



Computational exploration of *SLC14A1* genetic variants through structure modeling, protein-ligand docking, and molecular dynamics simulation

Tamanna Sultana, Sadia Islam Mou, Dipankor Chatterjee, Md. Omar Faruk **, Md. Ismail Hosen *

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

ARTICLE INFO

Keywords:

SLC14A1 gene
Urea transporter type B1 (UT-B1)
Bladder cancer
Single nucleotide polymorphism (SNP)
Molecular dynamics

ABSTRACT

The urea transporter UT-B1, encoded by the *SLC14A1* gene, has been hypothesized to be a significant protein whose deficiency and dysfunction contribute to the pathogenesis of bladder cancer and many other diseases. Several studies reported the association of genetic alterations in the *SLC14A1* (UT-B1) gene with bladder carcinogenesis, suggesting a need for thorough characterization of the UT-B1 protein's coding and non-coding variants. This study used various computational techniques to investigate the commonly occurring germ-line missense and non-coding SNPs (ncSNPs) of the *SLC14A1* gene (UT-B1) for their structural, functional, and molecular implications for disease susceptibility and dysfunctionality. *SLC14A1* missense variants, primarily identified from the ENSEMBL genome browser, were screened through twelve functionality prediction tools leading to two variants D280Y (predicted detrimental by maximum tools) and D280N (high global MAF) for rs1058396. Subsequently, the ConSurf and NetSurf tools revealed the D280 residue to be in a variable site and exposed on the protein surface. According to I-Mutant2.0 and MUpro, both variants are predicted to cause a significant effect on protein stability. Analysis of molecular docking anticipated these two variants to decrease the binding affinity of UT-B1 protein for the examined ligands to a significant extent. Molecular dynamics also disclosed the possible destabilization of the UT-B1 protein due to single nucleotide polymorphism compared to wild-type protein which may result in impaired protein function. Furthermore, several non-coding SNPs were estimated to affect transcription factor binding and regulation of *SLC14A1* gene expression. Additionally, two ncSNPs were found to affect miRNA-based post-transcriptional regulation by creating new seed regions for miRNA binding. This comprehensive *in-silico* study of *SLC14A1* gene variants may serve as a springboard for future large-scale investigations examining *SLC14A1* polymorphisms.

1. Introduction

SLC14A1 (Solute carrier family 14 member 1) gene encodes Urea transporter type B (UT-B) which is one of the major types of urea transporter in mammals. Urea transport across various cell membranes depends on the facilitated diffusion through this specialized transporter, since the strong dipole moment of urea prevents its passage through the non-polar lipid bilayer [1]. According to some recent large-scale genome-wide association studies (GWAS) of urothelial bladder cancer, variants in the *SLC14A1* gene were found to be associated with human bladder carcinogenesis [2–4].

Bladder, a unique urine depository, deals with a high urine concentration and is responsible for the maintenance of urea concentration,

cell homeostasis, and nitrogen balance [5]. Urea transporter UT-B1 is more abundantly expressed compared to UT-B2 isoform in the bladder than any other tissue and is located primarily in the basolateral membrane of the urothelial umbrella cells [6]. In the UT-B knock-out mice, the lack of the 'urea scavenger' gene has resulted in significant apoptosis and DNA damage in the urothelial cells, where the quantity of urea is nine times greater than that in wild-type mice, leading to the development of bladder carcinogenesis [7].

Again, increased urea concentrations brought on by UT-B deficiency can lead to aberrant arginine metabolism [8] and elevated amounts of NO [5], both of which can induce DNA damage and cell death in many ways. Hence, the buildup of urea caused by UT-B dysfunction or *SLC14A1* gene suppression in the urothelium may produce intracellular

* Corresponding author. Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh.

** Corresponding author. Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh.

E-mail addresses: faruk.bmb@du.ac.bd (Md.O. Faruk), ismail.hosen@du.ac.bd (Md.I. Hosen).

metabolic abnormalities that interact with conventional UBC (urothelial bladder cancer) pathways [9].

Additionally, UT-B urea transporters can be found in a wider variety of other tissues, including erythrocytes, brain, lung, heart, pancreas, colon, small intestine, prostate, kidney, skeletal muscle, bone marrow, etc. [10]. Thus, any influence or obstruction of the urea transport function of UT-B may result in the pathophysiology of several other diseases affecting various organs or tissues as well as blood pressure, digestive system, bone metabolism, male reproductive function, brain astrocyte activity, and cardiac function [8,11].

Single nucleotide polymorphisms (SNPs) are the most common types of genetic variations in the human genome. Among these variations, non-synonymous (nsSNPs)/missense SNPs in the coding regions are crucial and account for residual alteration that may exert a harmful

effect on the encoded protein [12]. Missense mutations account for almost half of all DNA variants related to genetic illnesses as causal or susceptibility factors [13]. These genetic changes may cause protein structural destabilization, improper gene regulation, protein hydrophobicity changes, charge disturbances, changes in geometry, dynamics, translation, protein-protein interactions, and protein integrity loss [14].

Despite the reported association of *SLC14A1* variants with bladder carcinogenesis and potential impact on other body functions, the mutational characterization of the *SLC14A1* gene (UT-B1) and the molecular mechanism of its coding and non-coding variants has not been fully explored. Recent research has seen an uptick in the usage of in silico methods for predicting how nsSNPs may affect protein structure and function [15,16]. Therefore, the current work aims to examine the structural and functional impact of some common coding

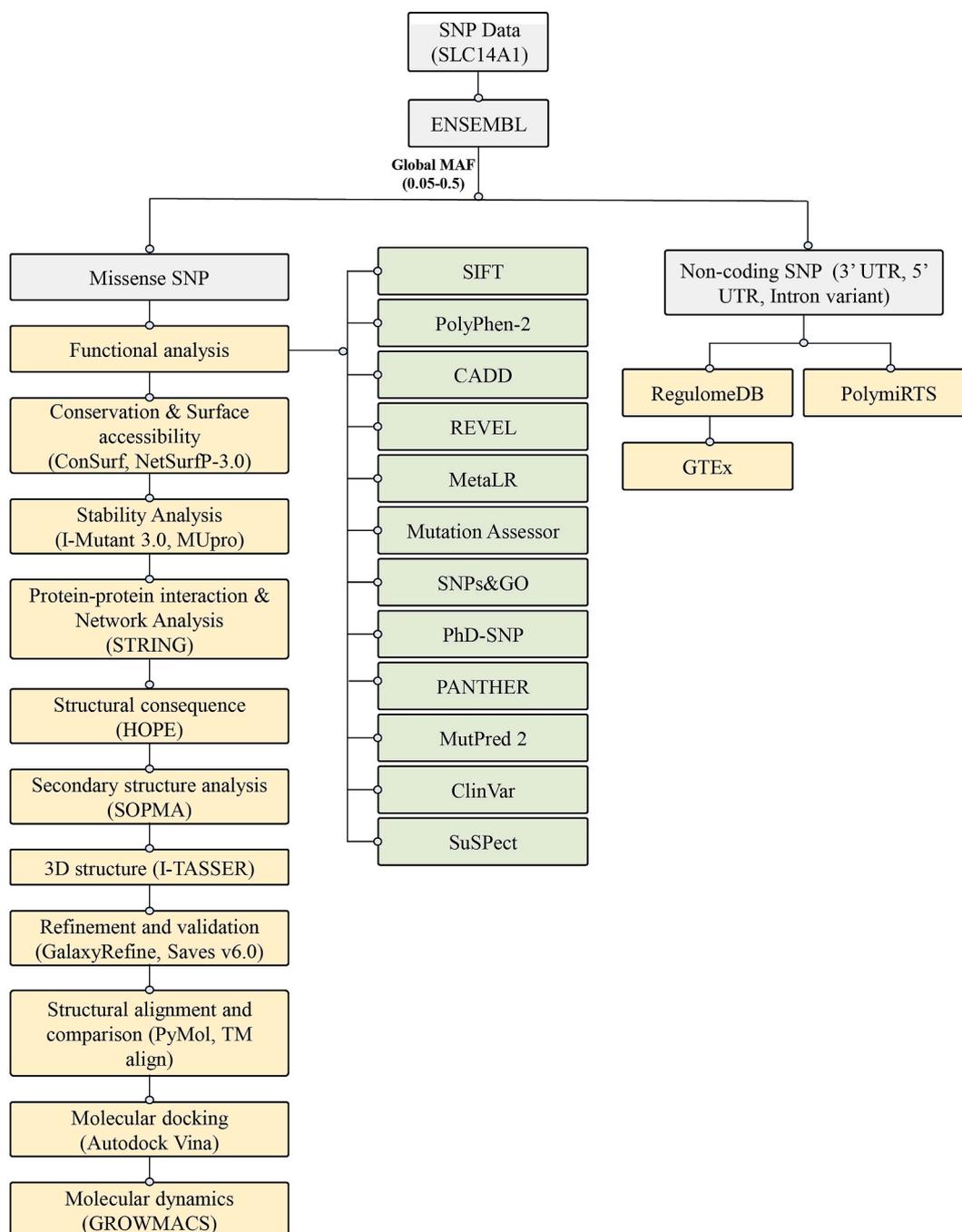


Fig. 1. Representation of the methodology in the flowchart.

non-synonymous and non-coding SNPs of the *SLC14A1* gene on the UT-B1 protein and its dynamic nature by utilizing a variety of publicly available online computational tools. The analyses of this study can potentially enhance the investigation of UT-B1 protein and the advancement of targeted therapies and drugs.

2. Materials and methods

The step-by-step investigation, the techniques, and the tools employed in this study to evaluate the impact of some common SNPs of the *SLC14A1* gene are outlined in Fig. 1.

2.1. Retrieval of SNP data

The study involves the investigation of the *SLC14A1* gene for its genetic variants in the ENSEMBL genome browser (<https://www.ensembl.org/index.html>) [17]. Among the 20 transcripts (splice variants) of this gene, a transcript encoding the canonical size of the UT-B1 protein (389 amino acids) was chosen for SNP data retrieval. Moreover, the sequence of the UT-B1 protein was retrieved from UniProt (<https://www.uniprot.org/>) [18].

2.2. Identification of the most frequent missense SNPs

Genetic variants of the selected transcript of the *SLC14A1* gene were presented in the tabular form enabling to identification of the most frequently occurring missense SNPs (nsSNPs) by setting the consequence type as missense variant and the global MAF (minor allele frequency) range of 0.05–0.5.

2.3. Analysis of functional consequence

The filtered nsSNPs were subjected to screening through several bioinformatics tools, including SIFT, PolyPhen2, MutPred2, etc., and finally, two selected variants were analyzed utilizing other web-based tools to predict the effect of the SNP on the function of the encoded protein.

SIFT-Sorting Intolerant from Tolerant (<https://sift.bii.a-star.edu.sg/>) algorithm predicts the functional impact of an amino acid substitution based on sequence homology and the altered amino acid's physical characteristics. The threshold value of the SIFT score is 0.05, below which substitutions are referred to as "deleterious," whereas all others are "tolerated" [19].

PolyPhen-2 (Polymorphism Phenotyping v2) (<http://genetics.bwh.harvard.edu/pph2/>) uses simple physical and comparative factors to foretell the functional impact. The PolyPhen score ranges from 0 to 1, based on which the SNPs are categorized into three classes: probably damaging, possibly damaging, and benign [20].

CADD (The Combined Annotation Dependent Depletion) (<https://cadd.gs.washington.edu/>) evaluates the deleteriousness of SNPs and indel variants by combining numerous annotations into a single measure. CADD score ranges from 1 to 99 with no threshold value. SNP with a higher CADD score are more likely to be deleterious [21].

REVEL (Rare Exome Variant Ensemble Learner) (<https://sites.google.com/site/revelgenomics/>) estimates the pathogenicity of missense variants by combining the results from several tests, including FATHMM v2.3, VEST 3.0, SIFT, PROVEAN, MutationTaster, LRT, etc. Scores vary from 0 to 1, and the likelihood of pathogenicity increases with increasing scores [22].

MetaLR predicts the deleteriousness of an SNP using a logistic regression-based ENSEMBL method and scores between 0 and 1 with a cut-off of 0.5 [23].

