.nl/services/CPORT>) identified the active and passive residues of CYLD and ubiquitin protein. PRODIGY [52] (<https://wenmr.science.uu.nl/prodigy/>) web server performs the calculation of the binding affinity between protein–protein dock

complexes. BIOVIA Discovery Studio [53] was used to perform docking complex analysis along with image generation.

2.4 Identification of the cancer-associated nsSNPs

The web tool mutation 3D [54] (<http://www.mutation3d.org/>) enables users to easily identify the cancer causing mutation clusters collected from 975,000 somatic mutations recorded in 6811 cancer sequencing studies. This tool applies 3D clustering approach to find out amino acid substitution of a protein that can cause cancer when a target protein along with its mutations inserted as a query. We used this tool to look at the nsSNPs that can predispose to cancer.

2.5 Molecular dynamics simulation

YASARA [55] simulation software uses AMBER14 [56] as a force field to analyze the changing outcome of wild-type and mutant dock complex by allowing them to interact for a fixed period. The simulating chamber was permitted to contain 20 Å around the protein that is filled with water at 0.997 g/ml density. Initially, protein–protein dock complex was cleaned along with the H-bond network optimization. The steepest descent method was used to minimize the energy of the protein–protein complex. In order to evaluate the short-range Coulomb and van der Waals interaction, the cut-off radius was limited to 8 Å. The PME (particle mesh Ewald) method was utilized to assess the long-range electrostatic interactions. Simulations were accomplished under constant pressure in water, and Berendsen thermostat process controlled the simulation temperature at 298 K. Counterions (Na or Cl) were introduced to maintain a concentration of 0.9% (NaCl) to neutralize the system at pH 7.4. Simulation was executed for 100 ns under constant temperature and pressure by maintaining a time-step interval of 2.5 femtoseconds (fs). This tool provides following types of data such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration, total number of hydrogen bonds, and helix, sheet, turn, and coil values after the completion of simulations.

2.6 Principal component analysis

Principal component analysis was performed to determine the dynamics of biological system by reducing data complexity and retrieving the coordinated movements found in the simulations. A correlation matrix was built to represent variations detected in MD trajectories and offers the prediction of the first two principal components based on the calculation of the eigenvectors and eigenvalues [57, 58]. We performed principal component analysis (PCA) by considering bond distances, bond angles, dihedral angles, planarity, van der Waals energies,

and electrostatic energies and thus analyze the structural and energy changes of the wild CYLD-ubiquitin complex and mutant CYLD-ubiquitin complexes (H827R, R830K). Minitab software (Minitab 19, Minitab Inc., State College, PA, USA), a multivariate data analysis tool, performed principal component analysis (PCA) to signify variations among different groups by analyzing 100-ns MD simulation data.

2.7 Protein–protein interacting network analysis

STRING [59] (<https://string-db.org/>), online database, helped with better understanding of the molecular interactions of CYLD with other proteins. STRING produced the data in simple interaction format (SIF) or GML format visualized by Cytoscape [60, 61], a freely accessible java-based software. The overall workflow is represented by Fig. 1.

3 Result

3.1 Retrieval of nsSNPs

We retrieved all the reported SNPs found in *CYLD* from NCBI dbSNP database. This database contains a total of 13,653 SNPs, where 13,111 SNPs were in the intronic

region, 2066 SNPs were in the noncoding region, and 658 SNPs were in the coding region. In case of coding region, 413 SNPs were missense, and 245 SNPs were synonymous. A total of 446 missense variants have been found as some reference SNP ID (rsID) contain multiple SNPs at a single position. We considered all missense variants for our further analysis.

3.2 Identification of damaging nsSNPs

All missense variants obtained from dbSNP database were subjected to seven different deleterious SNP prediction tools, namely: PANTHER, PROVEAN, PredictSNP, PolyPhen-2, Phd-SNP, PON-P2, and SIFT, to determine their damaging consequences on the structure and function of CYLD protein. Each server predicted different amount of pathogenic nsSNPs. Finally, we targeted 18 common nsSNPs which were predicted to be deleterious by all the 7 in silico tools out of 446 nsSNPs (Table 1).

3.3 Domain identification of CYLD

InterPro was used to conduct functional analysis of the CYLD protein by categorizing them into protein families and identifying the active sites and domain

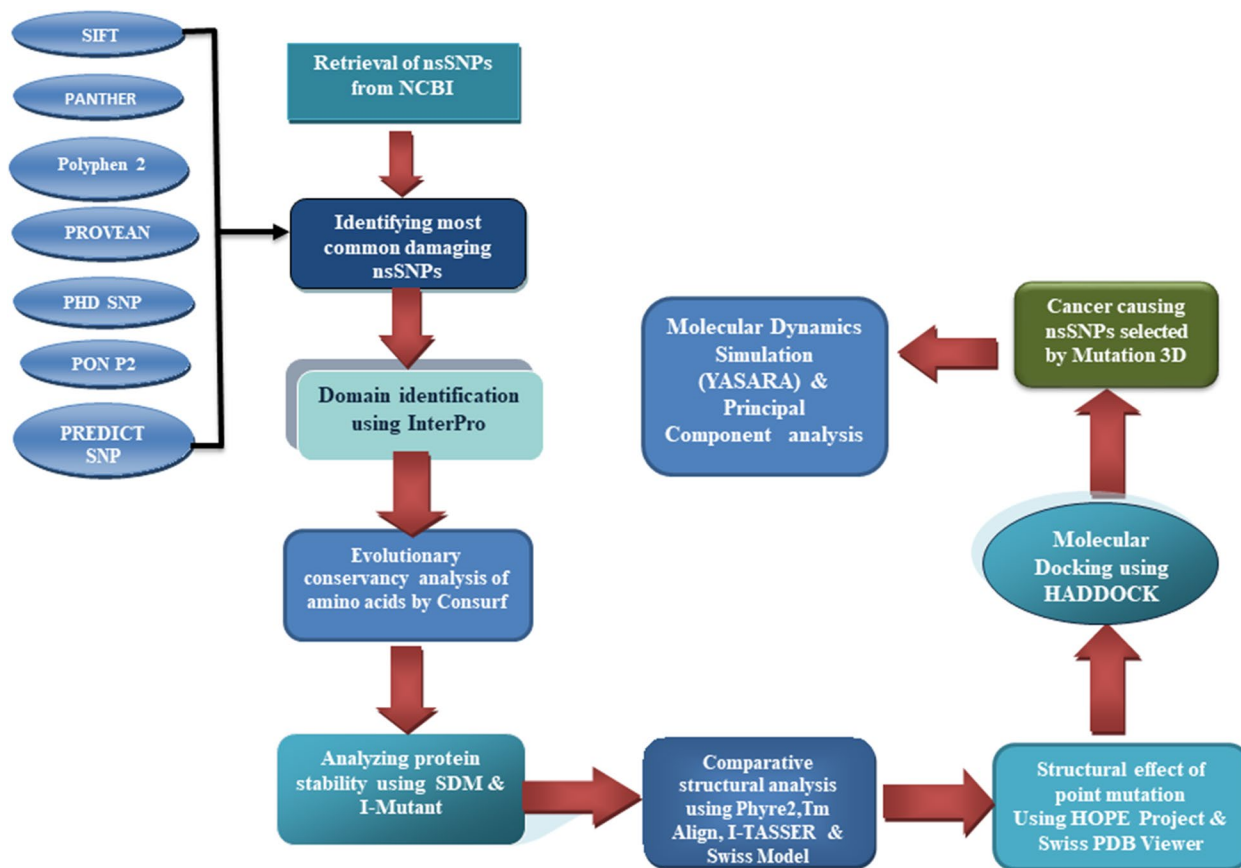


Fig. 1 A workflow representing all the in silico tools utilized for the computational analysis

