Malaysian Applied Biology
https://doi.org/10.55230/mabjournal.v54i4.3563

Research Article

Targeting PCSK9: Bioinformatics Analysis Reveals Functionally Damaging Missense Variants

Loo Keat Wei[1,2], Huey Chyi Loh[2], Lyn R. Griffiths[3]

1. Centre for Biomedical and Nutrition Research (CBNR), Universiti Tunku Abdul Rahman, Bandar Barat, 31900 Kampar, Perak, Malaysia
2. Department of Biological Science, Faculty of Science, Universiti Tunku Abdul Rahman, Bandar Barat, 31900 Kampar, Perak, Malaysia
3. Centre for Genomics and Personalised Health, Genomics Research Centre, School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, 60 Musk Ave, Kelvin Grove, QLD 4059, Australia

*Corresponding author: lookw@utar.edu.my

ABSTRACT

Proprotein convertase subtilisin/kexin type 9 (PCSK9) modulates cholesterol homeostasis by targeting low-density lipoprotein receptor (LDLR) for lysosomal degradation. Genetic polymorphisms in *PCSK9* can alter its autocatalytic processing, secretion, or binding affinity to LDLR. Reduce binding efficiency between PCSK9 and LDLR leads to elevated low-density lipoprotein cholesterol (LDL-C) level, thereby promoting atherosclerotic plaque formation and increasing the risk of ischemic stroke. The objective of this study was to identify the most functionally significant non-synonymous single-nucleotide polymorphisms (nsSNPs) in *PCSK9* via an integrated *in silico* analysis combining functional prediction tools (PROVEAN, SIFT, PolyPhen-2, SNAP2), protein stability and disease-association predictors, ligand-binding assessment, and post-translational modification analysis. A total of 4,979 *PCSK9* variants were retrieved from Ensembl GRCh37/hg19, and HGMD. Functional annotation using PROVEAN, SIFT, PolyPhen-2, and SNAP2 identified 253 nsSNPs, with PolyPhen-2 predicting the largest subset. Upon filtering through the protein stability, disease association, ligand binding, and post-translational modification, five nsSNPs (W156R, H226L, H229R, G337R, and G394V) emerged as the most deleterious, with potential to disrupt secondary autocatalytic processing and significantly impair LDLR-PCSK9 interactions. These findings highlight novel candidate variants that may serve as diagnostic biomarkers and therapeutic targets in dyslipidemia and cardiovascular disease.

Key words: Bioinformatics, low-density lipoprotein receptor (LDLR), proprotein convertase subtilisin/kexin type 9 (PCSK9), single nucleotide polymorphisms (SNPs),cardiovascular disease

INTRODUCTION

Over the recent years, hypercholesterolemia patients have consumed statins to limit LDL cholesterol production in the liver. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial--Lipid Lowering Trial (ALLHAT-LLT) involving 2867 individuals with moderate hyperlipidemia and hypertension has suggested that pravastatin possesses a non-significant benefit for the primary prevention of cardiovascular diseases (Han *et al*., 2017). Alternatively, proprotein convertase subtilisin-kexin type 9 (PCSK9) inhibitor such as alirocumab and evolocumab, can significantly reduce low-density lipoprotein cholesterol (LDL-C) level, which has emerged as the latest treatment for hypercholesterolemia (Chaudhary *et al*., 2017). Statins reduce hepatic cholesterol synthesis but also increase PCSK9 levels, which can limit their effectiveness. In contrast, PCSK9 inhibitors specifically block PCSK9, preventing LDL receptor (LDLR) degradation, increasing LDL uptake, and lowering the risk of atherosclerosis and ischemic stroke (Chaudhary *et al*., 2017).

PCSK9 protein (~74 kDa) is the ninth member of the proprotein convertase family, encoded by *PCSK9* gene located at chromosome 1p32.3. This protein undergoes two cleavage processes for maturation, activation and secretion. The PCSK9 protein (UniProt ID: Q8NBP7) consists of a signal peptide (residues 1 - 30) which directs the nascent protein into the secretory pathway, an N-terminal domain (residues 31 - 152) required for proper folding and self-inhibition of the enzyme, a subtilisin-like catalytic domain (residues 153 - 425) responsible for autocatalytic cleavage and interaction with the low density lipoprotein receptor (LDLR), and a C-terminal domain (residues 426 - 692) which contributes to LDLR binding and regulates PCSK9 stability and activity. In addition to its structural organization, PCSK9 undergoes posttranslational modifications, most notably phosphorylation at multiple serine residues (e.g., Ser47, Ser666, Ser668, and Ser688). These modifications, mediated by kinases such as Golgi casein kinase-like kinase, have been reported to influence PCSK9 stability, secretion, and proteolysis protection in a cell type-dependent manner, although they are not strictly required for its activity on LDLR degradation. PCSK9 protein mediates cholesterol homeostasis by regulating the hepatic cell surface LDLR. Specifically, PCSK9 binds to the epidermal growth factor-like repeat A (EGF-A) domain of LDLR, a critical site for this interaction, and subsequently redirects the receptor

Article History
Accepted: 2 December 2025
First version online: 30 December 2025
Cite This Article:
Wei, L.K., Loh, H.C. & Griffiths, L.R. 2025. Targeting PCSK9: Bioinformatics analysis reveals functionally damaging missense variants. Malaysian Applied Biology, 54(4): 176-185. https://doi.org/10.55230/mabjournal.v54i4.3563

to lysosomal degradation via two pathways (Abifadel *et* al., 2003; Au *et* al., 2015; Chaudhary *et al*., 2017). In the external route, where PCSK9 binds to LDLR on the cell surface, whereas in the internal route, PCSK9 binds to LDLR in the luminal secretory compartment and redirects the receptor to lysosomal degradation (Au *et al*., 2017).

