

RESEARCH

Open Access



# In-silico analysis of deleterious non-synonymous SNPs in the human AVPR1a gene linked to autism

Md. Delowar Kobir Jibon<sup>1</sup>, Md. Asadul Islam<sup>1</sup>, Md. Eram Hosen<sup>2</sup>, Md. Omar Faruque<sup>3</sup>, Rashed Zaman<sup>1</sup>, Uzzal Kumar Acharjee<sup>1</sup>, Biswanath Sikdar<sup>1</sup>, Yewulsew Kebede Tiruneh<sup>4\*</sup>, Md. Khalekuzzaman<sup>1\*</sup>, Motasim Jawi<sup>5</sup> and Magdi E. A. Zaki<sup>6\*</sup>

## Abstract

Single nucleotide polymorphisms are the most prevalent type of DNA variation occurring at a single nucleotide within the genomic sequence. The *AVPR1a* gene exhibits genetic polymorphism and is linked to neurological and developmental problems, including autism spectrum disorder. Due to the difficulties of studying all non-synonymous single nucleotide polymorphisms (nsSNPs) of the *AVPR1a* gene in the general population, our goal is to use a computational approach to identify the most detrimental nsSNPs of the *AVPR1a* gene. We employed several bioinformatics tools, such as SNPnexus, PROVEAN, PANTHER, PhD-SNP, SNP & GO, and I-Mutant2.0, to detect the 23 most detrimental mutants (R85H, D202N, E54G, H92P, D148Y, C203G, V297M, D148V, S182N, Q108L, R149C, G212V, M145T, G212S, Y140S, F207V, Q108H, W219G, R284W, L93F, P156R, F136C, P107L). Later, we used other bioinformatics tools to perform domain and conservation analysis. We analyzed the consequences of high-risk nsSNPs on active sites, post-translational modification (PTM) sites, and their functional effects on protein stability. 3D modeling, structure validation, protein-ligand binding affinity prediction, and Protein-protein docking were conducted to verify the presence of five significant substitutions (R284W, Y140S, P107L, R149C, and F207V) and explore the modifications induced due to these mutants. These non-synonymous single nucleotide polymorphisms can potentially be the focus of future investigations into various illnesses caused by *AVPR1a* malfunction. Employing in-silico methodologies to evaluate *AVPR1a* gene variants will facilitate the coordination of extensive investigations and the formulation of specific therapeutic approaches for diseases associated with these variations.

**Keywords** Polymorphism, Mutation, NsSNP, Docking, Autism, *AVPR1a*, Missense SNPs, In-silico

\*Correspondence:

Yewulsew Kebede Tiruneh  
Kebede@bdu.edu.et  
Md. Khalekuzzaman  
kzaman63@ru.ac.bd  
Magdi E. A. Zaki  
mezaki@imamu.edu.sa

<sup>1</sup> Professor Joardar DNA and Chromosome Research Laboratory, Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi 6205, Bangladesh

<sup>2</sup> Biomedical Science and Molecular Biology, College of Medicine and Dentistry, James Cook University, Townsville, Australia

<sup>3</sup> Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh

<sup>4</sup> Department of Biology, Biomedical Sciences Stream, Bahir Dar University, P.O.Box=79, Bahir Dar, Ethiopia

<sup>5</sup> Department of Basic Medical Sciences, College of Medicine, University of Jeddah, Jeddah, Saudi Arabia

<sup>6</sup> Department of Chemistry, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia



© The Author(s) 2025, corrected publication 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Globally, the human genome is approximately 99.9% identical, with individual genetic variances making up the remaining 0.1%. These genetic differences arise from random mutations [1]. The most ubiquitous kind of genetic variation in humans is represented by single-nucleotide polymorphisms (SNPs), an invaluable resource for deciphering complicated genetic features [2]. Missense mutations, also known as non-synonymous single nucleotide polymorphisms (nsSNPs), have the potential to induce phenotypic diversity in humans through modifications in protein expression [3]. Prior research suggests that non-synonymous single nucleotide polymorphisms (nsSNPs) contribute to around 50% of the mutations linked to different genetic disorders [4]. Substituting amino acids in conserved regions can affect the structure, stability, and function of proteins. Non-synonymous single nucleotide polymorphisms (nsSNPs) have the potential to alter the function of proteins, which in turn can elevate susceptibility to human diseases [5]. Autism spectrum disorder (ASD) is a severe neuropsychiatric illness that has strong hereditary underpinnings. Nevertheless, the genetic variables that contribute to autism are quite diverse, with several loci fulfilling distinct functions in various individuals [6].

Autism is a neurodevelopmental condition caused by several genes, with more than 90% of cases being influenced by genetics [7]. Arginine vasopressin (AVP) is an endogenous ligand that spontaneously binds to and stimulates *AVPR1 A* receptors in both the peripheral and central nervous systems. The *AVPR1 A*, or arginine vasopressin receptor 1A, has a profound influence on behaviors such as forming pair bonds, providing parental care, displaying aggression, and managing stress [8–11]. This receptor plays a crucial function in brain signaling. Pharmacological approaches and the examination of various animal models have demonstrated the benefits of understanding the role of *AVPR1 A* in behavior [12, 13]. AVP receptors have seven transmembrane domains and are categorized as G-protein-coupled receptors. At least three types of vasopressin receptors (V1R/V1a, V2R, and V3R/V1b) have been found in humans. *AVPR1a*, located on chromosome 12q14–15, is especially relevant to human behavioral research. This is because the specific patterns of V1a receptor gene expression in the brain play a significant role in the observed variations in social and reproductive behavior within and between species. The Vole model has demonstrated this relationship [14–16].

Preclinical research has demonstrated that arginine vasopressin (AVP) enhances some social behaviors, such as association and connection, through interacting with the V1a receptor (*AVPR1 A*) in the brain. The effects of AVP on behavior and the location of the V1a receptor

in the brain differ significantly among different mammalian species [17]. This suggests that the *AVPR1a* gene is a probable candidate for susceptibility to autism [18]. Previous studies investigating familial ties have demonstrated a strong association between the *AVPR1 A* gene and autism [19]. The presence of two microsatellite polymorphisms, RS1 and RS3, in the vicinity of the promoter region of *AVPR1 A*, which codes for the receptor subtype primarily responsible for regulating behavior, has been linked to autism and behavioral traits [20, 21]. The severity of autistic traits can be significantly influenced by a single nucleotide polymorphism (SNP) of the *AVPR1a* gene [22]. The *AVPR1a* gene encodes the vasopressin V1a receptor, one of the primary receptors for arginine vasopressin (AVP). A low arginine vasopressin (AVP) concentration level in cerebrospinal fluid (CSF) is an indicator of social impairment in monkeys with low social behavior and autistic children [23]. An extensive association study was conducted involving 3 microsatellites and twelve tag single nucleotide polymorphisms (SNPs) situated within and near the *AVPR1 A* gene in 205 Finnish families. This was followed by an assessment of the gene's promoter, which revealed a significant correlation with autism [24]. A study was undertaken in the Korean population to evaluate the relationship between autism spectrum disorder and changes in the *AVPR1 A* promoter region. The study used a family-based association test (FBAT) for this purpose. The results suggest that alterations in the *AVPR1 A* promoter region may have a role in the development of ASD and the regulation of *AVPR1 A* expression [25]. Here, we explored several computational approaches to pin down non-synonymous polymorphisms in the human *AVPR1 A* gene.

## Materials and methods

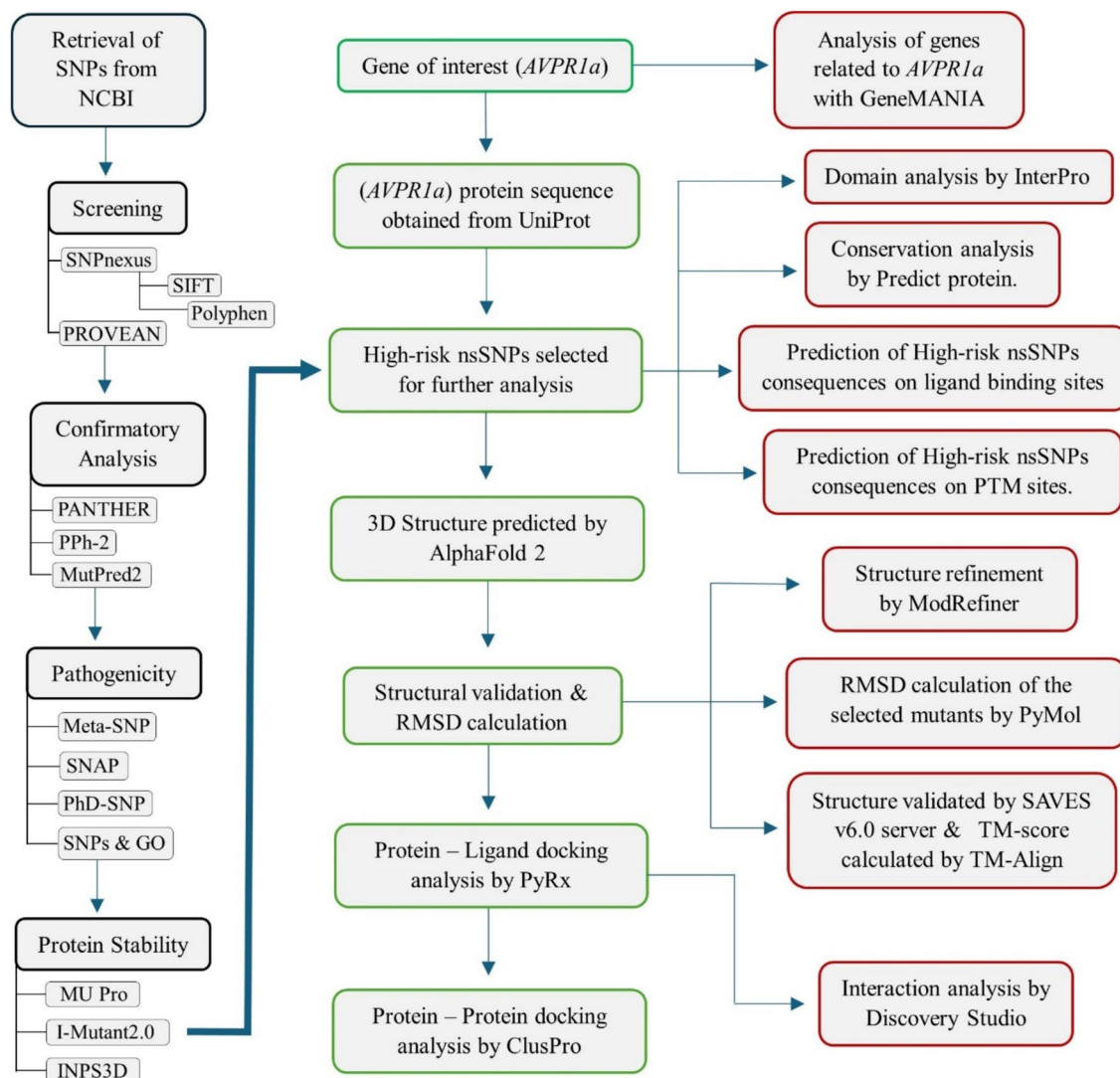
The overall workflow of this project is shown in Fig. 1.

### Retrieval of SNPs

A total of 402 nsSNPs associated with the human *AVPR1a* gene were retrieved from the dbSNP database (<https://www.ncbi.nlm.nih.gov/>). We collected information on SNPs, including SNP ID, protein accession number, location, residue alteration, and global minor allele frequency (MAF) [26]. The *AVPR1a* gene sequence was sourced from Uniprot (<https://www.uniprot.org/>). Studies investigated the harmful effects of missense SNPs on the *AVPR1a* gene.

### GeneMANIA to understand *AVPR1a* interactions with other genes

GeneMANIA (<https://genemania.org/>) was used to investigate the relationship between the *AVPR1a* gene and other genes based on pathways, expression,



**Fig. 1** Project workflow

localization, genetics, and protein interaction. This tool confirms the connective network between the *AVPR1a* gene and other genes [27].