Mutation Assessor (<http://mutationassessor.org/>) predicts the functional impact of amino acid changes based on evolutionary conservation in protein homologs and assigns a score ranging from 0 to 1 to each mutation, with higher scores indicating a higher likelihood of

functional impact [24].

SNPs&GO (<https://snps.biofold.org/snps-and-go/snps-and-go.html>) analyzes protein structure, function, as well as biological pathways and processes to determine if an SNP is disease-associated or neutral. If the disease probability score is > 0.5, the SNP is predicted disease associated [25]. Analysis for **PhD-SNP** [26] and **PANTHER** [27] is also performed by the SNPs&GO server.

MutPred2 (<http://mutpred.mutdb.org/>) classifies amino acid replacements as pathogenic or benign. Unlike SIFT and PolyPhen2, it predicts variant pathogenicity and the chemical causes behind it, along with a prediction score [28].

ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) compiles data on genetic variation and its impact on human health as well as descriptions of related features from various sources like Office of Rare Diseases, GeneReviews, OMIM, and the Human Phenotype Ontology (HPO), etc. [29].

SuSpect (<http://www.sbg.bio.ic.ac.uk/suspect/>) is an SAV (support vector machine) based method that offers scores between 0 and 100 with a cutoff value of 50 and predicts the phenotypic impact of missense variants using information from sequence-, structure- and system biology-based features [30].

2.4. Conservation analysis and surface accessibility prediction

ConSurf (<https://consurf.tau.ac.il/>) was used for the analysis of the evolutionary conservation levels of amino acids in protein sequences. By comparing homologous sequences from related animals, the program assigns color-coded conservation scores to each amino acid that shows how consistently an amino acid is used throughout a family of proteins. It also uses the NACSES algorithm to inform if a particular amino acid residue is exposed or buried [31].

NetSurfP-3.0 (<https://services.healthtech.dtu.dk/services/NetSurfP-3.0>), a free web server, uses protein language models and deep learning approach to accurately and quickly predict surface accessibility, secondary structure, disorder, and phi/psi dihedral angles of the amino acids in a protein sequence. Amino acids with RSA greater than 25% were considered exposed, while others were buried in the protein's 3D structure [32].

2.5. Protein stability analysis

I-Mutant2.0 (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi>), an SVM (support vector machine) tool, was employed to foretell how altering an amino acid would affect protein stability based on protein sequence or structure. Using empirical and machine learning methodologies, it estimates amino acid substitution-induced Gibbs free energy (ΔG) variations. Protein stability status is expressed by a value of G that is negative or greater than zero, indicating decreased or increased stability correspondingly [33].

MUpro (<https://mupro.proteomics.ics.uci.edu/>) is also a web-based tool for the prediction of protein stability change resulting from amino acid alteration using the Gibbs free energy (ΔG), which additionally assigns a confidence level between -1 and 1 [34].

2.6. Protein-protein interactions by STRING v11.5 database

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (<https://string-db.org/>), a biological database and web resource covering a wide range of model and non-model organisms, was used to predict gene ontology and the interaction profile of the UT-B1 protein with other proteins which can be either direct (physical) interaction or indirect (functional) association. The website uses computational methods to infer functional and physical relationships, and it links data from several databases to shed light on the proteins' interconnected networks [35].

2.7. Analysis of structural consequence

HOPE (Have Our Protein Explained) (<https://www.projecthope.org/>) is a freely accessible next-generation approach that analyzes the impact of point mutation on the 3D structure of the encoded protein. It gathers data from a variety of sources using WHAT IF Web services and describes the biological basis of a disease-related phenotype of proteins in a fully automated manner. It informs us about the change in size, charge, hydrophobicity, spatial structure, and bond differences brought on by alterations in human proteins. HOPE performs multiple sequence alignment (MSA), and based on conservancy level, it gives a MetaRNN score that ranges between 0.0 and 1.0. The higher the score, the more likely it is to be pathogenic [36].

2.8. Secondary structure analysis

SOPMA (Self-Optimized Prediction Method with Alignment) (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html) is a tool for secondary structure prediction of any given amino acid sequence of protein based on sequence homology with a known protein. An amino acid sequence that is short and homologous will typically create comparable secondary structures. SOPMA has an accuracy of about 70–80% in predicting the secondary structure (helix, sheet, turn, and coil) of proteins [37].

2.9. 3D modeling of proteins

An online platform called **I-TASSER** (Iterative Threading ASSEMBly Refinement) (<https://zhanggroup.org/I-TASSER/>) was employed to automatically generate a high-quality 3D model of the UT-B1 protein and its variants. I-TASSER is a 3D modeler based on the threading approach [38], also known as fold recognition, as it searches for specific secondary conformation of a known protein structure instead of an entire sequence of a homologous protein [39]. Templates having the highest Z-score for structure modeling are identified from the PDB library by LOMETS, a meta-server threading approach. C-score refers to a confidence score that normally falls between [−5,2], with a higher C-score denoting a model with a high level of confidence and vice versa [38].

2.10. Refinement and quality assessment of the 3D structures

The predicted models from I-TASSER were checked for Ramachandran favored region and ERRAT score by **SAVES v6.0** (<https://saves.mbi.ucla.edu/>), a comprehensive toolkit for evaluation of the stereochemical parameters of the models [40]. The Ramachandran plot displays a model's overall geometry along with the verified score and outcome, including favored, allowed, and forbidden regions.

Models having lower Rama-favored regions were refined using **GalaxyRefine** (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). It is a web server for protein structure refinement that excels at enhancing the quality of local structures by refining loop or terminal sections through ab initio modeling [41].

Further evaluation of the three predicted tertiary structures for wild-type protein and its two variants was done using tools like the Swiss-model structure assessment (<https://swissmodel.expasy.org/>), ERRAT, and ProSA (Protein Structure Analysis) (<https://prosa.services.came.sbg.ac.at/prosa.php>) [42]. These servers forecasted the results for the critical metrics used to assess the modeled structures, including Mol-Probity, Ramachandra plot, QMEAN, Z scores, and ERRAT scores.

2.11. Structural alignment and comparison between the variants and wild-type protein

The variant structures were assessed using Pymol [43] and TM-align (<https://zhanggroup.org/TM-align/>) [44] to estimate root mean square deviation (RMSD) and template modeling scores (TM score),

respectively. RMSD stands for the amount of deviation between the native structure and the predicted model. For, identical structures, the RMSD value is 0, and it progressively increases as the dissimilarity between the two structures grows. So, a higher RMSD score means that the mutant structure is more deviated from the native structure [45]. TM score is opposite to RMSD ranging between 0 and 1. TM-score of 0.5–1.00 designates that the aligned proteins are in about the same fold. An exact match between superimposed wild-type and variant structures receives a score of 1. Hence, the backbone C alpha coordination discrepancy between wild-type and mutant forms increases with decreasing TM score [46].

2.12. Molecular docking

The wild type, as well as two variant structures (D280Y and D280N), were caused to undergo molecular docking with four ligands, including cholesteryl hemisuccinate, octyl beta-D-glucopyranoside, tetraethylene glycol, and urea to observe the impact of variation on the ligand binding property of the protein. The first three small molecules were reported as ligands for urea transporter UT-B1 on PDB (Protein Data bank) (<https://www.rcsb.org/>) since they were found associated with the X-ray Crystallography structure of Human Urea Channel *SLC14A1/UT1* [47]. Urea was selected since UT-B1 is a urea transporter and is thought to interact with the transporter protein while being transported [48]. The 3D structures of these ligands were obtained from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) [49]. PyRx-embedded AutoDock Vina tool was used for this docking purpose [50]. Utilizing AutoDock Vina wizard, blind docking was performed between the protein (macromolecule) and selected ligands by setting the grid box to maximum (Dimensions X: 50.0819 Å, Y: 52.2146 Å, Z: 66.2067 Å) to cover the entire protein structure and other parameters were kept at default state. Up to nine conformers were taken into account for each ligand throughout the docking procedure. The conformations exhibiting the most favorable (lowest) free binding energy were chosen for the analysis of the interactions between the target receptor and ligands using Discovery Studio Visualizer [51] and Pymol [43].

2.13. Molecular dynamics simulation

The GROMACS (version 2020.6) simulation software was used to run 100 ns (nanosecond) molecular dynamics (MD) simulations on the wild-type and variant models of the UT-B1 protein [52]. The force field utilized in the simulation to describe the macromolecular system was GROMOS96 43a1. The spc216 water model was used to create a water box with edges 0.5 nm from the protein surface. Following energy minimization, isothermal-isochoric (NVT) equilibration, and isobaric (NPT) equilibration of the system, a 100 ns molecular dynamic simulation was performed using periodic boundary conditions. The snapshot interval for assessing the trajectory data was set to 100 ps (picosecond).

The root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), and solvent accessible surface area (SASA) analyses were carried out following the completion of the simulation using the rmsd, rmsf, gyrate, and sasa modules built into the GROMACS software. Each of these studies' plots was created in RStudio using the ggplot2 package [53].

2.14. Analysis of non-coding SNPs

Some particular non-coding SNPs (ncSNPs) of the *SLC14A1* gene were filtered out from the ENSEMBL genome browser [17] and were subjected to the RegulomeDB database (<https://regulomedb.org/>) to investigate the regulatory influence of the non-coding variants. The database assigns a rank between 1 and 7 to each variant, along with a regulomeDB score based on the type and strength of the evidence supporting its functional effect [54].

The Genotype-Tissue Expression (GTEx) (<https://gtexportal.org>)

/home/), a research project that aims to create an atlas of gene expression and regulation in multiple human tissues, offers public access to data such as gene expression, QTLs, and histology images [55]. The ncSNPs screened through RegulomeDB were further checked using the GTEx portal for their effect on gene expression in various tissues.

PolymiRTS (Polymorphism in microRNAs and their TargetSites) (<https://compbio.uthsc.edu/miRSNP/>) database was used to find out ncSNPs that occur at the seed region and target site of miRNA and affect mRNA-based post-transcriptional regulation. Several sources, such as miRecords, TarBase, and miTarBase, were compiled with experimentally verified miRNA-target interactions from both low-throughput and high-throughput research [56].

3. Result

3.1. Retrieval of SNP data of *SLC14A1* gene

The SNP data for the *SLC14A1* gene was extracted from the ENSEMBL genome browser. A total number of 7977 variants was associated with the transcript ID ENST00000321925.9 of the *SLC14A1* gene, which was classified into various types such as missense variants (4.11%), 3'UTR variants (8.14%), 5'UTR variants (0.45%), intron variants (81.87%), synonymous variants (1.47%) and others (3.96%) (Fig. 2). Missense or non-synonymous SNPs (nsSNP) alter the amino acid sequence of a protein and tend to affect its function. Therefore, nsSNPs were targeted for this study.

3.2. Identification of the most frequent missense SNPs in the *SLC14A1* gene

Frequently occurring missense variants were filtered out from the variant table of the selected *SLC14A1* gene transcript (Supplementary Table 1). Two SNPs, rs2298720 (E44K) and rs1058396 (D280N) were identified as having a global MAF of 0.235 and 0.411, respectively. However, rs1058396 had two other variants, D280H and D280Y. Therefore, all four variants were selected for analysis of their functional impact.

3.3. Analysis of functional consequence of the selected SNPs

Twelve web-based tools, including SIFT, PolyPhen-2, CADD, REVEL, MetaLR, Mutation assessor, SNPs&GO, PhD SNP, PANTHER, MutPred2, ClinVar and SuSPect were employed for the evaluation of the functional impact of the four variants (D280N, D280Y, D280H and E44K) on the UT-B1 protein. Among them, D280Y was predicted deleterious or

disease-associated by maximum tools, whereas D280H and E44K were forecasted as deleterious by 4 and 3 tools, respectively. All twelve tools refer D280N variant as benign, having no impact on the protein function (Table 1). Hence D280Y variant was preferred for further analysis along with D280N, as D280N is the most commonly occurring variant in the global population among all three variants of rs1058396.