PCSK9 is one of the three genes implicated in familial hypercholesterolemia through gain-of-function mutations (Chaudhary *et al*., 2017). It was first identified in 2003 by linkage analysis of French families with autosomal dominant hypercholesterolemia (ADH) who lacked mutations in *LDLR* and Apolipoprotein B (*APOB*). Subsequent sequencing revealed two causative mutations in *PCSK9* at the ADH locus. Since then, more than 50 gain-of-function mutations in *PCSK9* have been reported worldwide, most of which cluster in the catalytic and C-terminal domains, impair LDLR recycling, and markedly elevate plasma LDL-C levels. PCSK9 encodes neural apoptosis-regulated convertase 1 (NARC-1), a subtilisin-like serine protease that is highly expressed in the liver and plays a central role in cholesterol homeostasis (Abifadel *et* al., 2003). With the increasing number of gain-of-function polymorphisms detected in *PCSK9*, little is known about the effect of nsSNPs in the coding regions of the gene on protein stability (Abifadel *et* al., 2003; Au *et* al., 2017). Hence, the objective of this study was to determine the functional nsSNPs in the *PSCK9* gene.

METHODOLOGY

PCSK9 gene variants and protein information

Genomics information (accession numbers NG_009061.1 and NM_174936.3) of *PCSK9* was retrieved from the National Center for Biotechnology Information (NCBI) (Wheeler *et al*., 2007). Using transcript ID ENST00000302118.5 as the reference, *PCSK9* variants were compiled from Ensembl GRCh37/hg19 (Howe *et al*., 2021), and Human Gene Mutation Database (HGMD) (Stenson *et al*., 2003). Proteomic data containing UniProt ID (Q8NBP7) and amino acid sequence were retrieved from UniProtKB (Boutet *et al*., 2016). The PCSK9 crystal structure (PDB ID: 2P4E), solved by X-ray diffraction at 1.98 Å resolution (R-work = 0.200, R-free = 0.250), was obtained from the Protein Data Bank (PDB) (Protein Data Bank, 1971).

Predicting the deleterious impact of nsSNPs

PCSK9 nsSNPs were subjected to deleterious and damaging effect prediction using the Protein Variation Effect Analyser (PROVEAN) (Choi & Chan, 2015) and Screening for Non-Acceptable Polymorphisms (SNAP2) (Bromberg & Rost, 2007). Subsequent analyses were performed with Sorting Intolerant from Tolerant (SIFT) (Vaser *et al*., 2016) and Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei *et ak*., 2013). The inputs for PROVEAN, SIFT and PolyPhen-2 were chromosome location, wild- or mutant-type allele, while the input for SNAP2 was FASTA amino acid sequence NP_777596.2 of PCSK9 protein. Moreover, three additional computational tools, i.e. Predicting human Deleterious SNPs in the human genome (PhD-SNPg) (Capriotti & Fariselli, 2017), Position-Specific Evolutionary Preservation (PANTHER) (Thomas *et al*., 2003), and MutPred2 (Pejaver *et al*., 2017), were employed to evaluate the potential deleterious effects of nsSNPs. For PhD-SNPg, the chromosomal positions of the wild-type and mutant amino acids were provided as input (Capriotti & Fariselli, 2017), while the amino acid sequences of the nsSNPs were used for MutPred2 analysis (Pejaver *et al*., 2017). To minimize false-positive predictions, the MutPred2 threshold was set at 0.8 (Pejaver *et al*., 2017). In addition, the analysis incorporated complementary predictions obtained from both PhD-SNPg (Capriotti & Fariselli, 2017) and PANTHER (Thomas *et al*., 2003). Furthermore, the web servers and the corresponding cut-off values used for the functional prediction of nsSNPs are detailed in Supplementary Table 1.

Evolutionary conservation analysis of nsSNPs

The evolutionary conservation of amino acid residues in PCSK9 was evaluated using ConSurf-DB, which provides pre-calculated conservation profiles for proteins of known PDB structure. Homologous sequences were collected and aligned with HMMER and MAFFT, and residue conservation scores were calculated using the Rate4Site algorithm with Bayesian inference, accounting for phylogenetic relationships among aligned sequences (Pejaver *et al*., 2017). The conservation profile was mapped onto the PCSK9 crystal structure (PDB ID: 2P4E) to visualize functionally important regions.

nsSNPs on protein stability

Protein stability of the deleterious nsSNPs was predicted using PDB ID 2P4E. The effects of protein stability were predicted using four established computational tools, namely, Cologne University Protein Stability Analysis Tool (CUPSAT) (Parthiban *et al*., 2006), Prediction of Protein Stability Changes due to Single Amino Acid Mutations (ProSMS) (Wang & Sauer, 2010), Prediction of Protein Stability Changes for Single Site Mutations from Sequences (MUpro) (Cheng *et al*., 2006), and DUET with a complementary approach of site directed mutator and cutoff mutation scanning matrix (Pires *et al*., 2014). The cut-off value of 0 was used for protein stability prediction (Supplementary Table 1).

Disease-associated nsSNPs

Disease-associated nsSNPs were predicted using PMut (Ferrer-Costa *et al*., 2005) and SNPs&GO (Capriotti *et al*., 2013). For PMut batch analysis, the wild-type and mutant amino acids of nsSNPs identified in the human PCSK9 gene were used as input (Ferrer-Costa *et al*., 2005). For SNPs&GO, the corresponding amino acid sequences of the nsSNPs were submitted for analysis (Capriotti *et al*., 2013). Details of the web servers employed for disease-association prediction are provided in Table 1.

Ligand Binding Site

Ligand binding sites were observed from RaptorXBinding via Deep Learning algorithm (Peng & Xu, 2011), COACH that combines the predictions from binding-specific substructure comparison (TM-SITE) and sequence profile alignment (S-SITE) algorithms (Yang *et al*., 2013), COFACTOR (Zhang *et al*., 2017a) as well as ATPbind that assembles multiple support vector machine models (Hu *et al*., 2018). The amino acid sequences of PDB ID 2P4E were used as input for the aforementioned web

servers.