Table 1 List of highly deleterious nsSNPs screened by seven computational programs

	PANTHER	PROVEAN	PredictSNP	PolyPhen-2	PhD-SNP	PON-P2	SIFT
V478A	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
R489H	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
L610F	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
I644T	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
L648R	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
P698T	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
P698S	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
P698L	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
E747G	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
L781P	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
H827P	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
H827R	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
R830K	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
V864F	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
I867K	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
I867R	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
H871Q	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious
R894W	Probably damaging	Deleterious	Deleterious	Probably damaging	Disease	Pathogenic	Deleterious

of the corresponding protein. This domain identification analysis revealed that CYLD contains two functional domains which are Cap-Gly domain (127–540 amino acid) and USP domain (592–950 AA) (S1 Fig.). Among 18 nsSNPs, 16 nsSNPs were positioned at the USP domain, whereas 2 nsSNPs were in the Cap-Gly domain.

3.4 Evolutionary conservation analysis of deleterious nsSNP in CYLD

Evolutionary conserved amino acid residues in a protein play specific roles in various functional biological cascades. Point mutation in such conserved residues results in a protein's aberrant structural and functional properties. ConSurf web server facilitates the analysis of the evolutionary conservancy and solvent accessibility of the amino acid residues of CYLD protein (Fig. 2). Among 18 high-risk nsSNPs, it predicted 16 highly conserved amino acid residues with conservation score 9 (S1 Table), whereas 2 nsSNPs (L610F, L781P) were conserved and average conserved respectively. These conserved residues are classified as structural or functional depending on their location in the structure of a protein. Amino acid residues exposed in the surface of a protein are considered functional, whereas buried residues are predicted to be structural [62–64]. Therefore, these findings further highlight the importance of the deleterious effects of nsSNPs situated at those buried and exposed residues of the CYLD protein.

3.5 Prediction of changing structural stability

Amino acid substitutions are thought to have damaging impacts on protein stability. Our selected 18 nsSNPs were subjected to the I-Mutant and SDM tools to analyze the changes in stability of CYLD protein due to point mutations. I-Mutant calculated the free energy change values $\Delta\Delta G$ and reliability index value (RI). It predicted 14 nsSNPs that decreased the stability, whereas 4 nsSNPs (R894W, P698L, P698S, L781P) increased the stability of CYLD (Table 2). SDM tool identified 4 nsSNPs (R894W, P698L, P698S, P698T) as stabilizing and 14 nsSNPs as destabilizing that are specifically responsible for the protein instability and dysfunction (Table 2). Both tools confirmed three common variants: R894W, P698L, and P698S, which increased the stability of the protein after mutation. Therefore, we excluded these three missense variants from further analysis.

3.6 Comparative structural analysis of wild-type and mutant CYLD

We exploited three computational algorithms: Phyre2, I-TASSER, and SWISS-MODEL to perform comparative structural analysis of wild-type and 15 mutants of CYLD. These tools were used to generate 3D structures of wild-type protein and mutated proteins as the whole structure of CYLD is not available in Protein Data Bank. Phyre2 utilized 2VHF and 1IXD as template for the 3D protein structure modeling for USP and CAP-Gly domain of CYLD, respectively.

ConSurf Results



Fig. 2 Conservation analysis of amino acid residues of CYLD using ConSurf server

Table 2 Alterations in the structural stability profile of the CYLD protein by I-Mutant and SDM tool

ID	Substitution	I-Mutant	RI	DDG-free energy change value (kcal/mol)	SDM predicted stability change $\Delta\Delta G$ (kcal/mol)	Prediction
rs971330819	V478A	Decrease	7	-0.63	-0.06	Destabilizing
rs1363261645	R489H	Decrease	9	-1.79	-0.82	Destabilizing
rs758765051	L610F	Decrease	7	-1.25	-0.41	Destabilizing
rs747739683	I644T	Decrease	4	-0.56	-0.9	Destabilizing
rs1404766168	L648R	Decrease	9	-1.46	-0.7	Destabilizing
rs964054055	P698T	Decrease	5	-0.08	1.48	Stabilizing
rs964054055	P698S	Increase	2	-0.07	0.67	Stabilizing
rs1449388332	P698L	Increase	3	0.58	2.07	Stabilizing
rs121908389	E747G	Decrease	7	-0.78	-0.94	Destabilizing
rs886040889	L781P	Increase	1	-0.15	-2.29	Destabilizing
rs773384548	H827P	Decrease	2	-0.2	-1.87	Destabilizing
rs773384548	H827R	Decrease	0	-0.03	-0.52	Destabilizing
rs940896803	R830K	Decrease	1	-0.4	-0.34	Destabilizing
rs772100161	V864F	Decrease	3	0.14	-0.5	Destabilizing
rs763179319	I867K	Decrease	8	-0.4	-2.69	Destabilizing
rs763179319	I867R	Decrease	4	0.15	2.91	Destabilizing
rs754721077	H871Q	Decrease	9	-1.56	-1.36	Destabilizing
rs1597094057	R894W	Increase	1	-0.95	0.31	Stabilizing

Then, we determined variations between native and mutant 3D protein structure with the help of TM-align tool. Two variants, V478A and R489H located in CAP-Gly domain with RMSD value 0, indicated no dissimilarities between these two variants when compared with wild-type protein. On the other hand, comparison between native and 13 missense variants exhibits significant TM-score and RMSD value (S2 Table). Missense variants located on USP domain showed greater RMSD values, and among them, H871Q, I867K, and H827P variants had the highest TM-score. We also used I-TASSER for an additional structural study of 13 nsSNPs to verify the relevance of these findings. This server generated the top 5 reliable superimposed models of mutants over the wild-type protein based on minimum confidence score (C-score) along with significant TM-score and RMSD value. To conduct comparative analysis of all atom of a protein, we incorporated SWISS-MODEL. This homology modeling server utilized 2vhf.2 as a template to build the structure of CYLD and its mutant. SWISS-MODEL server also determined solvation, torsion, QMEAN, and C β value for both native and mutants which are shown in the S3 Table.

3.7 Analysis of structural effect of point mutation on CYLD protein

The project HOPE server analyzed the physiochemical alterations of CYLD protein structure as a result

of amino acid substitutions (S4 Table). We observed functional CYLD mutations that significantly change size and hydrophobicity in all mutant residues. L610F, L648R, H827R, V864F, I867K, and I867R mutant residues are larger, whereas I644T, P698T, E747G, L781P, H827P, R830K, and H871Q mutant residues are smaller when compared to wild-type residues. Repulsion was generated between the mutant residue and neighboring residues when a charge was added in H827R (Fig. 3), I867K, and I867R position due to mutation. On the other hand, protein-folding problems can arise in L648R missense variant and empty space formed in the core of the protein when mutation occurs in I644T and R830K (Fig. 3) position. Moreover, H871Q, I644T, L781P, E747G, H827P, and R830K nsSNPs also resulted in the loss of interactions.

