Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Genomic Diversity to Molecular Mechanism: Computational Prediction and Structural Insights into Gain-of-Function PCSK9 Variants for Precision Cardiovascular Medicine

Akanksha Kulshreshtha
Netaji Subhas University of Technology
Yatender Kumar
Netaji Subhas University of Technology
Sonika Bhatnagar
sbhatnagar@nsut.ac.in

Netaji Subhas University of Technology

Research Article

Keywords: Structure activity relationships, PCSK9 LDLR complex, Gain or Loss of function, cardiovascular disease, precision medicine

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11799797/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title Page

From Genomic Diversity to Molecular Mechanism: Computational Prediction and Structural Insights into Gain-of-Function PCSK9 Variants for Precision Cardiovascular Medicine

Akanksha Kulshreshtha[a], Yatender Kumar[b] and Sonika Bhatnagar[a*]

[a] Computational and Structural Biology Laboratory,
Department of Biological Sciences and Engineering,
Netaji Subhas University of Technology,
Dwarka, New Delhi 110078

[b]Mammalian Cell Culture Laboratory,
Department of Biological Sciences and Engineering,
Netaji Subhas University of Technology,
New Delhi 110078, India

*Corresponding author
Email: sbhatnagar@nsut.ac.in
Ph.: +91 11 25000110
Fax: +91 1125099022

Abstract

Context

Cardiovascular disease (CVD) is the leading cause of death worldwide, and PCSK9 regulates plasma LDL-C levels via degrading LDLR. Gain-of-function (GOF) variants of PCSK9 increase cardiovascular risk, whereas loss-of-function (LOF) variants have a beneficial effect.

Methods

The Indigenome database was mined for non-synonymous single nucleotide polymorphisms of PCSK9 using five different tools. Consensus deleterious predictions from three methods were selected. Further, structural stability was studied using I-Mutant2.0 and PremPSv1.0.1. Missense3Dv2 and HOPEv1.1.1 delineated altered residue properties. Modeling and MD Simulation (MDS) of GOF variants of the PCSK9–LDLR complex was done using GROMACS2021.1. Protein structural networks (PSNs) were constructed using WebPSNv2.0. Two known (R357C and R496W) and three novel (S89W, Q413R and R680W) GOF variants of PCSK9 were mined. One was unique to the Indian population (Q413R). MDS of the GOF mutants showed lower residue flexibility with increased LDLR interfacial area comprised of more contacting residues and hydrogen bonds in comparison with wild type. Two known and three predicted GOF variants formed fewer clusters with broadened FEL, lower MMPBSA binding energies and remodeled PSN edges. These findings suggest that GOF associated effects are caused by changes in long range residue connectivity and the dynamic organization of the PCSK9–LDLR complex, providing mechanistic insights into variant specific functional consequences. Comparison of GOF structure, dynamics and PSN with LOF PCSK9 can be used to distinguish the two states. Based on the insights gained into structure activity relationships from this study, annotation of PCSK9 functional variants can advance precision medicine strategies for cardiovascular disease.

Keywords:

Structure activity relationships; PCSK9 LDLR complex; Gain or Loss of function; cardiovascular disease; precision medicine

Abbreviations:

Single nucleotide polymorphisms, SNPs; Nonsynonymous SNP, nsSNPs; Molecular Dynamics Simulations, MDS; Wild Type, WT; Mutant Type, MT; Root Mean Square Deviation, RMSD; Proprotein convertase subtilisin/ kexin type 9, PCSK9; Low density lipoprotein receptor, LDLR; Wild Type PCSK9-LDLR complex, WT PCSK9-LDLR; C-terminal Cysteine Histidine Rich

domain, CHRD; Loss of Function, LOF; Gain of Function, GOF; Principal Component Analysis, PCA; Free Energy Landscape, FEL; Genome-Wide Association Study, GWAS; Protein structural network,PSN; low-density lipoprotein cholesterol,LDL-C

1.Introduction

Single Nucleotide Polymorphisms (SNPs) account for roughly 90% of sequence diversity in the human genome [1]. Due to their high frequency and extensive distribution, SNPs are useful molecular markers for identifying genomic areas linked to various diseases in genome-wide association studies (GWAS) [2]. Nonsynonymous (ns) SNPs that cause amino acid alterations are particularly important because they can directly affect protein structure and function. Both coding and non-coding SNPs alter molecular and cellular processes, increasing disease vulnerability [3].

High-throughput sequencing and GWAS have identified millions of SNPs, with variant data catalogued in repositories such as dbSNP [4], 10000 Genomes [5], GWAS Catalog [6] and Indigenome [7]. Despite these advances, most allele frequency databases and computational prediction models are based predominantly on European populations [8]. The Indian population shows exceptional genetic diversity and distinct variant spectra [9-11]. Equity aware variant interpretation emphasizes the need to integrate local allele frequency reference datasets with calibrated functional evidence to minimize ancestry related bias and improve variant classification across diverse populations [8]. National initiatives such as the GenomeIndia Project and the IndiGen Programme are generating population specific allele frequency resources that improve variant interpretation, reduce ancestry related bias, and support equitable implementation of precision medicine in the Indian population.

Structure function studies have converted SNP data into mechanistic explanations underlying diseases. Complementary computational approaches and Molecular dynamics simulations (MDS) [12-15] have been extensively utilized to elucidate the conformational flexibility of proteins, protein ligand binding mechanisms and complex biomolecular interactions [16-20]. In silico approaches integrating deleterious prediction, disease association analysis, evolutionary conservation, protein stability prediction, and structural characterization identified eight potentially deleterious nsSNPs in SLC30A8, which may change the protein stability, structure, and molecular interactions associated with the development of type 2 diabetes and colorectal cancer [21]. Similarly, pathogenic nsSNPs in vascular endothelial growth factor A were studied by integrated computational approach of deleterious variant prediction, evolutionary conservation analysis, MDS, principal component analysis (PCA), protein–protein interaction analysis and binding free-energy calculations, which revealed mutation induced changes in dimer interface interactions and impair receptor binding, contributing to CVD [22].

Proprotein convertase subtilisin/kexin type 9 (PCSK9) is one such key molecule regulating plasma LDL-cholesterol (LDL-C) levels. PCSK9 binding to low density lipoprotein receptor (LDLR) orients it towards lysosomal degradation instead of endosomal recycling, thus reducing its surface expression and hepatic LDL-C clearance. Thus, it is involved in the pathophysiology of atherogenesis [23, 24] leading to subclinical atherosclerosis [25] and CVD. The therapeutic potential of PCSK9 was further demonstrated by the FDA approval of enlicitide (Lipfendra) in July 2026, the first oral PCSK9 inhibitor. Enlicitide binds to the region I of PCSK9 which comprises residues in the catalytic domain including Thr377, Phe379 and Ser381, and blocks PCSK9 mediated LDLR degradation [26].

X-ray crystallographic studies have showed that mature PCSK9 has a prodomain, a subtilisin-like catalytic domain and a C-terminal cysteine-histidine-rich domain (CHRD) [27, 28]. The prodomain is still attached to the catalytic domain after autocatalytic processing, thus blocking the catalytic site, but allowing the protein to bind to LDLR and promote its degradation. The structure of PCSK9 in complex with the epidermal growth factor (EGF) precursor

homology domain of LDLR encompassing the EGF-A, EGF-B, β-propeller, and EGF-C subdomains has been determined (PDB id:3M0C). MDS of PCSK9-LDLR complex has shown three distinct regions of interaction namely: a) Region I between Catalytic domain of PCSK9 and EGF-A domain of LDLR, b) Region II between Catalytic domain of PCSK9 and EGF-B of LDLR and; c) Region III between Prodomain of PCSK9 and β-propeller domain of LDLR [29].

Loss of function (LOF) mutations in PCSK9 reduce the binding of PCSK9 to LDLR, correlating with a 15–28% drop in circulating LDL-C leading to an estimated 47–88% decrease in lifetime myocardial infarction risk [30-32]. Conversely, PCSK9 variants with Gain-of-function (GOF) mutations enhance the degradation of LDLR, leading to reduced receptor recycling, increased plasma LDL-C levels, and an elevated risk of early-onset coronary heart disease[33, 34]. The GOF mutant D374Y and showed that D374Y GOF mutant binds the receptor 25 times more tightly than wild-type (WT) PCSK9, and is associated with hypercholesterolemia and early onset CVD [27, 35-37]. Similarly, the S127R and R496W mutations have been extensively characterized using mutagenesis, binding assays, and cellular studies [38, 39]. MDS and MM/PBSA established increased intermolecular binding affinity in the S127R, D374Y PCSK9 GOF variants [40]. A rationally designed T9 peptide inhibits the PCSK9 D374Y–LDLR interaction, leading to improved LDLR function [41]. Structure based design has been used for design of small molecules to disrupt the PCSK9–LDLR interface directly [42] and at a distant allosteric site [43]. However, few studies have utilized the genome data for prediction of PCSK9 GOF variants in the Indian population. Also, the effects of the variants on protein structure, dynamics, and LDLR binding remain poorly understood.