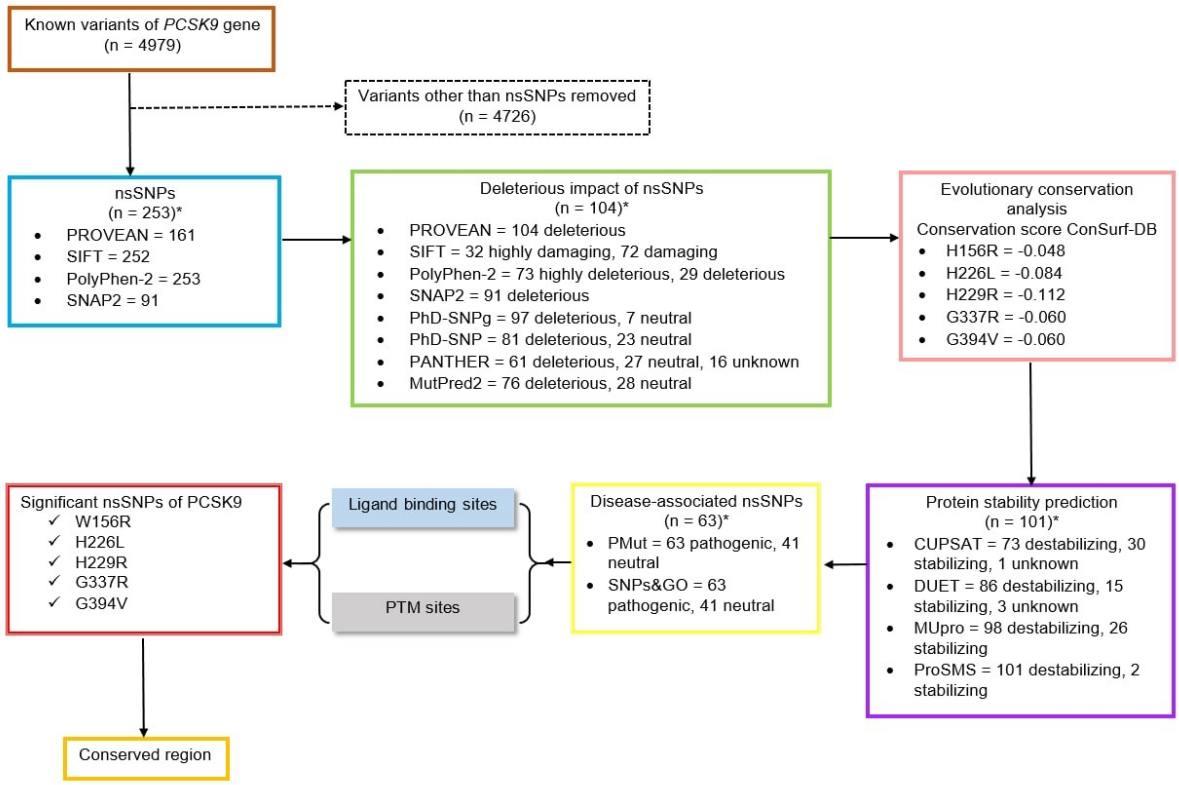
Post-Translational Modification Sites

The sites for post-translational modification (PTM) were predicted from PROSITE (Hulo *et al*., 2006) and a sequence-based predictor ModPred (Pejaver *et al*., 2014). PTM involved major modifications such as palmitoylation, methylation, O-glycosylation and N-glycosylation. Palmitoylation sites were predicted using CSS-Palm (Zhou *et al*., 2006). Meanwhile, the methylation sites were predicted using GPS-MSP Methyl-group Specific Predictor 1.0 (Deng *et al*., 2017). N- and O-linked glycosylation sites were predicted using the NetOGlyc 4.0 server (Zhang *et al*., 2024) to evaluate potential post-translational modifications. To ensure predictive reliability, only residues with scores above the default confidence threshold (>0.5) were retained. Consistent with the ligand-binding site analysis, amino acid sequences corresponding to PDB ID 2P4E served as input. The predicted glycosylation sites were subsequently examined in relation to the five novel nsSNPs. All identified glycosylation residues were positioned more than 5 Å from the nsSNP locations, suggesting the absence of direct structural overlap or steric interference.

RESULTS and DISCUSSION

Functional characterisation of nsSNPs

A total of 4979 known variants encompassing the 5' untranslated regions (UTR) and 3'UTR of the *PCSK9* gene, as well as its intron and exon regions, were retrieved from Ensembl GRCh37/hg19 and HGMD. Retrieved variants were analysed with PROVEAN, SNAP2, SIFT and PolyPhen-2 to obtain nsSNPs with deleterious and damaging effects. Upon filtering, the nsSNPs from PROVEAN (n=161), SIFT (n=252), PolyPhen-2 (n=253) and SNAP2 (n=91) were compared (Figure 1). Likewise, PhD-SNPg, PANTHER, and MutPred2 classified 97, 61, and 76 nsSNPs, respectively, as pathogenic. Based on the results of functional annotation, there were 104 deleterious nsSNPs in PROVEAN, 32 highly damaging and 72 damaging nsSNPs in SIFT, 73 highly deleterious and 29 deleterious nsSNPs in PolyPhen-2, and 91 nsSNPs in SNAP2 shown to affect the *PCSK9* gene (Figure 1). Therefore, a total of 104 deleterious nsSNPs were used in the subsequent analyses.



*Based on the highest number of nsSNPs obtained from the web servers.

Fig. 1. Flow chart of the bioinformatics analysis of the *PCSK9* gene.

Evolutionary conservation analysis of nsSNPs

ConSurf analysis was performed to assess the evolutionary conservation of residues affected by the most significant nsSNPs in PCSK9. Five variants (W156R, H226L, H229R, G337R, and G394V) corresponding to dbSNP entries rs1031725741, rs762279506, rs1004968088, rs865848494, and rs376066497 were mapped onto the PCSK9 structure (PDB ID: 2P4E). All five residues showed negative conservation scores (ranging from -0.048 to -0.112), indicating evolutionary conservation. These findings suggest that substitutions at these positions may disrupt functionally important regions of PCSK9, consistent with their predicted pathogenic effects.

Protein stability, disease association, ligand binding, and post-translational modification of the nsSNPs

It has been shown that 12 exons of the *PCSK9* gene were translated to 692 amino acids, the P and A domains of the PCSK9 protein. Of which, protein stability analysis revealed that nsSNPs were categorised into destabilising, stabilising or unknown effect (Figure 1). Likewise, the disease-associated nsSNPs predicted from PMut, and SNPs&GO, were either categorised as pathogenic or neutral (Figure 1).

