IJBPAS, October, 2022, 11(10): 4812-4824

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

MOLECULAR MECHANICS STUDIES ON THE MUTATED AMINO ACID REGIONS AND MOTIFS OF LDLR (LOW-DENSITY LIPOPROTEIN RECEPTOR) AND THEIR 3D VISUALIZATION

M. AARTHI [*1], SHAKILA L[2], N. JANANI [3] AND M. AGALYA [4]

1, 3, 4: Research Scholar, P.G and Research Department of Zoology, Ethiraj College for Women, Chennai

2: Assistant Professor, P.G and Research Department of Zoology, Ethiraj College for Women, Chennai

***Corresponding Author: Aarthi Murugavel: E Mail: aarthi.4496@gmail.com**

Received 17th Dec. 2021; Revised 23rd Jan. 2022; Accepted 12th March 2022; Available online 1st Oct. 2022

https://doi.org/10.31032/IJBPAS/2021/11.10.6506

ABSTRACT

The most important genetic cause of Familial Hypercholesterolemia (FH) is mutations occurring in the *LDLR* gene encoding the Low-density Lipoprotein Receptor. Mutations in this domain account for 51.7% of the total missense variants observed in LDLR. The aim of the current *Insilico* research work is to find out the involvement of the mutated amino acids in the HTH motifs regions of LDLR protein 3D structure using *Insilico* protocols. At first, *Insilico* study was carried out using HTH (Helix-Turn-Helix) motif tool and the HTH motifs regions present in the LDLR protein sequence was identified. Secondarily, the 3D structure of the selected LDLR sequence was explored using an automated homology modelling server and was visualized using an advanced molecular visualization software in order to identify the HTH motifs regions present in the mutated amino acids. The results obtained from this study shows that the mutated amino acids are directly involved in the HTH motif regions. The overall results clearly elucidate that the mutated amino acids are present in the HTH motifs regions and play a vital role in the field of research related to Structure Based –Drug Designing and clinical Pharmacology.

Keywords: LDLR, Protein Modelling and HTH (Helix-Turn-Helix) motifs

INTRODUCTION

Familial Hypercholesterolemia (FH; MIM# 143890) is a common hereditary disorder that causes abnormally high levels of low density lipoprotein cholesterol (LDL-C) from birth. If left untreated, FH greatly raises cardiovascular risk **[1, 2]** and causes the early start of atherosclerosis, increasing the risk of a heart attack, stroke, and death **[3]**. Among all the proteins implicated in the pathogenesis of this disease, the low density lipoprotein receptor (LDLR) (MIM# 606945) is the most prevalent hereditary cause, accounting for 80–85 percent of FH cases **[4]**. Over 2600 LDLR variations have been described to date (ClinVar database).

The LDLR gene is found on the short arm of chromosome 19 (19p13.1–13.3) and has a length of about 45 kb, encoding 18 exons and 17 introns. LDLR is an 839-amino acid protein that is generated in the endoplasmic reticulum (ER) and folds and is partially glycosylated. The mature protein is then glycosylated further in the Golgi apparatus **[5]**. The LDLR is divided into five functional domains: the N-terminal ligand-binding domain, the EGF-precursor homology domain, the O-linked sugars containing domain, the trans-membrane domain, and the C-terminal cytosolic domain **[6]**.

Mutations in LDLR can impair LDLR activity at various levels, so they are classified as follows: class 1 (no protein synthesis), class 2 (partial or complete retention of LDLR in the ER), class 3 (defective binding to apolipoprotein B (apoB), class 4 (defective endocytosis), and class 5 (defective endocytosis) (diminished LDLR recycling capacity) **[7, 8]**. LDLR's biological activity is to transport lipoproteins into cells, most notably low - density lipoprotein (LDL) **[9]**. When LDL binds to LDLR, the ligand-receptor complex enters the cell via clathrin-mediated endocytosis. Endosome acidification then releases cargo, allowing LDLR recycling back to the cell membrane while LDL is destroyed in the lysosomes. The LDL-LDLR complex is lysosomally degraded when cargo release fails **[10]**.

METHODOLOGY

Target Protein Sequence Retrieval: The target LDLR protein sequence was selected based on clinico-genetic references and the amino acid sequence was retrieved from UniProt database in FASTA format and was manually converted to show the respective mutated positions.