3.4. Determining the conservation level and surface accessibility

All the amino acid residues of the UT-B1 protein were represented using the ConSurf algorithm in terms of their structural and functional conservation levels. The ConSurf result revealed that the amino acid residue targeted in this study was an exposed residue and had a conservation score of 1. (Supplementary Fig. 1). Lower conservation score represents the D280 residue located in a variable region.

Again, an examination of surface accessibility using NetSurfP-3.0 predicted relative surface accessibility of all the amino acid residues of the protein, along with a representative RSA score. Asp (D) residue at 280 position that belongs to an alpha-helical structure is exposed and accessible with an RSA of 50% (Supplementary Fig. 2).

3.5. Effect on UT-B1 protein stability

The impact of the two variants (D280Y and D280N) on the stability of UT-B1 protein was evaluated by screening through two web-based tools: I-mutant 3.0 and MUpro. Since different algorithms are used by I-Mutant 3.0 and MUpro, the predicted result or assigned $\Delta\Delta G$ value slightly varied from each other. However, both tools predicted the two variants to decrease the stability of the protein as the $\Delta\Delta G$ is less than zero in each case. $\Delta\Delta G$ represents the free energy balance between the folded and unfolded states of the protein structure, and a negative score implies a reduction in the stability of the structure [57]. I-Mutant 3.0 predicted $\Delta\Delta G$ value was -1.36 kcal/mol for the D280Y variant and -1.30 kcal/mol for the D280N variant. Again, the $\Delta\Delta G$ value was predicted -0.695 kcal/mol for D280Y and -1.068 kcal/mol by MUpro.

3.6. Exploring the interaction profile with other proteins

Assessment of protein interaction by STRING v11.5 server anticipated that the UT-B1 protein encoded by the *SLC14A1* gene exhibited interconnection with ten other proteins, including proteins encoded by *SLC28A1*, *SLC4A1*, *AQP1*, *AQP2*, *AQP3*, *RHCE*, *RHBG*, *KEL*, *ACP1* and *ZNF134* genes (Fig. 3A). The proteins encoded by these genes are mainly membrane-embedded and are involved either in the transportation of ions, small molecules and water across the biological membrane or in

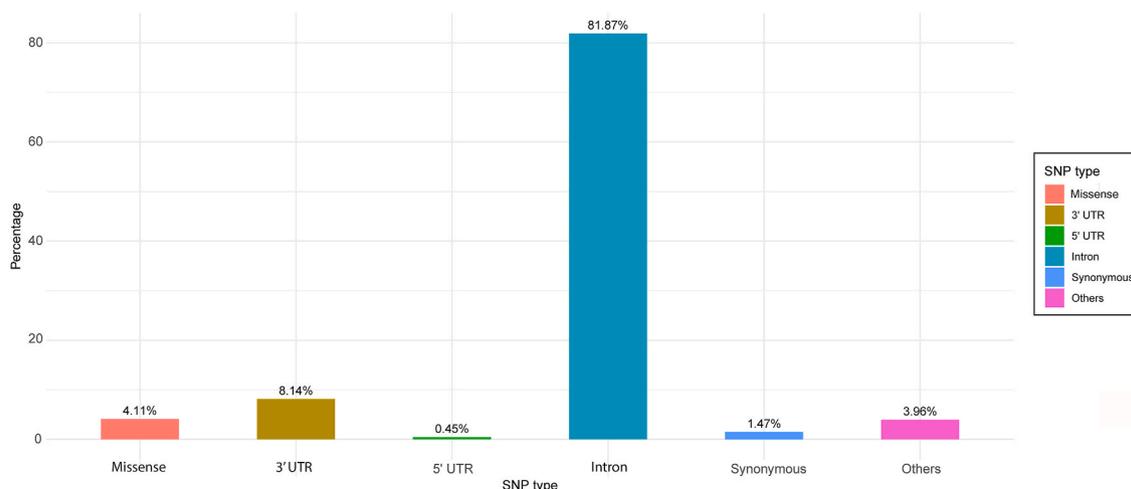


Fig. 2. Percentage of various SNP types of *SLC14A1* gene.

Table 1
Functional consequence prediction by twelve web-based tools.

Tools	rs1058396 (D280Y)		rs1058396 (D280N)		rs1058396 (D280H)		rs2298720 (E44K)	
	Score	Prediction	Score	Prediction	Score	Prediction	Score	Prediction
SIFT	0.01	Deleterious	0.65	Tolerated	0.02	Deleterious	0.02	Deleterious
PolyPhen	0.647	Possibly damaging	0.009	Benign	0.666	Possibly damaging	0.281	Benign
CADD	22	Likely benign	8	Likely benign	17	Likely benign	23	Likely benign
REVEL	0.212	Likely benign	0.061	Likely benign	0.173	Likely benign	0.158	Likely benign
MetaLR	0.169	Tolerated	0	Tolerated	0.168	Tolerated	0	Tolerated
Mutation assessor	0.712	Medium	0.187	Neutral	0.712	Medium	0.884	Medium
SNPs&GO	0.279	Neutral	0.111	Neutral	0.147	Neutral	0.194	Neutral
PhD SNP	0.618	Disease associated	0.329	Neutral	0.43	Neutral	0.581	Disease associated
PANTHER	0.435	Neutral	0.182	Neutral	0.326	Neutral	0.214	Neutral
MutPred2	0.698	Pathogenic	0.126	Likely benign	0.512	Likely Pathogenic	0.37	Likely benign
Clinvar	–	Not reported	–	Benign	–	Not reported	–	Not reported
SuSPect	32	Likely benign	23	Likely benign	29	Likely benign	12	Neutral

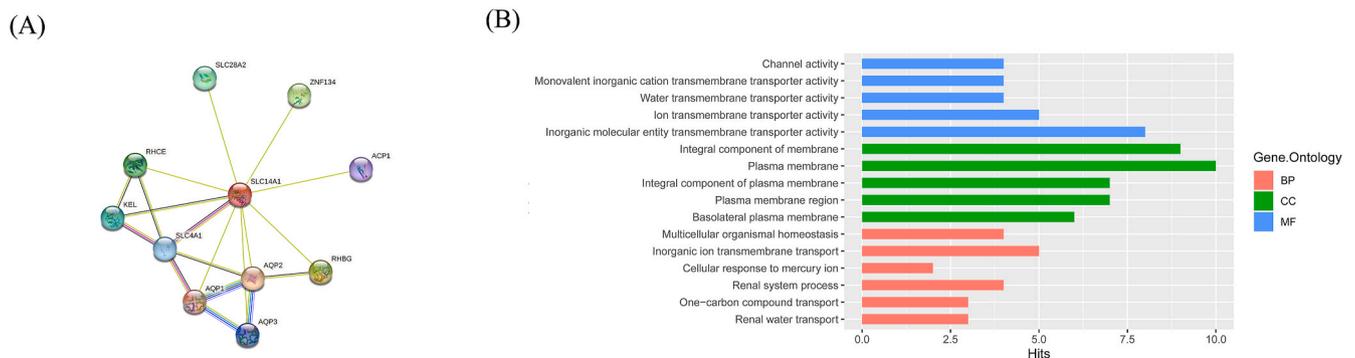


Fig. 3. (A) Interaction framework of *SLC14A1* with other cellular proteins predicted by STRING; (B) Significant gene ontology terms of *SLC14A1* (BP, Biological process; CC, Cellular components; MF, Molecular function).

blood group system. Moreover, gene ontology (GO) analysis characterized the appropriate biological processes, cellular components, and molecular functions linked with the target *SLC14A1* gene, which can be better identified and visualized in Fig. 3B.

3.7. Evaluation of the structural consequence

HOPE server discovered the underlying reasons for which variations in the amino acid sequence would possibly distort the protein structure and hamper its function. The server gave a clear view of how the wild-type residue was changed into the mutated residue in both variations

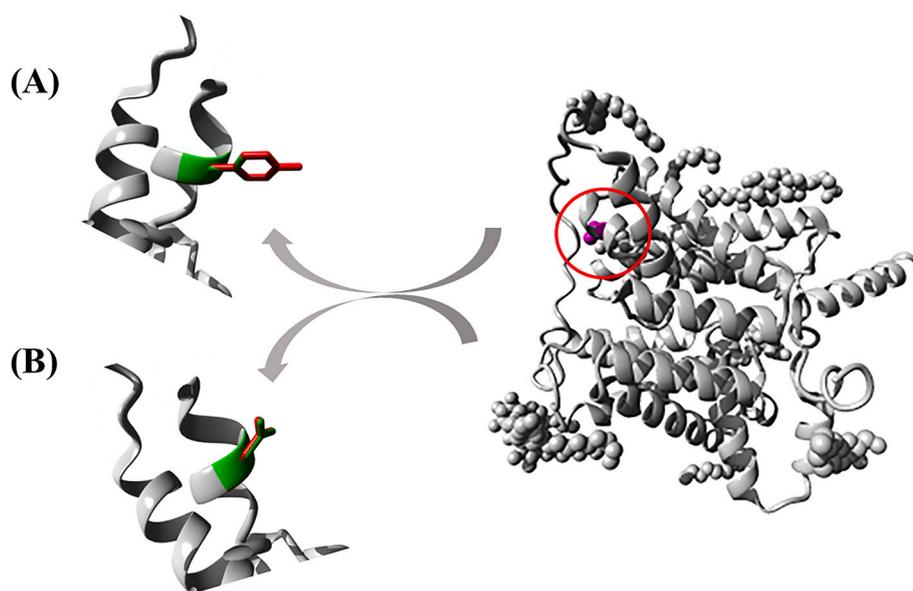


Fig. 4. Visualization of the native Aspartate residue alteration into variant Tyrosine (A) and Asparagine (B) through superimposition by Project HOPE. The image depicts the native residue in green, the variant residues in red, and the protein structure in grey. The encircled pink region is the site of alteration. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Fig. 4). In the case of the D280Y variant, the size, charge, and hydrophobicity were altered due to the mutation. The mutant residue Tyr (Y) was greater in size, neutral in charge compared to negatively charged Asp (D), and more hydrophobic in nature, while in the D280N variant, the Asn (N) residue only differed in charge from the wild-type Asp (D).

3.8. Differences in secondary structure

Relative secondary conformation, including helix, sheet, turn, coil, etc., was predicted by SOPMA for the wild-type UT-B1 protein as well as the mutant variants (D280Y and D280N). The prediction unveiled differences between the secondary structures of wild-type and mutant proteins. The wild-type UT-B1 protein was estimated to contain 46.79% alpha helix (182 residues), 13.79% extended strand (52 residues), 4.11% beta-turn (16 residues), and 35.73% random coil (139 residues). On the contrary, the D280Y variant protein had 49.61% alpha helix, 12.85% extended strand, 4.88% beta-turn, and 32.65% random coil, and D280N had 50.13% alpha helix, 12.34% extended strand, 4.37% beta-sheet and 33.16% random coil (Table 2).

3.9. 3D structure modeling and evaluation

I-TASSER provided the top five models for each of the native and mutant proteins. The best models having the highest C-score were refined using the GalaxyRefine server. The server again resulted in 5 refined models for corresponding proteins, which were then submitted to SAVESv.6 for quality check. Models with higher quality were chosen and further evaluated by PROCHECK, ERRAT, Swiss-model structure assessment, and ProSA (Fig. 5). Most of the residues (>90%) of the three predicted models fell in the Ramachandran favored region having very little (0.0%, 3.3% and 0.3%) in the disallowed region. The evaluation scores predicted by these tools are depicted in Table 3. The high ERRAT scores of 98.82%, 88.14%, and 92.93% for the wild-type, D280Y, and D280N variant proteins respectively indicated higher quality for non-bonded atomic interactions for three of the structures where the acceptable range is >50% [58]. Again, PROSA, a quality measure tool, assesses the divergence of the total energy of the structure with regard to an energy distribution produced from random conformations. The Z-scores below -6.00 predicted by ProSA for all three structures are indicative of the model's high quality as more negative Z-scores indicate better protein models [59]. Moreover, Molprobtity does a comprehensive examination of all-atom contacts, resulting in a quality score for the structure. The quality of the structure increases as the score gets closer to zero [60]. The "degree of nativeness" of a particular structure can be quantified using the QMEAN Z-score which compares the protein models to reference structures determined via X-ray crystallography. Low-quality models have QMEAN Z-scores below -4.0 [61]. The Mol-Probity scores of around 2 and QMEAN Z score of > -4.0 indicated overall good quality of the proteins.