After filtering 97 nsSNPs, ligand-binding predictions indicated that PCSK9 interacts with calcium, N-acetyl-D-glucosamine, angiotensin-converting enzyme, adenosine triphosphate (ATP), potassium, sulfate ions, magnesium, and 2′-deoxycytidine-5′-triphosphate (dCTP) at distinct residues (Figure 2). The confidence for calcium, N-acetyl-D-glucosamine, angiotensin-converting enzyme, and ATP interactions was high, whereas predictions for potassium, sulfate, magnesium (confidence 0.05 - 0.22), and dCTP (confidence 0.01 - 0.02) were less robust (Supplementary Table 1). Mapping these features revealed that ATP-binding pockets localize to residues 250 - 400, while cation- and nucleotide-binding sites are enriched in the N-terminal and central domains (Figure 2). Post-translational modification predictions identified consensus motifs for phosphorylation by protein kinase C and casein kinase II, as well as for N- and O-glycosylation and palmitoylation (Figure 2), suggesting strong regulatory control. Additional modifications, including sulfation, ADP-ribosylation, methylation, amidation, N-myristoylation, and multiple proteolytic cleavage sites, were also detected, albeit with less consistent support across algorithms (Figure 2, Supplementary Table 1).

When mapped onto the 686-residue sequence, distinct spatial patterns emerged. The N-terminal domain (residues 1 - 108) is enriched in calcium- and magnesium-binding motifs, together with predicted dCTP interactions (Figure 2), implicating it in cation coordination and nucleotide stabilization. The central domain (residues 218 - 400) harbors ATP-binding pockets (BS01 and BS02) and angiotensin-converting enzyme interaction sites (Figure 2), consistent with its catalytic role. By contrast, the C-terminal domain (residues 400 - 686) contains relatively few ligand-binding sites but a higher density of PTMs, including ADP-ribosylation, methylation, glycosylation, and palmitoylation (Figure 2), suggesting predominant roles in regulation, localization, and membrane association.

The PTM landscape emphasizes the regulatory complexity of PCSK9. Phosphorylation motifs distributed across the sequence provide multiple points for signal-dependent modulation, while lipid modifications and glycosylation may contribute to stability, trafficking, and membrane interactions. Moreover, the abundance of proteolytic cleavage sites suggests potential processing into smaller functional fragments (Figure 2).

By which, these findings indicate that ligand-binding motifs are concentrated in the N-terminal and catalytic core domains, whereas PTMs are more broadly distributed, particularly in the C-terminal region, where they likely fine-tune activity and interactions. This spatial partitioning highlights how PCSK9 integrates ligand-binding capacity with multilayered regulatory mechanisms to support its diverse biological functions.

The most significant nsSNPs of the *PCSK9* gene

Based on the results obtained from the 14 webservers, the top 5% of the pathogenic nsSNPs (e.g. W156R, H226L, H229R, G337R and G394V), were chosen as the most significant nsSNPs for *PCSK9* gene (Table 1). Table 1 shows the genomic location, functional characterisation, protein stability, disease association, and conservation score of the nsSNPs. Of which, W156R and G394V were located in the coil, H226L and H229R were situated in the alpha-helix, and G337R was in the beta-sheet.

Impact of novel PCSK9 mutations on protein structure and function

Each novel mutation was mapped onto the three-dimensional structure of PCSK9 (PDB ID: 2P4E) to assess its potential structural and functional consequences. Structural analyses focused on determining whether the variants were positioned near functionally critical regions, including catalytic residues, domain interfaces, and ligand-binding pockets, informed by ligand-binding site predictions and PTM annotations. The analysis indicated that W156R resides within the prodomain, adjacent to calcium-binding motifs, where it could interfere with cation coordination and compromise structural stability. Variants H226L and H229R are clustered within the catalytic core, in proximity to the ATP-binding site, suggesting potential effects on enzymatic activity. G337R is located within the catalytic domain, where the substitution of glycine with a bulkier arginine side chain may introduce steric hindrance within the substrate-binding cleft. Finally, G394V lies near the C-terminal domain interface, where it could disrupt interdomain communication and allosteric regulation. Collectively, these structural insights suggest that the identified mutations may impair PCSK9 function by destabilizing the protein, altering ligand recognition, or reducing catalytic efficiency.

A wide variety of computational tools and/or webservers were employed to understand the impact of *PCSK9* nsSNPs at the molecular level. As such, the results of 14 webservers indicated that the most significant nsSNPs affecting protein stability and disease association were W156R, H226L, H229R, G337R and G394V. The web servers are designed from various algorithms, by which the combinatory analysis can provide a more reliable prediction. For comparison, multiple computational tools were used in each analysis before drawing a solid conclusion. For example, PROVEAN (Choi *et al*., 2015), SIFT (Vaser *et al*., 2016), PolyPhen-2 (Adzhubei *et al*., 2013), and SNAP2 (Bromberg *et al*., 2007) were used to predict the deleterious nsSNPs within the *PCSK9* gene. PROVEAN uses sequence homology to generate the delta alignment score, which is not as specific as SIFT (Vaser *et al*., 2016), Polyphen-2 Arial and SNAP2 (Bromberg *et al*., 2007). This web server includes all types of mutations such as insertion, deletion, frameshift and substitution of all the synonymous SNPs and nsSNPs (Choi *et al*., 2015). The nsSNPs determined from PROVEAN (Choi *et al*., 2015) were compared with SIFT (Vaser *et al*., 2016), Polyphen-2 Arial and SNAP2 (Bromberg *et al*., 2007), as the latter only considered the coding SNPs. Thus, W156R, H226L, H229R, G337R, and G394V were determined as the most significant nsSNPs in the *PCSK9* gene.

Novel findings of this study included that both the H226L and W156R tended to affect the interaction between the LDLR EGF_A domain and the PCSK9 protein. These two mutations might serve as the main key players in reducing ischemic stroke and cardiovascular risks. EGF_A is located within the LDLR protein, which serves as the protein binding site for PCSK9 during the regulation of cholesterol metabolism (Chorba *et al*., 2018). It has been suggested that LDLR degradation is restricted following the presence of H226L and W156R mutations in the PCSK9 protein. Reduced LDLR degradation efficiency may enhance LDL-C

removal from the bloodstream. Thus, it decreases atherosclerotic plaque formation and ischemic stroke risk (Au *et al*., 2015).

