Screening of Pathogenic Missense Single Nucleotide Variants From *LHPP* Gene Associated With the Hepatocellular Carcinoma: An *In silico* Approach

Bioinformatics and Biology Insights
Volume 16: 1–9

DOI: 10.1177/11779322221115547

Malik Siddique Mahmood[1,2], Maryam Afzal[1], Hina Batool[3], Amara Saif[3], Tahreem Aqdas[1], Naeem Mahmood Ashraf[4] and Mahjabeen Saleem[1]

[1]School of Biochemistry & Biotechnology, University of the Punjab, Lahore, Pakistan. [2]Department of Biochemistry, NUR International University, Lahore, Pakistan. [3]Department of Life Sciences, University of Management and Technology, Lahore, Pakistan. [4]Department of Biochemistry & Biotechnology, University of Gujrat, Gujrat, Pakistan.

ABSTRACT: *LHPP* gene encodes a phospholysine phosphohistidine inorganic pyrophosphate phosphatase, which functions as a tumor-suppressor protein. The tumor suppression by this protein has been confirmed in various cancers, including hepatocellular carcinoma (HCC). *LHPP* downregulation promotes cell growth and proliferation by modulating the *PI3K/AKT* signaling pathway. This study identifies potentially deleterious missense single nucleotide variants (SNVs) associated with the *LHPP* gene using multiple computational tools based on different algorithms. A total of 4 destabilizing mutants are identified as *L22P, I212T, G227R*, and *G236R*, from the conserved region of the phosphatase. The 3-dimensional (3D) modeling and structural comparison of variants with the native protein reveals significant structural and conformational variations after mutations, suggesting disruption in the function of phospholysine phosphohistidine inorganic pyrophosphate phosphatase. The identified mutations might, therefore, participate in the cause of HCC.

KEYWORDS: Hepatocellular carcinoma, single nucleotide variants, *LHPP*, phospholysine phosphohistidine inorganic pyrophosphate phosphatase, tumor-suppressor protein

RECEIVED: March 14, 2022. **ACCEPTED:** June 11, 2022.

TYPE: Original Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was self-supported by the authors.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Mahjabeen Saleem, School of Biochemistry & Biotechnology, University of the Punjab, Quaid-i-Azam Campus, Lahore 54590, Pakistan. Email: mahjabeensaleem1@hotmail.com

Introduction

Hepatocellular carcinoma (HCC) is among the most assertive form of liver morbidities worldwide.[1] The mortality rates associated with the HCC makes it, the fourth leading cause of cancer deaths globally, with a significant load on the developing countries.[2,3] Hepatocellular carcinoma is a multifactorial disorder with variable etiological agents distributed in different geographical regions. The primary underlying factors include persistent liver cirrhosis, long-term liver diseases, diabetes mellitus, obesity, aflatoxins, hepatitis B, and C viruses. Despite the remarkable advances in therapeutics, the late diagnosis of this condition results in the continuous expansion in a disease incidence and mortality.[4] Therefore, detailed knowledge of HCC initiation and prognosis is essential for developing the early diagnosis and treatment strategies for this type of cancer. Recently, the contribution of many genetic loci in the incidence of HCC has been well established, including 1q21, which harbors potential tumor-suppressor and oncogenes. In addition, many genetic elements are known to implicate in various cancers.[5] *LHPP* is a tumor-suppressor gene encoding inorganic phosphatase, which negatively correlates with the cell cycle and metastasis. The overexpression of *LHPP* suppresses the expression of oncogenes, which suggests an anti-cancerous effect of this gene. The down-regulation of *LHPP* is therefore reported to induce hepato-carcinogenesis in humans.[6-8] Till now, the single nucleotide variations in *LHPP* are well characterized with multiple disorders, including rs35936514 for major depressive disorders (MDDs), rs34997829 for alcohol-dependent risky sexual behaviors, and rs201982221 for oropharyngeal carcinomas.[9-12]

The single nucleotide variants (SNVs) ensuring in the coding regions of the proteins affect the functional integrity of the protein and, therefore, increase the susceptibility toward many diseases, including cancer.[13] The screening of SNVs associated with specific phenotypes is a point of concern as it requires comprehensive testing of the mutated gene. A possible solution is prioritizing the mutations based on their functional characteristics using computational tools.[14] The *in silico* approaches offer significant advantages over the experimental methods in terms of speed, reliability, convenience, and cost.[15]

This study is designed for the computational screening of the missense SNVs from the *LHPP* gene, regulating the structure, function, and stability of its protein. Furthermore, the impact of pathogenic mutations on the structure of the protein is evaluated via geometrical simulations. This study would be a significant addition to the existing literature by revealing the association between *LHPP* mutations and HCC. The study

*Naeem Mahmood Ashraf is now affiliated to School of Biochemistry & Biotechnology, University of the Punjab, Lahore, Pakistan

would, therefore, contribute to the development of early diagnostic and management strategies for this disease.

