ISSN No:-2456-2165

EGR-1 Gene Polymorphisms Unveiled: Identifying Key Harmful Mutations for Targeted Therapeutic Strategies for Human Genetic Diseases

(Short Title: Forecasting Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) within the Human EGR-1 Gene)

Yamini Sri Sekar; Kuralayanapalya Puttahonnappa Suresh; Uma Bharathi Indrabalan; Madhumitha B.; Swati Rani; Sharanagouda S. Patil
ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru-560064,India.

Azhahianambi Palavesam
Tamil Nadu Veterinary and Animal Sciences University, Chennai, Tamil Nadu-600051,India

Nagendra Nath Barman
College of Veterinary Science, Assam Agricultural University (AAU), Guwahati 781022, India

Corresponding author: Dr. Kuralayanapalya Puttahonnappa Suresh, Ph.D. (NIMHANS)
Principal Scientist-Biostatistics Spatial Epidemiology lab ICAR-NIVEDI, Post Box 6450 Yelahanka, Bengaluru-560064

Abstract:- The EGR-1 gene encodes the EGR-1 protein in humans, featuring three zinc finger domains crucial for DNA binding and transcriptional regulation. While comprehensive investigations into the EGR-1 gene have occurred, a research gap persists in the systematic analysis of single nucleotide polymorphisms (SNPs) associated with this gene. This study aimed to fill this void by identifying and compiling a systematic collection of damaging non-synonymous SNPs (nsSNPs) within the EGR-1 gene, seeking a better understanding of mutational changes. From 2,712 SNPs in the dbSNP database, 455 non-synonymous missense substitutions were selected for analysis. These substitutions were then used to identify dominant variants, assess potential consequences, and conduct mutation analyses based on subPSEC scores. Gibbs free energy changes affected by dominant substitutions were examined, distinguishing between neutral and effect variants, and evaluated functional and phenotypic effects.Further analysis involved modeling substitutions to assess their structural impact on the EGR-1 protein, elucidating disturbances in interaction with closely related genes and pinpointing changes in binding affinity and hydrogen bonds. Our analysis identified five nsSNPs as the most hazardous substitutions, with rs201213506 (H334N) recognized as the most detrimental mutation in the DNA binding region of the EGR-1 gene. This investigation effectively revealed the significant impact of the H334N mutation on the gene's regulatory capabilities. Further exploration and understanding of these genetic variations could lead to the identification of innovative therapeutic markers for various inherited human diseases.

Keywords:- *Early growth response(EGR)-1 gene; nsSNP; Computational analysis; protein-protein interaction; Docking.*

I. INTRODUCTION

The zinc finger transcription factor EGR-1 member of the EGR family, also known as zif268 NGFI-A, Krox24, and TIS8[1]. The human chromosomal region 5q23–31 comprises the EGR-1 gene. Three zinc finger motifs can be seen in the Egr-1 DNA-binding domain [2]. One intron and two exons make up the straightforward structure of the EGR-1 gene. The first exon codes the first 99 amino acids of the deduced protein, while the second exon codes the remaining proteins, which are found in the three tandem zinc finger motifs. [3]. The presence of five serum response elements (SREs) in the human EGR-1 promoter is particularly significant. SRE-mediated activity requires two distinct classes of transcription factors, the ternary complex factor and the the serum response factor (SRF). Elk1, Sap1, or Sap2 are ternary complex factors that, in order to be biologically active, must contact DNA and attach to SRF [4].

The EGR-1 gene is substantially conserved in mice, rats, chickens, zebrafish, chimps, dogs, cows, and humans, among other species[5]. Most notably, EGR-1 is recognized for its functions in the adult neurological system in humans, where it controls vital processes that underlie neuronal activity, including neurotransmission, synaptic plasticity, higher-order processes like learning and memory, in addition to the reward and stress responses[6]. EGR-1 dysregulation has been linked to a variety of additional disease conditions, including cardiovascular illness, ischemia-reperfusion injury, acute lung injury, and atopic dermatitis, among others. [7][8][9][10].

A SNP is a genomic variability in the DNA that occurs at a single base location (i.e., occurrences that exceed 1% of the time in the general population)). Non-synonymous SNPs (nsSNPs), also known as missense variants present in the coding region of the gene, are significant in part because they add amino acid variations into their encoded proteins. Due to selection against the functional disruptions caused by amino acid variation, nsSNPs are proportionally less

common than synonymous SNPs that do not impact protein sequence [11]. The consequences of nsSNPs on protein function continue to be considered to account for a significant amount of the molecular functional variation in the human population.

For instance, nsSNPs and their associated amino acid variations can alter the kinetic parameters of enzymes, the DNA-binding capabilities of transcription-controlling proteins, the signal transduction functions of transmembrane receptors, and the structural functions of structural proteins [12]. Splicing, transcription, translation, post-translational modification, and protein stability are the five key stages at which bioinformatics techniques forecast the functional implications of SNPs. The majority of the bioinformatics tools available today exclusively look at the effects of SNPs in relation to a single biological function. The others, however, offer a thorough analysis of SNP function based on various methodologies, resources, and data [13][14][15]

To confirm their alleged functional impact, all the SNPs identified in the EGR-1 gene were examined using several composite and singleton techniques. Prioritizing the SNPs that were found to have functional effects including pathogenicity and phenotypic effects, amino acids substitutions that reduced protein stability, and polymorphisms that alter protein-protein interactions of the EGR-1. In order to find causative variations, the current work entails sorting through a list of SNPs. Utilizing the in-silico method to estimate the harmful SNPs in the EGR-1 gene and their considerable pathogenic effects on the functions and structure.

II. METHODS AND MATERIALS

A. *Data retrieval*

The National Centre for Biological Information (NCBI) (https://www.ncbi.nlm.nih.gov/) and UniProtKB (https://www.uniprot.org/) were utilized to compile data on the human EGR-1 gene and its protein sequence (FASTA format). The Short Genetic Variation database (dbSNP) (https://www.ncbi.nlm.nih.gov/snp/) was utilized to retrieve the gene's SNPs . dbSNP is conceivable as a catalog of any short variations in human nucleotide sequence [16]. Only the nsSNPs (missense SNPs) were filtered and considered for further exploration.