### Screening of deleterious nsSNPs

We employed two different bioinformatics tools to evaluate the likely impact of genetic variations extracted from the dbSNP databases. The tools mentioned above used: SNPnexus (<https://www.snp-nexus.org>) includes Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen) [28]. SIFT predicts harmful nsSNPs by examining protein homology sequences and natural nsSNP alignments. A score below 0.05 indicates that SIFT considers the nsSNPs to have a deleterious effect on protein function [29]. PolyPhen-2 predicts the functional

impact of amino acid substitutions on protein structure and function using sequence-based characterization [30]. PolyPhen generates a position-specific independent count (PSIC) score for each amino acid variant. Differences in PSIC scores for variants indicate their direct functional impact [31, 32]. PROVEAN is another tool we used to screen deleterious nsSNPs. PROVEAN predicts the functional impact of variants. A threshold value of  $\geq -2.5$  indicates a deleterious nsSNP [33].

### Confirmatory analysis of the deleterious nsSNPs

We cross-checked our screened nsSNPs with another three bioinformatics tools to reconfirm the level of severity and deleterious nature. The biological and evolutionary information for every protein-coding gene

is compiled in PANTHER (<http://pantherdb.org>) [34]. PPh-2 (<http://genetics.bwh.harvard.edu/pph2>) predicts how point mutations affect protein expression [35]. Mutpred2 (<http://mutpred.mutdb.org/>) is used to assess, using molecular and biological data, the possible structural consequences of nsSNPs arising from alterations in proteins [36].

### Screening of disease-associated SNPs

To examine the association of screened nsSNPs with a disease, PhD-SNP, SNPs&GO, and Meta-SNP were performed. In order to categorize an SNP's effect as either disease-related or neutral, the PhD-SNP tool (<https://snps.biofold.org/phdsnp/phd-snp.html>) generates an accuracy index score from 36,000 benign and harmful SNVs. It was developed and verified using the ClinVar dataset [37]. SNPs&GO (<https://snps-and-go.biocomp.unibo.it/snps-and-go>) assesses changes in amino acids at a particular location in a protein [38]. SNPs&GO and PhD-SNP are pivotal approaches based on machine learning that leverage comparative conservation scores derived from multiple sequence alignments [39]. In Meta-SNP (<https://snps.biofold.org/meta-snp>), the outputs from individual predictors are combined as input, and disease occurrence is predicted if mutations surpass a threshold of 0.5 [40].

### Functional effects of SNPs on protein stability

To determine the changes in protein stability, we used three different tools: MUpro, I-Mutant 2.0, and INPS3D. Protein stability assessment is commonly conducted using the MUpro server (<http://mupro.proteomics.ics.uci.edu>). This web server is built using two machine learning techniques: Support Vector Machines (SVM) and Neural Networks. These techniques assess how single-site changes in amino acids affect the stability of proteins and display the results as a rise or fall, denoted by positive or negative scores [41]. The neural network technique is employed by the I-Mutant 2.0 web server (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>). It is applied to predict potential changes in protein stability after mutations. A reliability index (RI) of 0 to 10, with 10 denoting the maximum dependability, is used to make predictions. The server also assesses the degree of protein instability and gives a free energy change number ( $\Delta\Delta G$ ) that shows if stability will rise or fall. A protein stability decrease is indicated by a  $\Delta\Delta G$  value less than 0, whereas an increase in protein stability is suggested by a value greater than 0 [42]. Protein stabilities can be predicted for both wildtype and mutant variants using a recently developed tool named INPS3D. The INPS-MD (Impact of Non-synonymous mutations on Protein Stability—Multi Dimension) web server (<https://inpsmd.biocomp.unibo.it/inpsSuite/default/index3D>) was utilized for this purpose. This tool takes into account several variables, including the molecular weights and hydrophobicities of the native and mutated amino acids, the alignment score difference, the likelihood of the original residue undergoing mutation, the relative solvent accessibility (RSA) of the original amino acid, and the local energy difference between the wildtype and altered protein structures [43].

unibo.it/inpsSuite/default/index3D) was utilized for this purpose. This tool takes into account several variables, including the molecular weights and hydrophobicities of the native and mutated amino acids, the alignment score difference, the likelihood of the original residue undergoing mutation, the relative solvent accessibility (RSA) of the original amino acid, and the local energy difference between the wildtype and altered protein structures [43].

### Domain analysis of *AVPR1a*

We utilized a widely used computational tool, InterPro (<https://www.ebi.ac.uk/interpro/>), to identify the functional domains of our desired protein (*AVPR1a*) [44]. This application uses a database of protein families, domains, and functional sites to find motifs and domains of proteins and, in turn, determine their functional characterization [45].

### Conservation analysis

In order to evaluate the amino acid conservation pattern within the protein sequence, we made use of the predict protein server (<https://predictprotein.org>). The *AVPR1a* protein's single-letter amino acid sequence was submitted for evaluation. More than thirty tools are integrated with this service, including ConSurf and other techniques for finding functional areas. Evolutionary conservation was analyzed using Bayesian empirical inference [46].

### Predictions of ligand binding sites

The meta-server program COACH (<http://zhanglab.ccmb.med.umich.edu/COACH/>) used two comparison techniques, TM-SITE and S-SITE, to find ligand binding templates from the BioLiP protein function database in order to predict protein-ligand binding sites. Additionally, sequence feature correlations and binding-specific sub-structure were used. In order to anticipate ligand binding sites (LBS), COACH employs a consensus approach by combining the predictions from several algorithms, including TM-SITE, S-SITE, COFACTOR, FINDSITE, and ConCavity. Cluster size, PDB hits, ligand names, consensus binding residues, and downloadable complex structures are the factors used by the COACH server to select the top ten models. Each model is then given a C- scores. The expected reliability is shown by the C-score, which has a range of 0 to 1. Higher scores correspond to higher reliability [47].

### Prediction of post translational modification (PTM) site

The neural network-based and frequently used program which called NetPhos 3.1 (<https://services.healthtech.dtu.dk/services/NetPhos-3.1/>) was used to estimate the probable phosphorylation sites of the *AVPR1a* protein. If a threshold score is more than 0.5, it suggests that a



certain location is probably phosphorylated [48]. In order to forecast probable locations of MHC-binding sites, we employed GPS-MBA 1.0 (<https://mba.biocuckoo.org/>) [49]. To identify potential SUMOylation and ubiquitylation sites, we used GPS-SUMO (<https://sumosp.biocuckoo.org/>) and GPS-Uber (<http://gpsuber.biocuckoo.cn/wsresult.php>) [50, 51].

### 3D modeling

The native structure of the AVPR1a protein was downloaded from the AlphaFold protein structure database (AlphaFold DB, <https://alphafold.ebi.ac.uk/>) [52]. AlphaFold2 predicted the rest of the mutant protein structure [53]. The protein sequences of mutants were modified according to the substitution of amino acid positions. In order to minimize steric clashes, obtain precise side-chain locations, and eliminate distracting stereochemical violations without compromising accuracy, we employ gradient descent in the Amber force field through AlphaFold2 to predict relaxed structure [54]. The ModRefiner tool (<http://zhanglab.ccmb.med.umich.edu/ModRefiner>) was utilized to refine the predicted structure [55].

### Structural validation and RMSD calculation

The selected structural model was validated using the widely accepted server SAVES v6.0 (<https://saves.mbi.ucla.edu>). This site offers tools like PROCHECK and ERRAT to assess the overall quality of the 3D model [56]. Furthermore, the RAMACHANDRAN plot produced by PROCHECK was used to evaluate the model's quality [57]. The alignment of a protein's tertiary structure with its primary structure is evaluated by 3D verification [58]. We utilized the pyMOL tool (<https://pymol.org/2>) to compute the root-mean-square deviation (RMSD) by superimposing the native and mutant protein structures, representing the difference between the two compared models. A higher RMSD value indicates a greater deviation between the two structures. On a scale ranging from 0 to 1, the TM-score evaluates the structural similarity of two models; a score of 1 denotes total similarity, while lower values suggest growing dissimilarity [59]. Afterward, the template modeling score (TM-score) was calculated by comparing the wild-type protein structure with mutant protein structures using TM-align (<https://zhanglab.ccmb.med.umich.edu/TM-align>) [60].

### Protein-ligand interaction analysis

We conducted docking of all chosen ligands with AVPR1a using the PyRx program (<https://pyrx.sourceforge.io>) [61]. Virtual ligand screening was carried out using the Lamarckian genetic algorithm (LGA), which combines AutoDock and AutoDock Vina [62]. By applying AutoDock tools to convert PDB files to

Pdbqt format and ascertain binding affinities. The grid size was modified as per the center (XYZ axis). The grid box center remained at coordinates X: -7.4813, Y: 4.9867, Z: 10.6823, with dimensions set to X: 109.4936, Y: 98.4684, and Z: 130.9739 Å [63]. Stronger ligand binding ability with the target receptor is indicated by negative values of the binding affinities of the ligands to the receptors, which were computed in kcal/mol [64]. Discovery Studio (<https://discover.3ds.com/discovery-studio-visualizer-download>) was utilized to visualize 2D and 3D interactions between ligands and proteins. It depicted the position and size of binding sites, non-bonding interactions, bonding angles and lengths of a docked ligand [65].

### Analyze docking results of protein-protein complex by ClusPro

We utilized the ClusPro web server (<https://cluspro.org>) to conduct protein-protein docking analysis. This tool is extensively employed for studying protein-protein docking interactions. ClusPro offers various sophisticated options to tailor the search procedure, such as removing unstructured protein regions, applying attraction or repulsion forces, considering pairwise distance constraints, producing homo-multimers, incorporating data from small-angle X-ray scattering (SAXS), and locating heparin-binding sites. Based on the type of protein, six different energy functions are accessible. Ten models, each with a center of densely packed clusters of low-energy docked structures, are produced by docking with each set of energy parameters [66].

### Molecular dynamics (MD) simulation

The protein-ligand complexes were subjected to MD simulations using GROMACS [67] and the WebGro server (<https://simlab.uams.edu/>). The ligand topology files were generated using the PRODRG Server [68], with a triclinic simulation box employed for system setup. The complexes were solvated using the SPC water model, and the system was neutralized by adding 0.15 M NaCl. The simulations were performed using the GROMOS96 43a1 force field. An initial energy minimization was carried out with 5000 steps of the steepest descent algorithm. Subsequently, the system was equilibrated under NVT and NPT ensembles with standard parameters, maintaining a temperature of 300 K and a pressure of 1.0 bar. Using the Leap-frog MD integrator, the MD trajectories were generated over a 200 ns timescale, with trajectory snapshots taken every 0.1 ns, yielding 2000 frames for analysis. The trajectory snapshots were subsequently analyzed to determine Rg, RMSD, RMSE, and SASA [67, 69].

## Results

### Download datasets of interest

The SNPs of the *AVPR1a* gene were acquired from the dbSNP database, which is widely considered the most actively utilized and comprehensive database currently accessible. According to the NCBI dbSNP database, the human *AVPR1a* gene displayed a sum of 4190 single nucleotide polymorphisms (SNPs). Among the entire collection, there were 402 non-synonymous SNPs (nsSNPs) (Table S1), 177 synonymous SNPs, 1625 SNPs placed in the 3' UTR, 168 SNPs in the 5' UTR, and 893 SNPs in intronic regions. The remaining SNPs were classified into various categories (Fig. 2). Only the non-synonymous single nucleotide polymorphisms (nsSNPs) were selected for this investigation.