HTH motifs (Helix-turn-Helix) prediction: The LDLR amino acid sequence

was analysed using GYM motif tool in order to identify the HTH motifs regions present in LDLR.

Protein Modelling and Validation: The selected LDLR sequence was converted into 3D structure using an automated protein homology modelling server called Swiss Model server. After modelling the 3D structure, it was validated using proCheck server for 3D structure quality assessment.

3D Structure Visualization: The modelled protein structure of LDLR was viewed using the advanced molecular visualization software, Discovery Studio software.

RESULTS

Figure 1 shows the FASTA format of Normal amino acids content of LDLR with amino acid positions (D:302,C:303,R:350,V:468,R:471) highlighted in yellow.

Figure 2 shows the 3D view of the normal protein structure of LDLR shown in secondary structure colour model with amino acid label (D: 302, C: 303, R: 350, V: 468, R: 471), visualized using Discovery Studio Software.

Figure 3 shows the Mutated amino acids sequence of LDLR with amino acid positions(G:302,Y:303,P:350,I:468,G:471) highlighted in yellow.

Figure 4 shows the 3D view of the Mutated protein structure of LDLR shown in Space fill model with amino acid (G:302,Y:303,P:350,I:468,G:471), visualized using Discovery Studio Software.

Assessment of ramachandran plot for the predicted mutated protein sequence of the modeled LDLR **(Figure 5)**.

Figure 6 represents the H-T-H (Helix –Turn-Helix) motifs regions present in the LDLR protein.

Figure 7 represents the H-T-H (Helix –Turn-Helix) motif regions present in the mutated LDLR protein. (Motif 1: 495-516 and 456-474 amino acids positions indicated in yellow colour.)

Figure 8 represent the 3D secondary structure colour view of mutated LDLR protein with Helix-Turn-Helix (HTH) motifs and amino acids regions (495-516) using Discovery studio software.

Figure 9 represents the 3D secondary structure colour view of mutated LDLR protein with Helix-Turn-Helix (HTH) motifs and amino acids regions (456-474) using Discovery studio software.

Figure 10 represent the 3D secondary structure colour view of mutated LDLR protein with Helix-Turn-Helix (HTH) motifs and amino acids regions (456-474) using Discovery studio software. The yellow region indicates the mutated amino acid G 471.

10 20 30 40 50
MGPWGWKLRW TVALLLAAAG TAVGDRCERN EFQCQDGKCI SYKWVCDGSA
60 70 80 90 100
ECQDGSDESQ ETCLSVTCKS GDFSCGGRVN RCIPQFWRCD GQVDCDNGSD
110 120 130 140 150
EQGCPPKTCS QDEFRCHDGK CISRQFVCDS DRDCLDGSDE ASCPVLTCGP
160 170 180 190 200
ASFQCNSSTC IPQLWACDND PDCEDGSDEW PQRCRGLYVF QGDSSPCSAF
210 220 230 240 250
EFHCLSGECI HSSWRCDGGP DCKDKSDEEN CAVATCRPDE FQCSDGNCIH
260 270 280 290 300
GSRQCDREYD CKDMSDEVGC VNVTLCEGPN KFKCHSGECI TLDKVCNMAR
310 320 330 340 350
DCRDWSDEPI KECGTNECLD NNGGCSHVCN DLKIGYECLC PDGFQLVAQR
360 370 380 390 400
RCEDIDECQD PDTCSQLCVN LEGGYKCQCE EGFQLDPHTK ACKAVGSIAY
410 420 430 440 450
LFFTNRHEVR KMTLDRSEYT SLIPNLRNVV ALDTEVASNR IYWSDLSQRM
460 470 480 490 500
ICSTQLDRAH GVSSYDTVIS RDIQAPDGLA VDWIHSNIYW TDSVLGTVSV
510 520 530 540 550
ADTKGVKRKT LFRENGSKPR AIVVDPVHGF MYWTDWGTPA KIKKGGLNGV
560 570 580 590 600
DIYSLVTENI QWPNGITLDL LSGRLYWVDS KLHSISSIDV NGGNRKTILE
610 620 630 640 650
DEKRLAHPFS LAVFEDKVFW TDIINEAIFS ANRLTGSDVN LLAENLLSPE
660 670 680 690 700
DMVLFHNLTQ PRGVNWCERT TLSNGGCQYL CLPAPQINPH SPKFTCACPD
710 720 730 740 750
GMLLARDMRS CLTEAEAAVA TQETSTVRLK VSSTAVRTQH TTTRPVPDTS
760 770 780 790 800
RLPGATPGLT TVEIVTMSHQ ALGDVAGRGN EKKPSSVRAL SIVLPIVLLV
810 820 830 840 850
FLCLGVFLLW KNWRLKNINS INFDNPVYQK TTEDEVHICH NQDGYSYPSR
860
QMVSLEDDVA