B. *Utilizing SIFT for Sequence Homology*

Sorts intolerant from tolerant (SIFT)(https://sift.bii.a-star.edu.sg/) is a sequence homology-based tool that identify variation in protein function caused by the modification in amino acid sequence [17]. 74% of non-sysnonymous SNPs in protein sequence databases were close enough to homologs, according to the SNP Consortium, to allow for SIFT prediction.. The SNPs were classified based on SIFT score (cut off ≤ 0.05) tolerated nsSNPs and (cut off ≥ 0.05) deleterious nsSNPs.

C. *Polyphen2 for predicting the alteration in the protein sequence*

Polymorphism Phenotyping v2.0 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/) based on sequence-based characterisation, predicts the functional effects owing to amino acid variation on the structure and activities of proteins. [18]. Straightforward comparative and physical considerations were inclined to forsee the possible impact of the variations of an amino acid on the function and structure of a human protein. The score of PolyPhen-2 ranges from 0.0 to 1.0 [18]. A value that is close to 0 implies that the change is not harmful(benign), whereas values that are closer to 1.0 suggests that the substitution is highly likely to be damaging.

D. *PANTHER tool for investigation of protein function and stability based on evolution*

The Protein ANalysis THrough Evolutionary Relationships (PANTHER) tool (http://www.pantherdb.org/tools/csnpScore.do) involves in comparing the protein sequence with a evolutionally related protein sequence [19]. Position-specific evolutionary conservation (subPSEC) scores, which are produced from the alignment of several proteins with evolutionary relationships, are used to analyze substitution [20]. IThe nsSNP is viewed as harmful if the subPSEC score is ≥ 0.5 [19].

E. *I MUTANT for protein stability prediction*

To assure the structural and stability integrity of a protein during engineering, it is crucial to evaluate any mutations' effects. Therefore, utilized I-Mutant 3.0(http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) for this purpose. The query protein sequence and the amino acid variant were presented as an entry to I-Mutant. For single point mutations, the change in Gibbs free energy change (G) and its sign were assessed using sequence information [21] .

F. *SNPs & GO to identify the impact of the change*

To assess the complications of variations in the EGR-1 protein, the SNPs&GO, a internet server that employs a support vector machine was used. The algorithm computed functional information, such as biological processes, cellular components, and molecular functions, which were then categorized by the Gene Ontology (GO) database [22]. With an accuracy rate of 81%, SNPs&GO can forecast the connection between SNPs and diseases. A probability score of ≥ 0.5 reveals that the SNPs are linked to medical issues [23].

G. *MUTATION ASSESSOR to detect the potential deleterious SNPs*

An online tool, Mutation Assessor (http://mutationassessor.org/r3/), evaluates the functional consequences of amino acid alterations by considering evolutionary conservation and validation through disease-related Online Mendelian Inheritance in Man and the polymorphic database [24]. The resulting amino acid shows three different functional impacts such as low, medium, and high.

H. SNAP2 to assess the solvent accessibility of original and mutant proteins

SNAP2 (Screening for Nonacceptable Polymorphisms)(https://rostlab.org/services/snap2web/) utilizes multiple sequence and variant features to distinguish between effect and neutral variants. The tool takes protein sequences in FASTA format as input and generates a score ranging from -100 (indicating a strong neutral prediction) to +100 (indicating a strong effect prediction). This score provides a reliable estimation of the probability that a particular mutation will modify the nature of the native protein [25] .

I. SUSPECT to analysis the influence of nsSNPs

Predicts phenotypic effects of nsSNPs using a support vector machine (SVM) method integrating sequence, structure and systems biology-based features[26][27]. SuSPect (http://www.sbg.bio.ic.ac.uk/suspect/) generates a score table ranging from 0 to 100, which is color-coded to indicate the predicted deleteriousness of a variant (blue for neutral and red for disease-causing. A score of 50 is suggested as the threshold between neutral and disease-causing variants, with higher or lower scores indicating more confident predictions.

J. Homology modelling

The inquiry progresses by employing the Phyre2(http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) homology modelling tool to examine the structural impact of these harmful nsSNPs. The Phyre2 modelling tool was employed to ascertain the confidence score and sequence identity, which were utilized to pick the templates [28] . The Mod Refiner was implemented to refine the protein structures.and there similarity is checked based on TM-Align and RMSD score. The Swiss Model (https://swissmodel.expasy.org/) server offers the Ramachandran plot's favoured region, QMEAN, and Molprobity score and saves v6.0 (https://saves.mbi.ucla.edu/) server offers ERRAT and PROCHECK to assess the validation of predicted structures.

K. Identifying the phenotypic effects

The changes in 14 different biophysical properties were used by MutPred 2 (http://mutpred.mutdb.org/) to analyse the impact of disease-associated SNPs. The tool provided p-values by assessing the potential addition or loss of certain features. P-values less than 0.05 and 0.01 were labelled as significant and highly significant results, respectively [17].

Have (y)Our Protein Explained (HOPE) (https://www3.cmbi.umcn.nl/hope/) server is employed to examine the structural and functional implications of point mutations [29]. HOPE being a next-generation web application, acts as an automated mutant analyser generates a comprehensive report on each mutation, outlining its impact on the protein's size, charge, bonding pattern, and interaction with other molecules.

L. Protien-protien interaction

The Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/cgi/input?sessionId=bMu4JixV5iRk) renders unprecedented coverage and ease of access to both experimental and predicted interaction information. STRING interactions are given a confidence score, and additional information such as protein domains and 3D structures are made accessible, all within a stable and consistent identifier space [30].

M. Molecular Docking

Using Discovery Studio[31], we formed a suitable target protein from the Alpha fold structure of EGR-1. The docking approach exploits peptide sequences from mutated EGR-1 proteins containing nsSNPs as ligands. We performed molecular docking with the PyRx virtual screening tool to determine the influence of disruptive point mutations on EGR-1 binding affinity with the NAB2 protein [32]. NAB2 can bind to and enhance the performance of EGR-1, leading to increased proliferation of EGR-1 target genes [33] . The UCSF chimera was employed to illustrate the docking result as well as the binding interaction between ligand and receptor proteins.

III. RESULTS

A. Data retrieval

The EGR-1 gene possessed 2,712 SNPs in the dbSNP collection. Among them, 455 SNPs were classified as non-synonymous (missense). Recognizing that non-synonymous SNPs frequently change the encoded amino acid, the current research focused solely on these SNPs for further examination. P18146 is the UniProt ID for this protein. The protein contains 543 amino acids.