### GeneMANIA to understand *AVPR1a* interactions with other genes

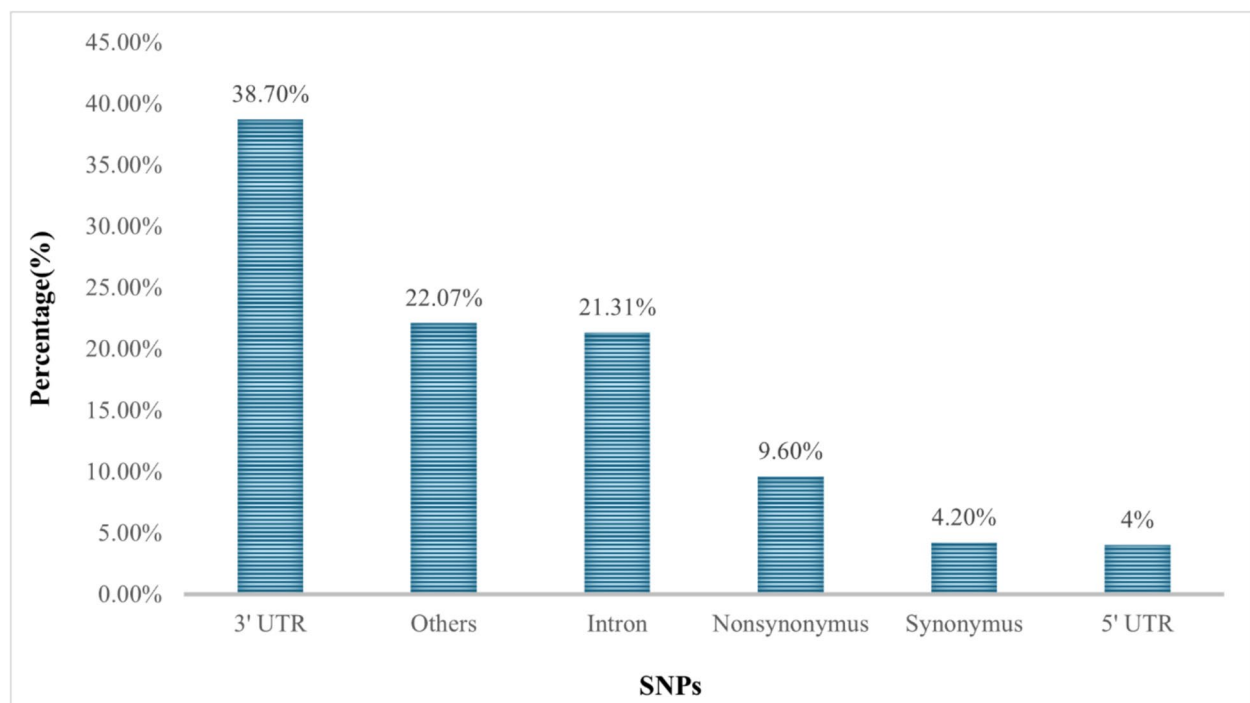
GeneMANIA efficiently analyzes the other genes related to the *AVPR1a* gene. The graphical representation of the analysis is illustrated in Fig. 3. These findings suggest that the *AVPR1a* gene may have a functional connection to the co-expressed genes and could be involved in common biological pathways. So, if any mutation occurs in the *AVPR1a* gene, it may also affect the overall gene network interactions among all the related genes.

### Screening of deleterious nsSNPs

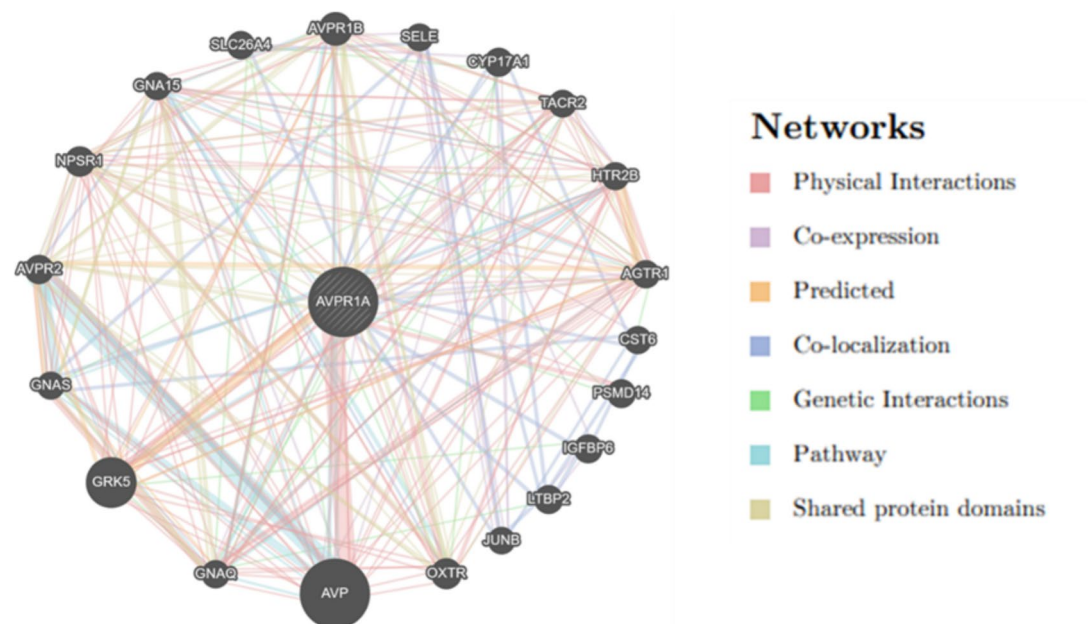
SIFT and Polyphen score initially indicates deleterious SNPs based on score. SIFT score 0 and polyphen score 1 denote the most deleterious nature of nsSNPs. The range of the Polyphen and SIFT output scores is 0 to 1. We specifically chose common non-synonymous single nucleotide polymorphisms (nsSNPs) that received a score of 0 in the SIFT algorithm and a score of 1 in the PolyPhen algorithm. This selection criteria ensures that only the most harmful SNPs are included in our study. The use of the PROVEAN tool facilitates the identification of the most harmful SNPs through further investigation. The threshold value of this tool is  $-2.5$ . The PROVEAN tool classified nsSNPs as harmful when the result was less than  $-2.5$ . Conversely, a score greater than  $-2.5$  is anticipated to be neutral. Finally, we have identified a total of 23 nsSNPs that met the specified criteria. These nsSNPs have been classified as having a high likelihood of impacting protein function (Table 1).

### Confirmatory analysis of the deleterious nsSNPs

Another three computational tools were applied to reconfirm the detrimental nature of initially screened nsSNPs to maintain the required accuracy. The combined prediction result of PANTHER, PPh-2, and Mutpred2 contributes to finalize the number of deleterious SNPs for further analysis (Table 2). The PANTHER tool makes



**Fig. 2** SNP distribution graph of *AVPR1a* gene. 3'UTR = 38.7 %, Others = 22.07%, Intron = 21.31%, Nonsynonymous = 9.6%, Synonymous = 4.20%, 5' UTR = 4%



**Fig. 3** Other genes related to the *AVPR1a* gene. Different colors of the legend indicated the parameters of constructed networks

**Table 1** Risky nsSNPs screened by SIFT, Polyphen, and PROVEAN

Variant ID	Nucleotides	AA Variations	SIFT Score	Prediction	Polyphen Score	Prediction	PROVEAN Score	Prediction
rs1260022270	C/T	R85H	0	De	1	P D	−4.557	De
rs1267958616	C/T	D202 N	0	De	1	P D	−4.955	De
rs1321994497	T/C	E54G	0	De	1	P D	−6.152	De
rs1325662981	T/G	H92P	0	De	1	P D	−9.126	De
rs1337643184	C/A	D148Y	0	De	1	P D	−8.744	De
rs1338176647	A/C	C203G	0	De	1	P D	−11.893	De
rs1369668995	C/T	V297M	0	De	1	P D	−2.867	De
rs1377891669	T/A	D148 V	0	De	1	P D	−8.744	De
rs1417441306	C/T	S182 N	0	De	1	P D	−2.931	De
rs1424280726	T/A	Q108L	0	De	1	P D	−6.529	De
rs1440280008	G/A	R149 C	0	De	1	P D	−7.772	De
rs1449556252	C/A	G212 V	0	De	1	P D	−8.918	De
rs369710823	A/G	M145 T	0	De	1	P D	−5.810	De
rs376518166	C/T	G212S	0	De	1	P D	−5.946	De
rs745458336	T/G	Y140S	0	De	1	P D	−8.344	De
rs748572296	A/C	F207 V	0	De	1	P D	−6.937	De
rs754449459	T/A	Q108H	0	De	1	P D	−4.663	De
rs758567125	A/C	W219G	0	De	1	P D	−12.884	De
rs767540299	G/A	R284 W	0	De	1	P D	−7.070	De
rs772227542	G/A	L93 F	0	De	1	P D	−3.650	De
rs773269527	G/C	P156R	0	De	1	P D	−8.794	De
rs776488571	A/C	F136 C	0	De	1	P D	−7.712	De
rs780705756	G/A	P107L	0	De	1	P D	−9.216	De

De Deleterious, P D Probably Damaging

**Table 2** Confirmatory assessment of screened deleterious nsSNPs through PANTHER, PPh-2, and Mutpred2

Variant ID	Nucleotides	AA Variations	PANTHER	PPh-2 Score	Prediction	MutPred2 Score
rs1260022270	C/T	R85H	P D	1	P D	0.534
rs1267958616	C/T	D202 N	P D	1	P D	0.841
rs1321994497	T/C	E54G	P D	1	P D	0.807
rs1325662981	T/G	H92P	P D	1	P D	0.898
rs1337643184	C/A	D148Y	P D	1	P D	0.945
rs1338176647	A/C	C203G	P D	1	P D	0.967
rs1369668995	C/T	V297M	P D	1	P D	0.646
rs1377891669	T/A	D148 V	P D	1	P D	0.936
rs1417441306	C/T	S182 N	P D	1	P D	0.756
rs1424280726	T/A	Q108L	P D	1	P D	0.877
rs1440280008	G/A	R149 C	P D	1	P D	0.930
rs1449556252	C/A	G212 V	P D	1	P D	0.867
rs369710823	A/G	M145 T	P D	1	P D	0.882
rs376518166	C/T	G212S	P D	1	P D	0.781
rs745458336	T/G	Y140S	P D	1	P D	0.852
rs748572296	A/C	F207 V	P D	1	P D	0.931
rs754449459	T/A	Q108H	P D	1	P D	0.802
rs758567125	A/C	W219G	P D	1	P D	0.926
rs767540299	G/A	R284 W	P D	1	P D	0.757
rs772227542	G/A	L93 F	P D	1	P D	0.820
rs773269527	G/C	P156R	P D	1	P D	0.942
rs776488571	A/C	F136 C	P D	1	P D	0.823
rs780705756	G/A	P107L	P D	1	P D	0.854

P D Probably Damaging

a prediction about how the nsSNPs will affect the way the protein functions. Every screened nsSNP in PPh-2 was predicted to be harmful (PSIC >0.5); these variants were expected to be extremely harmful, with a PSIC score of 1. The MutPred2 score shows how likely it is that a change in an amino acid will impact the function of the protein. Pathogenicity is predicted to be using a score threshold of 0.5. The higher the score, the more probable it is that an amino acid substitution is linked to a particular disease.

### Screening of disease-associated SNPs

It is very crucial to identify nsSNPs related to disease for further analysis. The Meta-SNP algorithm detected 22 nsSNPs that were associated with disease, excluding D202 N (rs1267958616). The G212S (rs376518166) mutation was classified as neutral by SNAP, but the remaining mutations were deemed to be associated with disease. The PhD-SNP analysis identified a total of 3 nsSNPs that were determined to be neutral: E54G, V297M, and R284 W and the rest of the nsSNPs were confirmed to be disease-causing. The SNPs & GO software identifies a total of 23 nsSNPs that are associated with disease. The

comprehensive forecast outcomes are succinctly outlined in Table 3.

### Impact of screened nsSNPs on the stability of proteins

According to the MU Pro tool, Q108L and P107L make the protein more stable, whereas the remaining 21 nsSNPs were projected to make less stable, which would reduce the protein activity. I-Mutant2.0 detected S182 N and Q108L increased, and the remaining were decreased the the protein stability. The indicated structural effect of 23 possible nsSNPs was acquired from INPS3D. The outputs of the protein stability evaluation are presented in Table 4.

### Detection of nsSNPs on the AVPR1a domains

InterPro predicted two functional domains of AVPR1a, which are (GPCR\_Rhodpsn\_7 TM) Seven-transmembrane rhodopsin-like G protein-coupled receptors domain (from amino acid 68 to 348), (V1R\_C) Conserved C-terminal domain of Vasopressin V1 receptors (from amino acid 372 to 418) (Fig. 4). This domain analysis result indicated that 22 out of 23 nsSNPs are positioned in the large GPCR\_Rhodpsn\_7 TM domain.