The structure models for two variants, D280Y and D280N, were superimposed separately with the wild-type model and visualized in PyMol. Alignment in PyMol and TM-align estimated that the D280Y variant slightly deviated from the predicted native structure with a PyMol RMSD score of 0.458 and TM-align score of 0.90745 whereas the D280N deviation was of PyMol RMSD score of 0.395 and TM-align score 0.91084.

Table 2

Secondary conformation prediction of the wild-type and variant UT-B1 proteins using SOPMA.

Motif	Wild Type		D280Y		D280N	
	Residue	Percentage	Residue	Percentage	Residue	Percentage
Alpha helix	182	46.79%	193	49.61%	195	50.13%
Extended Strand	52	13.37%	50	12.85%	48	12.34%
Beta Turn	16	4.11%	19	4.88%	17	4.37%
Random Coil	139	35.73%	127	32.65%	129	33.16%

3.10. Ligand binding property analysis

The impact of single nucleotide polymorphism on the ligand binding feature of protein was determined by docking using the PyRx-embedded AutoDock Vina tool. When the wild-type and variant models were docked with four ligands (cholesteryl hemisuccinate (Y01), octyl beta-D-glucopyranoside (BOG), tetraethylene glycol (PG4), and urea), an increase in the binding free energy was observed for each ligand binding with the variant models compared to the wild-type protein. Lower binding energy indicates strong binding affinity and with increasing binding energy, the affinity towards the ligand becomes weaker [62].

Cholesteryl hemisuccinate (Y01) displayed the highest affinity among the tested compounds. Interactions between Y01 and the wild-type UT-B1 protein involve multiple amino acids leading to the binding energy of -9.1, which rose to -8.0 and -7.7 for D280Y and D280N variants respectively. Similarly, other ligands such as octyl beta-D-glucopyranoside (BOG), tetraethylene glycol (PG4), and urea exhibited distinct binding patterns and affinities, with the mutations generally reducing binding affinity (higher binding free energy) across all ligands (Table 4). Binding free energy for Protein-BOG interaction was -6.8 for UT-B1 (wild-type) and -6.4 for both variants. For Protein-PG4 binding, it was -4.4, -4.1, and -3.5 for native UT-B1, D280Y, and D280N variants respectively. Again Protein-Urea interaction had binding free energy of -3.8 for wild-type UT-B1 and -3.3 for both variants.

Following the protein-ligand docking, the complexes were visualized in Discovery Studio for their interaction investigation. Due to the introduction of single nucleotide variation in the native structure, alterations in the interacting residues, as well as a binding site for a particular ligand, were observed (Figs. 6 and 7). In comparison with the wild-type UT-B1, both of the variants exhibited either a different or decreasing number of amino acid residues involved in protein-ligand interaction, yielding varying degrees of hydrogen bonds and other bonds (Table 4). The change in bonding pattern and number ultimately caused the declining affinity of the variant proteins. The lower binding affinity (higher binding free energy) of the variant proteins underscores the importance of residue 280 in UT-B1 protein-ligand interactions and provides valuable insights into the molecular mechanisms underlying ligand recognition and binding.

3.11. Molecular dynamic simulation

The physical basis of the structure and biological function of macromolecules can be better understood with the help of molecular dynamics (MD) simulations. For this reason, MD simulation was performed for the wild UT-B1 protein and its variants (D280Y and D280N) and compared the outcome.

The Root Mean Square Deviation (RMSD) was calculated to assess the systems' stability. A higher RMSD value indicates that the protein is more unstable [63]. After 25 ns, the RMSD for the wild-type UT-B1 stabilized fairly soon. Its value has stayed constant at 0.35–0.4 nm since then. In contrast, the RMSD values for both UT-B1 variants (D280Y and D280N) were higher (around 0.45 nm and above) reflecting the instability of the variant protein structure compared to the wild-type protein (Fig. 8A).

The regional flexibility of the protein was evaluated using the Room Mean Square Fluctuation (RMSF) method. Regions with higher RMSF

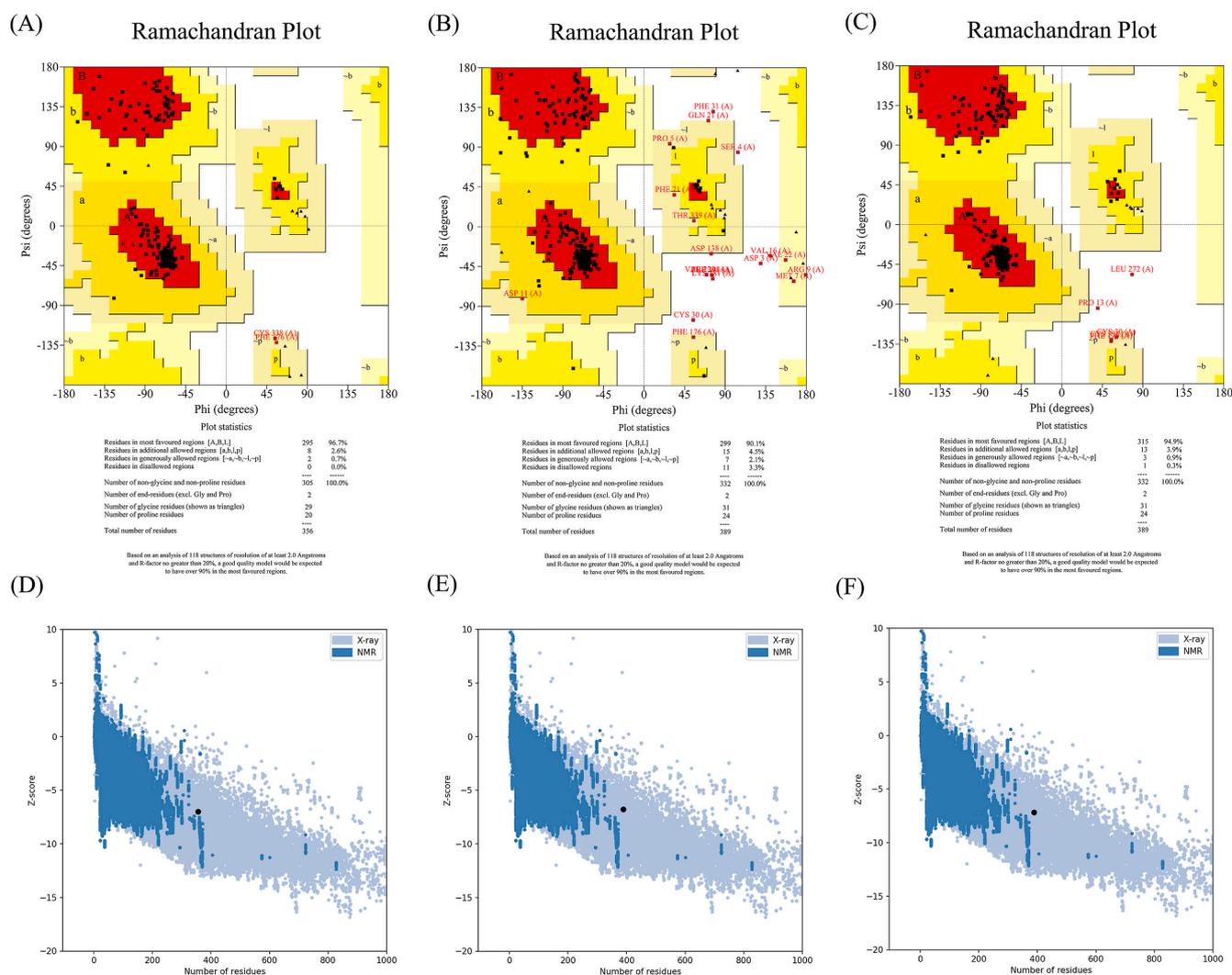


Fig. 5. Three-dimensional model assessment of the predicted protein structures. Ramachandran plot of the wild type (A) as well as its D280Y (B) and D280N (C) variants displaying Rama-favored region in red, the additionally allowed region in yellow, the generously allowed region in faint yellow, and the disallowed region in white color. The ProSA plot is showing the Z-score (black point) of the three proteins: wild-type (D), D280Y variant (E), and D280N (F). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3

Evaluation scoring of the refined models by ERRAT, PROCHEK, ProSA, and Swiss model assessment tools.

3D Model	Ramachandran favored region	ERRAT score	MolProbity score	ProSA Z score	Qmean Z score
Wild Type	96.70%	98.8166	1.12	-6.99	-1.75
D280Y	90.10%	88.1402	2.32	-6.76	-3.79
D280N	94.90%	92.9348	1.44	-7.19	-2.41

values typically correspond to flexible regions within secondary structures such as coils or loops, whereas regions with lower RMSF values generally align with more rigid secondary structures like alpha helices or beta sheets. In comparison to the natural protein, the variations of UT-B1 were less flexible around the 250th and 300–350th residues as their RMSF value was relatively lower [63] (Fig. 8B).

The degree of compactness was measured using the radius of gyration. Protein folding is stable when the radius of gyration is generally constant. The radius of gyration fluctuation indicates protein unfolding [63]. The wild-type UT-B1 reached a stable value immediately after 20 ns indicating stable folding or compactness of the protein. On the other hand, the higher and fluctuated radius of gyration represented the

unfolded structure of the variants (Fig. 8C).

SASA was employed in MD simulations to forecast the protein hydrophobic core stability. The likelihood of protein instability due to solvent accessibility increases with increasing SASA score [63]. When compared to the wild UT-B1, the variations showed higher SASA levels (Fig. 8D).

3.12. Non-coding SNPs of SLC14A1 gene analysis

A total number of 179 non-coding SNPs of SLC14A1 transcript ID ENST00000321925.9 was filtered from the ENSEMBL genome browser that had a global MAF range of 0.05–0.5. These SNPs include ten 3'UTR variants, two 5'UTR variants, and 167 intron variants (Supplementary Table 2).

3.12.1. Analysis by RegulomeDB and GTEx portal

All of these variants were submitted to RegulomeDB for the analysis of their effect on the regulatory mechanism of the gene. Out of 179 variants, about 103 variants were ranked by the server, which included 2c (1 SNP), 3a (8 SNPs), 4 (25 SNPs), 5 (45 SNPs), 6 (7 SNPs), and 7 (17 SNPs) (Fig. 9). The likelihood that a mutation is located in a functional region increases with lower ranks and with higher ranks, there is less or

Table 4
Ligand interactions and binding affinity of the predicted models for the four selected ligands.

Ligand	UT-B1 Protein	Binding Affinity	Interacting Amino Acids (AA)	Conventional H-bonded AA	Carbon H-bonded AA
Cholesteryl Hemisuccinate	Wild Type	-9.1	K43, R64, L116, S119, L121, P174, V175, L364	R64	K43, S119
	D280Y	-8	L132, F136, V203, I204, I206, T208, L285, F288, C338	T208	-
	D280N	-7.7	A46, Q68, N73, T368	Q68, N73, T368	-
Octyl beta-D-glucopyranoside	Wild Type	-6.8	Q68, N73, L116, L121, Y122, P174, V175, L364	N73	-
	D280Y	-6.4	Q68, F71, N73, L121, P174, V175, L364, V367	N73	V367
	D280N	-6.4	L121, S169, P174, L364, V367	-	S169
Tetraethylene Glycol	Wild Type	-4.4	W286, N289, A230, G323, A337	W286, N289, A337	A320, G323
	D280Y	-4.1	Q232, Y280, N289, G323	Q232, N289, G323	Y280
	D280N	-3.5	N73, S119, V367	N73, V367	S119
Urea	Wild Type	-3.8	S137, F198	S137, F198	-
	D280Y	-3.3	L285, N289	L285, N289	-
	D280N	-3.3	T39, D41, M42, K43, Y369, E372	T39, D41, M42, K43, Y369, E372	-

no evidence of the transcription factor binding [54].