B. Functional analysis

The SIFT algorithm was employed to predict the deleterious SNPs. All 455 missense SNPs and their associated rsIDs were inputted into a text file for the procedure. After analyzing 455 missense SNPs, 18 SNPs were identified as deleterious (Not Tolerant), as their Tolerance Index was identified to be ≥0.05. These 18 subsitutions were taken for further analysis.

Ployphen-2 forecasts the negative consequences using position-specific independent count (PSIC)scores. As an input query, the location, the substitutes, the order,were provided. 13 SNPs (4 possibly damaging and 9 probably damaging) were identified as deleterious among 455 SNPs.

SNAP2 identified nsSNPs that could change the protein's natural structure. The sequence was used as an input query, and 17 of the mutations were determined to be non-neutral. These 17 alterations were labeled as 'effect,' while others were labeled as neutral.

A mutation assessor is also used to examine the impact of subsitutions.When the UniProt ID of the protein with the mutations is submitted, the resulting amino acid substitutions possessed three distinct functional impacts: low, medium, and high.Only SNPs with a medium to high effect on protein were chosen for the studies, and only 9 SNPs fell into this category.

C. Pathogenicity analysis

The disease-causing genetic changes were identified using SNPs and GO. The FASTA sequence of the native EGR-1protein and its list of polymorphisms were used as

input, and the resulting output revealed 13 SNPs linked with the disease, each with a probability value greater than 0.5.

The SUSPECT has a broad range of scores and is also color-coded according to deleteriousness ranging from blue at one end and red at the other end representing the neutral and deleterious effects respectively. Through input sequence nine of these mutations were found to be above the cut-off value of 50, suggesting disease-causing.

PANTHER identifies potentially harmful SNPs that can substantially alter the structure or function of the EGR-1 protein. The query input was the protein FASTA sequence and substitutions. Ten SNPs with SubPSEC scores ≥0.5 were determined to be damaging.

D. Protein Stability analysis

The I Mutant tool was utilized to foresee variations in EGR-1 protein stability. The same protein sequence and polymorphisms were used as input, and the resultIs indicated that 16 out of 18 had decreased stability based on RI and free energy change values. (DDG).

H334N residue was shown to be harmful after being examined using eight bioinformatics tools. Seven tools acknowledged Y32D, P129L, S529L, and T534N residues as potentially harmful, thus those also examined further. (Table 1 &Table 2)

E. Homology Modelling

The potential 3D model of the EGR-1 protein was found using the Pyre2 internet server and also the subsituted models that are highly deleterious variations detected using the eight tools were generated separately. On the c1g2dC template, the models were built.. Then the subsituted models were refined using a ModRefiner. TM- align and values for RMSD were also examined for the detection of structural similarity between the normal and variant models of EGR-1 protein (Table 3). The projected normal and variant models of the EGR-1 protein's structural validity and stereochemical properties were performed by the Swiss Model Server and Saves. Ramachandran favoured region, QMean, MolProbity, ERRAT, and the threshold value is applied to analyze PROCHECK scores. (Table 4).

F. Phenotypic Effect analysis

When the mutations were analyzed by the MutPred2 server, 3 out of 5 ns SNPs scored more than 0.5 which proves that they had high pathogenic properties (Table 5). Y32D showed a loss of sulfation. P129L was crucial and showed a loss in both ADP-ribosylation and O-linked glycosylation along with altered transmembrane protein and ordered interface. In addition, Y32D, P129L, and H334N variations caused an altered disordered interface. The nsSNPs were run through project HOPE which aims at providing knowledge of how a mutation affects structure including charge, hydrophobicity, amino acid change special structure, and functions (Table 6). There is a change in the charge of Y32D from neutral to negative. 3 of the five (Y32D, P129L, H334N) showed a MetaRNN score ≥ 0.5 indicating pathogenicity. The residue H334N is present in the DNA binding region. The mutated residue is located on the surface of a domain with an unknown function.

G. Protein-Protein Interaction

From the STRING network, the 6 major interacting molecules with EGR-1 protein were revealed which include Transcription factor jun-D, Transcription factor jun-B, Proto-oncogene c-Fos, Fosb proto-oncogene, NGFI-A-binding protein 1 and NGFI-A-binding protein 2 (Figure1). Since the polymorphisim were predicted to interfere with the interaction of EGR-1 with other molecules, the function of these interacting molecules is affected by these amino acid alterations as indicated in Table 7.

H. Molecular Docking

The binding affinity of the altered EGR-1 with the NAB2 Protein was shown by molecular docking experiments. Three out of five nsSNPs (Y32D, P129L, and H334N) significantly lowered the binding affinity of the NAB2 protein. The UCSF Chimera software was employed to examine the binding affinity (kcal/mol) shown in Table 8 and bonding interaction patterns of H334N and P129L docked complexes (Figure 2 -Figure 3). The wild-type peptide sequence of Y32D forms 1 hydrogen bond with a binding energyof -5.7 kcal/mol whereas the mutant forms 4 hydrogen bonds with a reduced binding energy of -5.4 kcal/mol. The wild-type peptide sequence of P129L forms no hydrogen bond with a binding energy of -7.1 kcal/mol whereas the mutant forms 3 hydrogen links with a reduced binding energy of -5.0 kcal/mol. The wild-type peptide sequence of H334N forms two hydrogen links with a binding energy of -6.5 kcal/mol whereas the mutant forms 1 hydrogen bond with a reduced binding energy of -6.3 kcal/mol**.**

IV. DISCUSSION

Three zinc finger motifs can be found in the EGR-1 DNA-binding domain [2], SNPs in the DNA binding region may lead to disturbing the straightforward structure of the *EGR-1* gene. To produce a solid and accurate prediction with higher reliability, many different techniques were integrated for the first screening of the most harmful nsSNPs. Some tools, such as SIFT and Mutation Assessor, based their predictions on parameters such as sequence homology and amino acid physical properties. Others, such as SNAP2 and PolyPhen2, used machine learning to predict the structural and functional impact of changes. SNP&GO, SuSPect, and PANTHER were also used in the investigation to determine whether the polymorphisms were linked with pathogenicity. Protein stability is critical for a protein's structural and functional activity [34]. We identified the harmful nsSNPs that may influence the stability of the EGR-1 protein using the I-Mutant program.

In total, 5 nsSNPs were chosen as the most harmful since they were anticipated to be high-risk by the SNP prediction algorithms used in this work. This method found five nsSNPs as highly pathogenic: rs199627951, rs199793033, rs201213506, rs202136756, and rs371059427. This discovery eliminates the massive amount of laboratory work necessary to screen pathogenic nsSNPs.