**Table 3** Prediction of disease-associated SNPs by Meta-SNP, SNAP, PhD-SNP, and SNPs & GO

Variant ID	Nucleotides	AA Variations	Meta-SNP		SNAP		PhD-SNP		SNPs & GO	
rs1260022270	C/T	R85H	0.659	Di	0.590	Di	4	Di	7	Di
rs1267958616	C/T	D202 N	0.481	Ne	0.720	Di	5	Di	3	Di
rs1321994497	T/C	E54G	0.766	Di	0.740	Di	3	Ne	9	Di
rs1325662981	T/G	H92P	0.900	Di	0.775	Di	5	Di	10	Di
rs1337643184	C/A	D148Y	0.932	Di	0.840	Di	8	Di	9	Di
rs1338176647	A/C	C203G	0.905	Di	0.850	Di	5	Di	9	Di
rs1369668995	C/T	V297M	0.809	Di	0.755	Di	1	Ne	9	Di
rs1377891669	T/A	D148 V	0.928	Di	0.815	Di	8	Di	10	Di
rs1417441306	C/T	S182 N	0.727	Di	0.735	Di	3	Di	9	Di
rs1424280726	T/A	Q108L	0.568	Di	0.530	Di	5	Di	9	Di
rs1440280008	G/A	R149 C	0.911	Di	0.805	Di	5	Di	9	Di
rs1449556252	C/A	G212 V	0.742	Di	0.525	Di	4	Di	8	Di
rs369710823	A/G	M145 T	0.773	Di	0.590	Di	4	Di	9	Di
rs376518166	C/T	G212S	0.653	Di	0.310	Ne	3	Di	6	Di
rs745458336	T/G	Y140S	0.841	Di	0.705	Di	6	Di	9	Di
rs748572296	A/C	F207 V	0.721	Di	0.690	Di	3	Di	9	Di
rs754449459	T/A	Q108H	0.649	Di	0.550	Di	2	Di	8	Di
rs758567125	A/C	W219G	0.839	Di	0.825	Di	6	Di	9	Di
rs767540299	G/A	R284 W	0.751	Di	0.695	Di	0	Di	6	Di
rs772227542	G/A	L93 F	0.765	Di	0.710	Di	2	Ne	7	Di
rs773269527	G/C	P156R	0.900	Di	0.815	Di	0	Di	9	Di
rs776488571	A/C	F136 C	0.825	Di	0.615	Di	5	Di	9	Di
rs780705756	G/A	P107L	0.547	Di	0.640	Di	7	Di	8	Di

Di Disease, Ne Neutral

Polymorphism in the domain area could significantly alter the activity of protein.

### Conservation analysis

The research conducted by the Conservancy demonstrated a significant level of preservation in both the structure and function of all *AVPR1a* residues. Predict protein server provides 3 types of conservation scores: 1–3 (Minimal), 4–6 (Mid-level), 7–9 (high). We focused solely on the 7–9 scoring residues corresponding to the residues in the highly conserved region (Figure S2). We noticed that all 23 detected nsSNPs were present in the highly conserved region (Table 5). Prior research has demonstrated that essential amino acids, which play an active role in multiple biological functions, are important. All of these are situated within a protein's conserved area. Therefore, it may be inferred that non-synonymous single nucleotide polymorphisms (nsSNPs), which exhibit a high degree of conservation, have a significant detrimental impact on both the structural and functional characteristics of the *AVPR1a* protein.

### High-risk nsSNPs consequences on ligand binding sites

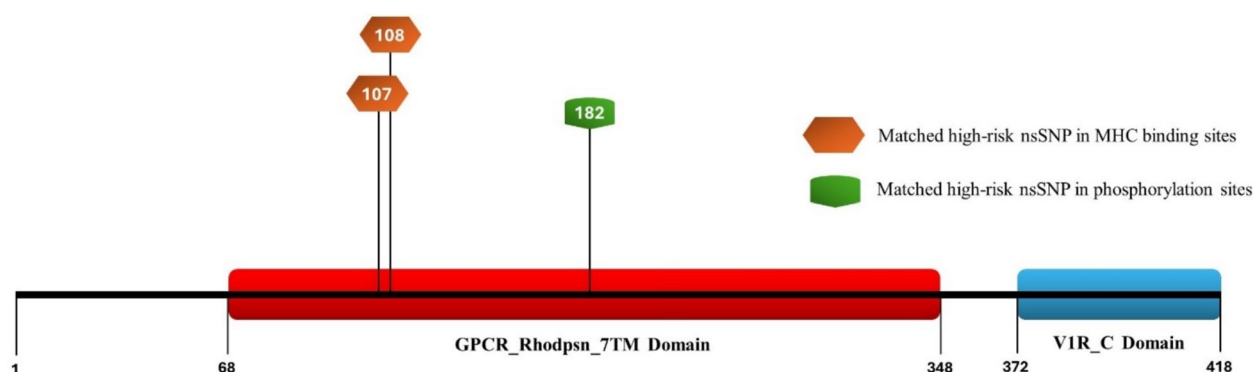
We employed the COACH server to forecast the ligand binding location of the *AVPR1a* protein. The COACH server utilizes a combination of programs from TM-SITE, S-SITE, COFACTOR, FINDSITE, and ConCavity to estimate the combined output. The predicted binding site residues are E54, Q108, W111, Q131, V132, M135, F136, D202, C203, W204, F207, Q209, K128, W304, F307, F308, M220, I224, S338, A334, Q311. We noticed that E54, Q108, F136, D202, C203, and F207 positions matched our screened highly deleterious nsSNPs. Hence, we can conclude that mutation of these positions can significantly alter the function of the protein due to active site modifications.

### High-risk nsSNPs consequences on post-translational modification (PTM) site

NetPhos 3.1 tool predicted probable 43 phosphorylation sites in the *AVPR1a* protein (Table S2). The positions are S4, S16, T23, T28, S29, T61, S70, T79, T83, S84, S94, T114, S138, Y150, S167, S182\*, T183, Y186, S190, T206, S213, T234, S253, S256, S278, S281, S283, T289, T323, S338, S341, S380, S382, T386, Y388, S389, S393, S397,

**Table 4** Results of protein stability changes due to the nucleotide polymorphisms predicted by MU Pro, I-Mutant2.0, and INPS3D

Variant ID	Nucleotides	AA Variations	MU Pro	Score	I-Mutant2.0	Score	INPS3D	Score
rs1260022270	C/T	R85H	Decrease	-1.0977846	Decrease	8	Decrease	-0.229538
rs1267958616	C/T	D202 N	Decrease	-0.43786389	Decrease	5	Decrease	-0.497514
rs1321994497	T/C	E54G	Decrease	-1.6185617	Decrease	9	Decrease	-0.792783
rs1325662981	T/G	H92P	Decrease	-0.88200575	Increase	6	Decrease	-1.14777
rs1337643184	C/A	D148Y	Decrease	-0.95038532	Decrease	3	Decrease	-0.1177
rs1338176647	A/C	C203G	Decrease	-1.9037191	Decrease	8	Decrease	-3.26869
rs1369668995	C/T	V297M	Decrease	-0.60384266	Decrease	8	Decrease	-0.434777
rs1377891669	T/A	D148 V	Decrease	-0.88691156	Decrease	2	Increase	1.16615
rs1417441306	C/T	S182 N	Decrease	-0.32708786	Increase	3	Decrease	-0.452437
rs1424280726	T/A	Q108L	Increase	0.36395533	Increase	2	Increase	0.0731143
rs1440280008	G/A	R149 C	Decrease	-0.54728157	Decrease	6	Decrease	-0.0546847
rs1449556252	C/A	G212 V	Decrease	-0.39457627	Decrease	6	Decrease	-0.851292
rs369710823	A/G	M145 T	Decrease	-2.6523221	Decrease	8	Decrease	-1.91913
rs376518166	C/T	G212S	Decrease	-0.92599754	Decrease	9	Decrease	-0.119486
rs745458336	T/G	Y140S	Decrease	-1.845721	Decrease	6	Decrease	-1.85495
rs748572296	A/C	F207 V	Decrease	-0.82949321	Decrease	8	Decrease	-1.16988
rs754449459	T/A	Q108H	Decrease	-0.51365487	Decrease	8	Decrease	-0.262685
rs758567125	A/C	W219G	Decrease	-1.9396669	Decrease	9	Decrease	-2.75964
rs767540299	G/A	R284 W	Decrease	-1.0222554	Decrease	4	Decrease	-0.438376
rs772227542	G/A	L93 F	Decrease	-0.91229984	Decrease	8	Decrease	-1.7408
rs773269527	G/C	P156R	Decrease	-0.63828646	Decrease	6	Decrease	-0.475707
rs776488571	A/C	F136 C	Decrease	-0.23947682	Decrease	7	Decrease	-0.928974
rs780705756	G/A	P107L	Increase	0.083468873	Decrease	4	Decrease	-0.41095

**Fig. 4** Schematic representation of the matched high-risk nsSNPs in the MHC binding and phosphorylation sites (PTM sites) into the domain area of the *AVPR1a* protein

S404, S407, S408, S410, S417 (Figure S3). Among them, S182 position matched our predicted highly damaging nsSNPs. To unveil the positions of the MHC binding sites of *AVPR1a* protein, we employed GPS-MBA 1.0 tool (Table S3). The position ranges are 22–33, 26–34, 29–37, 96–104, 100–108, 154–162, 253–261, 260–268, 273–281. In the position range 100–108, we found 107 and 108 positions which match our screened risky nsSNPs. The presence of highly damaging nsSNPs of these PTM sites

clearly indicates that mutation of those positions can significantly affect protein activities (Fig. 4). We also predict SUMOylation and ubiquitylation sites of *AVPR1a* protein, but none was found in our screened risky nsSNPs.

### 3D modeling

We predicted the 3D structures of all 23 mutants using AlphaFold2. We utilized the platform AlphaFold2 colab to perform structure prediction. We used relax number

**Table 5** Conservation analysis results of identified 23 mutants evaluated by predict protein server

Variant ID	AA substitution	Predict Protein	
		Score	Prediction
rs1260022270	R85H	7–9	Highly conserved
rs1267958616	D202 N	7–9	Highly conserved
rs1321994497	E54G	7–9	Highly conserved
rs1325662981	H92P	7–9	Highly conserved
rs1337643184	D148Y	7–9	Highly conserved
rs1338176647	C203G	7–9	Highly conserved
rs1369668995	V297M	7–9	Highly conserved
rs1377891669	D148 V	7–9	Highly conserved
rs1417441306	S182 N	7–9	Highly conserved
rs1424280726	Q108L	7–9	Highly conserved
rs1440280008	R149 C	7–9	Highly conserved
rs1449556252	G212 V	7–9	Highly conserved
rs369710823	M145 T	7–9	Highly conserved
rs376518166	G212S	7–9	Highly conserved
rs745458336	Y140S	7–9	Highly conserved
rs748572296	F207 V	7–9	Highly conserved
rs754449459	Q108H	7–9	Highly conserved
rs758567125	W219G	7–9	Highly conserved
rs767540299	R284 W	7–9	Highly conserved
rs772227542	L93 F	7–9	Highly conserved
rs773269527	P156R	7–9	Highly conserved
rs776488571	F136 C	7–9	Highly conserved
rs780705756	P107L	7–9	Highly conserved

5 for all of the mutants to predict the proper relaxed structure. We used modified protein sequences for all 23 mutants according to their mutation position changes. The predicted protein structures were downloaded in PDB format. We got wild-type protein sequence from the UniProt database and downloaded wild-type structure from the AlphaFold Protein Structure Database (Figure S1).

**Structural validation and RMSD calculation**

The modeled structures were validated by the SAVES v6.1 server, and the evaluation of the secondary structure was conducted using the RAMACHANDRAN plot. The RAMACHANDRAN plot revealed that a significant proportion of the residues of amino acids in the projected structures occupied a region with a significant level of favorability. The comprehensive validation results, RMSD values, and TM scores for all mutants are presented in Table 6. We calculated the RMSD of all 23 mutants using PyMOL (Fig. 5) and nominated 5 mutants (R284 W, Y140S, P107L, R149 C, and F207 V) based on the maximum RMSD value (Fig. 6). TM scores of all screened mutants also indicate the structural similarity

and dissimilarity between the native and mutant protein models.