Figure 1: Protein Sequence of LDLR - UniProt Database

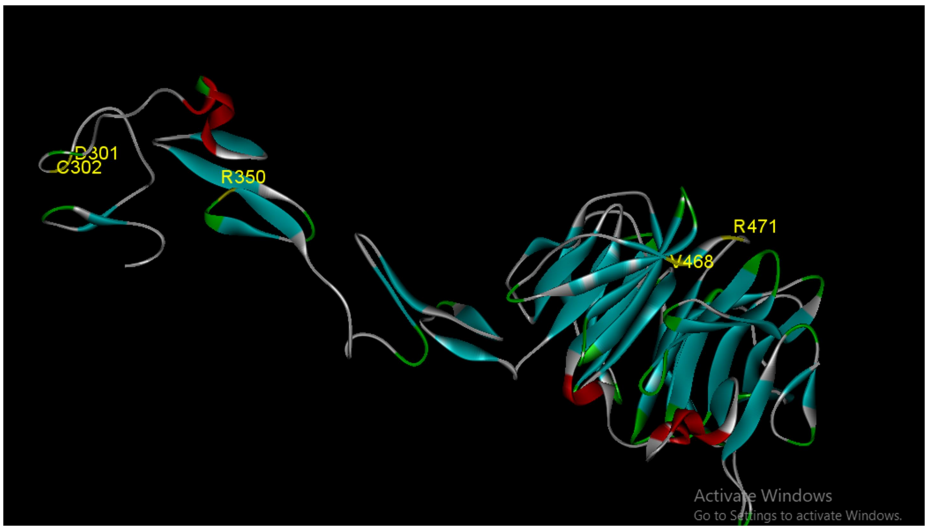

Figure 2: Protein Modelling: 3D structure of LDLR

10 20 30 40 50
MGPWGWKLRW TVALLLAAAG TAVGDRCERN EFQCQDGKCI SYKWVCDGSA
60 70 80 90 100
ECQDGSDESQ ETCLSVTCKS GDFSCGGRVN RCIPQFWRCD GQVDCDNGSD
110 120 130 140 150
EQGCPPKTCS QDEFRCHDGK CISRQFVCDS DRDCLDGSDE ASCPVLTCGP
160 170 180 190 200
ASFQCNSSTC IPQLWACDND PDCEDGSDEW PQRCRGLYVF QGDSSPRSAF
210 220 230 240 250
EFHCLSGECI HSSWRCDGGP DCKDKSDEEN CAVATCRPDE FQCSDGNYIH
260 270 280 290 300
GSRQCDREYD CKDMSDEVGC VNVTLCEGPN KFKCHSGECI TLDKVCNMAR
310 320 330 340 350
GYRDWSDEPI KECGTNECLD NNGGCSHVCN DLKIGYECLC PDGFQLVAQP
360 370 380 390 400
RCEDIDECQD PDTCSQLCVN LEGGYKCQCE EGFQLDPHTK ACKAVGSIAY
410 420 430 440 450
LFFTNRHEVR KMTLDRSEYT SLIPNLRNVV ALDTEVASNR IYWSDLSQRM
460 470 480 490 500
ICSTQLDRAH GVSSYDTIIS GDIQAPDGLA VDWIHSNIYW TDSVLGTVSV
510 520 530 540 550
ADTKGVKRKT LFRENGSKPR AIVVDPVHGF MYWTDWGTPA KIKKGGLNGV
560 570 580 590 600
DIYSLVTENI QWPNGITLDL LSGRLYWVDS KLHSISSIDV NGGNRKTILE
610 620 630 640 650
DEKRLAHPFS LAVFEDKVFW TDIINEAIFS ANRLTGSDVN LLAENLLSPE
660 670 680 690 700
DMVLFHNLTQ PRGVNWCERT TLSNGGCQYL CLPAPQINPH SPKFTCACPD
710 720 730 740 750
GMLLARDMRS CLTEAEAAVA TQETSTVRLK VSSTAVRTQH TTTRPVPDTS
760 770 780 790 800
RLPGATPGLT TVEIVTMSHQ ALGDVAGRGN EKKPSSVRAL SIVLPIVLLV
810 820 830 840 850
FLCLGVFLLW KNWRLKNINS INFDNPVYQK TTEDEVHICH NQDGYSYPSR
860
QMVSLEDDVA