Methodology

Retrieval of protein sequence and SNVs

The amino acid sequence of the *LHPP* was retrieved from the National Center for Biotechnology and Information (NCBI).[16] The missense SNVs reported for this protein were retrieved using Ensemble dbSNP.[17]

Screening of the potentially deleterious missense SNVs through different in silico *algorithms*

To evaluate the functional impact of selected missense SNVs, a diverse set of 19 prediction tools based on sequence homology, machine learning, sequence to structure, and consensus-derived algorithms, were used. The selection of multiple algorithms and multiple computational tools from each algorithm helped to screen the potentially deleterious SNVs involved in the prognosis of HCC. The missense mutations commonly marked as deleterious by computational tools based on distinct algorithms were considered, for the downstream analysis, excluding all other SNVs.

Sequence homology-based approaches. All missense mutations, retrieved from Ensemble dbSNP, were first evaluated using 4 sequence homology-based tools; SIFT, PROVEAN, Mutation Accessor, and PANTHER. Computational tools based on sequence homology algorithm identify the significantly pathogenic mutations based on their alignment with the known pathogenic mutations. For screening, the pathogenic mutations from SIFT, a prediction score of less than 0.05, was applied. Likewise, in PROVEAN and PANTHER delta alignment score $\leqslant -2.5$ and substitution position-specific evolutionary conservation (subPSEC) score $\geqslant -3$ was considered.[18-22] The missense mutations, commonly marked as deleterious via 4 homology-based computational tools, were selected for further analysis.

Machine learning–based approaches. The pathogenic missense mutations identified using sequence homology-based computational tools were further evaluated by the 7 computational tools based on the machine learning algorithm. These computational tools include; SNAP2, SAAP, MutPred, SusPect, PMut, SNP&GO, and PhD-SNP servers. All these computational tools utilize random forest, artificial neural network (ANN), and support vector machine (SVM) to classify the nonsynonymous single nucleotide variants (nsSNVs) into deleterious or tolerant substitutions. In SNAP2, the prediction score ranging from +1 to +100 indicates the deleterious missense mutations. Similarly, the prediction scores of >50, was applied in SusPect and SNP & GO. The screening of missense mutations from the PMut and MutPred was carried out at the cut-off value of >0.5 and 0.8, respectively.[23-29] Again, the missense SNVs, commonly marked as deleterious via supervised learning approaches, were selected for the next step of the analysis.

Sequence to structure-based approaches. PolyPhen-2, Site-Directed Mutator (SDM), PoPMuSiC, and Fold-X are the 4 computational tools, which were used for the evaluation of all missense SNVs selected from the sequence homology and machine learning–based algorithms. These tools consider the sequence or structural parameters for screening the disease-causing variations. Among these, PolyPhen-2 makes binary predictions, with 0 indicating the neutral substitutions and 1 to the deleterious substitutions.[30-33]

Consensus-based approaches. The further assortment of nsSNVs was performed, using 4 consensus-based computational tools, that is, Condel, Meta-SNP, PON-P2, and Predict-SNP methods. These computational tools integrate multiple algorithms to determine the potentially deleterious point mutations from neutral mutations.[34-37]

Evolutionary conservation analysis of deleterious missense SNVs. The amino acid substitutions in the evolutionarily conserved regions of the proteins can alter the protein stability, folding, and structure. Therefore, after the screening of missense mutations from the combination of sequence homology, machine learning, and consensus-based approaches, the mutations present in the evolutionarily conserved regions of the protein were traced via the ConSurf server. This tool identifies the evolutionarily conserved mutations by multiple sequence alignments with the homologous sequences. The conservation scores range from 1 (extremely variable) to 9 (highly conserved).[38]

Structural modeling and active site analysis

For structural comparison of the native and mutated protein models, the 3D structure of the LHPP protein was retrieved from protein data bank (PDB) using PDB-ID of 2X4D. However, the 3D models of the proteins with selected SNVs were built using Fold-X. The protein models were validated based on the Ramachandran plot that were designed using PROCHECK server.[39-41] As the mutations altering the amino acid residues in the active site of the protein are significantly deleterious, therefore, after structural modeling, the CASTp server was used to analyze whether the selected mutations are the part of active protein or not.[42]