In this study, SNPs of *PCSK9* gene were retrieved from the Indigenome database [7], leading to the identification of two known and three predicted GOF variants. Of the five GOF mutations classified in this study, R357C and R496W stand out as already well established PCSK9 GOF variants [39, 44-47]. Three novel mutations, namely S89W, Q413R and R680W were predicted GOF. Of these, Q413R is uniquely found only in the Indian population. All five GOF variants were stable in 500 ns MD studies and showed increased interfacial area, increased number of residue contacts, lower residue flexibility and higher intermolecular binding energies to LDLR relative to the WT PCSK9. Cluster analysis, PCA and Free energy landscape (FEL) further validated the structural stability of all five GOF variants R357C, R496W S89W, Q413R and R680W. Analysis of the Protein structural network (PSN) of each mutant in comparison with WT showed the plasticity of the intra and inter protein interaction networks due to the mutations. The predicted GOF variants S89W (0.2%) and R680W (0.24%) may be present in approximately 2.8 and 3.4 million individuals, respectively, while R357C and R496W (0.05%) may each occur in approximately 0.7 million individuals in India with adverse impact on cardiovascular health. The MD parameters can also be used to distinguish the GOF variants of PCSK9 from LOF ones. Systematic implementation of computational tools followed by MDS can delineate the PCSK9 gain or loss of function impacting cardiovascular health and disease.

2. Methodology

2.1 Retrieval of nsSNPs data

Genetic variants for the *PCSK9* gene were identified using the IndiGenome resource, which archives whole-genome sequences from over a thousand Indian participants [7]. For functional characterization, SNPnexus v5 [48] was used

to translate these nsSNPs into amino acid positions. Such mapping is a prerequisite for our downstream analysis of substitutions and their potential effects on the PCSK9 protein.

2.2 Assessment of the Functional Impact of nsSNPs

A computational pipeline as outlined in Supplementary Figure S1 was used to analyze the impact of nsSNPs. Our approach integrated five specialized tools: PredictSNP v1.0 [48], SIFT [49], PROVEAN v1.15[50], Snap 2.1 [51], and SNP & GO [52]. Only those nsSNPs demonstrating consensus deleterious predictions across a minimum of three independent algorithms were considered for further analysis. To determine whether a mutation would destabilize the protein fold, two complementary predictive tools were utilized: a) I-Mutant 2.0, a support vector machine based tool that predicts the protein's stability upon single site mutation, using its primary sequence and tertiary structure [53]; and b) PremPS v1.0.1 uses a random forest regression model to predict changes in stability by measuring change in Gibbs free energy of folding [54]. The potential structural effects of the variants were investigated using Missense3D v2[55] which predicted mutation induced structural changes using structure based algorithms, and Have (y) Our Protein Explained (HOPE) server v1.1.1 assessed changes in residue properties, including size, charge, polarity, and hydrophobicity [56].

2*.3 Structure preparation of Five selected nsSNPs*

The crystal structure of the human PCSK9–LDLR complex (PDB ID: 3M0C) was retrieved from the Protein Data Bank (PDB). Missing residues and atoms were modeled using the CHARMM-GUI server [57], ensuring the reconstruction of loops and unresolved regions while maintaining proper stereochemistry and side-chain orientations. The refined structure was energy-minimized and validated using PROCHECKv3.5.4 [58] and ERRAT is currently available online through the SAVES v6.1 structure-validation server hosted by the UCLA-DOE Institute [59] and the respective mutations were introduced using Swiss-PdbViewer [60].

2.4 MD Simulation of Five selected nsSNPs

To investigate the structural dynamics of the mutant (S89W, R357C, Q413R, R496W and R680W) PCSK9-LDLR complex, all atom MDS were carried out in triplicates using GROMACS 2021.1 and the CHARMM36 force field [61]. System preparation followed the standardized CHARMM-GUI protocols in accordance with previously established methodology for PCSK9-LDLR complex [29]. Trajectory frames were saved at every 10 ps intervals for subsequent analysis. Structural visualization and inspection were performed using VMD 2.0, Discovery studio and Swiss PDB Viewer.

2.5 Interfacial Area and Interaction Analysis

The PISA v1.52 (Proteins, Interfaces, Structures, and Assemblies) server [62] was used to evaluate the structural integrity of the protein-protein interfaces. This tool quantified the solvent accessible surface area (SASA) and the interfacial area for both the WT and MT PCSK9–LDLR complexes. Comparing these values allowed for a precise assessment of how specific mutations alter the interaction surface and overall complex stability. Furthermore, the specific residue pair interactions leading to the PCSK9–LDLR binding were mapped using PDBsum1 v1.0.1 [63]. This provided an exhaustive interfacial profile by identifying hydrogen bonds, salt bridges, and non-bonded contacts. This residue-level mapping highlights how specific mutations either strengthen or disrupt the binding interface.

2.6 Clustering, PCA and FEL Analysis

Following the simulations, Trajectory clustering, PCA, and FEL analysis were performed using the integrated GROMACS 2021.1 tools gmx cluster, gmx covar/gmx anaeig, and gmx sham, respectively. RMSD-based clustering with 4.5Å cutoff was applied to identifies dominant conformational states of each complex. Representative cluster structures were selected for further analysis. PCA was employed to capture the dominant collective motions and the FEL was generated by projecting the trajectory onto the first two principal components (PC1 and PC2) with the corresponding Gibbs free-energy values calculated to identify energetically favorable conformational states.

2.7 *Protein Structural Network (PSN)*

PSN is a representation of internal connectivity of a protein, the amino acid residues are nodes and the interaction between them is termed as edge. The PSN was constructed using webPSN v2.0 [64]. The webPSN v2.0 employs a mixed PSN and Elastic Network Model-Normal Mode Analysis (ENM-NMA), to investigate the structural communication in proteins. The shortest paths of the network between the variant residue and the interfacial residues involved in PCSK9-LDLR interaction were compared in WT versus MT.

3. Results and Discussion

The Indigenome database lists 1,189 SNPs in the *PCSK9* gene. SNP Nexus identified nsSNPs corresponding to 27 positions within the protein sequence. The consensus of three algorithms yielded five predicted nsSNPs as potentially damaging. Further, I-mutant and PremPS predicted decrease in stability for all five variants. The results of consensus prediction and stability change are shown in Supplementary Table S1. Additional evaluation employing Missense3D and HOPE yielded further insights into the potential structural and functional impact of the substituted amino acid residues. All the positions had high ConSurf conservation scores, emphasizing their critical role in protein structure and function. The results of Missense3D, HOPE and ConSurf are shown in Supplementary Table S2.

Of the five variants predicted as deleterious, two were known GOF mutations as confirmed through literature search, while three were novel. R357C and R496W are known GOF mutations and S89W, Q413R and R680W are novel variants reported in this study. Q413R was found to be unique to the Indian population. As shown in Supplementary figure S2, S89W is present in the prodomain, while R357C and Q413R are located in catalytic domain. R496W and R680W are in the CHRD of PCSK9. The known GOF and predicted variants were further analyzed using MDS to gain insights into the structure-function relationships.

3.1 Known and predicted GOF structure showed deviations in comparison with WT due to PCSK9 prodomain and CHRD movements

MDS of WT PCSK9–LDLR complex showed three regions of interaction between the two molecules. Region I is composed of residues in the binding cleft of the PCSK9 catalytic domain and residues of the EGFA domain of LDLR. Region II comprises the interactions between the modeled loop residues D213–A219 of the PCSK9 catalytic domain and LDL receptor EGFB domain. Region III encompasses the interactions between the PCSK9 prodomain and the β-propeller domain of the LDLR. These regions of interactions are lost in the MDS structures of the LOF PCSK9–LDLR variants [29].

All the five variants (S89W, R357C, Q413R, R496W and R680W) stabilized early in the 500ns MDS run as shown in Supplementary Figure S3a. While S89W, R357C, Q413R and R680W followed the trajectory of the WT closely, R496W exhibited significant deviation. The mean RMSD values of 500 ns MDS structure in triplicate were 8.85 ±

0.58Å for S89W**,** 9.58 ± 1.59Å for R357C, 11.82 ± 0.70Å for Q413R, 11.70 ± 2.00Å for R496W, and 11.83 ± 1.17Å for R680W between 1060 Cα residues (Supplementary figure S4). However, in all the five variants, the three regions of interaction between PCSK9 and LDLR were found to be intact.