Swiss-PDBViewer calculates energy state variations of a protein when the position of an atom or molecules changes. We determined the deviations in the energy minimization state of CYLD structure geometry in wild-type and 13 variants. The total energy of the wild-type protein was -20,130.191 kJ/mol, which decreased in case of L610F, L648R, P698T, and I867R variants after energy minimization. Other missense variants showed increase in total energy after energy minimization. Among them, H827R showed significant increase in total energy (-15,956.584 kJ/mol) after energy minimization (S5 Table). Structural changes in h-bond in R830K are shown in Fig. 4.

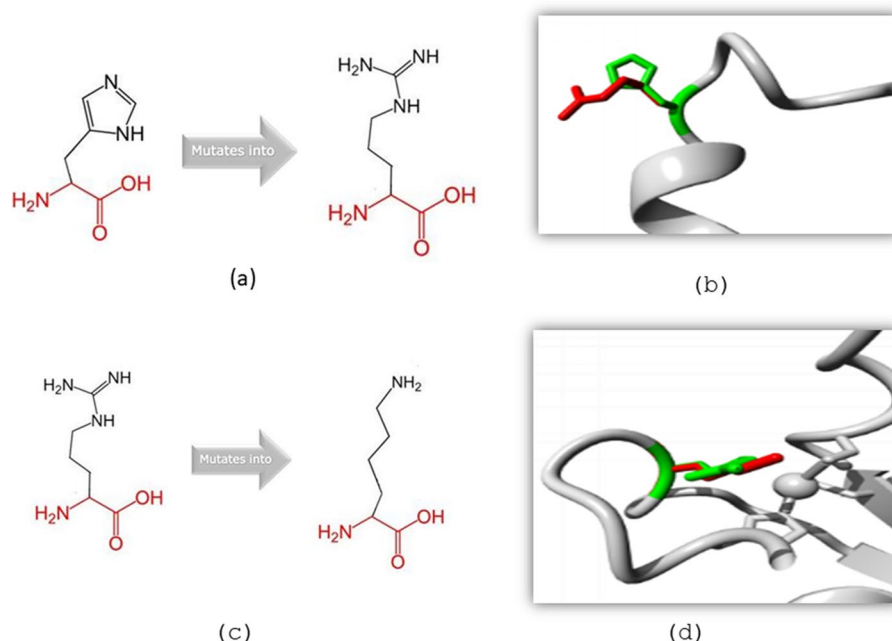


Fig. 3 Structural effect of the point mutation on variant H827R (**a** and **b**) and R830K (**c** and **d**) predicted by HOPE server. (Green color indicates wild, and red color indicates mutant residues)

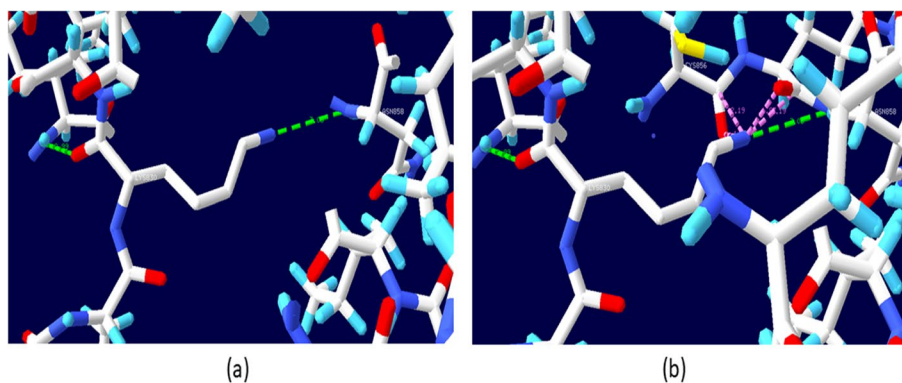


Fig. 4 Structural effect analysis of R830K by Swiss-PDBViewer (**a** represents R830 where R830 forms two H-bonds (2.99 Å, 3.08 Å) which are indicated by green discontinuous line; **b** represents 830 K where K830 clashes with C856 (1.43 Å, 1.90 Å, 2.03 Å) which are indicated by pink discontinuous line along with two H-bonds (green line))

3.8 Molecular docking analysis

Molecular docking analysis was performed between the CYLD (wild and mutants) and ubiquitin chain using HADDOCK to see how mutant interacts with ubiquitin compared to native CYLD protein. The PDB structure for USP domain of CYLD was taken from SWISS-MODEL using PDB ID: 2VHF (583aa-956aa) as a template as some residues were missing in the PDB structure. Ramachandran plot (S1 Fig.) verified the model where 92.47% amino acid residues are in favored region which assured the good quality of the model. On the other

hand, ubiquitin chain was derived from Protein Data Bank (PDB ID: 3WXE). Native and mutant structures were refined using refinement algorithm of HADDOCK. Active and passive residues of Ub and CYLD protein were determined by CPORT server that ensures binding of Ub protein in the appropriate binding site of CYLD. The binding affinity between native CYLD and ubiquitin was -14.8 kJ/mol. Among 13 high-risk nsSNPs, binding affinity increased in L610F, I644T, E747G, V864E, I867R, and H871Q, whereas binding affinity decreased in P698T, L781P, H827P, H827R, R830K, and I867K after mutation.

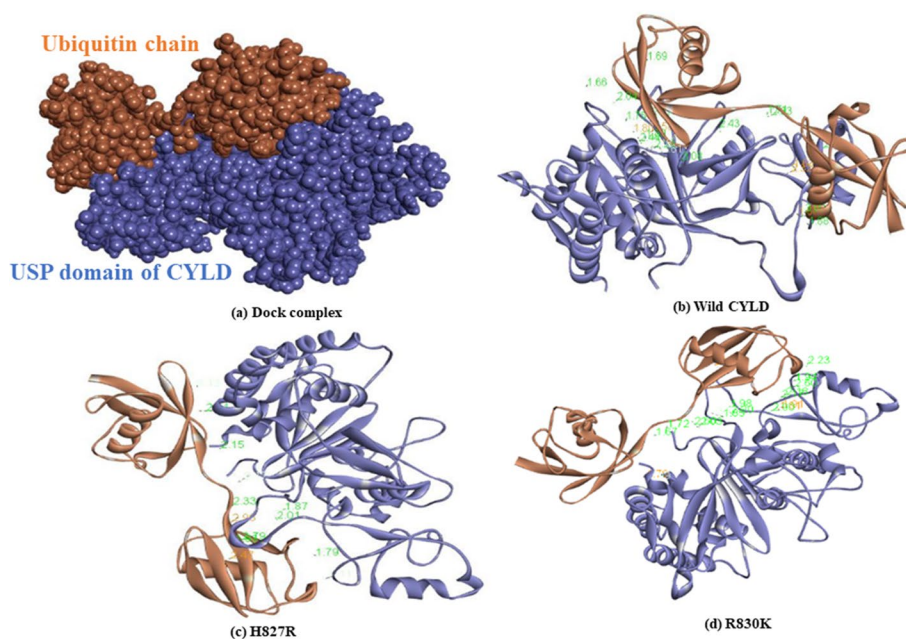


Fig. 5 Molecular docking analysis and visualization by BIOVIA Discovery Studio. (Blue indicates USP domain of CYLD, and chocolate color indicates ubiquitin. **a** indicates CYLD-ubiquitin docking complex, **b** indicates H-bond interactions between wild CYLD-ubiquitin dock complex, **c** represents H-bond interactions between mutant (H827R) CYLD-ubiquitin dock complex, **d** represents h-bond interactions between mutant (R830K) CYLD-ubiquitin dock complex)