**Protein-ligand interaction analysis**

As we know, our targeted gene is associated with autism, so for molecular docking analysis, we used 93 compounds with potential for autism treatment as references that are also available in the drug bank, including approved and clinical trial drugs category (Table S4). Our primary goal is to assess the variation in protein-ligand interactions between native and mutant proteins. We downloaded our reference compound structures in sdf format from the PubChem database. After molecular docking analysis, we selected the top 3 compounds for each target based on maximum binding affinity. The selected top 3 drug compounds represent lead molecules for each mutated variant and may have the potential to work properly against those reported deleterious SNPs (Table 7). We used Discovery Studio to visualize the 2D interactions (Fig. 7). We found significant differences between native and mutants in binding affinity and interacting residues responsible for hydrogen and hydrophobic bond formation (Table 7). In some cases, we got utterly new best-binding molecules compared to the native protein binding interaction profile. For example, Mutant F207 V exhibits an entirely new binding interaction profile compared to the native protein. Hence, we can conclude that polymorphisms can drastically alter the protein’s conformation.

**Analyze docking of protein-protein complex by ClusPro**

To assess the variation of the protein-protein docking results, we used native protein as a reference protein and evaluated the changes against the mutants. We perform protein-protein docking among 6 protein-protein complexes. These are Native-Native, Native-F207 V, Native-P107L, Native-R149 C, Native-R284 W, Native-Y140S. We noticed significant variation in the binding energy among the 6 complexes (Fig. 8). Thus, we can conclude that mutations can significantly alter the structural and functional characteristics of the protein.

**Assessment of MD simulation trajectories**

Molecular dynamics (MD) simulations are an essential computational method for studying the conformational flexibility, thermodynamic stability, and time-dependent behavior of biomolecular systems [64]. We performed a 200 ns molecular dynamics simulation to validate the docking outputs and the dynamic behaviour of the best resulting drug molecules against the five nominated mutated variants of the target gene in the cellular environment. After the simulation was completed, the dynamic trajectories were examined, and various metrics such as root-mean-square deviation (RMSD), radius of

**Table 6** Structural validation and RMSD calculation results of all the 23 High-risk nsSNP mutants by several rigorous analysis tools

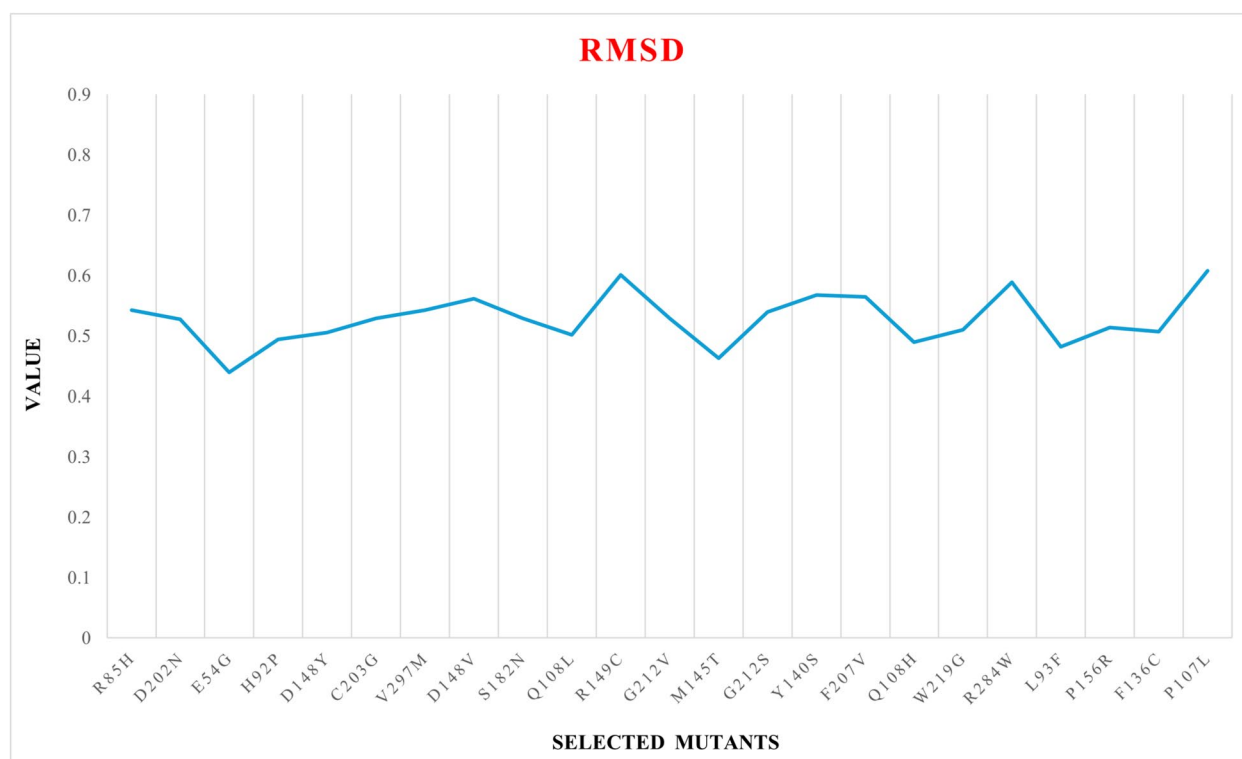
AA substitution	ERRAT Score	PROCHECK				Alpha Fold	PyMOL	TM Align
		Core	Allow	Generously	Disallowed	Relax Number	RMSD	TM score
R85H	91.8182	80.3%	14.2%	4.0%	1.6%	5	0.543	0.84317
D202 N	93.3131	81.4%	13.8%	4.3%	0.5%	5	0.528	0.82517
E54G	94.4444	80.5%	15.4%	2.7%	1.4%	5	0.440	0.78985
H92P	93.2927	80.2%	13.6%	3.5%	2.7%	5	0.494	0.83291
D148Y	94.7531	82.2%	12.4%	3.5%	1.9%	5	0.506	0.81940
C203G	90.8012	80.8%	14.1%	3.8%	1.4%	5	0.529	0.82390
V297M	93.0303	81.9%	13.0%	3.2%	1.9%	5	0.543	0.82911
D148 V	91.8429	80.3%	14.1%	4.1%	1.6%	5	0.562	0.81979
S182 N	90.7186	81.4%	13.6%	3.7%	1.3%	5	0.529	0.83106
Q108L	93.0091	80.5%	14.1%	3.8%	1.6%	5	0.502	0.82449
R149 C	92.1687	81.1%	13.2%	4.1%	1.6%	5	0.601	0.81542
G212 V	95.5836	80.5%	15.6%	2.5%	1.3%	5	0.529	0.82035
M145 T	93.1138	81.9%	13.0%	3.5%	1.6%	5	0.463	0.81857
G212S	95.4268	82.5%	12.1%	4.0%	1.3%	5	0.540	0.81713
Y140S	95.2681	81.4%	14.1%	3.2%	1.4%	5	0.568	0.82421
F207 V	91.4894	81.1%	13.8%	4.6%	0.5%	5	0.565	0.80465
Q108H	92.5	80.1%	14.2%	3.8%	1.9%	5	0.490	0.82186
W219G	94.1358	80.8%	13.8%	4.1%	1.4%	5	0.510	0.82389
R284 W	92.9664	82.2%	12.4%	3.5%	1.9%	5	0.589	0.82516
L93 F	92.8144	81.4%	12.4%	4.9%	1.4%	5	0.482	0.83330
P156R	94.3925	81.9%	13.5%	3.5%	1.1%	5	0.514	0.81200
F136 C	95.1515	82.7%	10.8%	4.1%	2.4%	5	0.507	0.81425
P107L	94.1935	80.8%	15.4%	2.0%	1.9%	5	0.608	0.80210

gyration (Rg), solvent-accessible surface area (SASA), and root-mean-square fluctuation (RMSF) were determined (Figs. 9 and 10). To understand the dynamic profiling of the native *AVPR1 A* protein, we simulated it. We considered it as a control (AVPR1 A\_APO) to evaluate the flexibility potential of the mutated protein form in complex with the best binding drug molecules (Y140S\_115237, R149 C\_115237, R284 W\_5073, P107L\_213046, and F207 V\_46200932). The PubChem CIDs of the studied compounds were retrieved as: Paliperidone (115237), Risperidone (5073), Lurasidone (213046), and Balovaptan (46200932).

The stiffness or flexibility of the ligand-protein complexes was evaluated based on their radius of gyration (Rg) values. Increased Rg values indicated greater conformational instability, while reduced Rg values reflected a more stable and rigid complex formation. Analysis of the Rg revealed that all protein-ligand complexes exhibited lower Rg values compared to the native (unbound) protein, suggesting increased structural compactness upon ligand binding (Fig. 9a). Among the complexes, R284 W\_5073 displayed the most rigid Rg profile, indicating enhanced conformational stability. Notably, while significant fluctuations were observed in most

complexes during the initial 0–100 ns simulation period, all complexes attained stability in the 101–200 ns time-frame, demonstrating sustained structural integrity over extended dynamics. To evaluate the firmness of the protein-ligand complexes, the root-mean-square deviation (RMSD) was measured and examined. Initial fluctuations (0–50 ns) were observed in all complexes except for the P107L\_213046 complex, which exhibited relatively stable behavior. Notably, four out of the five complexes (R149 C\_115237, R284 W\_5073, P107L\_213046, and F207 V\_46200932) achieved equilibrium after 51 ns and maintained stable conformations until the end of the simulation (200 ns). In contrast, the Y140S\_115237 complex displayed a fluctuating nature but still holds its potential as compared to the RMSD profile of the control (Fig. 9b).

To examine ligand-induced conformational changes and stabilization in the protein, Root Mean Square Fluctuation (RMSF) analysis was performed. Except Y140S\_115237, all the complexes displayed an overall similar range of fluctuations compared to the control, suggesting desirable protein residue flexibility. The Y140S\_115237 complex exhibited a bit higher fluctuations, particularly within the residue region 380–418 (Fig. 10a). An analysis of the Solvent Accessible Surface



**Fig. 5** RMSD line graph of selected deleterious mutants

Area (SASA) was conducted to assess protein folding, structural stability, and the influence of ligands on the protein's surface area. We noticed that all ligand-bound complexes exhibited a significantly reduced solvent-accessible surface area (SASA) profile compared to the control (AVPR1 A\_APO) over the entire 200 ns simulation trajectory. The persistently lower SASA values of all the complexes suggest enhanced structural compactness and stability, indicative of tighter packing and more favorable ligand-induced conformational dynamics (Fig. 10b). This consistency underscores the thermodynamic stability of the complexes, aligning with criteria for optimal binding interactions and reinforcing their potential as promising candidates for further investigation.