Figure 3: Mutated sequence of LDLRprotein - UniProt Database

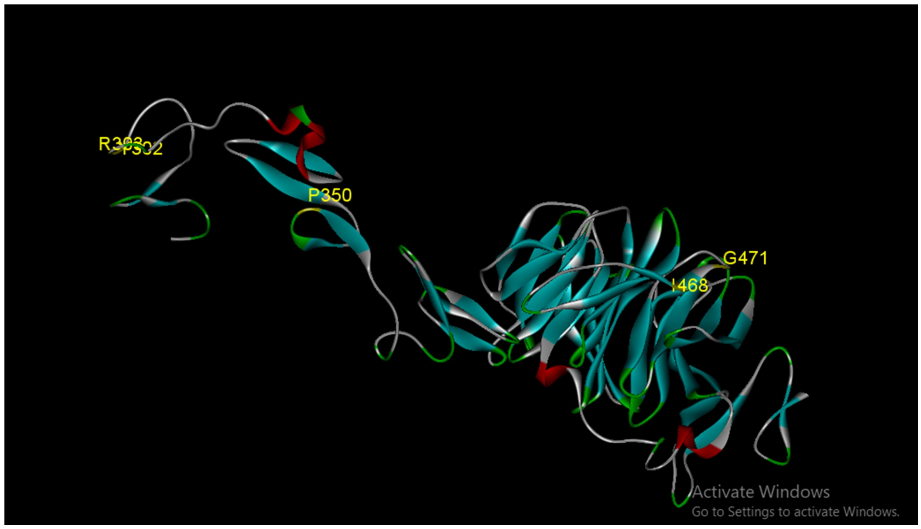

Figure 4: Protein Modelling: Mutated 3D structure of LDLR

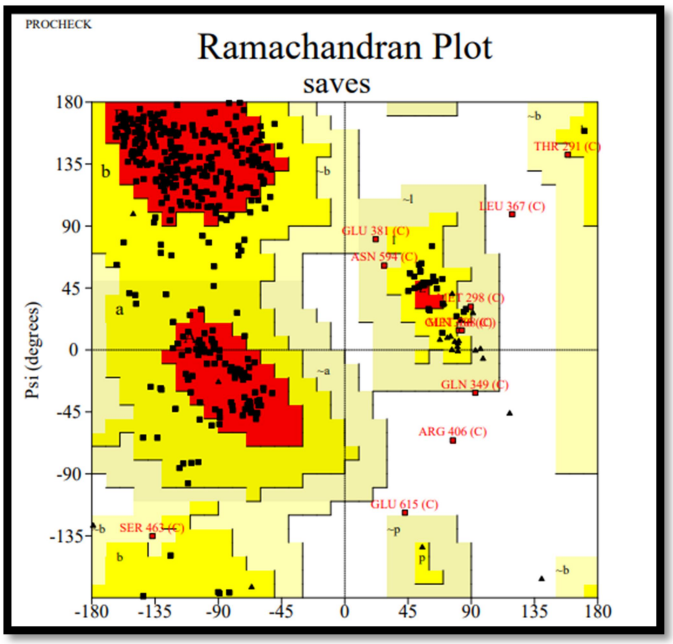

Figure 5: 3D structure Validation of LDLR

Plot statistics

Residues in most favoured regions [A,B,L]	306	79.7%
Residues in additional allowed regions [a,b,l,p]	67	17.4%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.8%
Residues in disallowed regions	4	1.0%
	----	------
Number of non-glycine and non-proline residues	384	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	32	
Number of proline residues	20	

Total number of residues	440	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

GYM 2.0: Results for Your Sequence

*** GYM Results Summary ***

Input Sequence with Highlighted HTH Motif Locations:

MGPWGWKLRWTVALLLAAAGTAVGDRCERNEFQCQDGKCISYKWVCDGSA
ECQDGSDESQ EETCLSVTCKSGDFSCGGRVNRCIPQFWRCDGQVDCDNG
SDEQGCPPKTCSQDEFRCHDGK CCISRQFVCDSDRDCLDGSDEASCPVL
TCGPASFQCNSSTCIPQLWACDNDPDCEDGSDEW PPQRCRGLYVFQGDS
SPCSAFEFHCLSGECIHSSWRCDGGPDCKDKSDEENCAVATCRPDE FFQ
CSDGNCIHGSRQCDREYDCKDMSDEVGCVNVTLCEGPNKFKCHSGECITL
DKVCNMAR DDCRDWSDEPIKECGTNECLDNNGGCSHVCNDLKIGYECLC
PDGFQLVAQRRCEDIDECQD PPDTCSQLCVNLEGGYKCQCEEGFQLDPH
TKACKAVGSIAYLFFTNRHEVRKMTLDRSEYT SSLIPNLRNVVALDTEV
ASNRIYWSDLSQRMICSTQLDRAHGVSSYDTVISRDIQAPDGLA VVDWI
HSNIYWTDSVLGTVSVADTKGVKRKTLFRENGSKPRAIVVDPVHGFMYWT
DWGTPA KKIKRGGLNGVDIYSLVTENIQWPNGITLDLLSGRLYWVDSKL
HSISSIDVNGGNRKTILE DDEKRLAHPFSLAVFEDKVFWTDIINEAIFS
ANRLTGSDVNLLAENLLSPEDMVLFHNLTQ PPRGVNWCERTTLSNGGCQ
YLCLPAPQINPHSPKFTCACPDGMLLARDMRSCLTEAEAAVA TTQETST
VRLKVSSTAVRTQHTTTRPVPDTSRLPGATPGLTTVEIVTMSHQALGDVA
GRGN EEKKPSSVRALSIVLPIVLLVFLCLGVFLLWKNWRLKNINSINFD
NPVYQKTTEDEVHICH NNQDGYS

Length of Sequence = 874

Predicted HTH Motif Locations:

Pick	Loc	LP	NPM	Score	Detected?	Motif
Best	511	5	15	47	+	LGTVSVADTKGVKRKTLFRENG
2nd	467	4	10	33	+	STQLDRAHGVSSYDTVISRDIQ

Figure 6: Motif Prediction: GYM Tool

10 20 30 40 50
MGPWGWKLRW TVALLLAAAG TAVGDRCERN EFQCQDGKCI SYKWVCDGSA
60 70 80 90 100
ECQDGSDESQ ETCLSVTCKS GDFSCGGRVN RCIPQFWRCD GQVDCDNGSD
110 120 130 140 150
EQGCPPKTCS QDEFRCHDGK CISRQFVCDS DRDCLDGSDE ASCPVLTCGP
160 170 180 190 200
ASFQCNSSTC IPQLWACDND PDCEDGSDEW PQRCRGLYVF QGDSSPCSAF
210 220 230 240 250
EFHCLSGECI HSSWRCDGGP DCKDKSDEEN CAVATCRPDE FQCSDGNCIH
260 270 280 290 300
GSRQCDREYD CKDMSDEVGC VNVTLCEGPN KFKCHSGECI TLDKVCNMAR
310 320 330 340 350
DCRDWSDEPI KECGTNECLD NNGGCSHVCN DLKIGYECLC PDGFQLVAQR
360 370 380 390 400
RCEDIDECQD PDTCSQLCVN LEGGYKCQCE EGFQLDPHTK ACKAVGSIAY
410 420 430 440 450
LFFTNRHEVR KMTLDRSEYT SLIPNLRNVV ALDTEVASNR IYWSDLSQRM
460 470 480 490 500
ICSTQLDRAH GVSSYDTVIS RDIQAPDGLA VDWIHSNIYW TDSVLGTVSV
510 520 530 540 550
ADTKGVKRKT LFRENGSKPR AIVVDPVHGF MYWTDWGTPA KIKKGGLNGV
560 570 580 590 600
DIYSLVTENI QWPNGITLDL LSGRLYWVDS KLHSISSIDV NGGNRKTILE
610 620 630 640 650
DEKRLAHPFS LAVFEDKVFW TDIINEAIFS ANRLTGSDVN LLAENLLSPE
660 670 680 690 700
DMVLFHNLTQ PRGVNWCERT TLSNGGCQYL CLPAPQINPH SPKFTCACPD
710 720 730 740 750
GMLLARDMRS CLTEAEAAVA TQETSTVRLK VSSTAVRTQH TTTRPVPDTS
760 770 780 790 800
RLPGATPGLT TVEIVTMSHQ ALGDVAGRGN EKKPSSVRAL SIVLPIVLLV
810 820 830 840 850
FLCLGVFLLW KNWRLKNINS INFDNPVYQK TTEDEVHICH NQDGYSYPSR
860
QMVSLEDDVA