Individual structural RMSD comparison with the MDS WT PCSK9-LDLR complex revealed that:

- In the S89W variant, the catalytic domain remained largely unchanged. The large RMSD is attributed to the movement of prodomain and CHRD. These alterations were accompanied by changes in orientation the EGF-B, β-propeller and the EGF-C domains of LDLR (Supplementary Fig. S4a).
- The R357C variant exhibited extensive changes in the orientation of the prodomain, catalytic domain, hinge region, and parts of the CHRD. In LDLR, deviations were primarily seen in the EGF-B/β-propeller interacting regions (Supplementary Fig. S4b).
- The Q413R catalytic domain also remained partly unchanged while there was extensive reorientation of the hinge region and CHRD. The entire LDLR molecule showed large deviations while the EGF-A domain remained relatively conserved (Supplementary Fig. S4c).
- The R496W variant also showed a relatively unchanged catalytic domain with extensive deviations in the prodomain, hinge region and CHRD. Deviations were observed throughout LDLR suggesting that R496W (Supplementary Fig. S4d).
- The R680W variation caused widespread deviations throughout the prodomain and the entire CHRD, despite the limited structural changes within the catalytic domain. Extensive deviation was observed in LDLR (Supplementary Fig. S4e).

The structural deviation analysis demonstrated that the CHRD and the prodomain of PCSK9 were the regions most susceptible to conformational perturbations across all five variants. In LDLR, the entire molecule undergoes reorientation to improve contacts at Regions I, II and III.

The MDS trajectory of three predicted variants with structural impact was similar to that of WT PCSK9-LDLR complex and two known GOFs. As the three predicted novel variants showed similar RMSD and stabilization patterns along with retention of the regions of interactions till the end of the simulation run, they were further explored as potential GOF variants. The additional aspects of structure and dynamic trajectory were further evaluated in comparison with the WT complex.

In order to strengthen the findings, the MD graphs, values and structures were systematically compared with that of known and predicted LOF PCSK9-LDLR structures reported previously[29]. The LOF variants exhibited greater RMSD fluctuations, with more peaks than the WT throughout the trajectory (Supplementary Figure S5a). Consistent with these findings, the 500ns structures of the LOF variants showed very little structural superposition.

3.2 Known and predicted GOF variants show increase in the interfacial area with increased residue contacts in region II and region III relative to the WT PCSK9-LDLR

The interface between PCSK9 and LDLR was characterized using interfacial area calculation, number of interacting residues, quantitation of hydrogen bonds of PCSK9 with LDLR and RMSD based clustering transitions. It was found that:

a. Increase in interfacial area between PCSK9 and LDLR was observed in two known and three novel predicted GOF variants as determined using the PISA server. In comparison, the interfacial area and interacting regions were observed to be lost in the LOF variants.

   In known and predicted GOF variants, as shown in Supplementary Table S3:

- Region III showed almost 2 times increase in the interfacial area in all the GOF variants (Known and predicted).
- The maximum increase in the interfacial area in region I and II is exhibited by S89W.
- The maximum increase in the interfacial area in region III is exhibited by GOF variant Q413R. PISA analysis revealed that the Pro–catalytic interface remained relatively conserved across all five variants with comparatively low variability. In contrast, the Region I+II interface displayed greater variation with R357C showing the largest interfacial area in this region. Q413R exhibited the highest Region III interfacial area among the five variants. The observed increase in interfacial areas was consistent with the increase in the number of contacts identified in each region, as R357C shows largest contacting residues in region I and II and Q413R shows largest contacting residues in region III as shown in Supplementary Table S4.

b. Further, the specific residue contacts contributing to the observed increase in all the variants were examined. In comparison with the WT, known and predicted GOF mutants generally exhibited an increased number of PCSK9–LDLR contact residues, indicating enhanced intermolecular interactions (Supplementary Table S4). R357C showed the greatest increase in total contacts of all the three regions, followed by R680W, Q413R, R496W and S89W.

c. All the variants except S89W demonstrated increased hydrogen bonding. The highest number of hydrogen bonds were seen in Q413R, followed by R680W, R357C and R496W. These findings suggest that in two known and three predicted GOF variants, strengthen the PCSK9–LDLR interface as shown in Figure 1 and Supplementary Table S5a and S5b.

d. RMSD-based clustering analysis was carried out to obtain conformational clusters and identify crucial protein transitions during the simulation. The representative structures of the three most populated clusters were selected and superimposed, as they accounted for over 75% of the trajectory frames for all proteins. As shown in Supplementary Figure S6 two known and three predicted GOF variants exhibited intact interactions across Regions I, II, and III maintaining a stable PCSK9–LDLR complex. The interfacial area was found to be enhanced in >70% representative structures from the clusters as shown in Supplementary Table S6. WT PCSK9-LDLR showed total 04 clusters, further all the GOF have showed cluster numbers closer to the WT S89W formed 06 clusters, R357C formed 04 clusters, Q413R formed 05 clusters, R496W formed 09 clusters and R680W formed 05 clusters. A smaller number of clusters indicates, less deviation and more stability among the GOF variants. In contrast LOF variants showed a large number of clusters ranging from 10 to 36 clusters in consonance with their structural instability. LOF variants also showed loss of contact residues in interacting regions along with the dissociation of PCSK9-LDLR complex.

Collectively, the expanded interfacial area in two known and three predicted GOF variants is accompanied by an increase in the number of contact residues across all the three contact regions between PCSK9 and LDLR, leading to

merging of regions I and II, with formation of a continuous interface (Figure 1). An enhanced hydrogen bonding network was observed for most of the variants except S89W. Along with the small number of RMSD clusters obtained during MDS, this provides structural evidence for a tighter and more stable PCSK9–LDLR interaction in these five variants, consistent with a GOF profile. Further the residue flexibility and energetics of the interactions were also probed.

3. 3 Known and predicted GOF variants show lower RMSF in comparison with the LOF variants of PCSK9

The average RMSF values reflect the extent of residue level fluctuations in the PCSK9–LDLR complex during the MD simulation. The RMSF profiles indicate that established and putative GOF variants exhibit residue level fluctuations similar to the WT PCSK9–LDLR complex (Supplementary Figure S3b). Higher flexibility was observed in LDLR region than PCSK9 region in all the known and predicted GOF variants as shown in the Supplementary figure S3b. In contrast, the LOF variants showed very high flexibility in almost all the residues of PCSK9-LDLR complex as shown in Supplementary Figure S5b.

3.4 Known and predicted GOF variants of PCSK9 show lower MMPBSA interaction energies with LDLR than the WT PCSK9-LDLR

The overall binding free energy was more negative in known and predicted GOF variants in comparison with the WT PCSK9–LDLR complex. The binding energy was observed in the order: WT<R357C<S89W<R680W<Q413R <R496W as indicated in the Supplementary Table S7. The decrease in binding free energy suggested increased affinity of interaction between PCSK9 and LDLR consistent with previous MD and MMPBSA analyses showing increased PCSK9–LDLR affinity for GOF variants such as D374Y and S127R [40]. D374Y has been shown to promote LDLR degradation, thus contributing to the GOF phenotype with elevated LDL-C levels [27, 65]. Conversely, the LOF variants demonstrated reduced overall binding energy in comparison with the WT PCSK9–LDLR complex, signifying fewer interactions between PCSK9 and LDLR.

The residue wise contributions to binding free energy are shown in Supplementary Table S8. As observed in Region I, II and III highlighted residues consistently show lower binding free energy in known and predicted GOF variants with respect to the WT. In PCSK9 R680W exhibited the lowest binding free energy in Region I, whereas R496W showed the lowest binding free energies in Regions II and III. For LDLR, Q413R exhibited the lowest binding free energies in Regions I and III, while R496W showed the lowest binding free energy in Region II. Further, a few additional interactions in Region III were also observed in known and predicted GOF variants like G130, E145, F150 and Q152.

The hotspot residues contributing to highest energy gain during LDLR interactions was found to vary in all the GOFs in comparison with the WT PCSK9-LDLR complex. However, additional residues gained in region II and III were different for each GOF variants. In Region I, no new residue interactions were observed in the GOF variants, but the energetic contributions of residues already involved in LDLR binding were enhanced. Regions II and III displayed increased energetic contributions from existing interacting residues, accompanied by the formation of additional GOF variant specific residue interactions, suggesting a variant specific change in the interacting residues that stabilizes the PCSK9–LDLR interaction.

Overall higher interactions were observed in established and putative GOF MT forms indicating that these residues contribute more strongly to the overall binding affinity of the complex and play a more significant role in stabilizing the interaction between PCSK9 and LDLR in the MT forms. This strongly confirmed the classification of the three novel variants as GOF. In contrast, the MMPBSA residue contributions toward complex stabilization were partially or completely lost in LOF variants.