Nine SNPs having a rank of 2c and 3a provided by RegulomeDB were further analyzed by the GTEx portal for their impact on tissue-specific gene expression and eQTL (Expression quantitative trait loci). All of the nine SNPs had multiple tissue eQTL, including skeletal muscle, esophagus, nerve, lung, thyroid, pituitary, heart, adipose tissue, and others. Among these, the two tissue sites with the highest impact on *SLC14A1* gene expression for each SNP were listed in Table 5, along with the P-value and normalized effect size (NES).

3.12.2. Analysis by PolymiRTS

A total of 179 filtered ncSNPs were screened through the PolymiRTS tool to determine whether or not any of the selected ncSNPs were in the miRNA seed region or miRNA binding site. It was found that two SNPs, rs1135979 (T > C) and rs1135980 (C > T), affect the miRNA binding site with a conservation score of 2. Both SNPs had known ancestral alleles, and the functional class 'C' denotes that they both create new miRNA sites without any support of experimental data denoted by N (Table 6). The conservation score represents the presence of the miRNA site in other vertebrate genomes, in addition to the genome being analyzed. According to the PolymiRTS database, the miRNA sites created by these two ncSNPs (rs1135979 and rs1135980) were also present in Rhesus and Chim along with the human genome.

4. Discussion

Urea transporter type B (UT-B), encoded by the *SLC14A1* gene, is implicated as a significant protein involved in the urinary concentrating mechanism and maintenance of cellular homeostasis. GWAS studies identified *SLC14A1* as a bladder cancer susceptibility gene and reported several associated loci [2–4]. Deficiency of UT-B1 leads to high urea and NO concentration in urothelial cells, suggesting a possible mechanism of bladder carcinogenesis [5,7,8]. Reportedly, genetic variations of the *SLC14A1* gene seem to impact the abundance of UT-B1 protein in the plasma membrane of urothelial cells and other tissues [64]. This is highly likely to affect the normal functioning of cells. Consequently, they hold significant implications for further investigation of *SLC14A1* variants to reveal the involvement of UT-B1 in human diseases, particularly different types of malignancies.

Screening through dbSNP and ENSEMBL genome browser identified rs1058396 as the most common genetic missense variant of the *SLC14A1* gene with a global minor allele frequency of 41.1%. rs1058396 is defined as ancestral allele G and altered allele A (D280N), C (D280H) and T (D280Y). Characterization of these variants for their probable structural and functional effect on UT-B1 protein involved a number of web-based tools and predicted D280Y (rs1058396; T allele) as deleterious by maximum prediction tools, while D280N (rs1058396; A allele), the common variant of the same SNP ID, as neutral. Since the level of

accuracy and prediction methods or algorithms of these tools differ from each other, and their web servers are being updated from time to time, the prediction results can vary from each other or the previously attained result using the same tool for the same SNP [65].

Keeping in mind the significant role of the conserved region for a protein's structure and function [66], an evolutionary conservation profile was generated for UT-B1 protein using ConSurf which revealed the D280 residue was located within the variable region while being exposed on the protein surface. The RSA (relative surface accessibility) of 50% and disease association level of 2% assigned by NetSurfP-3.0 to the D280 position indicated that the D280Y and D280N variants were likely to result in detrimental consequences. Exposed residues are more likely to be involved in interaction with other components within the cellular environment and alteration of solvent-accessible residues can cause significant changes in protein stability and interaction [57]. I-Mutant and MUpro servers calculated Gibbs free energy ($\Delta\Delta G$) of less than zero for both variants implying a reduction in the stability of the UT-B1 variant structures [57].

Moreover, the HOPE server analyzed the nsSNP to identify the general physiological and functional changes caused by the point mutation. In the case of the D280Y variant, the mutant residue (Tyr) resulted in increased size and hydrophobicity compared to wild-type Asp, while D280N causes no change in these properties. However, the negative charge of Asp is lost in both cases as Tyr and Asn are both uncharged amino acids. These changes unveil a possible effector mechanism of the variants because of the significant influence of amino acid characteristics in protein folding, stability, spatiotemporal dynamics of protein-protein interactions, and function of the protein, but they are not absolute and occasionally may be deceptive [57,67].

Besides UT-B1 protein, *SLC14A1* also encodes JK antigen (Kidd glycoprotein) and serves as the basis of the Kidd blood group system. The primary allelic variants, JkA (D280) and JkB (N280) result from a substitution at amino acid position 280, which constitutes a significant focus of the present study [68]. To have a deeper understanding of the importance of the *SLC14A1* gene, its interacting network was forecasted by the STRING server. *SLC14A1* (UT-B1) was found to interact with ten other proteins involved in some significant cellular systems, such as transportation across the cell membrane (Solute carrier proteins SLC28A1 and SLC4A1; Aquaporins AQP1, AQP2, and AQP3) and blood group system (RHCE, RHBG, and KEL). Gene ontology characterized the molecular function, cellular components, and biological processes of UT-B1 protein, supporting *SLC14A1*'s role in these important physiological processes. Interaction of membrane-embedded UT-B1 protein with other proteins may play a crucial role in maintaining their normal physiological functions and transporter activity. This network disruption or reduced protein stability due to D280Y and D280N variations can cause impairment in cellular homeostasis leading to cell death or other complications.

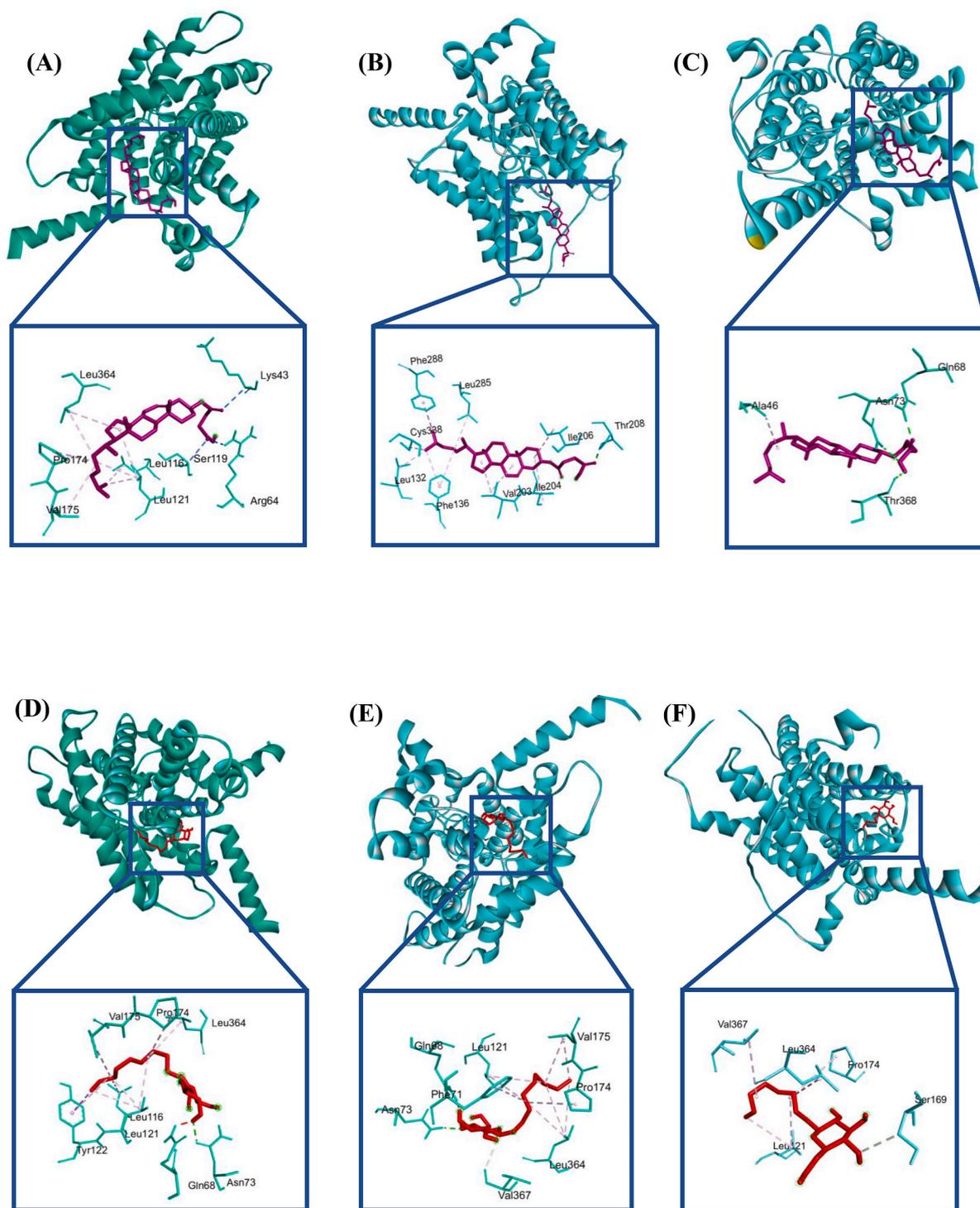


Fig. 6. Visualization of the protein-ligand complexes exploring the ligand binding sites, amino acid residues involved in the interaction, and their positions. (A) Wild-type + Y01, (B) D280Y + Y01, (C) D280N + Y01, (D) Wild-type + BOG, (E) D280Y + BOG, (F) D280N + BOG, (G) Wild-type + PG4, (H) D280Y + PG4, (I) D280N + PG4, (J) Wild-type + Urea, (K) D280Y + urea and (L) D280N + urea.

Furthermore, the structural effect of D280Y and D280N variants on the UT-B1 function can also be observed by their secondary and tertiary structure. Assessment via SOPMA unveiled changes in the relative secondary conformations between the native and variant structures. Compared to the wild-type structure, the variants exhibited a rise in alpha helix and beta-turn structures, accompanied by a decrease in the extended strand and random coil structures. These structural modifications suggest a potential influence of the variants at the D280 position on the protein's stability, flexibility, and functional and binding properties.

Further analysis is typically necessary to understand the specific effects of these changes on the protein's structure-function relationship.

The tertiary structure of the human urea channel *SLC14A1* (UT-B1 protein) available in PDB is incomplete, as some residues are missing in the structure. However, the variant structures needed to be predicted. When comparing the predicted variant structure to the empirically confirmed wild-type structure, there might be more divergence, which could possibly result in less accuracy than comparing two predicted structures. That's why the protein structure was modeled, instead of