Among them, significant reduction was observed in four nsSNPs, namely: R830K, H827R, P698T, and L781P with binding affinity -12.7 kJ/mol, -12.8 kJ/mol, -13.0 kJ/mol, and -13.4 kJ/mol, respectively. Binding affinity of these variants showed significant deviation compared to native protein. Binding affinity and dissociation constant

of all docking complexes were found from HADDOCK for both wild-type and mutant structures determined by PRODIGY server (S6 Table). We also determined the H-bond interaction between CYLD-ubiquitin dock complex applying BIOVIA Discovery Studio where wild-type CYLD formed 24-h bonds with ubiquitin. In case of

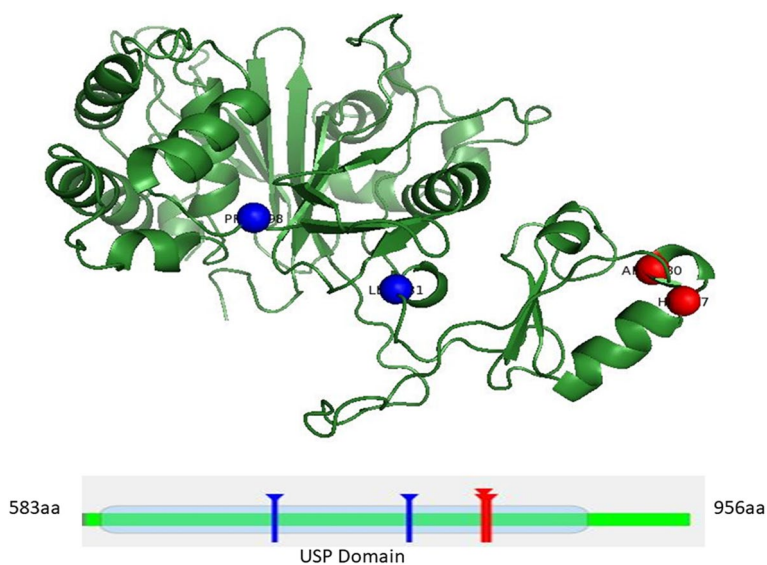


Fig. 6 Mutation 3D predicted the association of H827R and R830K (red mark) with cancer

R830K and H827R, total 22- and 13-h bonds were found in CYLD-ubiquitin dock complex respectively (Fig. 5) Interacting residues along with bond category and bond distance are represented in S7 Table.

3.9 Prediction of cancer causing nsSNPs

As CYLD is a tumor suppressor protein, loss of activity due to mutation can result in cancer. Mutation 3D is a server that predicts deleterious nsSNPs which are associated with human cancer. This analysis revealed the association of H827R and R830K with cancer (Fig. 6), and we considered these two nsSNPs for further analysis.

3.10 Molecular dynamic (MD) simulation

MD simulation was conducted to examine the deviation of the native and mutant CYLD-ubiquitin complex in relation to its initial conformation under physiological conditions. Trajectory analysis from the simulation enables the stability and flexibility of the system to be computed. The simulations were performed for 100 ns to investigate the structural flexibility, stability, and hydrogen bonding between the protein–protein complexes.

The overall changes in the protein stability due to the mutation were calculated by considering the root-mean-square deviation (RMSD) values. Mutant R830K and H827R complex exhibited great deviation in comparison to native CYLD-ubiquitin complex (Fig. 7). The average RMSD value for native complex was ~ 3.388 Å, which was increased in R830K and H827R mutant complex to ~ 5.278 Å and ~ 4.9575 Å, respectively (Fig. 7). The highest RMSD value for native complex was 4.418 Å at 39 ns;

meanwhile, the highest deviation was noticed for H827R complex with a 6.327 Å RMSD value at 79.75 ns compared with its initial structure. On the other hand, R830K complex showed deviation at 71 ns with 6.087 Å. Native complex showed mild deviation in RMSD value until 39 ns, and then the native complex stayed stable within the range of 2–4 Å for the rest of the time, indicating stability of the protein. On the other hand, mutant H827R showed an increasing tendency until 16.75 ns, and thereafter from 16.75 to 28 ns, RMSD value was decreased, and again, it started to increase at 29 to 100 ns at the range of 5–6 Å which is much higher than wild CYLD. Fluctuations that observed in this RMSD values indicate decreasing stability of the mutant H827R. In case of R830K complex, fluctuation rate is greater than mutant H827R. In R830K, the RMSD value started to increase at 11.5 ns and became unstable throughout the overall simulation period within the range of 5–6 Å which is higher than wild CYLD. Considerable fluctuations observed after 80 ns, and it continued up to 100 ns.

Furthermore, to determine the structural flexibility of the protein–protein complexes, we assessed the RMSF value (Fig. 7). This study revealed that R830K and native both complexes exhibited almost similar level of flexibility during the 100-ns simulation. However, some greater residual fluctuations also observed in case of R830K when compared with wild protein. The highest residual fluctuation for R830K was 8.22 Å observed at position GLN316 (899 aa of CYLD). On the contrary, H827R exhibited highest residual fluctuation 9.66 Å at position LYS179 (762 aa of CYLD) when compared to native and