Table 1. The most significant nsSNPs for *PCSK9* gene

Prediction	Webserver	nsSNPs				
		W156R g.55512262 T>C	H226L g.55518342 A>T	H229R g.55518351 A>G	G337R g.55523016 G>A	G394V g.55523709 G>T
		rs1031725741	rs762279506	rs1004968088	rs865848494	rs376066497
nsSNPs	PROVEAN (Choi *et al*., 2015)	-12.560	-11.000	-8.000	-7.530	-8.600
	SIFT (Vaser *et al*., 2016)	0.000	0.000	0.000	0.001	0.004
	PolyPhen-2 (Adzhubei *et al*., 2013)	1.000	1.000	1.000	1.000	1.000
	SNAP2 (Bromberg *et al*., 2007)	84.000	91.000	81.000	89.000	89.000
†	PhD-SNPg (Capriotti & Fariselli, 2017)	0.998	0.994	0.964	0.991	0.992
†	PANTHER (Thomas *et al*., 2003)	0.701	0.994	0.504	0.631	0.682
†	MutPred2 (Pejaver *et al*., 2017)	0.937	0.920	0.810	0.867	0.896
Protein stability	CUPSAT (Parthiban *et al*., 2006)	-4.590	0.130	-3.430	-2.270	7.000
	ProSMS (Wang & Sauer, 2010)	-2.120	-0.210	-1.500	-0.840	-0.310
	MUpro (Cheng *et al*., 2006)	-0.929	-0.176	-0.637	-0.152	-0.448
	DUET (Pires *et al*., 2014)	-0.959	-0.455	-2.002	-1.031	0.093
Disease-association	PMut (Ferrer-Costa *et al*., 2005)	0.806	0.878	0.789	0.878	0.878
	SNPs&GO (Capriotti *et al*., 2013)	0.708	0.781	0.643	0.794	0.779
Conservation score	Consurf-DB (Pejaver *et al*., 2017)	-0.048	-0.084	-0.112	-0.060	-0.060

[#]Variants were classified as deleterious/damaging according to the recommended thresholds for each prediction tool. [†]Variants exceeding the threshold values were considered deleterious or pathogenic; those below were classified as neutral/benign.[§]All outputs were aggregated and re-evaluated using our in-house method, which integrates multiple analyses beyond receiver operating characteristic (ROC). To increase confidence, variants consistently predicted as deleterious/damaging across several tools were prioritized for downstream analyses.

H226 is one of the serine protease catalytic triads that are located at the active site of the PCSK9 protein (Figure 3). Catalytic triad encompasses a set of three amino acids with acidic-basic-polar side chains. Our results indicate that the catalytic triads of PCSK9 protein, namely, S386 (polar), H226 (basic) and D186 (acidic), are potent for its protein maturation. Specifically, H226 is involved in the intramolecular cleavage of the PCSK9 protein. Apart from serving as the calcium ion binding site during autocatalytic processing, H226 also forms a hydrogen bond with Q152 residues at the secondary autocatalytic cleavage site (Cameron *et al*., 2008). Experimental studies have shown that substituting histidine at position 226 with leucine (H226L) impairs the calcium-dependent autocatalytic processing of PCSK9 in the endoplasmic reticulum, thereby blocking protein maturation and yielding a non-functional product (Chorba *et al*., 2018). This finding is consistent with our *in silico* analysis, which likewise identified H226 as a functionally critical residue, thereby reinforcing the reliability of our computational approach for detecting deleterious nsSNPs. Since the autocleavage processing of PCSK9 protein depends on the the integrity of its catalytic triad, the substitution of a basic amino acids (histidine) with a non-polar amino acids (leucine) is expected to impair the catalytic triad's, ultimately disrupting the protein's autocatalytic processing (Figure 4). An impaired PCSK9 protein reduces LDLR degradation, and thus enhances plasma LDL-C removal from the bloodstream. Such a phenomenon decreases the risks of hyperlipidemia, plaque formation, ischemic stroke occurrence and cardiovascular diseases (Au *et al*., 2015).