3.5 GOF variants show similar Rg and SASA profile in comparison with the WT PCSK9-LDLR

All the GOF variants showed a stable Rg profile, similar to that of the WT PCSK9-LDLR throughout the run suggesting overall protein structure fold maintenance (Supplementary Figure S3c). Further, all the variants showed similar SASA profile as that of the WT PCSK9-LDLR indicating compaction of the overall protein complex as shown in Supplementary Figure S3d. In contrast, the LOF variants have higher Rg profile relative to the WT PCSK9-LDLR complex and similar SASA profiles in comparison to the WT PCSK9-LDLR complex (Supplementary Figure S5c and Supplementary Figure S5d).

3.6 GOF variants showed displaced, distorted or compacted dimensions of protein motions and fluctuations in PCA

The PCA results illustrate the conformational space sampled by each system and indicate the dominant collective motions. PCA distinguishes the variations in conformational dynamics between WT PCSK9-LDLR and known and putative GOF complexes (Figure 2). WT PCSK9-LDLR complex spans PC1 (−10 to 30) and PC2 (−25 to +25), indicating a wide distribution across both components as shown as in Supplementary Table S9. The conformational flexibility was compared with the following variants:

- S89W motions show a slightly reduced spread relative to the WT. This suggests that S89W samples a somewhat restricted conformational space, indicating partial stabilization.
- R357C motions show compaction along the y-axis, thus indicating a restricted conformational space along PC2.
- Q413R spans a broader range than the WT distribution, along PC1 and PC2. However, the shape of the distribution changed considerably with a single indent in between.
- R496W motions show compaction along the y-axis, thus indicating a restricted conformational space along PC2, as seen in R357C.
- R680W motion show compaction along the y-axis, thus indicating a restricted conformational space along PC2, as seen in R357C and R496W. The compaction of the movements of GOF variants along PC1 and PC2 become more evident when compared with the LOF variants, that showed diffuse and spread-out spectra (Figure 2 and Supplementary Table S9).

3.7 GOF variants showed broadening and changing of the energy minima in comparison with the WT PCSK9-LDLR

FEL study further aids to the understanding of conformational stability and energy distribution of the system. All the variants showed more low energy metastable regions (blue color) in both the two-dimensional and three-dimensional FEL as shown in Supplementary Figure S7a and Supplementary Figure S7b respectively. Additionally, the structures corresponding to the metastable states were extracted for GOF variants along with the WT (Supplementary Figure S8). It was observed that all GOF variants retained interactions across Regions I, II, and III even in the low energy metastable states.

3.8 GOF mutations induce remodeling of the PSN characterized by altered interaction pathways

The number of residue contacts in PCSK9 and the interprotein contacts with LDLR were calculated by constructing PSNs of the WT and MT complexes. In comparison with the WT, all the five GOF variants have lower global connectivity but greater PCSK9–LDLR connectivity.

As shown in Supplementary Table S10, the total number of edges in the PSN for the WT PCSK9–LDLR complex were 1254. Thus, the reduced total connectivity in the GOF variants may reflect network edge redistribution gain of inter-protein interactions accompanied by loss of intra-protein interactions. Concomitantly, the total number of network links is reduced. In direct contrast to the GOF, all the LOF variants showed higher global connectivity but markedly reduced PCSK9–LDLR connectivity compared to the WT suggesting PSN connections were altered in a mutation dependent manner. GOF mutations induce mutation specific remodeling of the PSN, characterized by shortening of interaction pathways, emergence of new pathways and redistribution of network connectivity toward residues participating in the PCSK9-LDLR interface.

The shortest paths from each mutant residue to the interface interacting residues were traced by evaluating changes in network edges associated with the mutant residue relative to the corresponding WT network. The comparison of each of the MT interaction networks from the mutated residues to the nearest interface residues was traced and is shown in Supplementary Figure S9. In WT the interaction pathways traversed a larger network within the protein before reaching the interface (Supplementary Figure S9 1a). The minimal number of network linkages from the mutation site to the PCSK9-LDLR interface residues in the GOF variants are shown in Supplementary Table S11 and discussed below:

- S89W exhibited a longer communication pathway to P71 in Region III compared with WT, while showing shortened pathways to interface residues in Regions I and II. As shown in Supplementary Figure S9 1b, the minimum path connecting S89W to the interacting residues of region I and II (C378, F379, S381, R215 and H217) were redirected upstream in the interprotein network.
- R357C mutation elongates the pre-existing pathways to Region I (F379 and P155) and T214 in Region II, while simultaneously establishing new interaction pathways to F216 in Region II and P120 in Region III as shown in Figures S9 2a and 2b.
- Q413R significantly alters the PSN by increasing the path to F379 and shortening pathways connecting to D367 and to Region II interface residues (R215 and F216) while also establishing a new communication route to Region III (G106) as shown in Supplementary Figure S9 3a and 3b
- In the WT structure, residue R496 showed no detectable minimum path to the Region I and Region II interface residues, whereas the R496W mutant induces extensive PSN rewiring, with marked elongation of the path to D367 accompanied by the emergence of new communication routes toward Region II (F216) and Region III (L74 and M126) as shown in Supplementary Figure S9 4a and 4b
- R680W improves network connectivity to the PCSK9-LDLR interface, significantly shortening paths to Regions I (D367 and F379), II (F216), and III (L74 and M126) as shown in Supplementary Figure S9 5a and 5b.

3.9 Clinical implications of S89W PCSK9-LDLR, R357C PCSK9-LDLR, Q413R PCSK9-LDLR, R496W PCSK9-LDLR and R680W PCSK9-LDLR

The ClinVar database reports 1,548 SNP variations in the *PCSK9* gene. Further in the database 28 variations are classified as "pathogenic" and 14 as "likely pathogenic" variants in *PCSK9* gene [66]. As per Hummelgaard et al.,2023 47 GOF variants are known in *PCSK9* gene [46]. A few PCSK9 GOF variants and their reported effects are shown in Supplementary Table S12.

The integrated computational pipeline identified S89W, R357C, Q413R, R496W, and R680W as potential GOF variants in this study. Consensus-based predictions, together with ConSurf, Missense3D, and HOPE analyses, indicated that these variants show significant structural and functional impact due to their conserved positions and altered physicochemical properties. R357C and R496W were established GOF variants as discussed under:

- R357C PCSK9-LDLR is known to be associated with familial hypercholesterolemia in multi-ethnic populations [45, 67]. Western blot studies and PCSK9 expression in transfected HEK293 cells showed that R357C is a GOF mutation leading to increased LDL receptor binding affinity and decreased LDL uptake [44]. The Allele frequency of R357C (rs148562777) is 0.0005 (0.05%) corresponding to approximately ~ 1.4 million people in India carrying at least one copy of that variant allele. Thus, the prediction using integrated computational pipeline and MDS of R357C as reported in this study were validated by the findings from experimental and population-based studies.
- R496W PCSK9 variant is also strongly correlated with familial hypercholesterolemia, due to increased LDL receptor degradation and elevated plasma LDL-C levels. The R496W variant has been identified in multiple cohorts, including a Turkish familial hypercholesterolemia cohort with early onset coronary heart disease in seven of eight cases and a Saudi family-based study, where it is associated with elevated LDL-C levels [39, 47]. The first known case of the R496W variant in PCSK9 occurred in a patient also heterozygous for a LOF variant E228K in LDLR gene. The combined effect of both was severe hypercholesterolemia and early onset of cardiovascular illness [68]. The R496W variant was identified as one of the more severe PCSK9 GOF mutations in a multicenter familial hypercholesterolemia study, associated with elevated LDL-C levels [69]. Two different mechanisms of GOF mutation have been reported to operate in PCSK9. First, PCSK9 has a higher affinity for LDLR, resulting in increased receptor interaction and degradation as shown in this study. The second mechanism, based on previous experimental studies is impaired association of PCSK9 with LDL particles. Mutations S127R and R496W impair the interaction of PCSK9 with LDL, which may decrease the LDL binding with PCSK9, enhancing PCSK9 interaction with LDLR, therefore promoting the degradation of LDLR. These observations suggest that PCSK9 GOF can result from increased receptor binding affinity, but also from decreased LDL mediated inhibition of PCSK9 [70-72]. R496W has previously been reported to lead to PCSK9 GOF due to the loss of its binding to LDL particles. Interestingly, our study also suggested that R496W has increased binding affinity to LDLR, a potential dual mechanism of GOF by both decreased PCSK9–LDL association and increased PCSK9–LDLR interaction. The present finding gives a new structural insight into the molecular mechanism of the R496W GOF effect. The Allele frequency of R496W

(rs374603772) is 0.0005 (0.05%) corresponding to approximately ~ 1.4 million people in India carrying at least one copy of that variant allele.