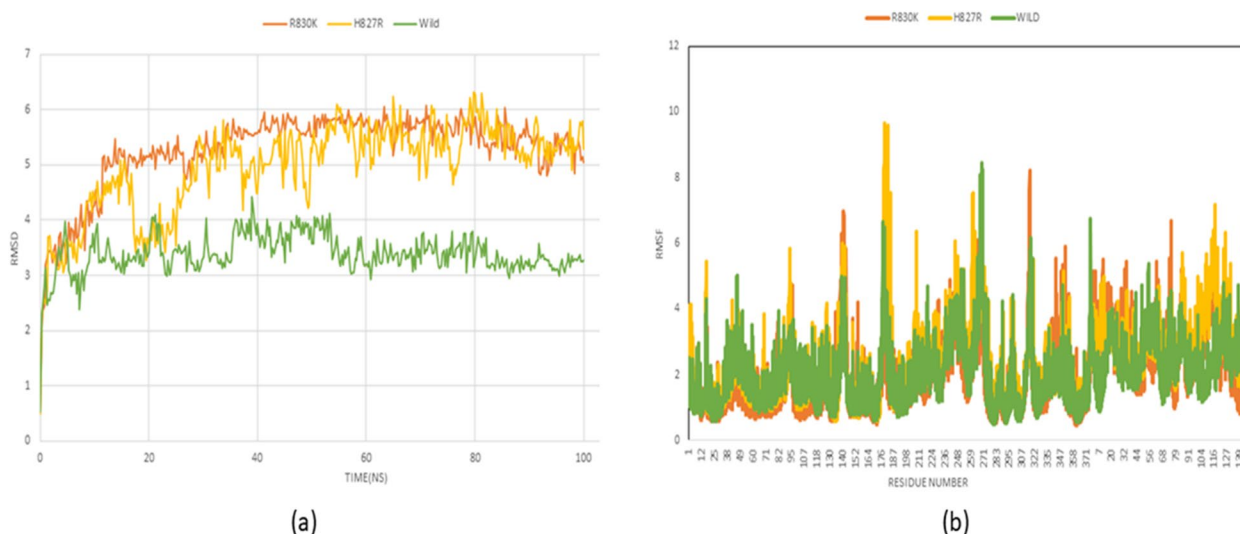


Fig. 7 Molecular dynamic simulation analysis performed by YASARA. **(a)** exhibits RMSD analysis of the Ca atoms of the structure of protein–protein complexes at 0 to 100 ns, whereas **(b)** represents RMSF analysis of the residues of the native and mutant CYLD protein with ubiquitin over the 100-ns simulation)



Fig. 8 Molecular dynamic simulation analysis performed by YASARA. **(a)** shows Rg analysis of the backbone structure of the protein–protein complexes over 100 ns, and **(b)** indicates H-bond analysis of the structure of protein–protein complexes over 100 ns

R830K complexes. Average fluctuation rate for native, mutant R830K and H827R was ~ 2.041 Å, ~ 2.085 Å, and ~ 2.425 Å, respectively. In terms of total residual portions, the RMSFs value of all mutant complexes differed significantly from the native complex structure.

From the radius of gyration (Rg) analysis, compactness and rigidity condition of protein–protein complex were determined. The Rg values of native protein complex ranged from 25.27 to 26.38 Å. In case of H827R and R830K, it ranged between 25.17 to 26.71 Å and 24.849 to 25.99 Å, respectively. In the average value for the native structure of CYLD and two mutants (H827R and R830K) were ~ 25.65 Å, ~ 25.70 Å, and ~ 25.37 Å, respectively (Fig. 8). It was observed that H827R complex had higher radius of gyration value than native and mutant R830K complexes, thus showing least compactness.

Following that, we studied the overall number of intramolecular hydrogen bonds present in the protein to assess the protein stability or the stability between proteins. Native complex of CYLD-ubiquitin displayed an average of ~ 392 H-bond throughout the 100-ns simulation. The average number of H-bond generated by CYLD-ubiquitin in mutant complexes R830K and H827R was estimated to be ~ 389 and ~ 386 , respectively, during the period of 100-ns simulation (Fig. 8). This analysis significantly depicts the impact of amino acid substitution on the backbone structure of the protein–protein complexes.

3.11 Principal component analysis

The principal component analysis model was constructed based on the analysis of the various structural and energy profile such as bond distances, bond angles, dihedral angles, planarity, van der Waals energies, and electrostatic energies getting from MD simulation analysis. Three training sets were taken into consideration for the further analysis. The first and second principal components (PC1 and PC2) of this PCA model cover 88.4% of the proportion variance. The score plot exhibits three different clusters for wild-type-ubiquitin complex (green), mutant H827R-ubiquitin complex (blue), and mutant R830K-ubiquitin complex (red) where PC1 covers 66.7% and PC2 covers 21.7% of the variance (Fig. 9). Different cluster formation for three training sets signifies fluctuations that occurred during MD simulation. Significant fluctuations were observed when wild type is compared with both mutant types. This analysis indicated that point mutations directed to the alterations of the structural and energy profile of the CYLD-ubiquitin complexes. Therefore, mutations in the 827th and 830th position of the CYLD resulted in the aberrant interaction pattern of the CYLD with ubiquitin.

3.12 Protein–protein network analysis

The functional interaction pattern of CYLD protein with other proteins in different biological pathways was predicted using the STRING database (Fig. 10). CYLD functionally interacts with TRAF2, TRAF6, IKBKG, RNF31, TNFRSF1A, DDX58, RIPK1, BIRC3, UBC, UBE2K, UBE2S, RPS27A, UBA52, RAD18, RPL8, RPS16, RPL19,

RPL35, RPS12, and RPL18A and thereby plays various significant biological roles (S9 Table). This interaction pattern of CYLD may be disturbed if any deleterious change occurs in CYLD protein. The data about degree of connectivity, average shortest path length, betweenness centrality, and closeness centrality of all the related protein of CYLD were predicted by Cytoscape (S8 Table). The highest number of interactions was seen with UBC (ubiquitin C) and UBA52 (ubiquitin A-52 residues ribosomal protein fusion product 1) with degree 20. Mutation can hamper all those interactions, highlighting the deleterious effect of nsSNPs of *CYLD*.

4 Discussion

CYLD is known as a deubiquitinase gene that exhibits tumor suppression activity in humans [2]. Mutation in *CYLD* is generally associated with many cancer types such as familial cylindroma, melanoma, salivary gland tumor, and breast cancer [2]. Investigation of the impact of point mutation on the structural and functional activity of *CYLD* protein is a difficult task. Application of various bioinformatics tools makes this analysis easier. In this study, we exploited the damaging consequences of nsSNPs of *CYLD* to study the effect on its structure and function using different computational approaches. We started our analysis by retrieving 446 nsSNPs recorded in NCBI database for *CYLD* gene. Subsequently, we examined these nsSNPs using seven different computational methods: PANTHER, PROVEAN, PREDICT SNP, PolyPhen-2, PhD-SNP, PON-P2, and SIFT for the screening out of high-risk nsSNPs. Each algorithm ranked nsSNPs based on their deleterious effect taking into consideration parameters such as sequence homology, structural homology, conservancy, and biological and physical characteristics of amino acids. The integration of different algorithms often serves as powerful tools to prioritize the functional SNP candidates [65]. Considering this, we focused on 18 significant nsSNPs of *CYLD* commonly predicted as deleterious by all the seven tools. InterPro, domain identification program, revealed that two nsSNPs were located at Cap-Gly domain required for the interaction with NEMO/IKK γ and TRAF2 [17], whereas the rest of the nsSNPs were positioned on the USP domain responsible for its deubiquitinase activity [18]. A study reported that mutation in conserved regions leads to the greater reduction in protein stability compared to non-conserved regions [66]. Therefore, we analyzed the conservation profile of our targeted nsSNPs; from there, we only considered highly conserved residues with the help of ConSurf server. Several studies demonstrate that alteration in protein stability due to SNP can cause degradation, misfolding, and coagulation in a protein leading to structural and functional impairments [67, 68], and we

have found 14 destabilizing residues among 18 nsSNPs in our study when we used I-Mutant and SDM tools.