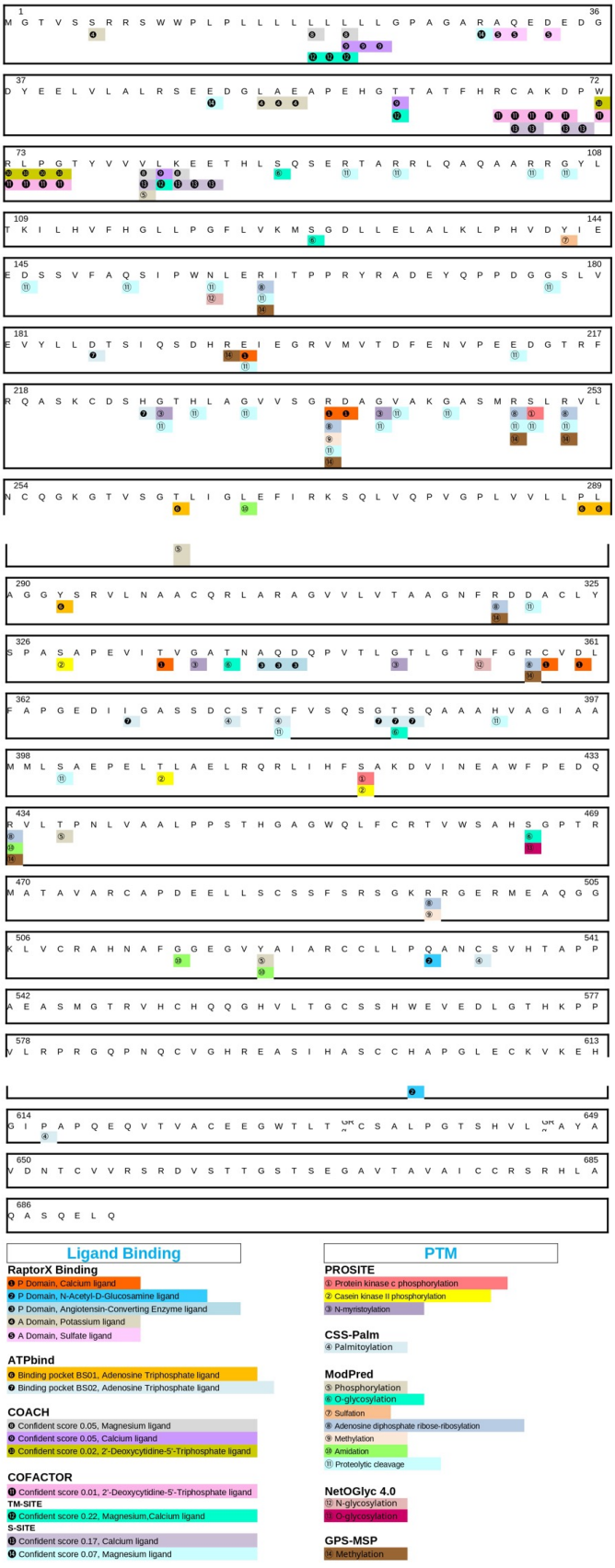


Fig. 2. Ligand binding sites and post translational modification sites.

Fig. 3. Location of the selected nsSNPs in the catalytic domain of PCSK9.

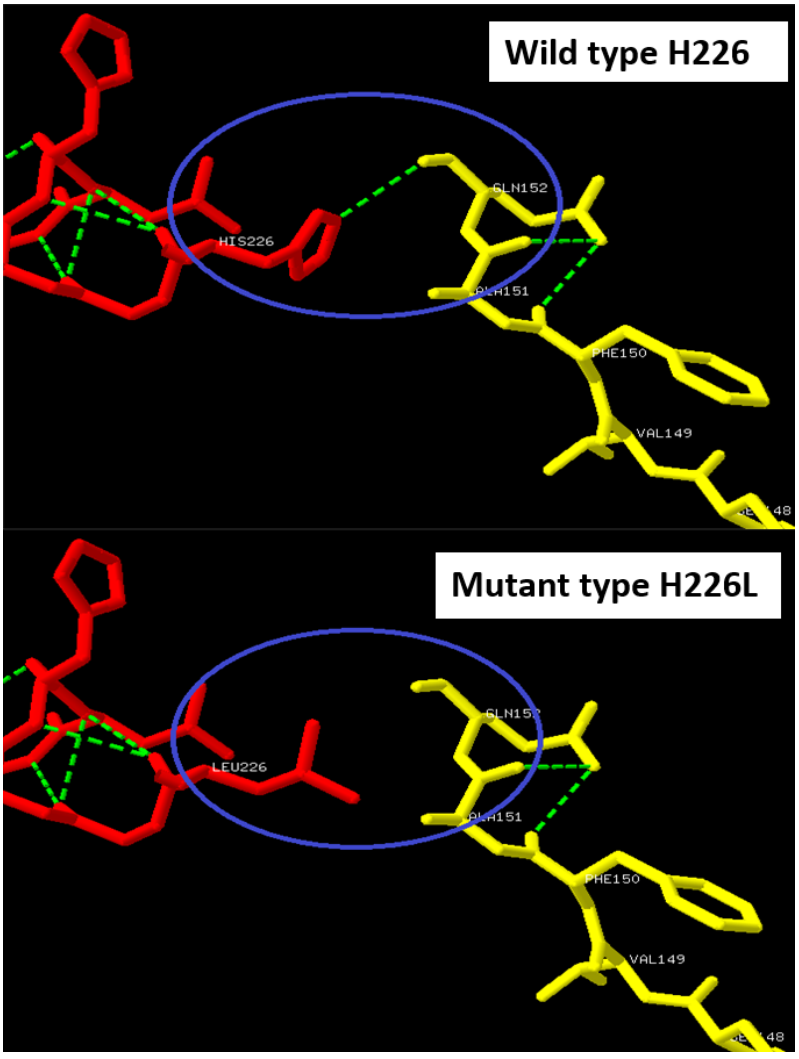


Fig. 4. Hydrogen bond interactions of H226 and Q152 residues.

H229R mutation is located within the same alpha-helix as similar to H226L [9], which may introduce torsion strain or steric clashes that cause the conformational change of an alpha-helix. Given its proximity to H226, the H229R mutation could also disrupt local hydrogen-bonding networks within the α-helix (Benjannet *et al*., 2012). Such structural perturbations may impair the hydrogen bond formation between H226 and Q152 residues, which further disrupts the autocatalytic cleavage of PCSK9 (Benjannet *et al*., 2012). Computational validation using FoldX is needed to further evaluate the potential impact of H229R on hydrogen bond formation and α-helix stability.

The present study reveals that the W156 residue, located within the P′ helix (residues 153–162, SIPWNLERIT), contributes to the autocatalytic cleavage site (VFAQ152′SIP) of PCSK9 (Weider *et al*., 2016). Structural studies have demonstrated that this region contains a highly conserved WNLxRI motif (residues 156–161), which is essential for stabilizing the helical conformation necessary for exposure of the EGF-A binding site [38,39]. Previous genetic analyses suggested that the W156R substitution is likely deleterious and pathogenic (Peloso *et al*., 2016), although no direct experimental investigation of this variant has been reported to date. Our *in silico* analysis further predicts that substituting tryptophan with arginine at this position may destabilize the P′ helix, thereby perturbing the presentation of the EGF-A binding site and reducing PCSK9-LDLR interaction. These findings should be interpreted as computational predictions supported by structural evidence (Zhang *et al*., 2017b), and will require experimental validation to confirm their functional consequences.