The present study predicted that S89W, Q413R and R680W represent novel GOF variants associated with higher circulating LDL-C leading to hypercholesterolemia. While, S89W and R680W are known SNPs, there is currently no available information regarding their impact in any population and Q413R SNP appears to be unique to the Indian population. The allele frequencies of the novel predicted GOF variants are discussed below in order to estimate their potential health impact in the Indian population.

- S89W (rs771978846) is 0.0020 (0.2%) corresponding to approximately 2.8–3.0 million individuals in India carrying at least one copy of that variant allele (based on India's population of ~1.43 billion).
- Q413R is 0.0005 (0.05%) corresponding to approximately ~ 1.4 million people in India carrying at least one copy of that variant allele.
- R680W (rs533555352) is 0.0024 (0.24%) corresponding to approximately 3.43 million people in India are likely to be affected. Identifying novel variants provides valuable insights into the individual predisposition or adaptation of diseases and highlights the genetic diversity present across various populations. Understanding these genetic predispositions enables development of personalized therapeutic strategies for more effective disease management.

4. Conclusion

Systematic SNP variant prediction, structural comparison and functional impact characterization studies are few. Such gaps may affect interpretation of genetic variation and its use in precision medicine. Therefore, detailed characterization of these variants can provide population specific insights and aid the development of better precision medicine approaches. The Indian population is diverse with many ethnic, linguistic and geographically distinct groups but is under represented in global genomic studies.

CVD remains a leading cause of morbidity and mortality worldwide and a major global health problem. PCSK9 is one of the major regulators of cardiovascular risk by maintaining plasma LDL-C levels through the regulation of LDLR degradation. Genetic variants of PCSK9 can alter its binding to LDLR and thus affect the metabolism of LDL-C, with implications for cardiovascular health and disease. MDS of the WT PCSK9-LDLR complex showed three interaction regions between the PCSK9 and LDLR. The structural basis of action of one known (R93C) and two predicted LOFs (N298D and A511V) was elucidated [29].

In the present study, the SNPs of PCSK9 were mined using Indegenome. The consensus of three out of five bioinformatics tools were used for the prediction of deleterious nsSNPs. Additional information about the stability of the mutation was evaluated using I-Mutant and PremPS. Further Missense3D and Hope was used for the information regarding the change in residue properties. Five of the potentially deleterious variants were further analyzed. Two were known variants (R357C and R496W) and three were predicted novel candidates (S89W, Q413R, and R680W). Notably, one predicted GOF variant Q413R was unique to the Indian population. S89W is present in the prodomain, R357C and Q413R in the catalytic domain, and R496W and R680W in the CHRD of PCSK9. The impact of the

mutation on the structure, dynamics and interaction with LDLR was assessed using 500ns MDS studies. Post-MDS analysis showed that:

a. All five variants stabilized early during the 500 ns MDS trajectory, with S89W, R357C, Q413R, and R680W closely following WT trajectory, whereas R496W showed deviation.
   - Analysis of the 500ns structure revealed that conformational perturbation in the PCSK9 CHRD and prodomain, while LDLR underwent reorientation to improve contacts across these regions.
   - Two known and three putative variants exhibited increased PCSK9-LDLR interfacial area particularly in Regions II and III, in comparison with WT. The maximum increase in the interfacial area in region I and II is exhibited by S89W. The maximum increase in the interfacial area in region III is exhibited by GOF variant Q413R.
   - Two known and three putative GOF variants showed enhanced interactions across Regions I–III. In comparison with the WT, GOF mutants generally exhibited an increased number of PCSK9–LDLR contact residues, indicating enhanced intermolecular interactions.
   - Except for S89W, all variants showed increased hydrogen bonding, further supporting strengthened PCSK9-LDLR interactions.
   - RMSD-based clustering analysis was carried out to identify conformational transitions during the simulation. Two known and three putative variants exhibited a small number of clusters comparable to WT. Each representative structure of the cluster maintained intact interactions across Regions I–III and indicating a stable PCSK9-LDLR complex.

b. RMSF profiles showed residue level fluctuations in two known and three putative variants comparable to the WT PCSK9–LDLR complex.

c. GOF variants exhibited more favorable overall binding energies than WT, following the order WT < R357C < S89W < R680W < Q413R < R496W. Among GOF variants, R680W showed the most favorable binding in PCSK9 Region I, while R496W was most favorable in Regions II and III; for LDLR, Q413R showed the lowest energies in Regions I and III, whereas R496W was most favorable in Region II. Additional Region III contacts involving E145, F150, and Q152 were also observed. This strongly indicated that the predicted variants as GOF.

d. All the GOF variants show a stable Rg profile and similar SASA profiles as of the WT PCSK9-LDLR indicating compaction of the overall protein complex.

e. PCA showed more compact conformational distributions in GOF variants suggesting restricted conformational sampling and greater structural stability. All the GOF variants showed more low energy regions in FEL.

f. GOF variants induce PSN remodeling characterized by altered interaction pathways. The WT PCSK9–LDLR complex contained 1,254 total PSN edges. All five GOF variants showed reduced global connectivity but increased PCSK9-LDLR connectivity relative to WT PCSK9-LDLR. GOF variants preferentially enhancing intermolecular links despite reduced overall network connectivity. Shortest communication paths from each mutation site to PCSK9-LDLR interface residues were traced by comparing network edges between mutant and WT networks, revealing mutation specific changes in the minimum number of links and distinct communication

pathways that differentiate GOF mutants from the WT and the LOF mutants, providing a useful PSN-based distinction between the WT, GOF and LOF mutants.

Unlike GOF variants, LOF variants showed distinct MDS profile, structural and interaction patterns:

i. LOF variant complexes exhibited higher RMSD fluctuations and reduced conformational stability. Interactions in Region I and Region II showed partial to complete loss (variant specific), while Region III exhibited complete loss of interactions in all the LOF variants. Consequently, the interfacial area was decreased in Regions I, II and III in all the LOF variants. All the LOF variants showed decreased hydrogen bonding, further supporting weakened PCSK9-LDLR interactions.

ii. RMSD-based clustering of LOF variants showed disrupted interactions across Regions I–III and higher cluster numbers than WT, indicating increased conformational heterogeneity and reduced stability of the PCSK9–LDLR complex.

iii. Overall RMSF was found to be more in case of all the three LOF variant PCSK9-LDLR complex.

iv. All the three LOF variants showed marked reduction in overall binding affinity.

v. PCA and FEL analyses revealed greater conformational heterogeneity and reduced stability of the PCSK9–LDLR complex in LOF variants.

vi. LOF variants induce PSN remodeling characterized by higher global connectivity but markedly reduced PCSK9-LDLR connectivity.

The study combines the MDS, binding-energy, interface and PSN analyses to demonstrate the mutation specific remodeling of the PCSK9-LDLR complex and PSN pathways. The presence of unique interaction patterns in GOF and LOF variants, including the predicted novel GOF variants, provides a mechanistic basis for understanding altered PCSK9 function. These findings may facilitate improved interpretation of PCSK9 variants, particularly those identified in the Indian population, by providing population specific insights into their structural and functional effects. The study explores the structural activity relationships of two known GOF mutations of PCSK9 leading to the functional impact on LDLR binding thus causing CVD. The structure, function and dynamics of the PCSK9-LDLR complex was investigated for three other predicted GOF variants.

All parameters of GOF mutations were systematically compared with previous LOF mutations. In all cases, the MDS and PSNs of GOF mutants were differentiated from that LOF mutants. This shows the feasibility of this computational approach for prediction of GOF and LOF variants of PCSK9. GOF variants of PCSK9 bind with LDLR with high affinity, leading to the lysosomal degradation of the complex and increased circulating LDL-C in the blood. Thus, GOF variants are associated with increased CVD risks. Taken together, the known and predicted GOF variants reported in this study are likely to affect the CVD risks in 6.9 million people in the Indian population. By distinguishing variants with differential effects on PCSK9 function, this study provides a basis for prioritization of CVD risk for future precision-medicine approaches.

Funding

The authors would like to acknowledge funding from Department of Biotechnology grant number BT/PR40197/BTIS/137/68/2023

Disclosures

The authors have no relevant financial or non-financial interests to disclose.

Ethical approval

This study lacks human and animal studies.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors did not use generative AI or AI-assisted technologies in the development of this manuscript.

Author contribution statement*:* AK- Conception, planning, data collection, analysis, and manuscript writing; YK- Conception, planning, and manuscript writing; SB- Conception, planning, manuscript correction and supervision.

Acknowledgement

The authors acknowledge computational support from the NSUT HPC facility.