Next, we approached to determine the profile of structural modifications caused by these destabilizing nsSNPs through the comparative structural analysis for both native and mutant protein models using Phyre2 homology model prediction server. TM-align tool determined the structural deviations of mutant models in comparison with the native protein. According to studies [46, 69], TM-score determines the topological similarity, whereas RMSD indicates average distance between α -carbon backbones of wild-type and mutant proteins. Greater RMSD value signifies greater deviation, and lower TM-score means higher dissimilarities between wild and mutant protein models. We furthered 13 nsSNPs based on higher RMSD value and lower TM score, and we found in one study [70] that they also selected nsSNPs based on higher RMSD. In our study, among 13 nsSNPs, highest RMSD value (2.21) was found in highly conserved H827R, and lowest TM-score (0.84714) was displayed by highly conserved R830K. I-TASSER generated confidence score by remodeling more reliable wild-type and mutant-type proteins. We also investigated relative terms in SWISS-MODEL such as solvation, torsion, QMEAN, and C β value comparing wild-type and mutant models. Project HOPE program provides deep insight on the detrimental effect of point mutation on the structural configuration of a protein. Analysis showed that wild-type residues replaced by smaller mutants result in the empty space formation due to the loss of significant interactions in case of R830K. Besides, misfolding and repulsion can cause when charge was added to H827R. The influence of deleterious nsSNPs on the energy minimization state of the *CYLD* protein determination is fundamental as protein achieves its stable conformation with lower energy after energy minimization according to a study [71]. On the contrary, structural changes due to mutation can restraint the protein to be stable easily. Findings showed that the total energy for the native *CYLD* protein was $-20,130.191$ kJ/mol after energy minimization. H827R mutant showed remarkable increase in energy $-15,956.584$ kJ/mol than wild type.

Furthermore, we performed molecular docking between *CYLD* PDB id:2VHF (583aa-956aa), and ubiquitin as binding interactions pattern among them has significant role in tumor suppressor activity of *CYLD* [10]. Studies showed that decreasing binding affinity due to mutation signifies impairment of the binding interaction pattern [72, 73]. Similarly, our analysis also revealed that 4 nsSNPs: L781P, P698T, H827R, and R830K mutant complex showed lowest binding affinity of -13.4 , -13.0 – 12.8 , and -12.7 kJ/mol, respectively, when compared with wild type (-14.6 kJ/mol). We observed higher

dissociation constant for these four nsSNPs (S6 Table) compared to the native CYLD which also substantiated weak binding interactions between ubiquitin and CYLD mutant protein complex. Mutation 3D specifically verified that mutation in H827 and R830 can have strong association with cancer, whereas no association was found for L781P and P698T.

We performed MD simulations to evaluate the dynamic behavior of our CYLD-ubiquitin complex in an aqueous environment for 100 ns. Simulation was executed with a time-step interval of 2.5 fs [74, 75]. This analysis mainly focused on the relative structural deviation of the H827R and R830K in comparison to wild-type CYLD protein. Mentionable variations in RMSD (root-mean-square deviation) value were observed in H827R and R830K compared with the wild-type protein. Wild-type CYLD exhibited variations in RMSD value up to 10.5 ns and then became stable within range between 2.9 and 4.4 Å during the simulation time frame. In case of H827R, we found highest peak at 80 ns with RMSD value 6.275 Å indicated that mutant H827R became unstable throughout the whole simulation period. On the other hand, we found highest peak at 71 ns with RMSD value 6.087 Å in case of R830K. Average RMSD value for mutants H827R (4.9575 Å) and R830K (5.278 Å) was much higher than the native CYLD (3.388 Å). These results indicated that H827R and R830K lead to the structural variation of the CYLD protein as higher RMSD value signifies structural distance of protein or protein complex. After that, we analyzed root-mean-square fluctuation (RMSF) of CYLD and its two mutants to evaluate mutation-causing fluctuations in structural part of a protein comparing with the actual structure of a protein. We observed higher residual fluctuation in H827R rather than R830K when referenced with native CYLD. We found highest RMSF 9.66 Å at positions LYS179 (762 aa of CYLD) for H827R. In case of R830K, higher RMSF value 8.22 was found at 316 (899 aa) residue.

Rg (radius of gyration) analysis determined the compactness of CYLD protein and thereby signified the folding rate as well as stability of that protein. We found that Rg values of CYLD protein complex ranged from 25.27 to 26.38 Å where as in case of H827R and R830K, Rg ranged between 25.17 to 26.71 Å and 24.849 to 25.99 Å, respectively. H827R mutant showed higher Rg value, and R830K showed less Rg value compared with wild type. From this, we can hypothesize that the compactness of the CYLD protein is probably affected by mutation at position 827 than at 830. Finally, H-bond analysis of CYLD was performed. A study revealed that the folding and stability of a protein can be affected by any change in H-bond formation [76]. Average H-bond

of native CYLD-ubiquitin displayed ~392 H-bonds, whereas the average H-bond generated by CYLD-ubiquitin in mutants R830K and H827R was calculated to be ~389 and ~386, respectively. Loss of H-bond in mutant complex signified its weak binding interaction with ubiquitin as well as its structural deformation. We found several studies [70, 77–79] where they did not performed molecular docking and MD simulations for observing changes in interaction pattern as well as stability of a protein after point mutation. Principal component analysis obtained from MD simulation hints at the aberrant structural and functional activity of CYLD due to the point mutation at 827 and 830 position. In several studies [58, 80], they found greater deviation in structure and energy profile by comparing wild and mutants, and we also found deviation in structure and energy profile by comparing wild-type CYLD with H827R and R830K mutants. We also examined the interacting partners of CYLD in various biological pathways through network analysis and suggested that H827R and R830K mutants can disturb those pathways.

CYLD performs its tumor suppressor activity by disassembling k-63 ubiquitin chain [20, 81] where interaction between C-terminal USP domain and ubiquitin chain is the prerequisite for this function. Mutation in USP domain can interrupt their Lys63-linked polyubiquitin cleavage activity resulting in cancer [1]. In our current study, we tried to short-list deleterious SNP disrupting the total catalytic activity and binding affinity of USP domain of CYLD and their strong association with cancer.

Throughout the study, a consistent workflow was developed for the reproducibility of this in silico deleterious SNPs prediction and multiple algorithms; tools were used to assess each step to increase the accuracy of the approaches by removing the artifacts from each tool. SNPs in the genome are thought to be critical in regard to functional and structural effects of proteins involving cellular metabolism, gene expression and disease susceptibility etc. This computational prediction-based approach would provide deep insights and faster outcomes for experimental validation.

In conclusion, mutation in tumor suppressor *CYLD* has been linked to a variety of cancers. Therefore, determining the effect of point mutations on the structural and functional activities of the CYLD protein is a challenging task. The use of numerous bioinformatics tools simplifies this assessment. In our present study, we employed multiple computational tools to investigate the harmful consequences of the mutant variant of CYLD on its structure and function. As mutant CYLD is associated with different cancer types, our results will be useful in the development of future diagnostic and research on

CYLD mutations. Our findings however required in vitro and in vivo experimental validation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s44342-024-00007-2>.