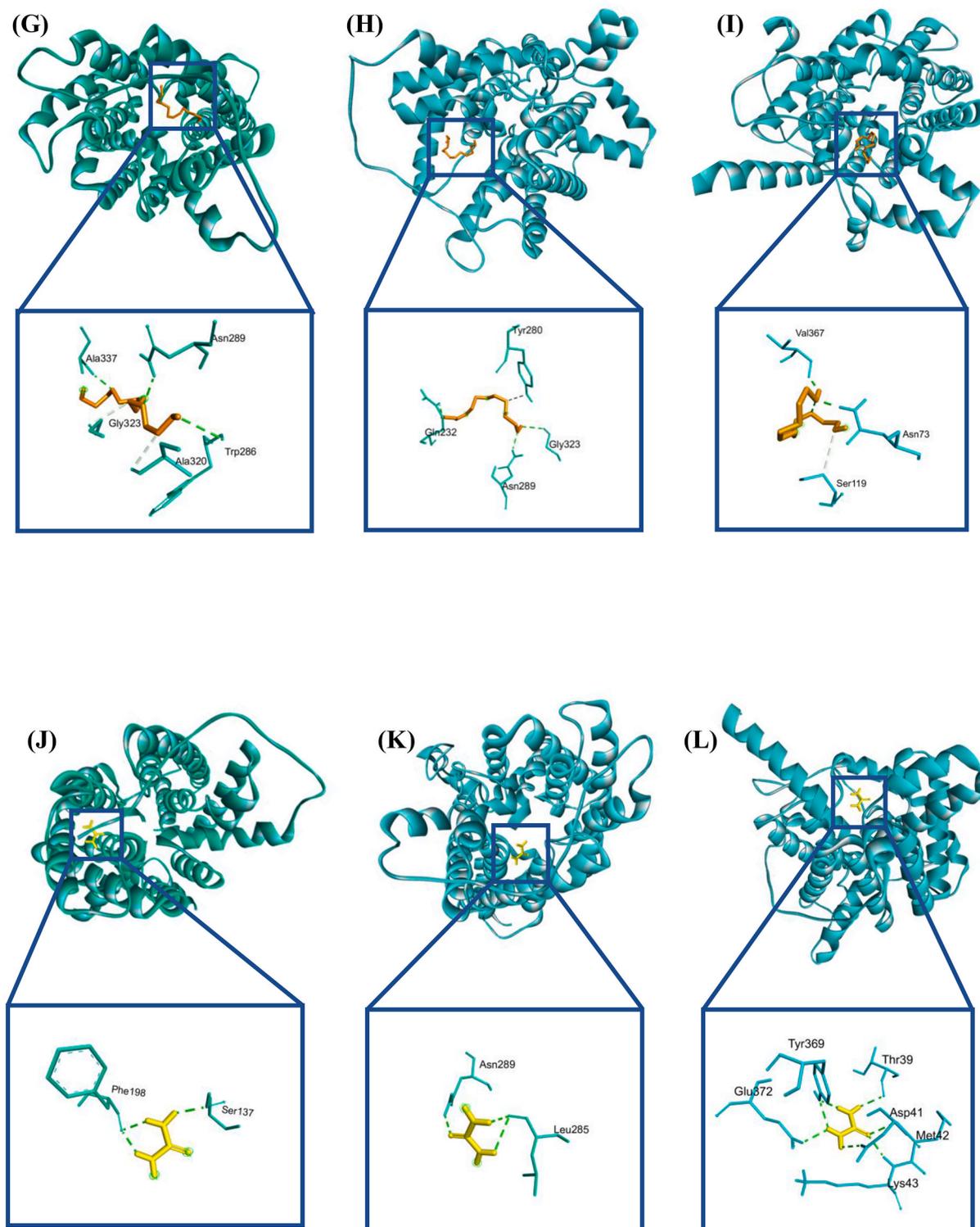


Fig. 7. Visualization of the protein-ligand complexes exploring the ligand binding sites, amino acid residues involved in the interaction, and their positions. (A) Wild-type + Y01, (B) D280Y + Y01, (C) D280N + Y01, (D) Wild-type + BOG, (E) D280Y + BOG, (F) D280N + BOG, (G) Wild-type + PG4, (H) D280Y + PG4, (I) D280N + PG4, (J) Wild-type + Urea, (K) D280Y + urea and (L) D280N + urea.

using the one from PDB. Nevertheless, several online tools are available such as I-TASSER [39] and AlphaFold [69] for 3D structural modeling. According to the assessment by the TM-align server (<https://zhanggrouop.org/TM-align/>), the I-TASSER generated structure was slightly more similar (TM value 0.99873 and RMSD value 0.24) to the established structure than the one from the Alpha fold (TM-align value 0.99853 and RMSD value 0.26). Therefore, I-TASSER was considered in

this study which was also employed in previous research to forecast the three-dimensional configuration of proteins [12,15].

The success of in silico research relies heavily on the reliability of 3D projected models. In some cases, the protein structure predicted from its sequence through template-based modeling may have less accurate side chains even with high sequence identity [70]. Therefore, energy minimization of the predicted models and fixing local and global

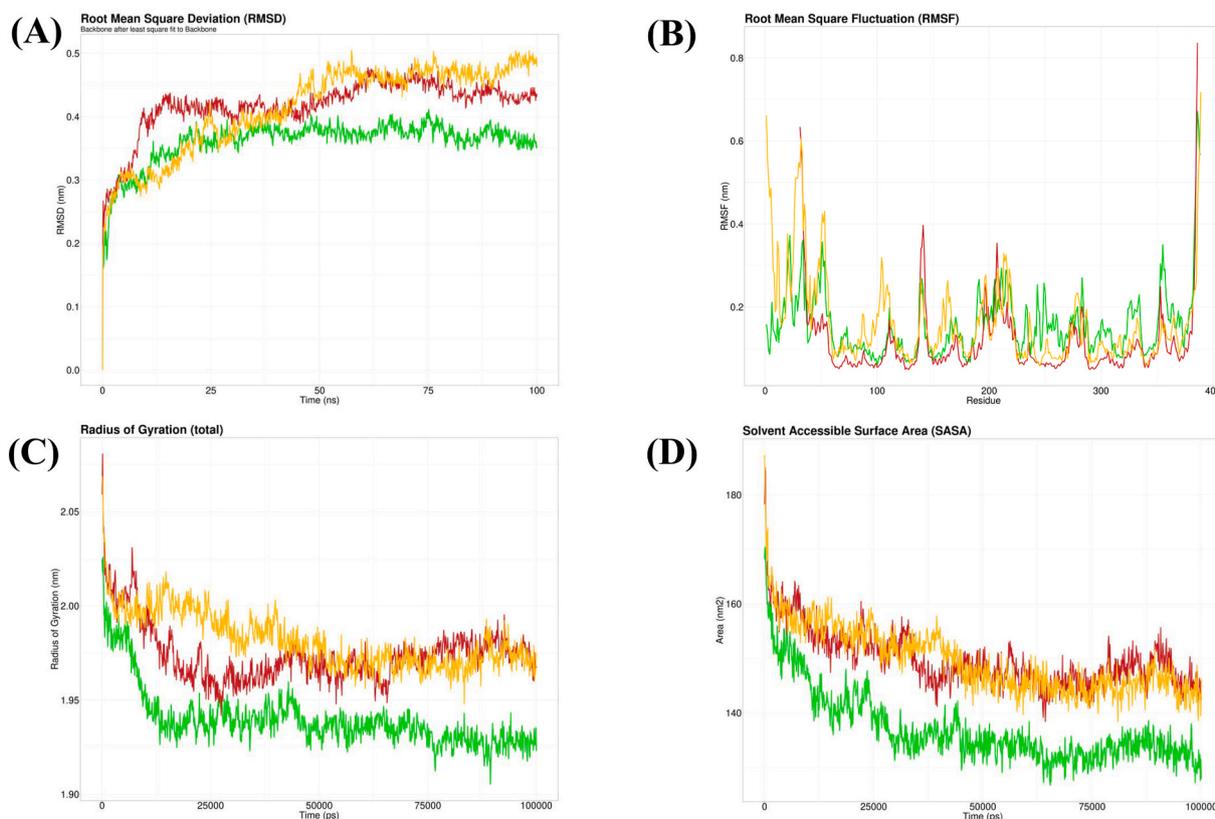


Fig. 8. RMSD (A), RMSF (B), Radius of gyration (C), and SASA (D) analysis of wild type UT-B1 (Green), D280Y variant UT-B1 (Red) and D280N UT-B1 (Yellow) proteins following dynamic simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

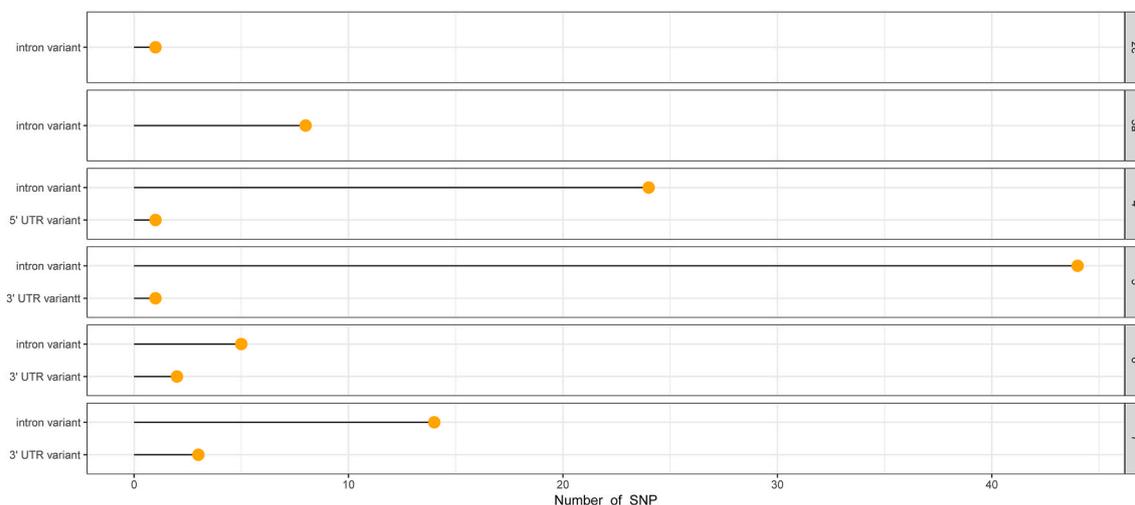


Fig. 9. Lollipop plot disclosing the number of various types of ncSNPs, categorized into different RegulomeDB ranks.

inaccuracies to obtain a more precise structure requires refining the 3D models, which was achieved by GalaxyRefine [41]. The quality of the refined protein structures was assessed to ensure high-quality structure to proceed with further analysis such as molecular docking which requires a high-resolution structure for more accuracy [71]. As the preference for a good quality model is about 90% residues in the most favored region of the Ramachandran plot [72], the 3D models of the UT-B1 proteins (>90% Rama favored region for wild-type and two variants) were of good quality. The overall quality of the refined models was additionally checked based on the ERRAT score (>85%),

Molprobrity score (around 2), Prosa Z score (below -6), and Qmean Z score (above -4) generated by ERRAT, ProSA and Swiss model assessment, all of which referred to the models as of overall good quality.

Structural alignment of the native protein and variant structures using TM-align and PyMOL estimated structural deviation based on the TM-score and RMSD values consecutively, though it was very insignificant. TM-score of 0.90745 for D280Y and 0.91084 for D280N variants indicated that they are not identical, but are in about the same fold with the wild-type protein [46]. On the contrary, the RMSD score is 0 for identical proteins and an increased score of 0.458 (D280Y) and 0.395

Table 5Top two tissue eQTL for the *SCL14A1* gene variants of Regulome rank 2c and 3a analyzed by GTEx portal.

Variant Id	ncSNP	P-Value	NES	Tissue
chr18_45726828_G_A_b38	rs7234310	7.3E-11 0.00000067	-0.24 -0.23	Muscle - Skeletal Esophagus - Muscularis
chr18_45727107_T_C_b38	rs10432193	7.3E-11 0.00000067	-0.24 -0.23	Muscle - Skeletal Esophagus - Muscularis
chr18_45728233_G_T_b38	rs9304322	0.00000067 0.0000016	-0.23 -0.17	Esophagus - Muscularis Nerve - Tibial
chr18_45728845_A_T_b38	rs2170974	7.3E-11 0.00000067	-0.24 -0.23	Muscle - Skeletal Esophagus - Muscularis
chr18_45728924_C_G_b38	rs9967412	7.3E-11 0.00000067	-0.24 -0.23	Muscle - Skeletal Esophagus - Muscularis
chr18_45731518_G_A_b38	rs10460035	0.00000015 0.000024	0.41 0.34	Adipose - Visceral (Omentum) Adipose - Subcutaneous
chr18_45731731_T_C_b38	rs8099449	1.2E-10 0.00000085	-0.23 -0.23	Muscle - Skeletal Esophagus - Muscularis
chr18_45741452_T_C_b38	rs12454680	1E-10 0.00000057	-0.23 -0.18	Muscle - Skeletal Lung
chr18_45744907_A_G_b38	rs568418	0.00000047 0.000014	0.42 0.32	Adipose - Visceral (Omentum) Lung

Table 6

Non-coding SNPs that affect the miRNA-based gene regulation detected by PolymiRTS.

ncSNP	Ancestral Allele	Altered Allele	miRSite	Function class	Conservation score	Experiment support
rs1135979	T	C	agctaCCTGGGAg	C	2	N
rs1135980	C	T	cCTGGGTGAcaag	C	2	N

(D280N) indicated slight deviation from the native structure [45].