Comparison of native and mutated protein models

The native and mutated protein models were compared using Discovery Studio that provides extensive insight into the protein structure and folding.[32,43]

Geometric simulations analysis

The deviation of a mutated confirmation from the native structure impacts the functional integrity of the respective protein. The 3D structures of native and mutated LHPP proteins were therefore compared using geometric structural simulations to identify the extent of fluctuations in the structural conformation

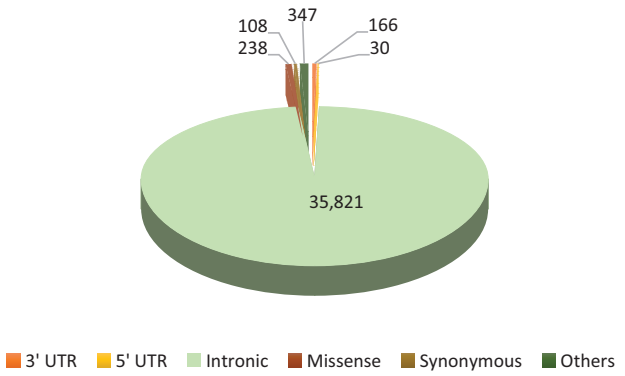


Figure 1. Distribution frequency of SNVs in the *LHPP* gene. Most variants are present in noncoding regions: 35,821 in introns, 166 in 3′ UTR, and 30 in 5′ UTR. In coding regions, most of the SNVs are missense (238), then comes synonymous (108). Other 347 mutations are present in a very low-distribution frequency including: splice acceptor and splice donor are in the same amount (166), along with the start lost (2) and stop-gained (13) variants. SNV indicates single nucleotide variant.

after mutation.[44] The geometric simulation approach, using NMSim, was used to predict the biologically related conformational transitions in the mutated LHPP protein models via different parameters including, root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and polar surface area (PSA).[45]

Molecular dynamics simulations

The molecular dynamics (MD) simulation analysis of the wild-type and mutated protein model was performed to assess the dynamic stability and structural features of the mutated proteins compared to the wild-type protein. The root mean square deviations of the mutated and wild-type proteins were checked for the period of 20 ns using VMD software.[13]

Results

Retrieval of protein sequence and SNVs

The protein sequence of *the* LHPP protein (Accession ID: NP_071409.3) contains 270 amino acids. For this protein, a total of 40,862 SNVs, are reported in the Ensemble dbSNP (Figure 1). Most of these variants were non-coding including; 35,821 variants in intron region, 30 in 5′-UTR, and 166 in 3′-UTR. Likewise, coding variants included 238 mutations as missense, 108 as synonymous, 166 splice acceptor, and 166 splice donor. In addition, there were 2 start-loss and 13 stop-gain variants (Figure 1). Since the missense SNVs are the main benefactors behind rare genetic disorders, therefore, this study considered only 238 missense SNVs for further analysis.[13]

Screening of potentially deleterious missense SNVs through different in silico *methods*

Out of 238 missense SNVs associated with *LHPP*, the pathogenic missense mutations were screened by using the combination of multiple *in silico* algorithms. The variants commonly marked as deleterious by all prediction methods were selected, ignoring the neutral substitutions at each step. The selected mutations were likely to affect the functions of the candidate protein. These computational algorithms are listed below in detail.

Sequence homology–based approaches. First, the evaluation of selected missense mutations via SIFT, PROVEAN, Mutation Assessor, and PANTHER, resulted in the screening of 87, 116, 140, and 152 variants as deleterious, respectively. Finally, a total of 52 missense mutations commonly marked as pathogenic by all these homology-based computational tools and therefore were selected for the next step (Supplemental File 1).

Machine learning–based approaches. The analysis of 52 variants, selected from the sequence homology algorithm, via the computational tools based on machine learning algorithms (SNAP2, SAAP, MutPred, SusPect, PMut, SNP & GO, and PhD-SNP), resulted in the exploration of 13 missense variants commonly by all these tools. Among these, SNAP2, SAAP, MutPred, SusPect, PhD-SNP, PMut, and SNP & GO individually predicted 42, 41, 45, 20, 39, and 29 missense variants as pathogenic, respectively. The selected 13 missense variants were common in the prediction results of all the above-mentioned tools (Supplemental File 2).

Sequence to structure-based approaches. PolyPhen-2, SDM, PoPMuSiC, and Fold-X are the protein sequence and structure-based protein stability predictors, which further evaluated the 13 missense SNVs. Among these, PolyPhen-2, SDM, and PoPMuSiC evaluated 12 variants as deleterious, whereas Fold-X predicted only 9 mutations as pathogenic. However, 6 missense mutations, that is, *L22P, G27R, L91P, I212T, G227R*, and *G236R* were nominated for the next step because of their combined predictions as deleterious by all the above-mentioned tools (Supplemental File 3).