Data Availability

Supplementary file is provided for further information

References

1. Rozman V, Kunej T (2018) Harnessing omics big data in nine vertebrate species by genome-wide prioritization of sequence variants with the highest predicted deleterious effect on protein function. OMICS 22:410-421. https://doi.org/10.1089/omi.2018.0046
2. Reza MN, Ferdous N, Emon MTH, Islam MS, Mohiuddin AKM, Hossain MU (2021) Pathogenic genetic variants from highly connected cancer susceptibility genes confer the loss of structural stability. Sci Rep 11:19264. https://doi.org/10.1038/s41598-021-98547-y
3. Nuckel H, Frey UH, Bau M, Sellmann L, Stanelle J, Durig J, et al (2007) Association of a novel regulatory polymorphism (-938C>A) in the BCL2 gene promoter with disease progression and survival in chronic lymphocytic leukemia. Blood 109:290-297. https://doi.org/10.1182/blood-2006-03-007567
4. Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28:352-355. https://doi.org/10.1093/nar/28.1.352
5. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al (2015) A global reference for human genetic variation. Nature 526:68-74. https://doi.org/10.1038/nature15393
6. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001-D1006. https://doi.org/10.1093/nar/gkt1229
7. Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, et al (2021) IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. Nucleic Acids Res 49:D1225-D1232. https://doi.org/10.1093/nar/gkaa923
8. Vijayasimha M (2026) Equity-aware variant interpretation needs local allele frequencies and calibrated functional evidence: comment on Bianco & Planello (2025). Mol Genet Genomics 301:21. https://doi.org/10.1007/s00438-025-02343-6
9. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461:489-494. https://doi.org/10.1038/nature08365
10. . Basu A, Sarkar-Roy N, Majumder PP (2016) Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. Proc Natl Acad Sci U S A 113:1594-1599. https://doi.org/10.1073/pnas.1513197113
11. Bhattacharyya C, Subramanian K, Uppili B, Biswas NK, Ramdas S, Tallapaka KB, et al (2025) Mapping genetic diversity with the GenomeIndia project. Nat Genet 57:767-773. https://doi.org/10.1038/s41588-025-02153-x
12. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99:1129-1143. https://doi.org/10.1016/j.neuron.2018.08.011
13. Hospital A, Goni JR, Orozco M, Gelpi JL (2015) Molecular dynamics simulations: advances and applications. Adv Appl Bioinform Chem 8:37-47. https://doi.org/10.2147/AABC.S70333
14. Ali AE, Li LL, Courtney MJ, Pentikainen OT, Postila PA (2024) Atomistic simulations reveal impacts of missense mutations on the structure and function of SynGAP1. Brief Bioinform 25:bbae458. https://doi.org/10.1093/bib/bbae458
15. Elnageeb ME, Elfaki I, Modawe GA, Elfaki AO, Alzahrani OR, Abuagla HA, et al (2025) Computational prediction of the pathogenic variants of arachidonate 5-lipoxygenase activating protein using molecular dynamics simulation. PLoS One 20:e0329126. https://doi.org/10.1371/journal.pone.0329126
16. Abdul Samad F, Suliman BA, Basha SH, Manivasagam T, Essa MM (2016) A comprehensive in silico analysis on the structural and functional impact of SNPs in the congenital heart defects associated with NKX2-5 gene—a molecular dynamic simulation approach. PLoS One 11:e0153999. https://doi.org/10.1371/journal.pone.0153999
17. Kulshreshtha A, Bhatnagar S (2024) Structural effect of the H992D/H418D mutation of angiotensin-converting enzyme in the Indian population: implications for health and disease. J Biomol Struct Dyn 1-18. https://doi.org/10.1080/07391102.2024.2321246

18. Singh S, Sharma K, Kulshreshtha A (2024) Comparative binding affinity and stability of PHA synthase from Pseudomonas putida and Pseudomonas aeruginosa with acetyl and acyl substrates. J Biosci 49.
19. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al (2025) PubChem 2025 update. Nucleic Acids Res 53:D1516-D1525. https://doi.org/10.1093/nar/gkae1059
20. Irfan M, Iqbal T, Hashmi S, Ghani U, Bhatti A (2022) In silico prediction and functional analysis of nonsynonymous SNPs in human CTLA4 gene. Sci Rep 12:20441. https://doi.org/10.1038/s41598-022-24699-0
21. Uddin MM, Hossain MT, Hossain MA, Ahsan A, Shamim KH, Hossen MA, et al (2024) Unraveling the potential effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on the protein structure and function of the human SLC30A8 gene on type 2 diabetes and colorectal cancer: an in silico approach. Heliyon 10:e37280. https://doi.org/10.1016/j.heliyon.2024.e37280
22. Islam R, Islam MJ, Jaman S, Hossen MA, Ahmed SS, Afrose SS, et al (2025) Investigating the structural and functional consequences of pathogenic SNPs on human VEGFA dimer: insights from molecular dynamics study. ACS Omega 10:52872-52883. https://doi.org/10.1021/acsomega.5c07106
23. Lagace TA (2014) PCSK9 and LDLR degradation: regulatory mechanisms in circulation and in cells. Curr Opin Lipidol 25:387-393. https://doi.org/10.1097/MOL.0000000000000114
24. Ugovsek S, Likozar AR, Levstek T, Podkrajsek KT, Zupan J, Antonic S, et al (2025) Gain of function variants in PCSK9 gene in high risk patients after myocardial infarction with increased lipoprotein (a) values and treated with statins. Sci Rep 15:26753. https://doi.org/10.1038/s41598-025-12432-6
25. Coggi D, Ward J, Macchi C, Gigante B, Amato M, Lyall DM, et al (2025) PCSK9 genetic variants, carotid atherosclerosis and vascular remodelling. Open Heart 12. https://doi.org/10.1136/openhrt-2025-003348
26. Anderer S (2026) FDA approves first oral PCSK9 inhibitor to lower LDL cholesterol. JAMA 336:633. https://doi.org/10.1001/jama.2026.9491
27. Cunningham D, Danley DE, Geoghegan KF, Griffor MC, Hawkins JL, Subashi TA, et al (2007) Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. Nat Struct Mol Biol 14:413-419. https://doi.org/10.1038/nsmb1235
28. Piper DE, Jackson S, Liu Q, Romanow WG, Shetterly S, Thibault ST, et al (2007) The crystal structure of PCSK9: a regulator of plasma LDL-cholesterol. Structure 15:545-552. https://doi.org/10.1016/j.str.2007.04.004
29. Kulshreshtha A, Bhatnagar S (2026) Impact of loss-of-function variations of PCSK9 on LDLR interaction and dynamics: implications for heart function. Biochem Biophys Res Commun 799:153224. https://doi.org/10.1016/j.bbrc.2025.153224
30. Mansfield BS, Bene-Alhasan Y, Ballantyne CM, Raal FJ (2026) The evolving therapeutic landscape of PCSK9 inhibition. Atherosclerosis 414:120670. https://doi.org/10.1016/j.atherosclerosis.2026.120670
31. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 354:1264-1272. https://doi.org/10.1056/NEJMoa054013
32. Kathiresan S, Myocardial Infarction Genetics C (2008) A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. N Engl J Med 358:2299-2300. https://doi.org/10.1056/NEJMc0707445
33. Horton JD, Cohen JC, Hobbs HH (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 32:71-77. https://doi.org/10.1016/j.tibs.2006.12.008
34. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, et al (2013) Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. Eur Heart J 34:3478-3490a. https://doi.org/10.1093/eurheartj/eht273
35. Ni YG, Di Marco S, Condra JH, Peterson LB, Wang W, Wang F, et al (2011) A PCSK9-binding antibody that structurally mimics the EGF(A) domain of LDL-receptor reduces LDL cholesterol in vivo. J Lipid Res 52:78-86. https://doi.org/10.1194/jlr.M011445