ISSN No:-2456-2165

The work continues with the Phyre2 homology modelling program to examine the structural consequences of these harmful nsSNPs. We constructed wild-type and mutant protein models for the EGR-1 protein using c1g2dC as a template. Furthermore, we used the Tm Align program to predict the RMSD and TM scores of these Phyre2-generated wildtype and mutant protein models. The TM-align tool was used to compare the structural similarities between wild-type and mutant structures. A low TM score and a high RMSD value imply structural dissimilarity, however, we discovered a high TM score (0.99120) and a low RMSD value (0.33); this shows that both structures are on the same fold. [35]. The wild-type structure is a snippet of the entire structure, and only one mutation in that region may be responsible for the resemblance between normal and variant structures.

Validation of experimental models is required in need to enhance targeted protein structure quality. SWISS-MODEL (RAMACHANDRAN PLOT, QMEAN, Molprobity) and SAVES (PROCHECK, ERRAT) were employed to make sure of this. The Ramachandran plot is the most prioritized verification matrix since it depicts the - torsion angles of the protein backbone of projected models. PROCHECK displays the stereochemical quality of a given protein structure by dividing the Ramachandran plot into distinct regions—core, allowed, generously allowed, and banned. In the important or most preferred region, more than 90% of the residues can be selected as a favorable structure. [36]. A QMEAN-Z score of -4.0 or below shows a model of low quality, whereas a higher value points to the favourable states of the structure, and a Molprobity score closer to zero represents a structure of higher quality [37][38]. The ERRAT score of 95% or greater often indicates a high degree of structure resolution [36].

These five highly harmful nsSNPs could have a negative impact on the protein's structure, according to predictions made by MutPred2 and the HOPE server. STRING is essential for filtering and assessing functional genomics data, as well as offering a simple platform for evaluating protein structural, functional, and evolutionary features [39]. This database was used in the current work to identify the connection of the EGR-1 protein with related proteins that could be a part of various pathways and disruption of these pathways may result in illnesses. STRING results represent the disturbance in protein-protein interaction of the EGR-1 protein due to the substitutions.

Docking research indicated that three of the five nsSNPs (D32, L129, and H334) have a lower binding affinity with NAB2 protein than the wild-type residues. The most notable alteration in binding affinity was identified in P129L, where there is a substantial loss of H-bond interactions in the binding pocket. Leu127 creates two H-bonds with Asp235 in the docking complex, and Lie131 forms an H-bond with Lue128. In short, the docking study demonstrated that the aforementioned variations had a considerable impact on the EGR-1 protein's functional activity.

V. CONCLUSION

As a transcriptional factor, the *EGR-1* gene products play vital roles in various cellular mechanisms and prevent uncontrolled cell growth and proliferation. Consequently, alteration of the EGR-1 gene has been related with upregulating different types of diseases including cardiovascular illness, ischemia-reperfusion injury, acute lung injury, atopic dermatitis, sepsis and cancers. nsSNP rs201213506 which corresponds to an H334N amino acid change, was predicted to be the most deleterious among 455 non-synonymous SNPs in EGR-1, which was confirmed by eight computational tools. The amino acid changes Y32D, P129L, S529L, and T534N were predicted to be equally deleterious by seven SNP computational tools. The complete failure of protein-protein interaction and alterations in hydrogen bonds during docking demonstrated that H334N(rs201213506) in the EGR-1 DNA binding region is an pertinent target for disrupting the EGR-1 gene. However, the predicted deleterious effect of rs201213506(H334N) warrants animal study to confirm the consequence of the mutation in the animals. As EGR-1 gene is highly conserved across laboratory mammals, a transgenic mice model with (EGR-1, H334N) mutation would answer the deleterious effect. The current study's findings would undoubtedly be useful in future prospects involving huge population-based investigations as well as drug discovery, particularly in generating personalised medical care.

ACKNOWLEDGEMENT

The authors are thankful to the institute ICAR-NIVEDI for providing necessary infrastructure, facility and guidance throughout the study.

- **Funding:** This research did not receive any funding.

DECLARATION

- **Ethics approval and consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Competing interests:** The authors declare that they have no conflict of interest.
- **Availability of data:** The data that support the findings of this study are available on NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp/).
- **Funding:** This research did not receive any funding.

AUTHOR INFORMATION

- **Authors Contributions:** YSS conceptualized the methodology, performed the analysis and drafted the manuscript; UBI carried out the protein interaction analysis. BM participated in phenotypic analysis. SR participated in docking. AP, NNB and SSP reviewed the paper. KPS supervised the work and thoroughly analysed and edited the manuscript. All authors read and approved the manuscript.

CORRESPONDING AUTHOR

Correspondence to Kuralayanapalya Puttahonnappa Suresh (suresh.KP@icar.gov.in)

AUTHORS DETAILS

[1]ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru-560064, India. [2]Tamil Nadu Veterinary and Animal Sciences University, Chennai, Tamil Nadu 600051,India. [3]College of Veterinary Science, Assam Agricultural University (AAU), Guwahati 781022, India.

ABBREVIATIONS

EGR-1: Early growth response gene 1,SNP: Single Nucleotide Polymorphism ,Elk1: ETS Like-1 protein, NCBI: National Center for Biotechnology Information, SIFT: Sorts intolerant from tolerant, PolyPhen:Polymorphism phenotyping, SNAP: Screening for Nonacceptable Polymorphisms, HOPE: Have (y)Our Protein Explained, STRING: Search Tool for the Retrieval of Interacting Genes, SVM:Support vector machine, RMSD: Root-mean-square deviation, TM:Template modeling, NAB2: NGFI-A-binding protein 2, nsSNP: Nonsynonymous single-nucleotide polymorphisms.

REFERENCES

[1]. V. P. Sukhatme *et al.*, "A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization," *Cell*, vol. 53, no. 1, pp. 37–43, 1988, doi: 10.1016/0092-8674(88)90485-0.

[2]. G. Thiel and G. Cibelli, "Regulation of life and death by the zinc finger transcription factor Egr-1," *J. Cell. Physiol.*, vol. 193, no. 3, pp. 287–292, 2002, doi: 10.1002/jcp.10178.