Consensus-based methods. Furthermore, the 6 missense SNVs with the consensus-based *in silico* tools, Condel, Meta-SNP, PON-P2, and Predict-SNP, also assorted the pathogenic missense mutations. Three of these tools, Condel, Meta-SNP, and Predict-SNP, labeled all 6 variants as deleterious while, PON-P2 marked 1 variant (*G27R*) as neutral (Supplemental File 4). Thus, the 5 missense mutations, that is, *L22P, G227R, L91P, I212T*, and *G236R* were nominated from the LHPP protein, as pathogenic mutations. All these mutations were likely to associate with the cause of HCC (Figure 2).

Evolutionary conservation analysis of deleterious missense SNVs. ConSurf web server highlights the evolutionarily conserved regions from the LHPP protein using the empirical Bayesian approach. This server marked 4 missense SNVs, that is, *L22P, I212T, G227R*, and *G236R*, to be present in the evolutionarily conserved regions of the LHPP protein. The 4

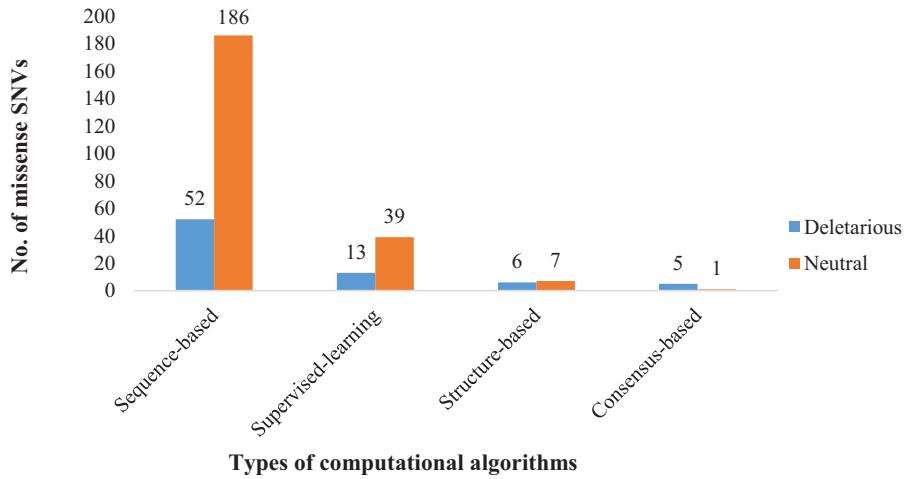


Figure 2. Evaluation of missense SNVs by sequence homology-based, supervised learning–based, sequence to structure-based, and consensus-based algorithms. First, 52 missense SNVs are labeled as pathogenic by sequence homology–based tools. Out of these 52 deleterious missense SNVs, 13 were predicted as damaging by supervised learning–based tools, 6 by sequence to structure-based tools, and 5 by consensus-based tools. SNV indicates single nucleotide variant.

Table 1. The evolutionarily conservation analysis of deleterious missense mutations from *LHPP* using ConSurf web server.

VARIANT ID	AMINO ACID POSITION	CONSERVATION SCORES	NORMALIZED VALUES	CONSURF PREDICTION
rs754022892	*L22P*	8	−0.814	Highly conserved
rs766371253	*L91P*	6	−0.427	Average
rs199534407	*I212T*	8	−0.770	Highly conserved
rs142386969	*G227R*	9	−1.172	Highly conserved
rs759928988	*G236R*	9	−1.356	Highly conserved

selected missense mutations were likely to have a substantial impact on the structure and function of the protein and therefore were filtered for further analysis (Table 1).

Structural modeling and active site analysis

The 3D structures of 4 mutants were built using Fold-X and compared with the native model of LHPP protein (PDB ID = 2X4D). The Ramachandran plot showed that more than 90% residues of each of these models reside in the most favored regions of the plot indicating the validity of these models (Additional file 5). The CASTp analysis showed that all 4 mutations were the part of active protein (Figure 3).

Comparison of native and mutated protein models

The comparative modeling suggested the altered profile of the molecular interactions among the native and mutated models. In *L22P, amino acid proline replaced the* nonpolar leucine at position 22. Likewise, in *I212T*, a polar amino acid threonine was substituting the nonpolar amino acid isoleucine at position 212. In 2 other variants, *G227R* and *G236R*, a positively charged arginine replaced a small nonpolar glycine at 227 and 236 positions, respectively. The amino acid substitutions in these mutants resulted in the significant alterations in the molecular interactions, which were responsible for the decreased stability of mutants and their contributions to the pathogenicity mechanism (Table 2).