As structural alteration has a profound influence on a protein's interaction feature, protein-ligand docking assessed how the variants altered the UT-B1 protein's molecular interaction. It has been demonstrated in a few studies that mutant protein has an altered binding affinity score than the wild-type protein. In most cases, a larger negative binding energy indicates a more favorable ligand orientation within the binding site [15,16]. When docked with ligands (cholesteryl hemisuccinate, octyl beta-D-glucopyranoside, tetraethylene glycol, and urea) using PyRx embedded AutoDock Vina tool, both the D280Y and D280N variants exhibited increased free binding energy for all four ligands, implicating weaker binding affinity. This may be due to deviated ligand orientation due to amino acid alteration and can lead to the decreased functional activity of the UT-B1 transporter protein and increased risk of disease pathogenesis. The interacting amino acid residues and their bonding patterns were also changed as a result of variations in the structure.

The behaviors of proteins and other biomolecules can be mimicked with atomic precision and with high temporal resolution using molecular dynamic (MD) simulations. The primary aim of this study was to evaluate intrinsic structural disparities between variant and wild-type proteins, thus concentrating solely on unbound proteins facilitated a direct contrast of inherent structural dynamics, free from the interference of ligand binding. This approach was also used in earlier studies to analyze the alteration of the dynamic nature of proteins owing to missense SNPs [16,63]. It has been proven useful in interpreting the functioning mechanisms of proteins and other biomolecules, as well as discovering the structural foundation for disease [73]. Through MD simulation (100 ns), four parameters, such as RMSD, RMSF, the radius of gyration (Rg), and SASA, were assessed to observe the physical basis of the structure-to-function relationship of variant proteins compared to the wild-type protein. Increasing RMSD scores for D280Y and D280N variants indicated protein instability, with D280N showing the highest instability. Reductions in RMSF values in specific regions suggest decreased flexibility in altered structures which justified the increased alpha helix and decreased coils motifs in the variant structures. Proteins require flexibility for proper function and interaction with other proteins or molecules [74] and alterations in flexibility can compromise

this. Fluctuations in the radius of gyration resemble unfolding in mutant structures, contrasting with the stable folding of the wild type. Destabilization of hydrophobic cores increases solvent accessibility, as measured by SASA. Molecular dynamics simulations reveal overall destabilization by mutations, likely disrupting protein function.

Along with the missense variants, several non-coding variants of the *SLC14A1* gene play a critical role in the expression of a gene at the transcriptional level or post-transcriptionally. To date, a plethora of approaches that center on regulatory areas of the genome have been created to focus on to emphasize disease-associated non-coding variations [75]. Screening through RegulomeDB website revealed nine non-coding SNPs (rs7234310, rs10432193, rs9304322, rs2170974, rs9967412, rs10460035, rs8099449, rs12454680, and rs568418) which are more likely to occur at the regulatory site, modulate transcription factor (TF) binding and might be associated with altered *SLC14A1* gene expression in different tissues [76]. GTEx portal showed the tissue locations where the *SLC14A1* gene expression might be significantly altered due to these specific polymorphisms. Two such tissue sites having the highest effect on *SLC14A1* expression were reported in this study which will support future investigations studying these genetic alterations and gene expression analysis of *SLC14A1* gene in those tissues.

Gene expression is also influenced by a different type of post-transcriptional control which is accomplished through the binding of miRNA to the 3' UTR region of the targeted mRNA [77]. With the use of the PolymiRTS server, two SNPs (rs1135979 and rs1135980) were discovered, both of which produce new seed regions. Seed areas are evolutionarily preserved and essential for mRNA complementarity which means miRNAs that were shown to have a connection with these two SNPs will lead to suppression of *SLC14A1* mRNA. In the recent past, numerous instances of these functional miRNA-binding site SNPs have been recognized for their potential use as cancer biomarkers [78].

Altogether, the data from several in silico analyses, molecular docking, and molecular dynamics simulations point to the negative consequences of D280Y and D280N variations of UT-B1 transporter protein. Protein instability and reduced flexibility leading to altered interaction properties implicated by the variants reveal the molecular mechanism of UT-B1 dysfunction in carcinogenesis [3] and other health

complications [8,11]. Additionally, several non-coding SNPs related to regulatory effect and miRNA-based gene suppression have been identified which needs further experimental validation.

4.1. Strengths and limitations

The strength of this study lies in the extensive investigations of these SNPs using multiple computational techniques and analyses that work based on different algorithms. Their collective interpretation provided a thorough understanding of the implications of these single nucleotide polymorphisms on disease susceptibility and dysfunctionality of protein. Therefore, this study can provide strong background support to further laboratory-based research such as mRNA level and protein level expression profiling of UT-B1 protein and functional assay to validate the association of UT-B1 protein variations with various diseases, specifically urothelial carcinoma.

However, there are certain limitations of our study. This study only addressed germline missense and non-coding (intronic, 3'UTR, and 5'UTR) variants of the *SLC14A1* gene (UT-B1 protein), solely depending on various computational tools and algorithms without any laboratory-based experiments. Again, additional mutations such as somatic SNVs and indels, previously linked to illnesses, are not examined in this in-silico analysis. For example, an in-frame deletion of exon 4 in UT-B1 has been identified as a significant factor in bladder cancer [79]. As understanding the relationship between a molecular and disease phenotype is very intricate, a thorough biological analysis at the molecular level is crucial for validating and reinforcing the study.

5. Conclusion

The D280N and D280Y variants of rs1058396 of the *SLC14A1* gene were characterized in this study by deciphering its functional implications, structural alterations, molecular interactions, dynamics features, and other attributes using an exhaustive in silico approach. This extensive analysis and characterization of rs1058396 can be helpful in the continuation of research pertaining to UT-B1 protein in clinical samples. It opens the door to the prospect of investigating the role of this SNP in disease pathogenesis and establishing a molecular biomarker for disease diagnosis and personalized treatment opportunities.

Funding

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

Acknowledgement

We acknowledge University of Dhaka for covering the article processing charge (APC) for this publication.

CRediT authorship contribution statement

Tamanna Sultana: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Sadia Islam Mou:** Visualization, Validation, Methodology, Investigation. **Dipankor Chatterjee:** Visualization, Validation, Investigation, Data curation. **Md. Omar Faruk:** Writing – review & editing. **Md. Ismail Hosen:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

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

Data availability

All data are available with the manuscript

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbrep.2024.101703>.

References

- [1] C.P. Smith, G. Rousselet, Facilitative urea transporters, *J. Membr. Biol.* 183 (1) (Sep 2001) 1–14, <https://doi.org/10.1007/s00232-001-0048-7>.
- [2] M. Garcia-Closas, et al., A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3, *Hum. Mol. Genet.* 20 (21) (Nov 2011) 4282–4289, <https://doi.org/10.1093/hmg/ddr342>.
- [3] V. Singh, P.K. Jaiswal, R.D. Mittal, Replicative study of GWAS TP63C/T, TERTC/T, and SLC14A1C/T with susceptibility to bladder cancer in North Indians: Equal contribution, *Urol. Oncol.: Seminars and Original Investigations* 32 (8) (Nov 2014) 1209–1214, <https://doi.org/10.1016/j.urolonc.2014.05.013>.
- [4] T. Rafnar, et al., European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene, *Hum. Mol. Genet.* 20 (21) (Nov 2011) 4268–4281, <https://doi.org/10.1093/hmg/ddr303>.
- [5] B. Yang, L. Bankir, A. Gillespie, C.J. Epstein, A.S. Verkman, Urea-selective concentrating defect in transgenic mice lacking urea transporter UT-B, *J. Biol. Chem.* 277 (12) (Mar 2002) 10633–10637, <https://doi.org/10.1074/jbc.M200207200>.
- [6] C. Walpole, A. Farrell, A. McGrane, G.S. Stewart, Expression and localization of a UT-B urea transporter in the human bladder, *Am. J. Physiol. Ren. Physiol.* 307 (9) (Nov 2014) F1088–F1094, <https://doi.org/10.1152/ajprenal.00284.2014>.
- [7] Z. Dong, et al., Urea transporter UT-B deletion induces DNA damage and apoptosis in mouse bladder urothelium, *PLoS One* 8 (10) (Oct 2013) e76952, <https://doi.org/10.1371/journal.pone.0076952>.
- [8] B. Yang, X. Li, L. Guo, Y. Meng, Z. Dong, X. Zhao, Extrarenal phenotypes of the UT-B knockout mouse, *Subcell. Biochem.* 73 (2014) 153–164, https://doi.org/10.1007/978-94-017-9343-8_10.
- [9] R. Hou, X. Kong, B. Yang, Y. Xie, G. Chen, SLC14A1: a novel target for human urothelial cancer, *Clin. Transl. Oncol.* 19 (12) (Dec 2017) 1438–1446, <https://doi.org/10.1007/s12094-017-1693-3>.
- [10] R.T. Timmer, et al., Localization of the urea transporter UT-B protein in human and rat erythrocytes and tissues, *Am. J. Physiol. Cell Physiol.* 281 (4) (Oct 2001) C1318–C1325, <https://doi.org/10.1152/ajpcell.2001.281.4.C1318>.
- [11] L. Yu, et al., Physiological functions of urea transporter B, *Pflügers Archiv* 471 (11–12) (Dec 2019) 1359–1368, <https://doi.org/10.1007/s00424-019-02323-x>.
- [12] E. Emadi, F. Akhondi, S.M. Kalantar, M. Emadi-Baygi, Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences, *BMC Genet.* 21 (1) (Dec 2020) 94, <https://doi.org/10.1186/s12863-020-00890-y>.
- [13] P.D. Stenson, et al., The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting, *Hum. Genet.* 139 (10) (Oct 2020) 1197–1207, <https://doi.org/10.1007/s00439-020-02199-3>.
- [14] R. Bhattacharya, P.W. Rose, S.K. Burley, A. Prlić, Impact of genetic variation on three dimensional structure and function of proteins, *PLoS One* 12 (3) (Mar 2017) e0171355, <https://doi.org/10.1371/journal.pone.0171355>.
- [15] Md S. Hossain, A.S. Roy, Md S. Islam, In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions, *Sci. Rep.* 10 (1) (Sep 2020) 14542, <https://doi.org/10.1038/s41598-020-71457-1>.
- [16] G.R.C. Pereira, A.N.R. Da Silva, S.S. Do Nascimento, J.F. De Mesquita, In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants, *J. Cell. Biochem.* 120 (3) (Mar 2019) 3583–3598, <https://doi.org/10.1002/jcb.27636>.
- [17] F.J. Martin, et al., Ensembl 2023, *Nucleic Acids Res.* 51 (D1) (Jan 2023) D933–D941, <https://doi.org/10.1093/nar/gkac958>.
- [18] A. Bateman, et al., UniProt: the universal protein knowledgebase in 2023, *Nucleic Acids Res.* 51 (D1) (Jan 2023) D523–D531, <https://doi.org/10.1093/nar/gkac1052>.
- [19] N.-L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, P.C. Ng, SIFT web server: predicting effects of amino acid substitutions on proteins, *Nucleic Acids Res.* 40 (W1) (Jul 2012) W452–W457, <https://doi.org/10.1093/nar/gks539>.
- [20] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, *Curr Protoc Hum Genet* 76 (1) (Jan 2013), <https://doi.org/10.1002/0471142905.hg0720s76>.
- [21] P. Rentzsch, D. Witten, G.M. Cooper, J. Shendure, M. Kircher, CADD: predicting the deleteriousness of variants throughout the human genome, *Nucleic Acids Res.* 47 (D1) (2019), <https://doi.org/10.1093/nar/gky1016>.
- [22] N.M. Ioannidis, et al., REVEL: an Ensemble method for predicting the pathogenicity of Rare missense variants, *Am. J. Hum. Genet.* 99 (4) (2016), <https://doi.org/10.1016/j.ajhg.2016.08.016>.
- [23] C. Dong, et al., Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies, *Hum. Mol. Genet.* 24 (8) (Apr 2015) 2125–2137, <https://doi.org/10.1093/hmg/ddu733>.
- [24] B. Reva, Y. Antipin, C. Sander, Predicting the functional impact of protein mutations: application to cancer genomics, *Nucleic Acids Res.* 39 (17) (Sep 2011) e118, <https://doi.org/10.1093/nar/gkr407>.
- [25] R. Calabrese, E. Capriotti, P. Fariselli, P.L. Martelli, R. Casadio, Functional annotations improve the predictive score of human disease-related mutations in proteins, *Hum. Mutat.* 30 (8) (2009), <https://doi.org/10.1002/humu.21047>.