36. Lagace TA, Curtis DE, Garuti R, McNutt MC, Park SW, Prather HB, et al (2006) Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. J Clin Invest 116:2995-3005. https://doi.org/10.1172/JCI29383
37. Tveten K, Str MT, Berge KE, Leren TP (2013) PCSK9-mediated degradation of the LDL receptor generates a 17 kDa C-terminal LDL receptor fragment. J Lipid Res 54:1560-1566. https://doi.org/10.1194/jlr.M034371
38. EEroglu Z, Tetik Vardarli A, Duzgun Z, Gunduz C, Bozok Cetintas V, Kayikcioglu M (2018) Case-control study on PCSK9 R496W (rs374603772) and D374Y (rs137852912) mutations in Turkish patients with primary dyslipidemia. Anatol J Cardiol 19:334-340. https://doi.org/10.14744/AnatolJCardiol.2018.86648
39. Kaya E, Kayikcioglu M, Tetik Vardarli A, Eroglu Z, Payzin S, Can L (2017) PCSK9 gain-of-function mutations (R496W and D374Y) and clinical cardiovascular characteristics in a cohort of Turkish patients with familial hypercholesterolemia. Anatol J Cardiol 18:266-272. https://doi.org/10.14744/AnatolJCardiol.2017.7654
40. Martin WR, Lightstone FC, Cheng F (2020) In silico insights into protein-protein interaction disruptive mutations in the PCSK9-LDLR complex. Int J Mol Sci 21. https://doi.org/10.3390/ijms21051550
41 Lammi C, Sgrignani J, Roda G, Arnoldi A, Grazioso G (2019) Inhibition of PCSK9(D374Y)/LDLR protein-protein interaction by computationally designed T9 lupin peptide. ACS Med Chem Lett 10:425-430. https://doi.org/10.1021/acsmedchemlett.8b00464
42. Wu X, Zhu X, Fang M, Qi F, Yin Z, Zhang JZH, et al (2025) Insights into binding mechanisms of potential inhibitors targeting PCSK9 protein via molecular dynamics simulation and free energy calculation. Molecules 30. https://doi.org/10.3390/molecules30142962
43. Mukhopadhyay S, Chakrabarty S (2025) Mixed-solvent molecular dynamics simulation reveals a druggable allosteric pocket in the PCSK9 C-terminal domain. J Phys Chem B 129:11737-11744. https://doi.org/10.1021/acs.jpcb.5c05530
44. Di Taranto MD, Benito-Vicente A, Giacobbe C, Uribe KB, Rubba P, Etxebarria A, et al (2017) Identification and in vitro characterization of two new PCSK9 gain-of-function variants found in patients with familial hypercholesterolemia. Sci Rep 7:15282. https://doi.org/10.1038/s41598-017-15543-x
45. Allard D, Amsellem S, Abifadel M, Trillard M, Devillers M, Luc G, et al (2005) Novel mutations of the PCSK9 gene cause variable phenotype of autosomal dominant hypercholesterolemia. Hum Mutat 26:497. https://doi.org/10.1002/humu.9383
46. Hummelgaard S, Vilstrup JP, Gustafsen C, Glerup S, Weyer K (2023) Targeting PCSK9 to tackle cardiovascular disease. Pharmacol Ther 249:108480. https://doi.org/10.1016/j.pharmthera.2023.108480
47. Shaik NA, Al-Shehri N, Athar M, Awan A, Khalili M, Al Mahadi HB, et al (2023) Protein structural insights into a rare PCSK9 gain-of-function variant (R496W) causing familial hypercholesterolemia in a Saudi family: whole exome sequencing and computational analysis. Front Physiol 14:1204018. https://doi.org/10.3389/fphys.2023.1204018
48. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C (2020) SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res 48:W185-W192. https://doi.org/10.1093/nar/gkaa420
49. . Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40:W452-W457. https://doi.org/10.1093/nar/gks539
50. Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745-2747. https://doi.org/10.1093/bioinformatics/btv195
51. Hecht M, Bromberg Y, Rost B (2015) Better prediction of functional effects for sequence variants. BMC Genomics 16:S1. https://doi.org/10.1186/1471-2164-16-S8-S1
52. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30:1237-1244. https://doi.org/10.1002/humu.21047

53. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306-W310. https://doi.org/10.1093/nar/gki375
54. Chen Y, Lu H, Zhang N, Zhu Z, Wang S, Li M (2020) PremPS: predicting the impact of missense mutations on protein stability. PLoS Comput Biol 16:e1008543. https://doi.org/10.1371/journal.pcbi.1008543
55. Khanna T, Hanna G, Sternberg MJE, David A (2021) Missense3D-DB web catalogue: an atom-based analysis and repository of 4M human protein-coding genetic variants. Hum Genet 140:805-812. https://doi.org/10.1007/s00439-020-02246-z
56. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics 11:548. https://doi.org/10.1186/1471-2105-11-548
57. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859-1865. https://doi.org/10.1002/jcc.20945
58. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477-486. https://doi.org/10.1007/BF00228148
59. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511-1519. https://doi.org/10.1002/pro.5560020916
60. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714-2723. https://doi.org/10.1002/elps.1150181505
61. Huang J, MacKerell AD Jr (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem 34:2135-2145. https://doi.org/10.1002/jcc.23354
62. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774-797. https://doi.org/10.1016/j.jmb.2007.05.022
63. Laskowski RA, Thornton JM (2022) PDBsum extras: SARS-CoV-2 and AlphaFold models. Protein Sci 31:283-289. https://doi.org/10.1002/pro.4238
64. Felline A, Seeber M, Fanelli F (2022) PSNtools for standalone and web-based structure network analyses of conformational ensembles. Comput Struct Biotechnol J 20:640-649. https://doi.org/10.1016/j.csbj.2021.12.044
65. Kwon HJ, Lagace TA, McNutt MC, Horton JD, Deisenhofer J (2008) Molecular basis for LDL receptor recognition by PCSK9. Proc Natl Acad Sci U S A 105:1820-1825. https://doi.org/10.1073/pnas.0712064105
66. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980-D985. https://doi.org/10.1093/nar/gkt1113
67. Guo Q, Feng X, Zhou Y (2020) PCSK9 variants in familial hypercholesterolemia: a comprehensive synopsis. Front Genet 11:1020. https://doi.org/10.3389/fgene.2020.01020
68. Pisciotta L, Priore Oliva C, Cefalu AB, Noto D, Bellocchio A, Fresa R, et al (2006) Additive effect of mutations in LDLR and PCSK9 genes on the phenotype of familial hypercholesterolemia. Atherosclerosis 186:433-440. https://doi.org/10.1016/j.atherosclerosis.2005.08.015
69. Hopkins PN, Defesche J, Fouchier SW, Bruckert E, Luc G, Cariou B, et al (2015) Characterization of autosomal dominant hypercholesterolemia caused by PCSK9 gain of function mutations and its specific treatment with alirocumab, a PCSK9 monoclonal antibody. Circ Cardiovasc Genet 8:823-831. https://doi.org/10.1161/CIRCGENETICS.115.001129
70. Pandit S, Wisniewski D, Santoro JC, Ha S, Ramakrishnan V, Cubbon RM, et al (2008) Functional analysis of sites within PCSK9 responsible for hypercholesterolemia. J Lipid Res 49:1333-1343. https://doi.org/10.1194/jlr.M800049-JLR200
71 Sarkar SK, Foo ACY, Matyas A, Asikhia I, Kosenko T, Goto NK, et al (2020) A transient amphipathic helix in the prodomain of PCSK9 facilitates binding to low-density lipoprotein particles. J Biol Chem 295:2285-2298. https://doi.org/10.1074/jbc.RA119.010221

72. Sarkar SK, Matyas A, Asikhia I, Hu Z, Golder M, Beehler K, et al (2022) Pathogenic gain-of-function mutations in the prodomain and C-terminal domain of PCSK9 inhibit LDL binding. Front Physiol 13:960272. https://doi.org/10.3389/fphys.2022.960272

Legends for Figures

Figure 1: PCSK9-LDLR structure: PCSK9 is represented as a ribbon colored: prodomain (cyan), catalytic domain (green), hinge region (dark blue) and C-terminal domain (red). LDLR is represented by grey ribbon. The interacting regions in PCSK9 and LDLR are represented in CPK style.

(a) WT: Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(b) S89W: S89W is marked by orange spheres, Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(c) R357C: R357C is marked by dark green spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(d) Q413R: Q413R is marked by salmon pink spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(e) R496W: R496W is marked by brown spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey

colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(f) R680W: R680W is marked by lilac spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

Figure 2: PCA of WT PCSK9-LDLR and MT PCSK9-LDLR during the 500 ns MD simulation. Plot of PC1 and PC2 is shown as follows: a) WT PCSK9-LDLR (black), b) S89W PCSK9-LDLR in yellow, c) R357C PCSK9-LDLR in green, d) Q413R PCSK9-LDLR in pink, e) R496W in brown and R680W in purple colour. Further LOF variants are g) R93C in Royal blue, h) N298D as Indigo and i) A511V as light blue colour.