Figure 7: Motif Prediction: GYM Tool

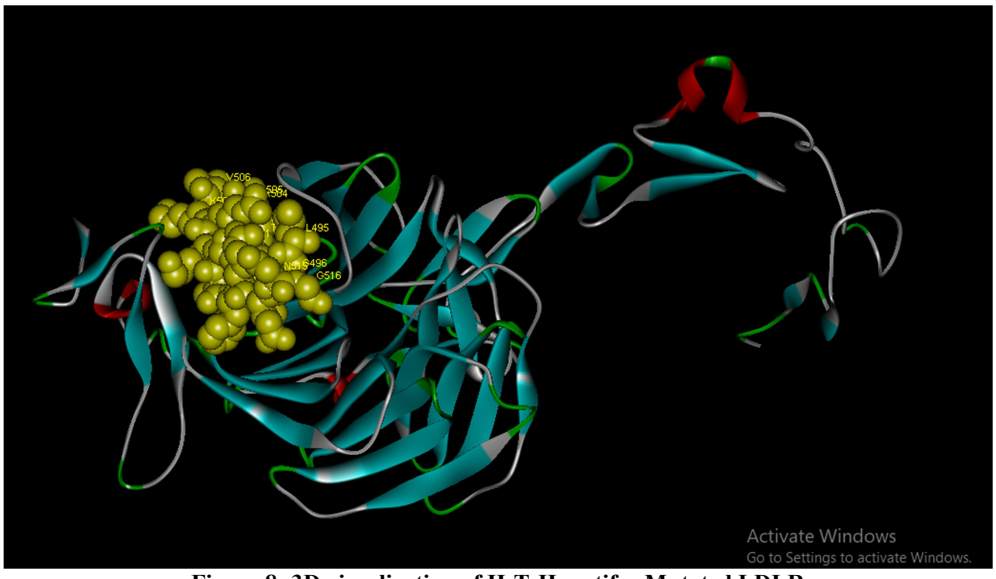

Figure 8: 3D visualization of H-T-H motifs –Mutated LDLR

Figure 9: 3D visualization of H-T-H motifs –mutated LDL

Figure 10: 3D visualization of H-T-H motifs –mutated LDLR

DISCUSSION

In this research investigation, the extract information of the protein responsible for Familial hypercholesterolemia, LDLR (UniProt: P01130). It is present in the 19th chromosome and the length of its gene sequence is 2583nt and the length of its protein sequence is 860aa. In the primary step of the investigation, we perform Helix-Turn-Helix motif analysis of LDLR using GYM motif tool. The normal amino acids sequence with respective positions are shown in **Figure 1** and the mutated mono acids sequence with respective position are shown in **Figure 3**. The mutated positions are changed manually corresponding to the normal amino acids by making use of Clinical literature **[11-15]**.

(Leitersdorf *et al*., 1990) **[16]** investigated the LDL receptor genes of 11

French Canadian FH homozygotes. Only 3 different LDLR haplotypes were identified, and the corresponding coding region of the allele was sequenced. Three different missense mutations were recognized. Assays were performed to identify each of these mutations and were directly applied to 130 FH heterozygotes from the greater Montreal area. The common deletion (606945.0025) which accounts foraround60% of cases (Hobbs *et al*., 1987) **[17]** and the smaller deletion (606945.0026) detected by (Ma *et al*., 1989) **[18]** and found in approximately 5% of French Canadians were also identified. LDL receptor mutations were identified by them in 76% of the subjects and 14% had 1 of the 3 missense mutations. In the Saguenay-Lac-Saint-Jean region of Quebec province, (De Braekeleer, 1991) **[19]** estimated the prevalence of Familial Hypercholesterolemia as 1/122, in comparison with the habitually estimated frequency of 1/500 among the people of Europe.