Nonetheless, there remains a lack of studies reporting on the novel mutation (i.e. G337R) in PCSK9. Based on PROSITE predictions, G337 may represent as a potential site for N-myristoylation, although this remains to be experimentally validated. N-myristoylation is a type of lipid modification that involves a covalent attachment of the myristoyl group to the alpha-amino group of N-terminal glycine residues via an amide bond (Udenwobele *et al*., 2017). Myristoylation at the histone tail enhances the formation of heterochromatin that suppresses *PCKS9* transcription activity. According to the PROSITE consensus motif (Hulo *et al*., 2006), the G337R substitution is predicted to interfere with the canonical recognition site for N-myristoylation, thereby reducing the likelihood of this modification in PCSK9. This prediction is based on the sequence requirements defined in the PROSITE documentation (PDOC00008) and is consistent with experimental studies of N-myristoylation (Glaser *et al*., 1988). Specifically, an N-terminal glycine is considered indispensable for recognition by N-myristoyltransferase, whereas replacement with a charged residue such as arginine is generally incompatible with enzymatic recognition. Although no direct experimental data are currently available for the PCSK9 G337R variant, this inference is supported by the well-established substrate specificity of the N-myristoylation machinery (Glaser *et al*., 1988). Such a phenomenon prevents the formation of bulky PCSK9 in histone modification and enhances its protein activity, leading to hypercholesterolemia.

In addition, this study highlights the G394V substitution in PCSK9, which is reported in ClinVar (Allele ID: 907416) as a variant of uncertain significance in relation to familial hypercholesterolemia. Although no experimental studies have yet investigated its functional consequences, its location raises the possibility that it could influence splicing efficiency and generate alternative isoforms. Further work will be required to validate this prediction.

Nevertheless, complete degradation mechanism between PCSK9 and LDLR may not be supportive to discuss other potential *PCSK9* nsSNPs (i.e. V81E, R160W, G227D, R251C, P327Q, C358Y, T385I and C654R), in which the literature is relatively scarce. For example, Supplementary Table 2 shows that the aforementioned nsSNPs can affect the EGF_A domain of the PCSK9 protein; however, most of the literature available to date has been focused on the common SNPs of *PCSK9* (i.e. S127R, R46L, Q152H, and D374Y). Hence, future studies incorporating molecular docking, structural modeling and molecular dynamics simulations, together with visualization tools such as PyMol and iCn3D, are warranted to elucidate the three-dimensional structural differences between wild-type and mutant-type PCSK9 protein.

CONCLUSION

The most significant nsSNPs W156R, H226L, H229R, G337R and G394V may disrupt the second autocatalytic processing of PCSK9 protein and interrupt the bonding interaction between LDLR and PCSK9 proteins significantly, thus serving as a new target for diagnostic biomarkers and drug development.

ACKNOWLEDGEMENT

The study was funded by Universiti Tunku Abdul Rahman Research Publication Scheme (UTARRPS 6251/L10) and Universiti Tunku Abdul Rahman Research Fund [IPSR/RMC/UTARRF/2021- C1/L07].

CONFLICT OF INTEREST

The authors declared no conflict of interest.

ETHICAL STATEMENT

Not applicable.

REFERENCES

Abifadel, M., Varret, M., Rabès, J.P., Allard, D., Ouguerram, K., Devillers, M. & Boileau, C. 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature Genetics, 34(2): 154-156. https://doi.org/10.1038/ng1161

Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics, 76(1): 7-20. https://doi.org/10.1002/0471142905.hg0720s76

Au, A. & Wei, L.K. 2017. The impact of APOA5, APOB, APOC3 and ABCA1 gene polymorphisms on ischemic stroke: evidence from a meta-analysis. Atherosclerosis, 265: 60-70. https://doi.org/10.1016/j.atherosclerosis.2017.08.018

Au, A., Griffiths, L.R., Cheng, K.K., Kooi, C.W., Irene, L. & Wei, L.K. 2015. The influence of OLR1 and PCSK9 gene polymorphisms on ischemic stroke: evidence from a meta-analysis. Scientific Reports, 5: 18224. https://doi.org/10.1038/srep18224

Benjannet, S., Hamelin, J., Chrétien, M. & Seidah, N.G. 2012. Loss-and gain-of-function PCSK9 variants cleavage specificity, dominant negative effects, and low density lipoprotein receptor (LDLR) degradation. Journal of Biological Chemistry, 287(40): 33745-33755. https://doi.org/10.1074/jbc.M112.399725

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A.J., ... & Xenarios, I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. In: Plant Bioinformatics: Methods and Protocols, pp. 23-54. Springer, New York. https://doi.org/10.1007/978-1-4939-3167-5_2

Bromberg, Y. & Rost, B. 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Research, 35(11): 3823-3835. https://doi.org/10.1093/nar/gkm238

Cameron, J., Holla, Ø.L., Laerdahl, J.K., Kulseth, M.A., Ranheim, T., Rognes, T. & Leren, T.P. 2008. Characterization of novel mutations in the catalytic domain of the PCSK9 gene. Journal of Internal Medicine, 263(4): 420-431. https://doi.org/10.1111/j.1365-2796.2007.01915.x

Capriotti, E. & Fariselli, P. 2017. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Research, 45(W1): W247-W252. https://doi.org/10.1093/nar/gkx338

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P.L., Altman, R.B. & Casadio, R. 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics, 14(3): S6. https://doi.org/10.1186/1471-2164-14-S3-S6

Chaudhary, R., Garg, J., Shah, N. & Sumner, A. 2017. PCSK9 inhibitors: a new era of lipid lowering therapy. World Journal of Cardiology, 9(2): 76-91. https://doi.org/10.4330/wjc.v9.i2.76

Cheng, J., Randall, A. & Baldi, P. 2006. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins: Structure, Function, and Bioinformatics, 62(4): 1125-1132. https://doi.org/10.1002/prot.20810

Choi, Y. & Chan, A.P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics, 31(16): 2745-2747. https://doi.org/10.1093/bioinformatics/btv195

Chorba, J.S., Galvan, A.M. & Shokat, K.M. 2018. Stepwise processing analyses of the single-turnover PCSK9 protease reveal its substrate sequence specificity and link clinical genotype to lipid phenotype. Journal of Biological Chemistry, 293(6): 1875-1886. https://doi.org/10.1074/jbc.M117.811315

Deng, W., Wang, Y., Ma, L., Zhang, Y., Ullah, S. & Xue, Y. 2017. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. Briefings in Bioinformatics, 18(4): 647-658. https://doi.org/10.1093/bib/bbw041