[3]. C. H. Tsai-Morris, X. Cao, and V. P. Sukhatme, "5' Flanking Sequence and Genomic Structure of Egr-1, a Murine Mitogen Inducible Zinc Finger Encoding Gene," *Nucleic Acids Res.*, vol. 16, no. 18, pp. 8835–8846, 1988, doi: 10.1093/nar/16.18.8835.

[4]. A. J. Whitmarsh, P. Shore, A. D. Sharrocks, and R. J. Davis, "Integration of MAP kinase signal transduction pathways at the serum response element," *Science (80-.).*, vol. 269, no. 5222, pp. 403–407, 1995, doi: 10.1126/science.7618106.

[5]. E. Havis and D. Duprez, "EGR1 Transcription Factor is a Multifaceted Regulator of Matrix Production in Tendons and Other Connective Tissues," pp. 1–26, 2020, doi: 10.3390/ijms21051664.

[6]. F. Duclot and M. Kabbaj, "The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders," vol. 11, no. March, pp. 1–20, 2017, doi: 10.3389/fnbeh.2017.00035.

[7]. H. Nishi, K. H. Nishi, and A. C. Johnson, "Early Growth Response-1 Gene Mediates Up-Regulation of Epidermal Growth Factor Receptor Expression during Hypoxia," pp. 827–834, 2002.

[8]. N. Ngiam, M. Post, and B. P. Kavanagh, "Early growth response factor-1 in acute lung injury," pp. 1089–1091, 2023, doi: 10.1152/ajplung.00265.2007.Perspectives.

[9]. L. M. Khachigian, "Early growth response-1, an integrative sensor in cardiovascular and inflammatory disease," *J. Am. Heart Assoc.*, vol. 10, no. 22, pp. 1–8, 2021, doi: 10.1161/JAHA.121.023539.

[10]. H. Yeo, "Disrupting the DNA Binding of EGR-1 with a Small-Molecule Inhibitor Ameliorates 2 , 4-Dinitrochlorobenzene-Induced Skin Inflammation," pp. 1851–1855, 2020, doi: 10.1016/j.jid.2020.12.029.

[11]. M. Cargill *et al.*, "Characterization of single-nucleotide polymorphisms in coding regions of human genes," *Nat. Genet.*, vol. 22, no. 3, pp. 231–238, 1999, doi: 10.1038/10290.

[12]. D. Chasman and R. M. Adams, "Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation," *J. Mol. Biol.*, vol. 307, no. 2, pp. 683–706, 2001, doi: 10.1006/jmbi.2001.4510.

[13]. P. Bhatti, D. M. Church, J. L. Rutter, J. P. Struewing, and A. J. Sigurdson, "Candidate single nucleotide polymorphism selection using publicly available tools: A guide for epidemiologists," *Am. J. Epidemiol.*, vol. 164, no. 8, pp. 794–804, 2006, doi: 10.1093/aje/kwj269.

[14]. A. D. Johnson, "Methods in genetics and clinical interpretation, Single-nucleotide polymorphism bioinformatics, A comprehensive review of resources," *Circ. Cardiovasc. Genet.*, vol. 2, no. 5, pp. 530–536, 2009, doi: 10.1161/CIRCGENETICS.109.872010.

[15]. L. Li and D. Wei, "Bioinformatics tools for discovery and functional analysis of single nucleotide polymorphisms," *Adv. Exp. Med. Biol.*, vol. 827, pp. 287–310, 2015, doi: 10.1007/978-94-017-9245-5_17.

[16]. S. T. Sherry, M. Ward, and K. Sirotkin, "dbSNP - database for single nucleotide polymorphisms and other classes of minor genetic variation," *Genome Res.*, vol. 9, no. 8, pp. 677–679, 1999, doi: 10.1101/gr.9.8.677.

[17]. P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nat. Protoc.*, vol. 4, no. 7, pp. 1073–1082, 2009, doi: 10.1038/nprot.2009.86.

[18]. I. A. Adzhubei *et al.*, "A method and server for predicting damaging missense mutations," *Nat. Methods*, vol. 7, no. 4, pp. 248–249, 2010, doi: 10.1038/nmeth0410-248.

[19]. L. R. Brunham, R. R. Singaraja, T. D. Pape, A. Kejariwal, P. D. Thomas, and M. R. Hayden, "Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene," *PLoS Genet.*, vol. 1, no. 6, pp. 0739–0747, 2005, doi: 10.1371/journal.pgen.0010083.

[20]. P. D. Thomas *et al.*, "Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools," *Nucleic Acids Res.*, vol. 34, no. WEB. SERV. ISS., pp. 645–650, 2006, doi: 10.1093/nar/gkl229.

[21]. E. Capriotti, P. Fariselli, and R. Casadio, "I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure," *Nucleic Acids Res.*, vol. 33, no. SUPPL. 2, pp. 306–310, 2005, doi:

10.1093/nar/gki375.
[22]. R. Calabrese, E. Capriotti, P. Fariselli, P. L. Martelli, and R. Casadio, "Functional annotations improve the predictive score of human disease-related mutations in proteins," *Hum. Mutat.*, vol. 30, no. 8, pp. 1237–1244, 2009, doi: 10.1002/humu.21047.
[23]. E. Capriotti, R. Calabrese, P. Fariselli, P. L. Martelli, R. B. Altman, and R. Casadio, "WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation.," *BMC Genomics*, vol. 14 Suppl 3, no. Suppl 3, 2013, doi: 10.1186/1471-2164-14-s3-s6.
[24]. M. Arifuzzaman, S. Mitra, R. Das, A. Hamza, N. Absar, and R. Dash, "In silico analysis of nonsynonymous single-nucleotide polymorphisms (nsSNPs) of the SMPX gene," *Ann. Hum. Genet.*, vol. 84, no. 1, pp. 54–71, 2020, doi: 10.1111/ahg.12350.
[25]. Y. Bromberg, G. Yachdav, and B. Rost, "SNAP predicts effect of mutations on protein function," *Bioinformatics*, vol. 24, no. 20, pp. 2397–2398, 2008, doi: 10.1093/bioinformatics/btn435.
[26]. T. Yasmin, "In silico comprehensive analysis of coding and non-coding SNPs in human mTOR protein," *PLoS One*, vol. 17, no. 7 July, pp. 1–23, 2022, doi: 10.1371/journal.pone.0270919.
[27]. E. Capriotti, R. Calabrese, and R. Casadio, "Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information," *Bioinformatics*, vol. 22, no. 22, pp. 2729–2734, 2006, doi: 10.1093/bioinformatics/btl423.
[28]. L. a Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "Europe PMC Funders Group The Phyre2 web portal for protein modelling , prediction and analysis," *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, 2015, doi: 10.1038/nprot.2015.053.The.
[29]. H. Venselaar, T. A. H. te Beek, R. K. P. Kuipers, M. L. Hekkelman, and G. Vriend, "Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces," *BMC Bioinformatics*, vol. 11, 2010, doi: 10.1186/1471-2105-11-548.
[30]. D. Szklarczyk *et al.*, "The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Res.*, vol. 39, no. SUPPL. 1, pp. 561–568, 2011, doi: 10.1093/nar/gkq973.
[31]. Q. Wang, J. He, D. Wu, J. Wang, J. Yan, and H. Li, "Interaction of α-cyperone with human serum albumin: Determination of the binding site by using Discovery Studio and via spectroscopic methods," *J. Lumin.*, vol. 164, pp. 81–85, 2015, doi: 10.1016/j.jlumin.2015.03.025.
[32]. J. E. Hempel, C. Hong, Charles, and C. H. Williams, *Chemical Biology Methods and Protocols*, vol. 1263, no. January. 2015.
[33]. P. Faculty, B. Ph, and B. Mgr, "Department of Biochemistry Regulation of Transcription by Proteins of the Early Growth Response and Myb families Ph . D . Thesis," 2013.
[34]. M. C. Deller, L. Kong, and B. Rupp, "Protein stability: A crystallographer's perspective," *Acta Crystallogr. Sect. Struct. Biol. Commun.*, vol. 72, pp. 72–95, 2016, doi: 10.1107/S2053230X15024619.
[35]. M. Arshad, A. Bhatti, and P. John, "Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study," *PLoS One*, vol. 13, no. 1, pp. 1–13, 2018, doi: 10.1371/journal.pone.0188143.
[36]. N. Zobayer and A. B. M. Aowlad Hossain, "In silico characterization and homology modeling of histamine receptors," *J. Biol. Sci.*, vol. 18, no. 4, pp. 178–191, 2018, doi: 10.3923/jbs.2018.178.191.
[37]. P. Benkert, M. Biasini, and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models," *Bioinformatics*, vol. 27, no. 3, pp. 343–350, 2011, doi: 10.1093/bioinformatics/btq662.
[38]. I. W. Davis *et al.*, "MolProbity: All-atom contacts and structure validation for proteins and nucleic acids," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 375–383, 2007, doi: 10.1093/nar/gkm216.
[39]. L. J. Jensen *et al.*, "STRING 8 - A global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, pp. 412–416, 2009, doi: 10.1093/nar/gkn760.

TABELS

Table 1: Details of variations in the EGR-1 gene predicted as high-risk SNPs out of 455 nsSNPs

SNP ID	Amino acid Variant	Single Nucleotide Variation	SIFT		PolyPhen 2	Panther
			score	Prediction		
rs199793033	Y32D	T/G	0.028	Deleterious	Probably damaging	Probably damaging
rs199627951	P129L	C/T	0.006	Deleterious	Probably damaging	Probably damaging
rs201213506	H334N	C/T	0.015	Deleterious	Probably damaging	Probably damaging
rs371059427	S529L	C/T	0.01	Deleterious	Probably damaging	Probably damaging
rs202136756	T534N	C/A	0.01	Deleterious	Possibly damaging	Possibly damaging

Table 2: Details of variations in the EGR-1 gene predicted as high-risk SNPs out +of 455 nsSNPs

SNP ID	Amino acid Variant	I Mutant		SNPs&GO	Mutation Assessor	SNAP2		Suspect
		DDG<0	Prediction	Prediction		Score	Prediction	
rs199793033	Y32D	-0.94	Decrease	Disease	low	85	Effect	50
rs199627951	P129L	-0.6	Decrease	Disease	low	63	Effect	59
rs201213506	H334N	-0.5	Decrease	Disease	Medium	53	Effect	51
rs371059427	S529L	-0.01	Decrease	Disease	Medium	71	Effect	43
rs202136756	T534N	-0.86	Decrease	Disease	low	75	Effect	63

Table 3: Structural similarity scores

MUTATION	TM- ALIGN	RMSD VALUE
Y32D	0.98798	0.38
P129L	0.98006	0.90
H334N	0.99120	0.33
S529L	0.98897	0.42
T534N	0.97884	0.99

Table 4: structural validation of wild and mutant proteins.

MODEL	SWISS-MODEL RAMACHANDRAN PLOT VALUE.	PROCHECK	QMEAN	MOL-PROBITY	ERRAT
NATIVE	97.33%	95.0%	-1.05	1.63	93.8272
Y32D	97.67%	93.7%	-1.80	1.96	92.5
P129L	100%	94.9%	-0.99	1.76	87.1795
H334N	100%	95.0%	-0.86	1.88	89.8734
S529L	98.84%	93.7%	-1.17	1.76	91.1392
T534N	100.00%	94.9%	-1.25	1.84	85.00

Table 5: Impacts of non-synonymous SNPs on structural & functionational properties of EGR1 determined.

Mutation	Probability of deleterious mutation	Structural & functional properties
Y32D	0.733	Altered Disordered interface (P= 0.38)
		Gain of B-factor (P=0.27)
		Gain of Ubiquitylation at K34 (P=0.20)
		Gain of Methylation at K34 (P=0.10)
		Loss of Sulfation at Y32 (P=0.09)
P129L	0.601	Altered Disordered interface (P=0.27)
		Altered Ordered interface (P=0.26)
		Altered Transmembrane protein (P=0.25)
		Loss of ADP-ribosylation at R127 (P=0.20)
		Loss of O-linked glycosylation at T126 (P=0.12)
H334N	0.525	Altered Disordered interface (P=0.27)
		Gain of B-factor (P=0.26)
		Gain of Methylation at K330 (P=0.12)

		Gain of Sulfation at Y338 (P=0.02)
S529L	0.404	-
T534N	0.451	-

Table 6: The consequence of nsSNPs on the structural features of the protein such as size, charge, hydrophobic nature, and MetaRNN scores is determined by HOPE.