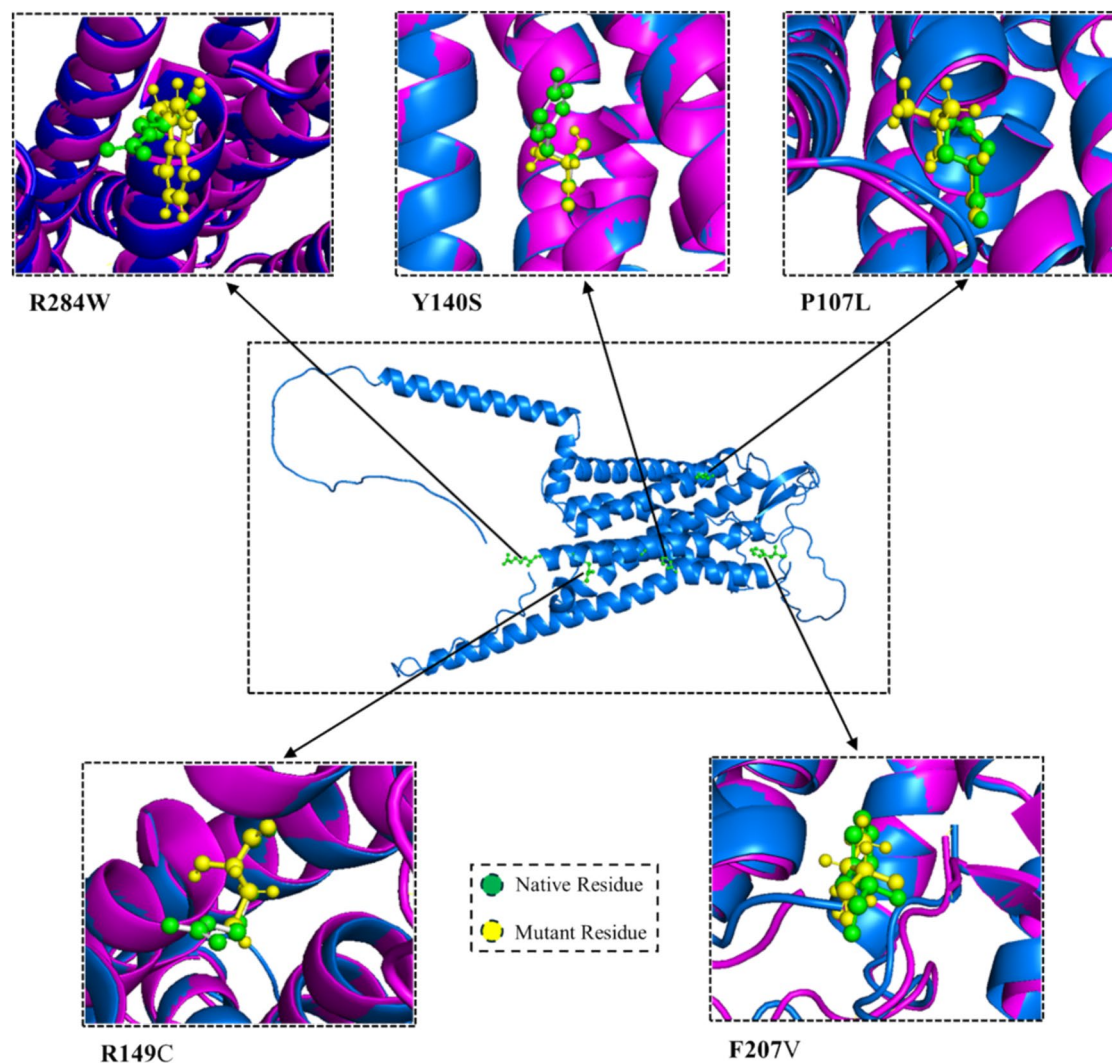
## Discussion

In this study, we utilized several *in silico* tools to identify the most detrimental missense mutations in the *AVPR1a* gene. We successfully identified 23 highly deleterious nsSNPs with a strong potential to impact *AVPR1a* gene activity drastically. To determine the most significant genetic alterations, an approach combining predictions from multiple tools was employed to examine the nsSNPs in the *AVPR1a* gene that have a high likelihood of impacting biological processes. AVP and *AVPR1a* have already been shown to be related to anxiety-like behavior

in multiple studies [70, 71]. *AVPR1a* gene methylation is directly associated with social behavioral changes [72]. A previous research study identified two microsatellite polymorphisms in the 5' flanking region of the *AVPR1a* gene in 115 autism trios. Furthermore, they successfully screened approximately 2 kb of the 5' flanking region and the coding region, identifying 10 single nucleotide polymorphisms (SNPs) [17]. Another study suggests that variations in the noncoding regions of the vasopressin 1a receptor gene (*AVPR1a*) are associated with a range of socioemotional traits in voles, chimps, and humans. These variations may influence behavioral changes by altering gene expression at specific sites [73]. The *AVPR1a* gene plays a key role in regulating social behaviors, including social interaction, social recognition, pair bonding, and aggression, primarily by encoding the vasopressin receptor 1A (V1aR). Due to a single nucleotide polymorphism (SNP) or point mutation in the *AVPR1a* gene, changes in gene expression occur, which could directly contribute to autism. For instance, the SNP of the *AVPR1a* gene (rs1042615) was identified in a previous study on autism susceptibility [18]. Altered vasopressin signaling may affect emotional processing and social memory in ASD.

Additionally, polymorphism in the *AVPR1a* gene is directly responsible for other health conditions, such





**Fig. 6** Comparison between native protein structure and nominated 5 mutant forms (R284W, Y140S, P107L, R149C, and F207V). The green color denotes native residues, while the mutated residues are yellow

as pain. The SNP of the *AVPR1a* gene (rs10877969) was previously identified as a candidate pain SNP, found in the promoter region of the *AVPR1a* gene on chromosome 12 [74]. The *AVPR1a* gene (12q14–15) has three microsatellite loci ((GT)<sub>25</sub>, RS1, and RS3) that are functionally significant in its promoter region [75]. The RS3 microsatellite is associated with altruism [76] and autism [17], while RS1 is responsible for Novelty Seeking and Harm Avoidance variation [20] and autism [21]. The role of the *AVPR1a* gene in regulating social behavior is supported by experimental research conducted in animal model: in particular, *AVPR1 A* antagonist led to a reduction in aggression [77], while decreased *AVPR1 A* resulted in reduced anxiety and social behavior deficits in voles [73]. The promoter region of *AVPR1 A* has

polymorphisms that may interact differently with specific transcriptional factors, affecting quantitative aspects like sociality in autistic children [25]. This volume of data persuades us that the *AVPR1a* gene should be a prime candidate for our rigorous investigation, aligning with the theme of this study.

We utilized the NCBI dbSNP database to identify all available SNPs for the *AVPR1a* gene. We then utilized GeneMANIA to assess the overall related gene network interaction types of the *AVPR1a* gene. After that, screening began with all accessible non-synonymous SNPs of the *AVPR1a* gene. Figure 2 shows the segmental SNP distribution graph of the *AVPR1a* gene. In this study, rs1260022270, rs1267958616, rs1321994497, rs1325662981, rs1337643184, rs1338176647, rs1369668995, rs1377891669, rs1417441306,

**Table 7** Binding affinities and Interacting residues of protein ligands of native *AVPR1a* and mutant R284 W, Y140S, P107L, R149 C, and F207 V that acquired through molecular docking and non-bond interaction analysis

Protein-Ligands	Binding affinities	Hydrogen bond Interactions	Hydrophobic Interactions	Protein-Ligands	Binding affinities	Hydrogen bond Interactions	Hydrophobic Interactions
Native- Lurasidone	-8.5	THR183	PHE187, PRO184, ALA215, TRP211, TRP219	P107L- Lurasidone	-8.3	ALA215	TRP211, TRP219, ALA215, VAL188
Native- Paliperidone	-8.2	ASN196, GLY6, PHE207	ALA5, ILE192, ALA9	P107L- Balovaptan	-8.2	ALA215	ILE208, PRO184, PHE187, VAL188, ALA215
Native- Bosentan	-8.0	SER213, SER314, GLN311	ALA5, SER4, ILE330	P107L- Paliperidone	-8.1	ALA215	ILE208, VAL188, ALA215
F207 V- Balovaptan	-8.2	ILE208, VAL207, PHE187	PHE187, TRP219, VAL188, ALA215, ILE208	R149 C- Paliperidone	-8.5	SER4, ARG116, ASN195, ASN196, SER325	TRP204, TRP111
F207 V- Naltrexone	-8.0	ARG214, VAL315	TRP322, ALA5	R149 C- Leucovorin	-8.4	TRP111, ARG116, TRP322, ALA5, GLY6, ASP8, ASP112, TYR115	ALA5
F207 V- Risperidone	-8.0	THR218	ALA215, VAL207, ILE208	R149 C- Lurasidone	-8.1	ASN196, SER4, ASP44	PRO7, VAL197
R284 W- Risperidone	-8.7	ARG116, ASN196, ASN195, SER325, ASP202	TRP204, ARG2, TRP111, VAL197	Y140S- Paliperidone	-8.4	SER4, ASP44, VAL321, ALA5, TRP322	ALA5, TRP204, VAL321
R284 W- Piperacillin	-8.4	ARG116, ASN195, GLU324, SER325, ARG2	PRO41, VAL197	Y140S- Lurasidone	-8.3	SER4, ASN196, ALA5, VAL194	PRO7
R284 W- Paliperidone	-8.3	SER213, ALA5, GLY6, PRO7, ASP8	TRP322, GLY6, PRO7, ALA5	Y140S- Brexpiprazole	-8.1	ARG214, TRP322, ALA5	SER325, TRP204, ALA5

rs1424280726, rs1440280008, rs1449556252, rs369710823, rs376518166, rs745458336, rs748572296, rs754449459, rs758567125, rs767540299, rs772227542, rs773269527, rs776488571, rs780705756 are considered most risky among total 402 non-synonymous SNPs of the *AVPR1a* gene according to our meta-analysis by different reliable computational methods. We initially screened deleterious nsSNPs and then performed confirmatory analysis and disease association studies for those nsSNPs. We predicted the probable impact of screened SNPs on the stability of the protein. Domain analysis identified two functional domains, and the output confirms the presence of 22 nsSNPs out of 23 in the functional domain area. Then, Conservation analysis was employed to identify the highly conserved regions of the target protein and pinpoint the screened risky nsSNPs within those regions of the protein. Post-translational modification (PTM) site prediction identified probable PTM sites, and the prediction of ligand binding sites pinpointed the active sites of our desired protein. We identified high-risk nsSNP consequences at both PTM sites and active sites of the *AVPR1a* protein, indicating that protein function may be significantly altered due to these polymorphisms in crucial sites of the protein. Homology modeling with AlphaFold2 was used to generate a relaxed 3D model of the protein sequences of mutants, and

the wild-type structure was downloaded from the AlphaFold Protein Structure Database. Structure validation was necessary to assess the accuracy level of the modeled structures, and we utilized the SAVES v6.1 server to validate these three-dimensional protein structures.

RMSD calculation measures the structural changes of protein due to mutations. Figure 5 presents the RMSD of all selected mutants in a line graph format. High RMSD mutants, including R284 W, Y140S, P107L, R149 C, and F207 V, were nominated as the 5 most detrimental variants. PyRx was used to assess the variation in protein-ligand binding affinities among native and mutant proteins. We docked 93 reference ligands to both wild-type and 5 nominated mutants. For the F207 V variant, Balovaptan, Naltrexone, and Risperidone were the top three compounds with the best binding energy profiles, suggesting potential efficacy against this mutant. Similarly, for P107L, the most promising compounds were Lurasidone, Balovaptan, and Paliperidone; for R149 C, Paliperidone, Leucovorin, and Lurasidone; for R284 W, Risperidone, Piperacillin, and Paliperidone; and for Y140S, Paliperidone, Lurasidone, and Brexpiprazole. These compounds exhibited strong binding affinities and may have therapeutic potential against their respective mutant variants. Furthermore, post-docking analysis was