Figures

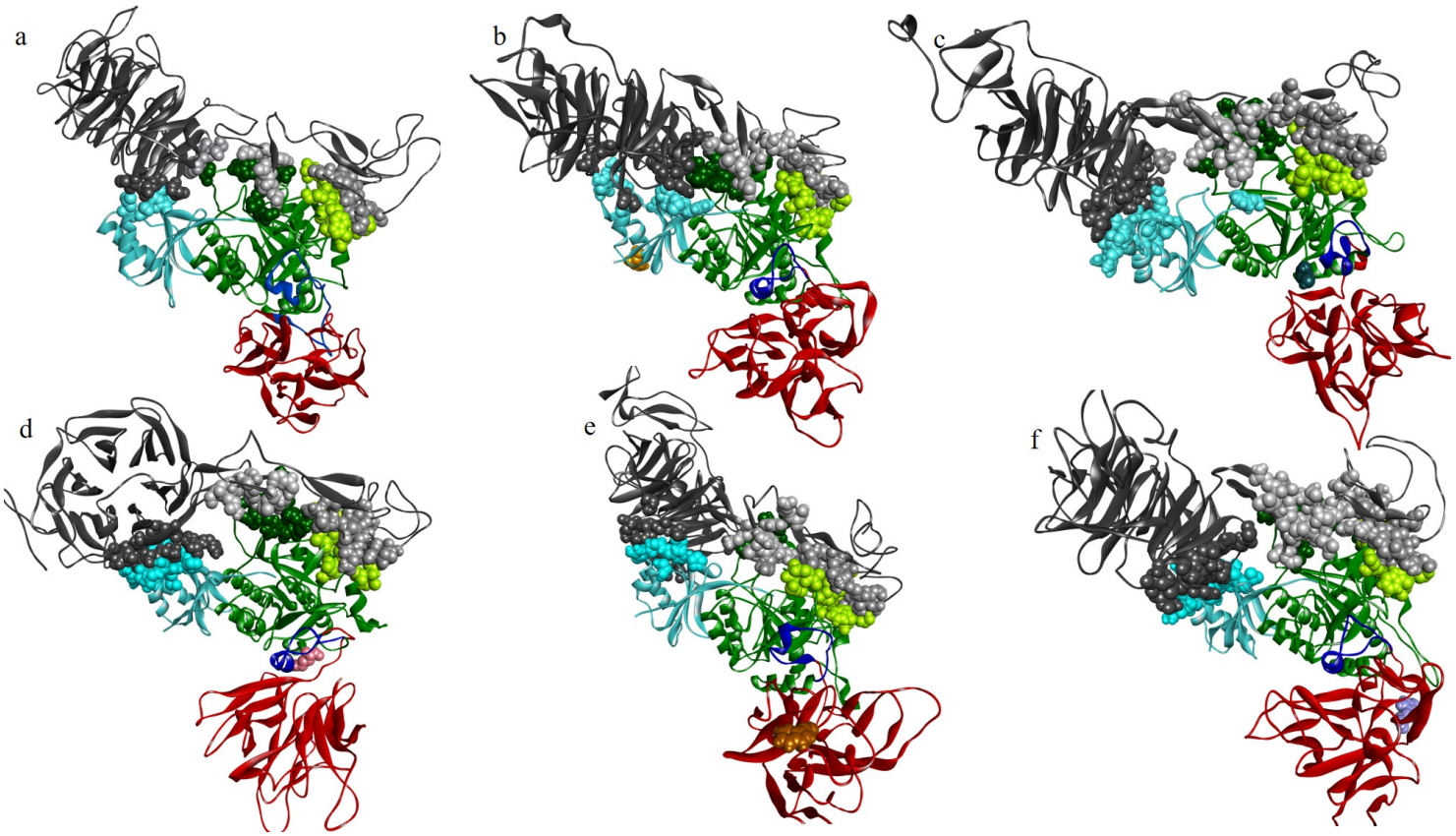

Figure 1

PCSK9-LDLR structure: PCSK9 is represented as a ribbon colored: prodomain (cyan), catalytic domain (green), hinge region (dark blue) and C-terminal domain (red). LDLR is represented by grey ribbon. The interacting regions in PCSK9 and LDLR are represented in CPK style.

(a) WT: Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(b) S89W: S89W is marked by orange spheres, Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(c) R357C: R357C is marked by dark green spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(d) Q413R: Q413R is marked by salmon pink spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(e) R496W: R496W is marked by brown spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

(f) R680W: R680W is marked by lilac spheres. Region I interactions are shown, with PCSK9 represented in lime green using CPK representation, and LDLR represented in dark gray. Region II interaction is shown in PCSK9 by Forest green (Dark green) colour and LDLR in light grey colour. Region III interaction are shown in PCSK9 by cyan colour and LDLR in charcoal colour in CPK representation.

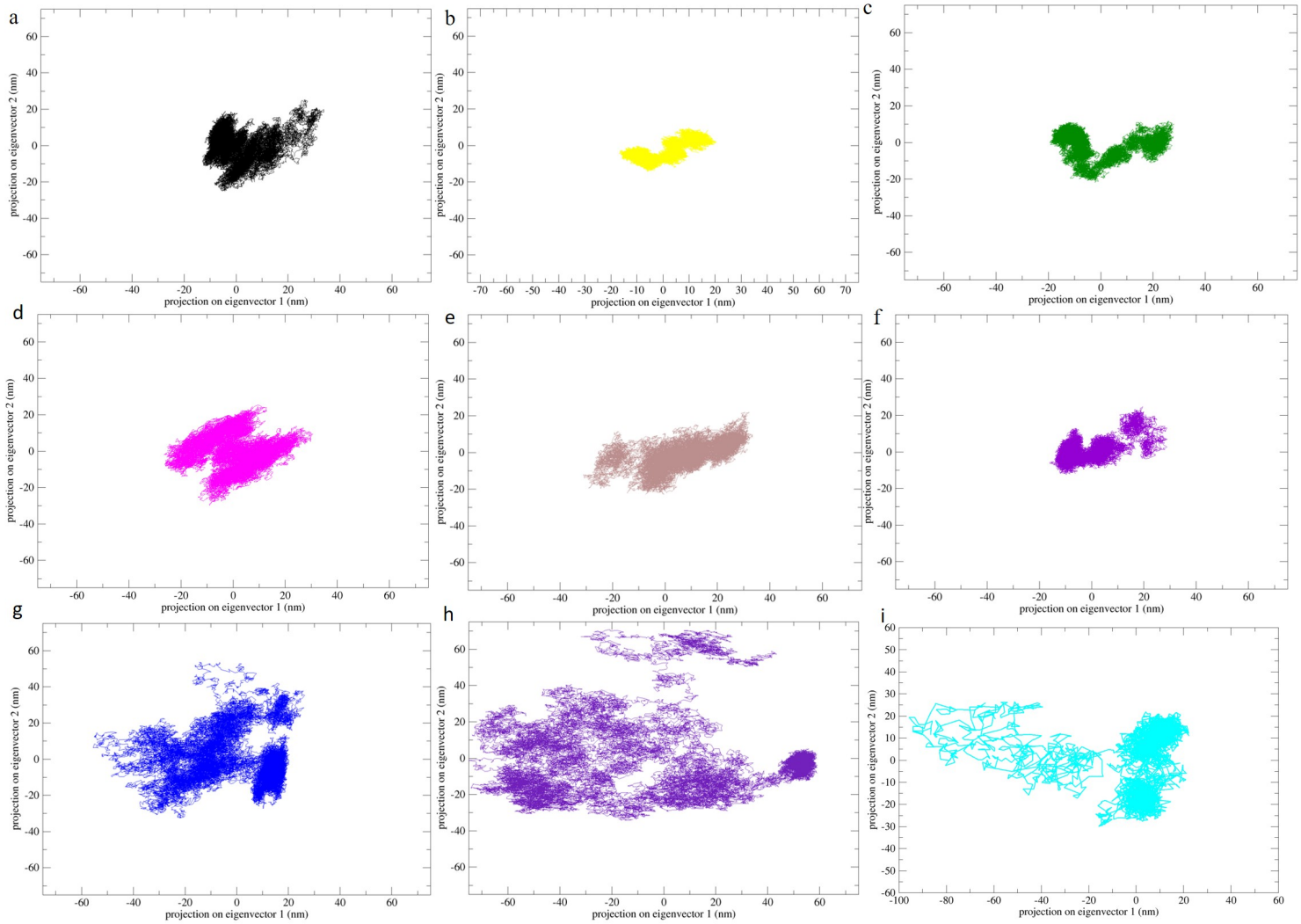

Figure 2

PCA of WT PCSK9-LDLR and MT PCSK9-LDLR during the 500 ns MD simulation. Plot of PC1 and PC2 is shown as follows: a) WT PCSK9-LDLR (black), b) S89W PCSK9-LDLR in yellow, c) R357C PCSK9-LDLR in green, d) Q413R PCSK9-LDLR in pink, e) R496W in brown and R680W in purple colour. Further LOF variants are g) R93C in Royal blue, h) N298D as Indigo and i) A511V as light blue colour.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile2026.docx