Additional file 1: S1 Fig. Domain identification of CYLD using InterPro server. CYLD contains two functional domains namely: Cap-Gly domain (127-540 Amino acid) and USP domain (592-950 AA). **S2 Fig.** The 3D structure of CYLD protein (USP domain) constructed by SWISS MODEL and structural quality assessment using Ramachandran plot. **S1 Table.** Prediction of known disease-associated variants. **S2 Table.** Conservancy analysis using ConSurf server. **S3 Table.** TM-score and RMSD value predicted by TM align tool. **S4 Table.** Swiss model Result. **S5 Table.** Structural effect of the point mutation predicted by HOPE project. **S6 Table.** Energy minimization state of wild and 13 mutant variants determined by Swiss PDB Viewer. **S7 Table.** Binding affinity and dissociation constant of all docking complex shown in the table. **S8 Table.** Hydrogen bond interactions between amino acid residues of CYLD (wild & Mutant variant)-Ubiquitin docking complexes given below. **S9 Table.** Protein-protein interaction analysis using Cytoscape. **S10 Table.** Functions of proteins interacting with CYLD in the PPI network.

Acknowledgements

We acknowledge the Department of Biotechnology and Genetic Engineering and Computational Biology and Chemistry Lab, Noakhali Science and Technology University, for providing the research work supports.

Authors' contributions

Conceptualization, NMB, MRA, HAR, and SG. Data curation, ASR and TF. Formal analysis, ASR, TF, MAM, DAS, NJA, and MKI. Methodology, NMB, MRA, HAR, and SG. Writing — original draft, ASR, TF, and HAR. Writing — review and editing, MRA, NMB, and MSH.

Declarations

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 22 November 2023 Accepted: 26 December 2023

Published online: 31 May 2024

References

- Bignell GR, Warren W, Seal S, Takahashi M, Rapley E, Barfoot R, et al. Identification of the familial cylindromatosis tumour-suppressor gene. *Nat Genet.* 2000;25(2):160–5.
- Massoumi R. CYLD: a deubiquitination enzyme with multiple roles in cancer. *Future Oncol.* 2011;7(2):285–97.
- Dosemeci A, Thein S, Yang Y, Reese TS, Tao-Cheng J-H. CYLD, a deubiquitinase specific for lysine63-linked polyubiquitins, accumulates at the postsynaptic density in an activity-dependent manner. *Biochem Biophys Res Commun.* 2013;430(1):245–9.
- Wu X, Fukushima H, North BJ, Nagaoka Y, Nagashima K, Deng F, et al. SCF β -TRCP regulates osteoclastogenesis via promoting CYLD ubiquitination. *Oncotarget.* 2014;5(12):4211.
- Bikker R, Christmann M, Preuß K, Welz B, Friesenhagen J, Dittrich-Breiholz O, et al. TNF phase III signalling in tolerant cells is tightly controlled by A20 and CYLD. *Cell Signal.* 2017;37:123–35.
- Sato Y, Goto E, Shibata Y, Kubota Y, Yamagata A, Goto-Ito S, et al. Structures of CYLD USP with Met1-or Lys63-linked diubiquitin reveal mechanisms for dual specificity. *Nat Struct Mol Biol.* 2015;22(3):222–9.
- Draber P, Kupka S, Reichert M, Draberova H, Lafont E, de Miguel D, et al. LUBAC-recruited CYLD and A20 regulate gene activation and cell death by exerting opposing effects on linear ubiquitin in signaling complexes. *Cell Rep.* 2015;13(10):2258–72.
- Stegmeier F, Sowa ME, Nalepa G, Gygi SP, Harper JW, Elledge SJ. The tumor suppressor CYLD regulates entry into mitosis. *Proc Natl Acad Sci.* 2007;104(21):8869–74.
- Reissig S, Hövelmeyer N, Tang Y, Weih D, Nikolaev A, Riemann M, et al. The deubiquitinating enzyme CYLD regulates the differentiation and maturation of thymic medullary epithelial cells. *Immunol Cell Biol.* 2015;93(6):558–66.
- Johari T, Maiti TK. Catalytic domain mutation in CYLD inactivates its enzyme function by structural perturbation and induces cell migration and proliferation. *BiochimBiophysActa Gen Subj.* 2018;1862(9):2081–9.
- Massoumi R, Chmielarska K, Hennecke K, Pfeifer A, Fässler R. Cylid inhibits tumor cell proliferation by blocking Bcl-3-dependent NF- κ B signaling. *Cell.* 2006;125(4):665–77.
- Takami Y, Nakagami H, Morishita R, Katsuya T, Hayashi H, Mori M, et al. Potential role of CYLD (cylindromatosis) as a deubiquitinating enzyme in vascular cells. *Am J Pathol.* 2008;172(3):818–29.
- O'Donnell MA, Perez-Jimenez E, Oberst A, Ng A, Massoumi R, Xavier R, et al. Caspase 8 inhibits programmed necrosis by processing CYLD. *Nat Cell Biol.* 2011;13(12):1437–42.
- Dubois A, Wilson V, Bourn D, Rajan N. CYLD GeneticTesting for Brooke-Spiegler syndrome, familial cylindromatosis and multiple familial trichoepitheliomas. *PLoS Curr.* 2015;7.
- Young A, Kellermayer R, Szigeti R, Teszas A, Azmi S, Celebi J. CYLD mutations underlie Brooke-Spiegler, familial cylindromatosis, and multiple familial trichoepithelioma syndromes. *Clin Genet.* 2006;70(3):246–9.
- Blake PW, Toro JR. Update of cylindromatosis gene (CYLD) mutations in Brooke-Spiegler syndrome: novel insights into the role of deubiquitination in cell signaling. *Hum Mutat.* 2009;30(7):1025–36.
- Saito K, Kigawa T, Koshiha S, Sato K, Matsuo Y, Sakamoto A, et al. The CAP-Gly domain of CYLD associates with the proline-rich sequence in NEMO/IKK γ . *Structure.* 2004;12(9):1719–28.
- Komander D, Lord CJ, Scheel H, Swift S, Hofmann K, Ashworth A, et al. The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Mol Cell.* 2008;29(4):451–64.
- Chen ZJ. Ubiquitin signalling in the NF- κ B pathway. *Nat Cell Biol.* 2005;7(8):758–65.
- Kovalenko A, Chable-Bessia C, Cantarella G, Israël A, Wallach D, Courtois G. The tumour suppressor CYLD negatively regulates NF- κ B signalling by deubiquitination. *Nature.* 2003;424(6950):801–5.
- Masoumi K, Shaw-Hallgren G, Massoumi R. Tumor suppressor function of CYLD in nonmelanoma skin cancer. *J Skin Cancer.* 2011;2011.
- Sun S. CYLD: a tumor suppressor deubiquitinase regulating NF- κ B activation and diverse biological processes. *Cell Death Differ.* 2010;17(1):25–34.
- Alameda J, Moreno-Maldonado R, Navarro M, Bravo A, Ramirez A, Page A, et al. An inactivating CYLD mutation promotes skin tumor progression by conferring enhanced proliferative, survival and angiogenic properties to epidermal cancer cells. *Oncogene.* 2010;29(50):6522–32.
- Ma F, Zhi C, Wang M, Li T, Khan SA, Ma Z, et al. Dysregulated NF- κ B signal promotes the hub gene PCLAF expression to facilitate nasopharyngeal carcinoma proliferation and metastasis. *Biomed Pharmacother.* 2020;125:109905.
- Bhagwat M. Searching NCBI's dbSNP database. *Curr Protocols Bioinform.* 2010;32(1):1.19. 1-1.8.
- Sim N-L, Kumar P, Hu J, Henikoff S, Schneider N, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012;40(W1):W452–7.
- Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zundulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol.* 2014;10(1): e1003440.
- Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 2016;44(D1):D336–42.

29. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protocols Human Genet.* 2013;76(1):7.20. 1–7. 41.
30. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015;31(16):2745–7.
31. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. 2012.
32. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics.* 2006;22(22):2729–34.
33. Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS ONE.* 2015;10(2):e0117380.
34. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 2012;40(D1):D306–12.
35. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 2005;33(suppl_2):W306–W10.
36. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* 2011;39(suppl_2):W215–W22.
37. Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 2017;45(W1):W229–35.
38. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics.* 2004;20(8):1322–4.
39. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinform.* 2010;11(1):1–10.
40. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis.* 1997;18(15):2714–23.
41. Johansson MU, Zoete V, Michielin O, Guex N. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics.* 2012;13(1):1–11.
42. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725–38.
43. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods.* 2015;12(1):7–8.
44. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 2015;43(W1):W174–81.
45. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845–58.
46. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005;33(7):2302–9.
47. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296–303.
48. De Vries SJ, Van Dijk M, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc.* 2010;5(5):883–97.
49. Van Zundert G, Rodrigues J, Trellet M, Schmitz C, Kastriitis P, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mole Biol.* 2016;428(4):720–5.
50. Bienert S, Waterhouse A, De Beer TA, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL repository—new features and functionality. *Nucleic Acids Res.* 2017;45(D1):D313–9.
51. de Vries SJ, Bonvin AM. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE.* 2011;6(3):e17695.
52. Xue LC, Rodrigues JP, Kastriitis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics.* 2016;32(23):3676–8.
53. Wang Q, He J, Wu D, Wang J, Yan J, Li H. Interaction of α -cyperone with human serum albumin: determination of the binding site by using Discovery Studio and via spectroscopic methods. *J Lumin.* 2015;164:81–5.
54. Meyer MJ, Lapcevic R, Romero AE, Yoon M, Das J, Beltrán JF, et al. mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome. *Hum Mutat.* 2016;37(5):447–56.
55. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins.* 2004;57(4):678–83.
56. Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR, et al. Lipid14: the amber lipid force field. *J Chem Theory Comput.* 2014;10(2):865–79.
57. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins.* 1993;17(4):412–25.
58. Shahinozzaman M, Ahmed S, Emran R, Tawata S. Molecular modeling approaches predicted 1, 2, 3-triazolyl ester of ketorolac (15K) to be a novel allosteric modulator of the oncogenic kinase PAK1. *Sci Rep.* 2021;11(1):1–19.
59. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2012;41(D1):D808–15.
60. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Data mining in proteomics: Springer;* 2011. p. 291–303.
61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
62. Zhang H, Zhu J, Wang C, Sun S, Zheng W-M, Bu D. Improving prediction of burial state of residues by exploiting correlation among residues. *BMC Bioinformatics.* 2017;18(3):165–75.
63. Malleshappa Gowder S, Chatterjee J, Chaudhuri T, Paul K. Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *Sci World J.* 2014;2014.
64. Gromiha MM, Ahmad S. Role of solvent accessibility in structure based drug design. *Curr Comput Aided Drug Des.* 2005;1(3):223–35.
65. Priya Doss CG, Chakraborty C, Chen L, Zhu H. Integrating in silico prediction methods, molecular docking, and molecular dynamics simulation to predict the impact of ALK missense mutations in structural perspective. *BioMed research international.* 2014;2014.
66. Kragelund BB, Poulsen K, Andersen KV, Baldursson T, Krøll JB, Neergård TB, et al. Conserved residues and their role in the structure, function, and stability of acyl-coenzyme A binding protein. *Biochemistry.* 1999;38(8):2386–94.
67. Deller MC, Kong L, Rupp B. Protein stability: a crystallographer’s perspective. *Acta Crystallogr F Struct Biol Commun.* 2016;72(2):72–95.
68. Hossain MS, Roy AS, Islam MS. In silico analysis predicting effects of deleterious snps of human rassf5 gene on its structure and functions. *Sci Rep.* 2020;10(1):1–14.
69. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci.* 2001;10(7):1470–3.
70. Arshad M, Bhatti A, John P. Identification and in silico analysis of functional SNPs of human TAGAP protein: a comprehensive study. *PLoS ONE.* 2018;13(1):e0188143.
71. Gautam B. Energy minimization. *Homology Molecular Modeling-Perspectives and Applications: IntechOpen;* 2020.
72. Islam MJ, Khan AM, Parves MR, Hossain MN, Halim MA. Prediction of deleterious non-synonymous SNPs of human STK11 gene by combining algorithms, molecular docking, and molecular dynamics simulation. *Sci Rep.* 2019;9(1):1–16.
73. Akter S, Roy AS, Tonmoy MIQ, Islam MS. Deleterious single nucleotide polymorphisms (SNPs) of human IFNAR2 gene facilitate COVID-19 severity in patients: a comprehensive in silico approach. *J Biomole Struct Dynamics.* 2021:1–17.
74. Hannan MA, Dash R, Sohag AAM, Moon IS. Deciphering molecular mechanism of the neuropharmacological action of fucosterol through integrated system pharmacology and in silico analysis. *Marine drugs.* 2019;17(11).
75. Vivar-Sierra A, Araiza-Macías MJ, Hernández-Contreras JP, Vergara-Castañeda A, Ramírez-Vélez G, Pinto-Almazán R, et al. In silico study of polyunsaturated fatty acids as potential SARS-CoV-2 spike protein closed conformation stabilizers: epidemiological and computational approaches. *Molecules.* 2021;26(3):711.

76. Chikalov I, Yao P, Moshkov M, Latombe J-C. Learning probabilistic models of hydrogen bond stability from molecular dynamics simulation trajectories. *BMC Bioinformatics*. 2011;12(1):1–6.
77. Kaur T, Thakur K, Singh J, Kamboj SS, Kaur M. Identification of functional SNPs in human LGALS3 gene by in silico analyses. *Egypt J Med Human Genet*. 2017;18(4):321–8–8.
78. Rozario LT, Sharker T, Nila TA. In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. *PLoS ONE*. 2021;16(6): e0252932.
79. Rajasekaran R, Sudandiradoss C, Doss CGP, Sethumadhavan R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. *Genomics*. 2007;90(4):447–52.
80. Brummelkamp TR, Nijman SM, Dirac AM, Bernards R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF- κ B. *Nature*. 2003;424(6950):797–801.
81. Trompouki E, Hatzivassiliou E, Tschirritzis T, Farmer H, Ashworth A, Mosialos G. CYLD is a deubiquitinating enzyme that negatively regulates NF- κ B activation by TNFR family members. *Nature*. 2003;424(6950):793–6.