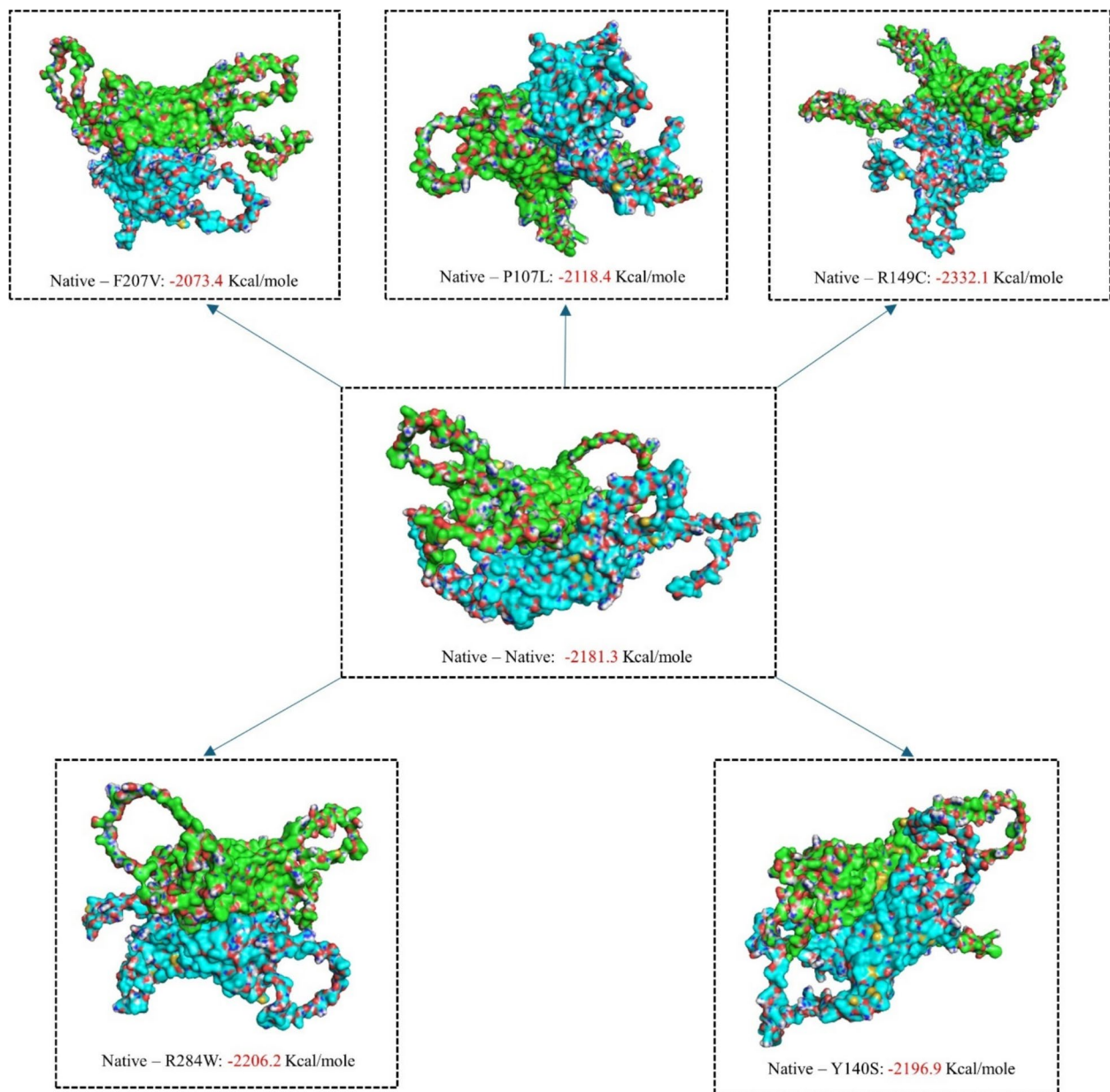


**Fig. 7** 2D visual representation of wild protein and mutant variants (R284W, Y140S, P107L, R149C, and F207V) with ligands including their binding residues

carried out using Discovery Studio to evaluate non-bond interactions. The purpose of docking is to examine how the activity of binding ligands correlates with 3D protein

structure and to identify potential compounds that are feasible for working against target mutant variants. To observe the disparity in protein-protein interaction

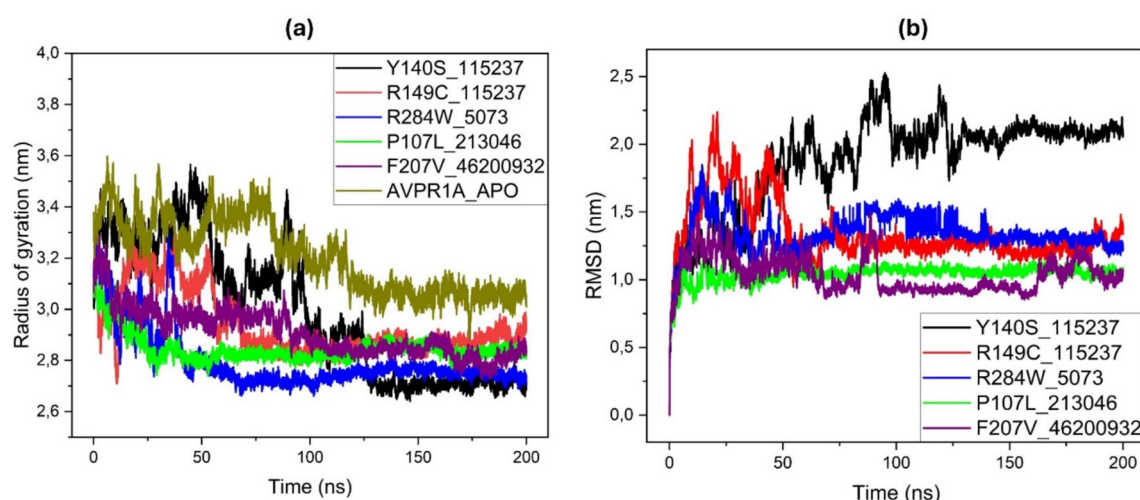




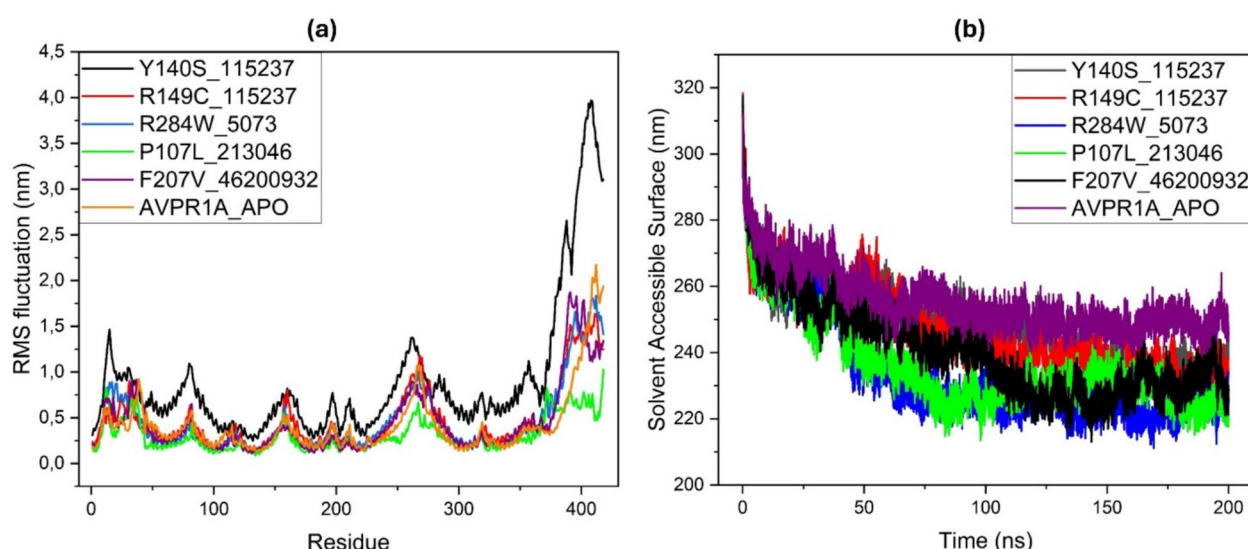
**Fig. 8** Binding affinities of 6 protein-protein complexes analyzed by ClusPro

levels, we utilized a widely used server called ClusPro. We observed significant variations in both protein-protein and protein-ligand docking outputs. Those significant variations suggest a noticeable impairment in the *AVPR1a* protein due to polymorphisms. We carried out a 200 ns molecular dynamics simulation to verify the docking outcomes and analyze the structural dynamics of the most promising drug molecules when bound to the five specified mutant forms of the target gene in an artificially produced relevant environment [78]. A detailed evaluation of dynamic properties was conducted by computing

critical metrics—including RMSD, RMSF, SASA, and Rg—from MD simulation data. The results collectively indicated stable structural conformations in most protein-ligand systems. All complexes showed considerable dynamic trends in MD simulations, reflecting stable binding affinities and confirming the structural robustness. The research methodology employed in this study is based on establishing a connection between the alterations and their molecular effects on the protein. When numerous programs or tools are used to achieve a single goal, the results are more dependable since each operates



**Fig. 9** **a** RG and **b** RMSD, analysis of the unligated control (AVPR1 A\_APO) and (Y140S\_115237, R149 C\_115237, R284 W\_5073, P107L\_213046, and F207 V\_46200932) complex. For each system, the MD simulations were performed for 200 ns



**Fig. 10** **a** RMSF and **b** SASA, analysis of the unligated control (AVPR1 A\_APO) and (Y140S\_115237, R149 C\_115237, R284 W\_5073, P107L\_213046, and F207 V\_46200932) complex. For each system, the MD simulations were performed for 200 ns

using a different algorithm. Even though our screened highly detrimental nsSNPs in this study were not tested under laboratory conditions, such as in vitro or other assays related to identifying the functional significance of mutations, the overall findings, obtained through rigorous meta-analysis using different computational approaches, highly prioritized those nsSNPs for further laboratory studies and clinical assessments. To further understand the specific function of these harmful nsSNPs on the *AVPR1a* gene, thorough wet lab research and trials on various model species may be beneficial. Future

genome association studies will be capable of identifying damaging SNPs associated with specific patients with autism and other health conditions based on the findings of this study.

## Conclusion

This study employed in-silico analysis to investigate the potential impact of nsSNPs on the structure, function, and stability of the *AVPR1a* protein. The presence of 23 mutations likely caused impairment in the structure and function of the *AVPR1a* protein, potentially



affecting its activity. We evaluated the influence of 23 non-synonymous single nucleotide polymorphisms on changes in protein stability and functional alterations. Additionally, we predict functional domains, probable active sites, and PTM sites of the *AVPR1a* protein and unveil the consequences of the presence of high-risk nsSNPs in these domain areas, ligand binding sites, and PTM sites. We propose 5 mutants based on high RMSD values. Then, we evaluated the variation in several interaction profiles between native and mutant proteins through analysis of protein-ligand and protein-protein docking interactions. It exhibits the effects of mutants on the protein's conformational changes, such as alterations in protein structural and functional properties. To fully understand and analyze these data on SNPs, it is necessary to conduct comprehensive clinical trials that include a diverse population. Additionally, experimental studies focusing on mutations are required to validate the findings.

#### Abbreviations

nsSNPs	Non-synonymous single nucleotide polymorphisms
ASD	Autism spectrum disorder
AVP	Arginine vasopressin
<i>AVPR1 A</i>	Arginine vasopressin receptor 1A
FBAT	Family-based association test
MAF	Minor allele frequency
SVM	Support Vector Machines
RSA	Relative solvent accessibility
LBS	Ligand binding sites
RMSD	Root-mean-square deviation
LGA	Lamarckian genetic algorithm
SAXS	Small-angle X-ray scattering

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11655-1>.

Supplementary Material 1.

#### Acknowledgements

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU). We also would like to thank the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh, for providing high-performance computers.

#### Mandated data types

Gene info: NCBI Gene ID: 552.

Sequencing data: Not applicable.

Genetic polymorphisms: The details about genetic polymorphisms of the *AVPR1a* gene were already stated in the S1 Table of the supplementary file.

Linked genotype and phenotype data: Not applicable.

Macromolecular structure: Not applicable.

Microarray data: Not applicable.

Crystallographic data for small molecules: Not applicable.

#### Clinical trial number

Not applicable.

#### Authors' contributions

M.D.K.J., M.A.I., and M.E.H. contributed to the conceptualization and design of the study. M.O.F., R.Z., and U.K.A. carried out data collection and initial analysis. B.S. and Y.K.T. contributed to methodology and validation. M.K. supervised the project and provided critical revisions. M.J. and M.E.A.Z. contributed to writing, review, and editing of the manuscript. All authors contribute significantly as well as they have read and approved the final manuscript.