Gym motif tools to identify the H-T-H motif **[20, 21]**. The HTH motif is present in several proteins which is responsible for gene expression regulation. In the results obtained from the study of helix-turn-helix motifs, the LDLR protein sequence shows an amino acid range of 456 to 474(LDRAHGVSSYDTVISRDIQ) and 495 to 506 (LGTVSV ADTKGVKRKTLFRENG. Of this, the R471G amino acid mutation appears to fall in the helix-turn-helix range of 456 to 474 **(Figure 6 and 7)**.

In this research study, SWISS-MODEL was used to convert the amino acid sequence of LDLR into 3D structure **(Figure 2 and 4)**. SWISS-MODEL **[22-25]** was used to analyse the molecular and structural details of LDLR in an elaborate manner for the purpose of docking. SWISS-MODEL is a server for automated comparative modelling of three-dimensional (3D) protein structures. (Waterhouse *et al*. 2018) **[23]** computed models by the SWISS-MODEL server homology modelling pipeline which is based on ProMod3, an in-house comparative modelling engine based on Open Structure. The modelled 3D protein was thoroughly estimated using ProCheck server **[26]** for assessment of Ramachandran Plot. After modelling, the 3D structure of the mutated protein was validated using proCheck server. **Figure 5** shows the assessment of Ramachandran Plot which confirms that there is no error (79.9 %) in the modelled protein **(Figure 2 and 4)**. All the above results were clearly elucidated. It is proved that the Helix-Turn-Helix motif falls within

the mutated amino acid range **(Figure 8, 9 and 10)**.

CONCLUSION

In proteins, a structural motif describes the relationship between secondary structural elements. An individual motif generally comprises of only a few elements, e.g., the 'helix-turn-helix' motif. In this research, we completely focus on the Helix – turn –Helix motifs regions involved in the mutated amino acids positions. Finally, we conclude that in the modelled mutated protein target, LDLR, amino acid mutation (R471G) occurs in the regions of the Helix-turn-Helix motifs. Hence, the entire results would play a pivotal role in the field of computer-aided drug designing and drug designing.

ACKNOWLEDGMENT

The authors acknowledge the help extended by Dr. Balaji Munivelan, Ph. D., CEO and Senior Bioinformatician (bioinfobalaji@gmail.com) ABS Geno-informatics, Chennai, for his contribution towards *Insilico* drug docking studies.

REFERENCES

[1] Hopkins, P. N., Toth, P. P., Ballantyne, C. M., and Rader, D. J. **Familial hypercholesterolemias: prevalence, genetics, diagnosis and screening recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia**. *Journal of clinical lipidology*, *5*(3), 2011, S9-S17.

[2] Talmud, P. J., Futema, M., and Humphries, S. E. **The genetic architecture of the familial hyperlipidaemia syndromes: rare mutations and common variants in multiple genes**. *Current opinion in lipidology*, *25*(4), 2014, pp: 274-281.

[3] Ference, B. A., Ginsberg, H. N., Graham, I., Ray, K. K., Packard, C. J., Bruckert, E., ... and Catapano, A. L. **Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel**. *European heart journal*, *38*(32), 2017, pp: 2459-2472.

[4] Benito-Vicente, A., Uribe, K. B., Jebari, S., Galicia-Garcia, U., Ostolaza, H., and Martin, C. **Familial hypercholesterolemia: the most frequent cholesterol metabolism disorder caused disease**. *International journal of*

molecular sciences, *19*(11), 2018**,** pp: 3426.

[5] Hobbs, H. H., Brown, M. S., and Goldstein, J. L. **Molecular genetics of the LDL receptor gene in familial hypercholesterolemia.** *Human mutation*, *1*(6), 445-466.

[6] GOLDSTEIN, J. L. (2009). BROWN Michael S. **The LDL receptor***, ArteriosclThrom Vas*, *29*(4), 1992, pp: 431-8.

[7] Hobbs, H. H., Russell, D. W., Brown, M. S., and Goldstein, J. L. **The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein.** *Annual review of genetics*, *24*(1), 1990, pp: 133-170.

[8] Tolleshaug, H., Goldstein, J. L., Schneider, W. J., and Brown, M. S. **Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia**. *Cell*, *30*(3), 1982, pp: 715-724.

[9] Jeon, H., Meng, W., Takagi, J., Eck, M. J., Springer, T. A., and Blacklow, S. C. **Implications for Familial Hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair**. *Nature structural biology*, *8*(6), (2001), pp: 499-504.