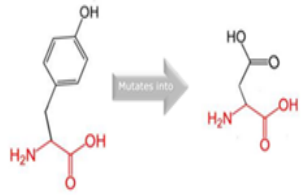
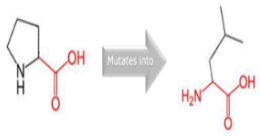
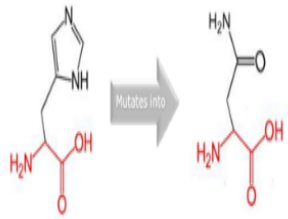
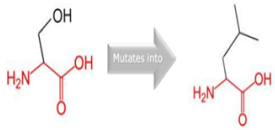
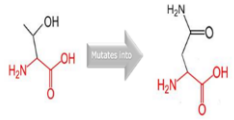
RESIDUE	STRUCTURE	PROPERTIES
Y32D		• The mutant residue is shorter than the wild-type residue. The mutant residue charge is NEGATIVE compared to the NEUTRAL charge of the wild-type residue. • The natural residue is more hydrophobic than the mutant residue. • This area and its function may be disturbed by the variations in amino acid characteristics. • This variant's MetaRNN score is 0.7630575
P129L		• The natural residue is smaller than the mutant residue. • Since prolines are notorious for being incredibly rigid, they result in a special backbone conformation that might be required in this position. • This specific conformation may be disturbed by the mutation. • This variant's MetaRNN score is 0.660638
H334N		• The residue has DNA interactions or is located in a DNA binding region. • In comparison to the wild-type residue, the mutant residue is smaller. • The different characteristics of amino acids can alter this area and its functionality. • The variant residue is present on the surface of a domain with an unknown function. This will cause a possible loss of external interactions. • This variant's MetaRNN score is 0.567263
S529L		• The mutant residue is larger than the residue of the natural type. • Comparing the two residues, the mutant residue is more hydrophobic. • This variant's MetaRNN score is 0.37630606
T534N		• The mutant residue is bigger than the wild-type residue. • The wild-type residue is more hydrophobic than the mutant residue. • This variant's MetaRNN score is 0.34320903.

Table 7: indicating the interaction between mutated EGR-1 and wild-type proteins

Template	JUND	JUNB	FOS	FOSB	NAB1	NAB2
EGR-1	✓	✓	✓	✓	✓	✓
EGR-1 **Y32D**	—	—	—	—	✓	✓
EGR-1 **P129L**	✓	✓	—	✓	—	—
EGR-1 **H334N**	—	—	—	—	✓	✓
EGR-1 **S529L**	✓	✓	—	✓	—	—
EGR-1 **T534N**	✓	✓	—	✓	—	—

Table 8: Docking results of EGR-1 proteins with NAB2 protein.

Residue	Binding affinity (kcal/mol)	Residue	Binding affinity (kcal/mol)
Y32	-5.7	D32	-5.4
P129	-7.1	L129	- 5
H334	-6.5	N334	-6.3
S529	-4.7	L529	-5.8
T534	-6.3	N534	-6.3

FIGURES

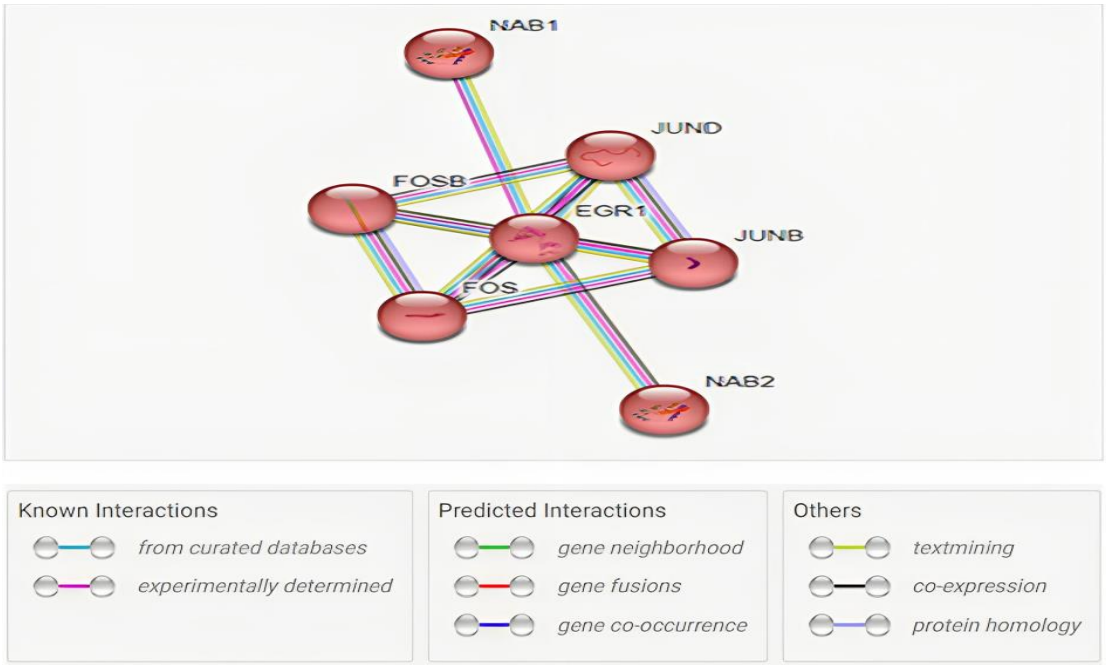


Fig. 1: The protein interaction network of the EGR-1 protein, shows NAB2, NAB1, FOSB, FOS, JUND, and, JUNB are closely related proteins.

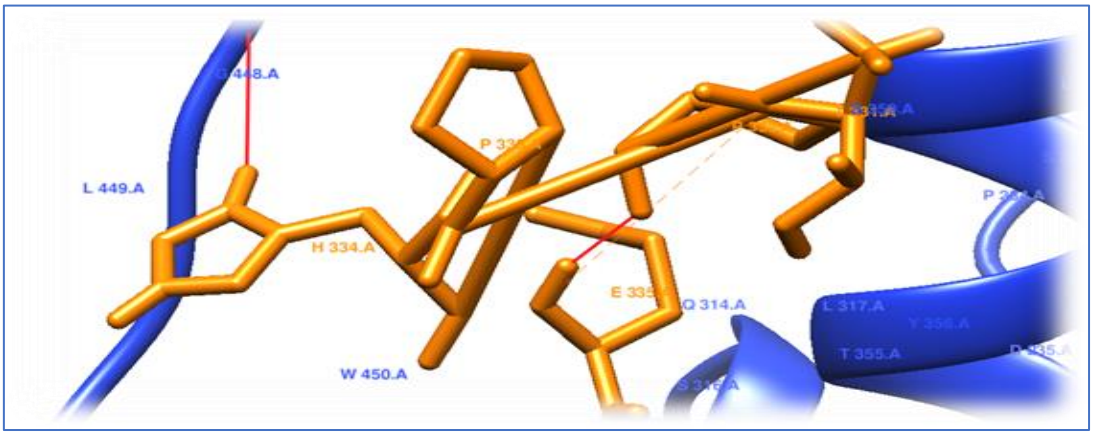


Fig. 2(a): With H334 as ligand hydrogen links are formed between His334- Gly448 and Glu335-Pro332.