#### Data availability

All relevant data generated or analysed during this study are included in this manuscript, and will be available upon reasonable request. The datasets retrieved during the current study are available in the NCBI (<https://www.ncbi.nlm.nih.gov/>) and Uniprot database (<https://www.uniprot.org/>). The NCBI Gene ID: 552 and Uniprot accession: P37288. The details about considered genetic polymorphisms for this study of the *AVPR1a* gene are listed in the S1 Table of the supplementary file. All relevant data generated through rigorous analysis are included in this manuscript.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 19 January 2025 Accepted: 29 April 2025

Published: 15 May 2025

#### References

1. Forsberg L, de Faire U, Marklund SL, Andersson PM, Stegmayr B, Morgenstern R. Phenotype determination of a common Pro-Leu polymorphism in human glutathione peroxidase 1. *Blood Cells Mol Dis*. 2000;26:423–6.
2. Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitzel NO, Hillier L, Kwok P-Y, Gish WR. A general approach to single-nucleotide polymorphism discovery. *Nat Genet*. 1999;23:452–6.
3. George Priya Doss C, Rajasekaran R, Sudandiradosh C, Ramanathan K, Purohit R, Sethumadhavan R. A novel computational and structural analysis of nsSNPs in CFTR gene. *Genomic Med*. 2008;2:23–32.
4. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebel MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins Struct Funct Bioinf*. 2010;78:365–80.
5. Jia M, Yang B, Li Z, Shen H, Song X, Gu W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. *PLoS One*. 2014;9:e104311.
6. Persico AM, Napolioni V. Autism genetics. *Behav Brain Res*. 2013;251:95–112.
7. Bailey A, Phillips W, Rutter M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *Autism*. 2013;17(2):159–96.
8. Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat Rev Neurosci*. 2011;12:524–38.
9. Koshimizu T, Nakamura K, Egashira N, Hiroshima M, Nonoguchi H, Tanoue A. Vasopressin V1a and V1b receptors: from molecules to physiological systems. *Physiol Rev*. 2012;92:1813–64.
10. Insel TR. The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. *Neuron*. 2010;65:768–79.
11. Egashira N, Mishima K, Iwasaki K, Nakanishi H, Oishi R, Fujiwara M. Role of vasopressin receptor in psychological and cognitive functions. *Nihon Yakurigaku Zasshi*. 2009;134:3–7.
12. Bielsky IF, Hu S-B, Szegda KL, Westphal H, Young LJ. Profound impairment in social recognition and reduction in anxiety-like behavior in

- vasopressin V1a receptor knockout mice. *Neuropsychopharmacology*. 2004;29:483–93.
13. Appenrodt E, Schnabel R, Schwarzberg H. Vasopressin administration modulates anxiety-related behavior in rats. *Physiol Behav*. 1998;64:543–7.
  14. Thibonnier M. Vasopressin receptors: molecular mechanisms in hypertension and cardiovascular diseases. *Mol Mech Hypertens*. 2006;173.
  15. Thibonnier M, Graves MK, Wagner MS, Chatelain N, Soubrier F, Corvol P, Willard HF, Jeunemaitre X. Study of V1-vascular vasopressin receptor gene microsatellite polymorphisms in human essential hypertension. *J Mol Cell Cardiol*. 2000;32:557–64.
  16. Hammock EAD, Young LJ. Variation in the vasopressin V1a receptor promoter and expression: implications for inter- and intraspecific variation in social behaviour. *Eur J Neurosci*. 2002;16:399–402.
  17. Kim S-J, Young LJ, Gonen D, Veenstra-VanderWeele J, Courchesne R, Courchesne E, Lord C, Leventhal BL, Cook EH Jr, Insel TR. Transmission disequilibrium testing of arginine vasopressin receptor 1A (AVPR1A) polymorphisms in autism. *Mol Psychiatry*. 2002;7:503–7.
  18. Wassink TH, Piven J, Vieland VJ, Pietila J, Goedken RJ, Folstein SE, Sheffield VC. Examination of AVPR1A as an autism susceptibility gene. *Mol Psychiatry*. 2004;9:968–72. <https://doi.org/10.1038/sj.mp.4001503>.
  19. Yirmiya N, Rosenberg C, Levi S, Salomon S, Shulman C, Nemanov L, Dina C, Ebstein RP. Association between the arginine vasopressin 1a receptor (AVPR1A) gene and autism in a family-based study: mediation by socialization skills. *Mol Psychiatry*. 2006;11:488–94.
  20. Meyer-Lindenberg A, Kolachana B, Gold B, Olsh A, Nicodemus KK, Mattay V, Dean M, Weinberger DR. Genetic variants in AVPR1A linked to autism predict amygdala activation and personality traits in healthy humans. *Mol Psychiatry*. 2009;14:968–75.
  21. Tansey KE, Hill MJ, Cochrane LE, Gill M, Anney RJL, Gallagher L. Functionality of promoter microsatellites of arginine vasopressin receptor 1A (AVPR1A): implications for autism. *Mol Autism*. 2011;2:1–8.
  22. Wilczyński KM, Auguściak-Duma A, Stasiak A, Cichoń L, Sieroń A, Janas-Kozik M. The role of single nucleotide polymorphisms within genes for oxytocin and vasopressin receptors in the presentation and severity of autistic traits. *Eur Psychiatry*. 2023;66:S102–S102.
  23. Talbot CF, Oztan O, Simmons SMV, Trainor C, Ceniceros LC, Nguyen DKK, Del Rosso LA, Garner JP, Capitanio JP, Parker KJ. Nebulized vasopressin penetrates CSF and improves social cognition without inducing aggression in a rhesus monkey model of autism. *Proc Natl Acad Sci*. 2024;121:e2418635121.
  24. Kantojärvi K, Oikarinen J, Kotala I, Kallela J, Vanhala R, Onkamo P, Järvelä I. Association and promoter analysis of AVPR1A in Finnish autism families. *Autism Res*. 2015;8:634–9.
  25. Yang SY, Kim SA, Hur GM, Park M, Park J-E, Yoo HJ. Replicative genetic association study between functional polymorphisms in AVPR1A and social behavior scales of autism spectrum disorder in the Korean population. *Mol Autism*. 2017;8:1–10.
  26. Bhagwat M. Searching NCBIs dbSNP database. *Curr Protoc Bioinformatics*. 2010;32:1–19.
  27. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38:W214–20.
  28. Hasnain MJU, Shoaib M, Qadri S, Afzal B, Anwar T, Abbas SH, Sarwar A, Talha Malik HM, Tariq Pervaz M. Computational analysis of functional single nucleotide polymorphisms associated with SLC26A4 gene. *PLoS One*. 2020;15:e0225368.
  29. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–4.
  30. Adzhubei IA, Schmidt L, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–9.
  31. Mahmud Z, Malik SUF, Ahmed J, Azad AK. Computational analysis of damaging single-nucleotide polymorphisms and their structural and functional impact on the insulin receptor. *Biomed Res Int*. 2016;2016:2023803.
  32. 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;9:16426.
  33. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7:e46688.
  34. 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:D336–42.
  35. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;76:7–20.
  36. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, Mort M, Cooper DN, Sebat J, Lakoucheva LM. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*. 2020;11:5918.
  37. 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;22:2729–34.
  38. Magesh R, George Priya Doss C. Computational pipeline to identify and characterize functional mutations in ornithine transcarbamylase deficiency. *3 Biotech*. 2014;4:621–34.
  39. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat*. 2009;30:1237–44.
  40. Leong IUS, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet*. 2015;16:1–13.
  41. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins Struct Funct Bioinf*. 2006;62:1125–32.
  42. 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:W306–10.
  43. Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics*. 2016;32:2542–4.
  44. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*. 2012;40:D306–12.
  45. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MDR. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res*. 2001;29:37–40.
  46. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigsmid P, Schafferhans A, Roos M, Bernhofer M. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res*. 2014;42:W337–43.
  47. Yang J, Roy A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 2013;29:2588–95.
  48. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4:1633–49.
  49. Cai R, Liu Z, Ren J, Ma C, Gao T, Zhou Y, Yang Q, Xue Y. GPS-MBA: computational analysis of MHC class II epitopes in type 1 diabetes. *PLoS One*. 2012;7:e33884.
  50. Zhao Q, Xie Y, Zheng Y, Jiang S, Liu W, Mu W, Liu Z, Zhao Y, Xue Y, Ren J. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res*. 2014;42:W325–30.
  51. Wang C, Tan X, Tang D, Gou Y, Han C, Ning W, Lin S, Zhang W, Chen M, Peng D. GPS-Uber: a hybrid-learning framework for prediction of general and E3-specific lysine ubiquitination sites. *Brief Bioinform*. 2022;23:bbab574.
  52. Varadi M, Bertoni D, Magana P, Paramval U, Piduchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*. 2024;52:D368–75.
  53. Skolnick J, Gao M, Zhou H, Singh S. AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J Chem Inf Model*. 2021;61:4827–31.

54. Yang Z, Zeng X, Zhao Y, Chen R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct Target Ther.* 2023;8:115.
55. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J.* 2011;101:2525–34.
56. Colovos C, Yeates TO. Verification of protein structures: patterns of non-bonded atomic interactions. *Prot Sci.* 1993;2:1511–9.
57. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26:283–91.
58. Mahfuz A, Khan MA, Deb P, Ansary SJ, Jahan R. Identification of deleterious single nucleotide polymorphism (SNP) s in the human TBX5 gene & prediction of their structural & functional consequences: an in silico approach. *Biochem Biophys Rep.* 2021;28:101179.
59. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci.* 2001;10:1470–3.
60. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005;33:2302–9.
61. Yuliana D, Bahtiar FI, Najib A. In silico screening of chemical compounds from roselle (*Hibiscus Sabdariffa*) as angiotensin-I converting enzyme inhibitor used PyRx program. *ARPN J Sci Technol.* 2013;3:1158–60.
62. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Chem Biol Methods Protoc.* 2015;1263:243–50.
63. Kawsar, Sarkar MA, et al. Chemical descriptors, PASS, molecular docking, molecular dynamics and ADMET predictions of glucopyranoside derivatives as inhibitors to bacteria and fungi growth. *Org Commun.* 2022;15(2):203.
64. Alom MW, Jibon MDK, Faruque MO, Rahman MS, Akter F, Ali A, Rahman MM. Integrated gene expression data-driven identification of molecular signatures, prognostic biomarkers, and drug targets for glioblastoma. *Biomed Res Int.* 2024;2024:6810200.
65. Kumer A, Chakma U, Rana MM, Chandro A, Akash S, Elseehy MM, Albogami S, El-Shehawi AM. Investigation of the new inhibitors by sulfadiazine and modified derivatives of  $\alpha$ -D-glucopyranoside for white spot syndrome virus disease of shrimp by in silico: quantum calculations, molecular docking, ADMET and molecular dynamics study. *Molecules.* 2022;27(4):3694.
66. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein–protein docking. *Nat Protoc.* 2017;12:255–78.
67. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1:19–25.
68. Schüttelkopf AW, Van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Biol Crystallogr.* 2004;60:1355–63.
69. Kadoura A, Salama A, Sun S. Switching between the NVT and NpT ensembles using the reweighting and reconstruction scheme. *Procedia Comput Sci.* 2015;51:1259–68.
70. Neumann ID, Landgraf R. Balance of brain oxytocin and vasopressin: implications for anxiety, depression, and social behaviors. *Trends Neurosci.* 2012;35:649–59.
71. Bielsky IF, Hu S-B, Ren X, Terwilliger EF, Young LJ. The V1a vasopressin receptor is necessary and sufficient for normal social recognition: a gene replacement study. *Neuron.* 2005;47:503–13.
72. Bodden C, van den Hove D, Lesch KP, Sachser N. Impact of varying social experiences during life history on behaviour, gene expression, and vasopressin receptor gene methylation in mice. *Sci Rep.* 2017;7:8719.
73. Barrett CE, Keebaugh AC, Ahern TH, Bass CE, Terwilliger EF, Young LJ. Variation in vasopressin receptor (*Avpr1a*) expression creates diversity in behaviors related to monogamy in prairie voles. *Horm Behav.* 2013;63:518–26. <https://doi.org/10.1016/j.yhbeh.2013.01.005>.
74. Roach KL, Hershberger PE, Rutherford JN, Molokie RE, Wang ZJ, Wilkie DJ. The AVPR1A gene and its single nucleotide polymorphism rs10877969: a literature review of associations with health conditions and pain. *Pain Manag Nurs.* 2018;19:430–44. <https://doi.org/10.1016/j.pmn.2018.01.003>.
75. Kazantseva AV, Kutlumbetova YY, Malykh SB, Lobaskova MM, Khusnutdinova EK. Arginine-vasopressin receptor gene (*AVPR1A*, *AVPR1B*) polymorphisms and their relation to personality traits. *Russ J Genet.* 2014;50:298–307.
76. Avinun R, Israel S, Shalev I, Gritsenko I, Bornstein G, Ebstein RP, Knafo A. AVPR1A variant associated with preschoolers' lower altruistic behavior. *PLoS One.* 2011;6:e25274.
77. Caldwell HK, Lee H-J, Macbeth AH, Young WS III. Vasopressin: behavioral roles of an "original" neuropeptide. *Prog Neurobiol.* 2008;84:1–24.
78. Du Y, Wang H, Chen L, Fang Q, Zhang B, Jiang L, Wu Z, Yang Y, Zhou Y, Chen B. Non-RBM mutations impaired SARS-CoV-2 spike protein regulated to the ACE2 receptor based on molecular dynamic simulation. *Front Mol Biosci.* 2021;8:614443.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.