- [26] E. Capriotti, P. Fariselli, PhD-SNP: a webserver and lightweight tool for scoring single nucleotide variants, *Nucleic Acids Res.* 45 (W1) (Jul 2017) W247–W252, <https://doi.org/10.1093/nar/gkx369>.
- [27] P.D. Thomas, D. Ebert, A. Muruganujan, T. Mushayama, L. Albou, H. Mi, PANTHER: making genome-scale phylogenetics accessible to all, *Protein Sci.* 31 (1) (Jan 2022) 8–22, <https://doi.org/10.1002/pro.4218>.
- [28] V. Pejaver, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, *Nat. Commun.* 11 (1) (2020), <https://doi.org/10.1038/s41467-020-19669-x>.
- [29] M.J. Landrum, et al., ClinVar: improvements to accessing data, *Nucleic Acids Res.* 48 (D1) (2020), <https://doi.org/10.1093/nar/gkz972>.
- [30] C.M. Yates, I. Filippis, L.A. Kelley, M.J.E. Sternberg, SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features, *J. Mol. Biol.* 426 (14) (2014), <https://doi.org/10.1016/j.jmb.2014.04.026>.
- [31] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res.* 38 (SUPPL. 2) (2010), <https://doi.org/10.1093/nar/gkq399>.
- [32] 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) (Jul 2022) W510–W515, <https://doi.org/10.1093/nar/gkac439>.
- [33] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.* 33 (Web Server) (Jul 2005) W306–W310, <https://doi.org/10.1093/nar/gki375>.
- [34] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins: Struct., Funct., Bioinf.* 62 (4) (Dec 2005) 1125–1132, <https://doi.org/10.1002/prot.20810>.
- [35] D. Szklarczyk, et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res.* 49 (D1) (2021), <https://doi.org/10.1093/nar/gkaa1074>.
- [36] H. Venselaar, T.A.H. te Beek, R.K.P. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, *BMC Bioinf.* 11 (2010), <https://doi.org/10.1186/1471-2105-11-548>.
- [37] C. Geourjon, G. Deléage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Bioinformatics* 11 (6) (1995) 681–684, <https://doi.org/10.1093/bioinformatics/11.6.681>.
- [38] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, *Nat. Methods* 12 (1) (Jan 2015) 7–8, <https://doi.org/10.1038/nmeth.3213>.
- [39] W. Zheng, C. Zhang, E.W. Bell, Y. Zhang, I-TASSER gateway: a protein structure and function prediction server powered by XSEDE, *Future Generat. Comput. Syst.* 99 (Oct 2019) 73–85, <https://doi.org/10.1016/j.future.2019.04.011>.
- [40] R.A. Laskowski, M.W. MacArthur, J.M. Thornton, PROCHECK: Validation of Protein-Structure Coordinates, 2012, pp. 684–687, <https://doi.org/10.1107/97809553602060000882>.
- [41] L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, *Nucleic Acids Res.* 41 (W1) (Jul 2013) W384–W388, <https://doi.org/10.1093/nar/gkt458>.
- [42] 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) (May 2007) W407–W410, <https://doi.org/10.1093/nar/gkm290>.
- [43] B.H.M. Mooers, Simplifying and enhancing the use of PyMOL with horizontal scripts, *Protein Sci.* 25 (10) (Oct 2016) 1873–1882, <https://doi.org/10.1002/pro.2996>.
- [44] Y. Zhang, TM-align: a protein structure alignment algorithm based on the TM-score, *Nucleic Acids Res.* 33 (7) (Apr 2005) 2302–2309, <https://doi.org/10.1093/nar/gki524>.
- [45] O. Carugo, S. Pongor, A normalized root-mean-square distance for comparing protein three-dimensional structures, *Protein Sci.* 10 (7) (Jul 2001) 1470–1473, <https://doi.org/10.1110/ps.690101>.
- [46] J. Xu, Y. Zhang, How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* 26 (7) (Apr 2010) 889–895, <https://doi.org/10.1093/bioinformatics/btq066>.
- [47] C. Shao, et al., Simplified quality assessment for small-molecule ligands in the Protein Data Bank, *Structure* 30 (2) (Feb 2022) 252–262.e4, <https://doi.org/10.1016/j.str.2021.10.003>.
- [48] M. Han, L.Y. Chen, Molecular dynamics simulation of human urea transporter B, *Mol. Simulat.* 47 (12) (Aug 2021) 1022–1028, <https://doi.org/10.1080/08927022.2021.1941944>.
- [49] S. Kim, et al., PubChem 2023 update, *Nucleic Acids Res.* 51 (D1) (Jan 2023) D1373–D1380, <https://doi.org/10.1093/nar/gkac956>.
- [50] S. Dallakyan, A.J. Olson, Small-Molecule Library Screening by Docking with PyRx, 2015, pp. 243–250, https://doi.org/10.1007/978-1-4939-2269-7_19.
- [51] Y.-D. Gao, J.-F. Huang, [An extension strategy of Discovery Studio 2.0 for non-bonded interaction energy automatic calculation at the residue level], *Dongwuxue Yanjiu* 32 (3) (Jun 2011) 262–266, <https://doi.org/10.3724/SP.J.1141.2011.03262>.
- [52] M.J. Abraham, et al., GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (Sep 2015) 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.
- [53] H. Wickham, ggplot2, *Wiley Interdiscip Rev Comput Stat* 3 (2) (Mar 2011) 180–185, <https://doi.org/10.1002/wics.147>.
- [54] A.P. Boyle, et al., Annotation of functional variation in personal genomes using RegulomeDB, *Genome Res.* 22 (9) (Sep 2012) 1790–1797, <https://doi.org/10.1101/gr.137323.112>.
- [55] J. Lonsdale, et al., The genotype-tissue expression (GTEx) project, *Nat. Genet.* 45 (6) (Jun 2013) 580–585, <https://doi.org/10.1038/ng.2653>.
- [56] A. Bhattacharya, J.D. Ziebarth, Y. Cui, PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways, *Nucleic Acids Res.* 42 (D1) (Jan 2014) D86–D91, <https://doi.org/10.1093/nar/gkt1028>.
- [57] Z. Zhang, M.A. Miteva, L. Wang, E. Alexov, Analyzing effects of naturally occurring missense mutations, *Comput. Math. Methods Med.* 2012 (2012) 1–15, <https://doi.org/10.1155/2012/805827>.
- [58] N. Chinthakunta, S. Cheemanapalli, S. Chinthakunta, C.M. Anuradha, S.K. Chitta, A new insight into identification of in silico analysis of natural compounds targeting GPR120, *Network Modeling Analysis in Health Informatics and Bioinformatics* 7 (1) (Dec 2018) 8, <https://doi.org/10.1007/s13721-018-0166-0>.
- [59] R. Prajapat, I. Bhattachar, A. Kumar, Homology modeling and structural validation of type 2 diabetes associated transcription factor 7-like 2 (TCF7L2), *Trends in Bioinformatics* 9 (1) (Jan 2016) 23–29, <https://doi.org/10.3923/tb.2016.23.29>.
- [60] C.J. Williams, et al., MolProbity: more and better reference data for improved all-atom structure validation, *Protein Sci.* 27 (1) (Jan 2018) 293–315, <https://doi.org/10.1002/pro.3330>.
- [61] R. Santhoshkumar, A. Yusuf, In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of *Curcuma longa*, *J. Genet. Eng. Biotechnol.* 18 (1) (Dec 2020) 24, <https://doi.org/10.1186/s43141-020-00041-x>.
- [62] S. Kwofie, et al., In silico screening of isocitrate lyase for novel anti-buruli ulcer natural products originating from Africa, *Molecules* 23 (7) (Jun 2018) 1550, <https://doi.org/10.3390/molecules23071550>.
- [63] R. Dash, et al., Computational SNP analysis and molecular simulation revealed the most deleterious missense variants in the NBD1 domain of human ABCA1 transporter, *Int. J. Mol. Sci.* 21 (20) (Oct 2020) 7606, <https://doi.org/10.3390/ijms21207606>.
- [64] E.S. Wester, et al., Erythroid urea transporter deficiency due to novel *JK* null alleles, *Transfusion (Paris)* 48 (2) (Feb 2008) 365–372, <https://doi.org/10.1111/j.1537-2995.2007.01532.x>.
- [65] J. Thusberg, A. Olatubosun, M. Vihinen, Performance of mutation pathogenicity prediction methods on missense variants, *Hum. Mutat.* 32 (4) (Apr 2011) 358–368, <https://doi.org/10.1002/humu.21445>.
- [66] L.H. Greene, E.D. Chrysin, L.I. Irons, A.C. Papageorgiou, K.R. Acharya, K. Brew, Role of conserved residues in structure and stability: tryptophans of human serum retinol-binding protein, a model for the lipocalin superfamily, *Protein Sci.* 10 (11) (Apr 2009) 2301–2316, <https://doi.org/10.1110/ps.22901>.
- [67] Y. Xu, H. Wang, R. Nussinov, B. Ma, Protein charge and mass contribute to the spatio-temporal dynamics of protein-protein interactions in a minimal proteome, *Proteomics* 13 (8) (Apr 2013) 1339–1351, <https://doi.org/10.1002/pmic.201100540>.
- [68] J.R. Hamilton, Kidd blood group system: outwardly simple with hidden complexity, *ISBT Sci. Ser.* 14 (1) (Feb 2019) 3–8, <https://doi.org/10.1111/vox.12458>.
- [69] J. Jumper, et al., Highly accurate protein structure prediction with AlphaFold, *Nature* 596 (7873) (Aug 2021) 583–589, <https://doi.org/10.1038/s41586-021-03819-2>.
- [70] Adiyaman and McGuffin, Methods for the refinement of protein structure 3D models, *Int. J. Mol. Sci.* 20 (9) (May 2019) 2301, <https://doi.org/10.3390/ijms20092301>.
- [71] P.C. Agu, et al., Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management, *Sci. Rep.* 13 (1) (Aug 2023) 13398, <https://doi.org/10.1038/s41598-023-40160-2>.
- [72] B.K. Ho, R. Brasseur, The Ramachandran plots of glycine and pre-proline, *BMC Struct. Biol.* 5 (1) (Dec 2005) 14, <https://doi.org/10.1186/1472-6807-5-14>.
- [73] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, *Neuron* 99 (6) (Sep 2018) 1129–1143, <https://doi.org/10.1016/j.neuron.2018.08.011>.
- [74] P. Craveur, et al., Protein flexibility in the light of structural alphabets, *Front. Mol. Biosci.* 2 (May 2015), <https://doi.org/10.3389/fmolb.2015.00020>.
- [75] S.S. Nishizaki, et al., Predicting the effects of SNPs on transcription factor binding affinity, *Bioinformatics* 36 (2) (Jan 2020) 364–372, <https://doi.org/10.1093/bioinformatics/btz612>.
- [76] Genetic effects on gene expression across human tissues, *Nature* 550 (7675) (Oct 2017) 204–213, <https://doi.org/10.1038/nature24277>.
- [77] D.P. Bartel, MicroRNAs, *Cell* 116 (2) (Jan 2004) 281–297, [https://doi.org/10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5).
- [78] C. Preskill, J.B. Weidhaas, SNPs in MicroRNA binding sites as prognostic and predictive cancer biomarkers, *Crit. Rev. Oncog.* 18 (4) (2013) 327–340, <https://doi.org/10.1615/CritRevOncog.2013007254>.
- [79] R. Hou, M. Alemozaffar, B. Yang, J.M. Sands, X. Kong, G. Chen, Identification of a novel UT-B urea transporter in human urothelial cancer, *Front. Physiol.* 8 (APR) (Apr 2017) 245, <https://doi.org/10.3389/fphys.2017.00245>.