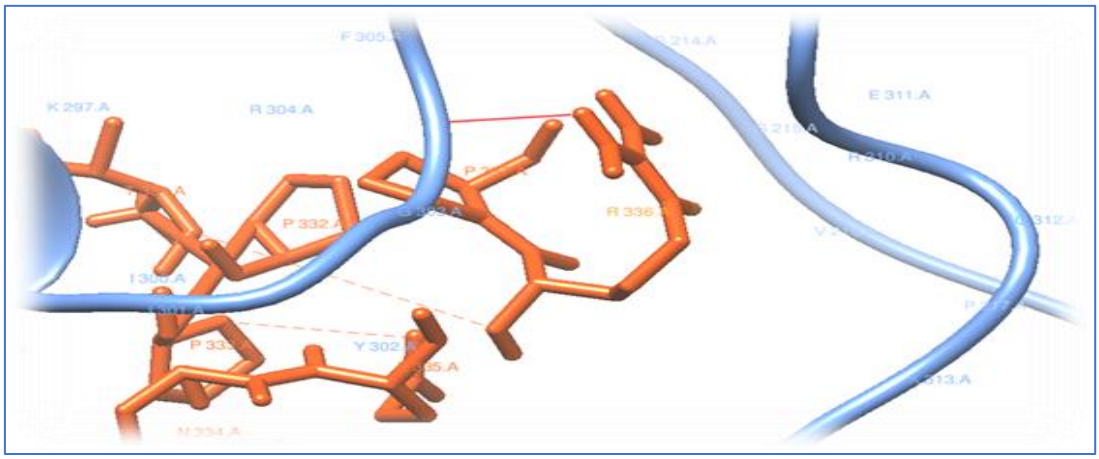


Fig. 2(b): With N534 as ligand hydrogen links are formed between Phe535-Arg296, Asn534-Phe535, Thr532-Phe535, Thr532-Asn534 and Met531- Asn534.

ISSN No:-2456-2165

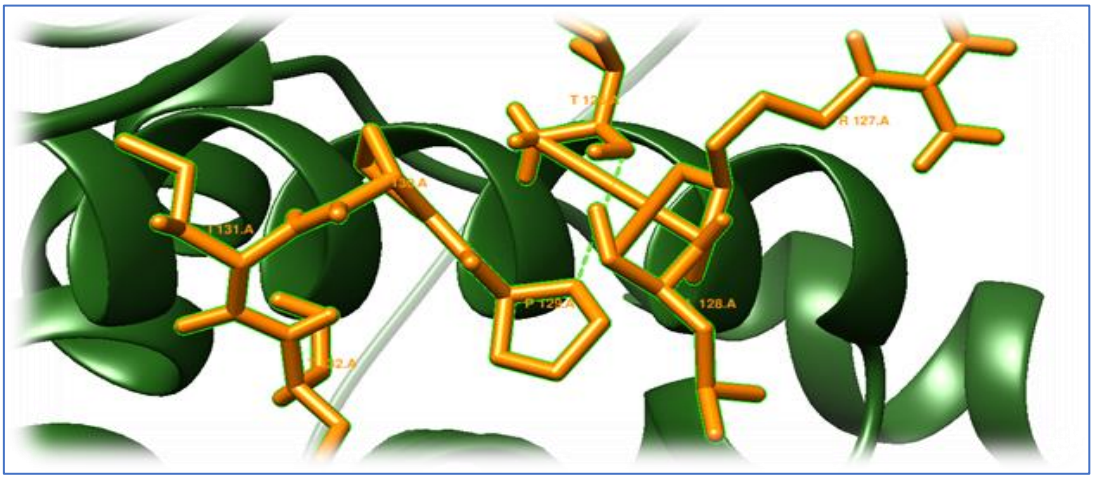


Fig. 3(a): With P129 as ligand no hydrogen bond are formed.

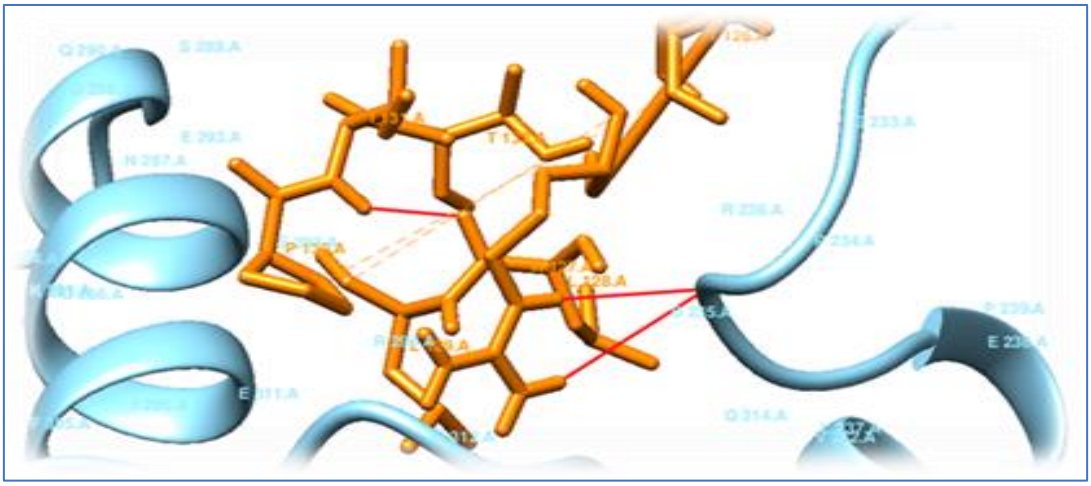


Fig. 3(b): With L129 as ligand hydrogen links are formed between Met531- Leu528,Thr526-Glu311 and Thr526-Lys313.