[10] Huang, S., Henry, L., Ho, Y. K., Pownall, H. J., andRudenko, G.**Mechanism of LDL binding and release probed by structure-based mutagenesis of the LDL receptor [S].** *Journal of lipid research*, *51*(2), 2010, pp: 297-308.

[11] Cenarro A., Jensen H.K., Casao E., Civeira F., Gonzalez-Bonillo J., Rodriguez-Rey J.C., Gregersen N., Pocovi M.Identification of recurrent and novel mutations in the LDL receptor gene in Spanish patients with familial hypercholesterolemia.*Hum. Mutat*: *11*, 1998,pp: 413-413.

[12] Webb, J. C., Sun, X. M., McCarthy, S. N., Neuwirth, C., Thompson, G. R., Knight, B. L., andSoutar, A. K. **Characterization of mutations in the low density lipoprotein (LDL)-receptor gene in patients with homozygous familial hypercholesterolemia, and frequency of these mutations in FH patients in the United Kingdom**. *Journal of lipid research*, *37*(2), 1996,pp**:** 368-381.

[13] Day, I. N. M., Whittall, R. A., O'dell, S. D., Haddad, L., Bolla, M. K., Gudnason, V., and Humphries, S. E. **Spectrum of LDL receptor gene mutations in heterozygous familial hypercholesterolemia**. *Human mutation*, *10*(2), 1997, pp: 116-127.

[14] Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., ...and Lander, E. S.. **Characterization of single-nucleotide polymorphisms in coding regions of human genes.** *Nature genetics*, *22*(3), 1999, pp: 231-238.

[15] Al-Khateeb, A., Zahri, M. K., Mohamed, M. S., Sasongko, T. H., Ibrahim, S., Yusof, Z., andZilfalil, B. A. **Analysis of sequence variations in low-density lipoprotein receptor gene among Malaysian patients with familial hypercholesterolemia**. *BMC medical genetics*, *12*(1),2011, pp: 1-11.

[16] Leitersdorf, E., Tobin, E. J., Davignon, J., and Hobbs, H. H. **Common low-density lipoprotein receptor mutations in the French Canadian population.** *The Journal of clinical investigation*, *85*(4), 1990,pp: 1014-1023.

[17] Hobbs, H. H., Brown, M. S., Russell, D. W., Davignon, J., and Goldstein, J. L. **Deletion in the gene for the low-density-lipoprotein receptor in a majority of French Canadians with familial hypercholesterolemia**. *New England Journal of Medicine*, *317*(12), 1987, pp: 734-737.

[18] Yuanhong, M. A., Bétard, C., Roy, M., Davignon, J., andKessling, A. M. **Identification of a second "French Canadian" LDL receptor gene deletion and development of a rapid method to detect both deletions**. *Clinical genetics*, *36*(4), 1989, pp: 219-228.

[19] De Braekeleer, M. **Hereditary disorders in Saguenay-Lac-St-Jean (Quebec, Canada)**. *Human heredity*, *41*(3), 1991, pp: 141-146.

[20] Narasimhan, G., Bu, C., Gao, Y., Wang, X., Xu, N., andMathee, K. **Mining protein sequences for motifs**. *Journal of Computational Biology*, *9*(5), 2002, pp: 707-720.

[21] Gao, Y., Mathee, K., Narasimhan, G., and Wang, X. **Motif detection**

in protein sequences. In *6th International Symposium on String Processing and Information Retrieval. 5th International Workshop on Groupware (Cat. No. PR00268)*, 1999, pp: 63-72. IEEE.

[22] Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., andSchwede, T. **The SWISS-MODEL Repository—new features and functionality**. *Nucleic acids research*, *45*(D1), 2017**,** D313-D319.

[23] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R. and Schwede, T. **SWISS-MODEL: homology modeling of protein structures and complexes**. *Nucleic acids research*, *46*(W1), 2018, W296-W303.

[24] Guex, N., Peitsch, M. C., andSchwede, T. **Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective.** *Electrophoresis*, *30* (S1), 2009, S162-S173.

[25] Studer, G., Tauriello, G., Bienert, S., Biasini, M., Johner, N., andSchwede, T. ProMod3—**A versatile homology modelling toolbox**. *PLoS computational biology*, *17*(1), 2021, e1008667.

[26] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. **PROCHECK: a program to check the stereochemical quality of protein structures**. *Journal of applied crystallography*, *26*(2), 1993, pp: 283-291.