Ferrer-Costa, C., Gelpí, J.L., Zamakola, L., Parraga, I., De La Cruz, X. & Orozco, M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics, 21(14): 3176-3178. https://doi.org/10.1093/bioinformatics/bti486

Glaser, L. 1988. The biology and enzymology of eukaryotic protein acylation. Annual Review of Biochemistry, 57: 69-99. https://doi.org/10.1146/annurev.bi.57.070188.000441

Han, B.H., Sutin, D., Williamson, J.D., Davis, B.R., Piller, L.B., Pervin, H., Pressel, S.L. & Blaum, C.S. 2017. Effect of statin treatment vs usual care on primary cardiovascular prevention among older adults: the ALLHAT-LLT randomized clinical trial. JAMA Internal Medicine, 177(7): 955-965. https://doi.org/10.1001/jamainternmed.2017.1442

Hu, J., Li, Y., Zhang, Y. & Yu, D.J. 2018. ATPbind: accurate protein–ATP binding site prediction by combining sequence-profiling and structure-based comparisons. Journal of Chemical Information and Modeling, 58(2): 501-510. https://doi.org/10.1021/acs.jcim.7b00397

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., ... & Sigrist, C.J. 2006. The PROSITE database. Nucleic Acids Research, 34(suppl_1): D227-D230. https://doi.org/10.1093/nar/gkj063

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., ... & Flicek, P. 2021. Ensembl 2021. Nucleic Acids Research, 49(D1): D884-D891. https://doi.org/10.1093/nar/gkaa942

Parthiban, V., Gromiha, M.M. & Schomburg, D. 2006. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Research, 34(suppl_2): W239-W242. https://doi.org/10.1093/nar/gkl190

Pejaver, V., Hsu, W.L., Xin, F., Dunker, A.K., Uversky, V.N. & Radivojac, P. 2014. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Science, 23(8): 1077-1093. https://doi.org/10.1002/pro.2494

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.J., Mort, M., Cooper, D.N., Sebat, J., Iakoucheva, L.M. & Mooney, S.D. 2017. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. BioRxiv, 134981. https://doi.org/10.1101/134981

Peloso, G.M., Lange, L.A., Varga, T.V., Nickerson, D.A., Smith, J.D., Griswold, M.E., Musani, S., Polfus, L.M., Mei, H., Gabriel, S. & Quarells, R.C. 2016. Association of exome sequences with cardiovascular traits among blacks in the Jackson Heart Study. Circulation: Cardiovascular Genetics, 9(4): 368-374. https://doi.org/10.1161/CIRCGENETICS.115.001384

Peng, J. & Xu, J. 2011. RaptorX: exploiting structure information for protein alignment by statistical inference. Proteins: Structure, Function, and Bioinformatics, 79(S10): 161-171. https://doi.org/10.1002/prot.23175

Pires, D.E., Ascher, D.B. & Blundell, T.L. 2014. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Research, 42(W1): W314-W319. https://doi.org/10.1093/nar/gku411

Protein Data Bank. 1971. Protein Data Bank. Nature New Biology, 233(223): 10-1038. https://doi.org/10.1038/newbio233223b0

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., ... & Cooper, D.N. 2003. Human gene mutation database (HGMD®): 2003 update. Human Mutation, 21(6): 577-581. https://doi.org/10.1002/humu.10212

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. & Narechania, A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Research, 13(9): 2129-2141. https://doi.org/10.1101/gr.772403

Udenwobele, D.I., Su, R.C., Good, S.V., Ball, T.B., Varma Shrivastav, S. & Shrivastav, A. 2017. Myristoylation: an important protein modification in the immune response. Frontiers in Immunology, 8: 751. https://doi.org/10.3389/fimmu.2017.00751

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. 2016. SIFT missense predictions for genomes. Nature Protocols, 11(1): 1-12. https://doi.org/10.1038/nprot.2015.123

Wang, L. & Sauer, U. 2010. Prediction of protein stability changes due to single amino acid mutations. BMC Bioinformatics, 11: 190. https://doi.org/10.1186/1471-2105-11-S1-S5

Weider, E., Susan-Resiga, D., Essalmani, R., Hamelin, J., Asselin, M.C., Nimesh, S., Ashraf, Y., Wycoff, K.L., Zhang, J., Prat, A. & Seidah, N.G. 2016. Proprotein convertase subtilisin/kexin type 9 (PCSK9) single domain antibodies are potent inhibitors of low density lipoprotein receptor degradation. Journal of Biological Chemistry, 291(32): 16659-16671. https://doi.org/10.1074/jbc.M116.717736

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., ... & Yaschenko, E. 2007. Database resources of the national center for biotechnology information. Nucleic Acids Research, 36(suppl_1): D13-D21. https://doi.org/10.1093/nar/gkm1000

Yang, J., Roy, A. & Zhang, Y. 2013. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics, 29(20): 2588-2595. https://doi.org/10.1093/bioinformatics/btt447

Zhang, C., Freddolino, P.L. & Zhang, Y. 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. Nucleic Acids Research, 45(W1): W291-W299. https://doi.org/10.1093/nar/gkx366

Zhang, L., Deng, T., Pan, S., Zhang, M., Zhang, Y., Yang, C., ... & Mi, J. 2024. DeepO-GlcNAc: A web server for prediction of protein O-GlcNAcylation sites using deep learning combined with attention mechanism. Frontiers in Cell and Developmental Biology, 12: 1456728. https://doi.org/10.3389/fcell.2024.1456728

Zhang, Y., Ultsch, M., Skelton, N.J., Burdick, D.J., Beresini, M.H., Li, W., Kong-Beltran, M., Peterson, A., Quinn, J., Chiu, C. & Wu, Y. 2017. Discovery of a cryptic peptide-binding site on PCSK9 and design of antagonists. Nature Structural & Molecular Biology, 24(10): 848-856. https://doi.org/10.1038/nsmb.3453

Zhou, F., Xue, Y., Yao, X. & Xu, Y. 2006. CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). Bioinformatics, 22(7): 894-896. https://doi.org/10.1093/bioinformatics/btl013