Geometric simulations analysis

The conformational and geometrical deviations among wild-type LHPP protein and its 4 potentially pathogenic mutants were analyzed by using a web-based server, NMSims. The comparison was based on RMSD, RMSF, the Rg, and PSA of the protein models.[45] The RMSD value of the native *LHPP* model was 3.1 Å, while *I212T, L22P, G227R*, and *G236R* mutants were having RMSD values of 2.7, 3.5, 3.29, and 3.0 Å, respectively. Mutant *L22P* showed maximum deviations from the native model (Figure 4A).

Moreover, the RMSF values for *L22P, I212T, G227R*, and *G236R* mutants were 4.05, 2.64, 4.7, and 3.2 Å, respectively, compared to the native *LHPP* conformation, 2.4 Å. The RMSF values indicated that *L22P* and *G227R* have significantly fluctuated models. *G227R* presented an extensive fluctuation range from 4.75 to 1.27 Å (Figure 4B).

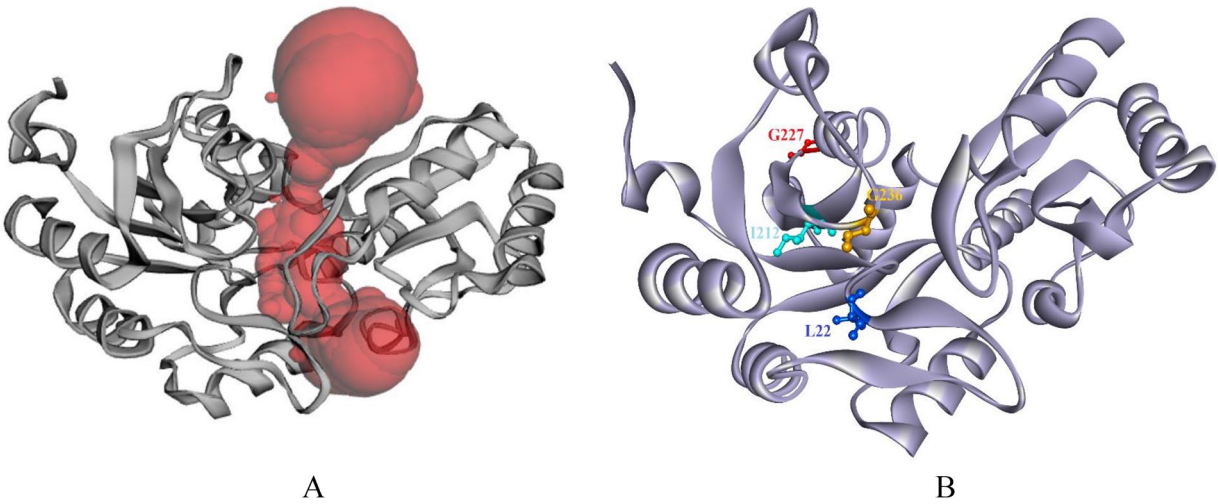


Figure 3. Active site analysis of LHPP protein using CASTp server: (A) the active portion of protein is indicated in red and (B) the G227, L22, I212, and G236 are the parts of active site of protein.

Furthermore, Rg evaluated the spatial packing, compactness, and geometrical size of the respective protein. The higher the Rg, higher would be the compactness of the mutated model. The calculated Rg values for *L22P, I212T, G227R*, and *G236R*, were 17.8, 17.3, 17.4, and 17.84 Å, respectively. However, the native protein was having Rg value of 17.92 Å, suggesting decreased compactness of all mutated models (Figure 4C).

Polar surface area is the sum of protein's surface area having polar atoms (O, H). The PSA analysis of the mutant structures revealed a slight deviation between the native and the mutants. However, *G227R* showed higher divergence (5625.41-6686.85 Å) in this property when compared with other variants. The values for *L22P, I212T*, and *G236R* lied within the range of native ensemble (Figure 4D).

MD simulations

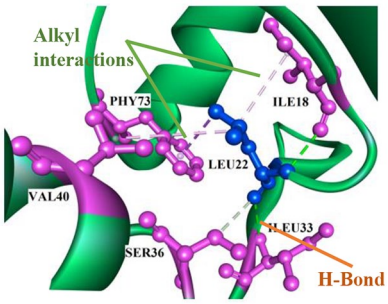
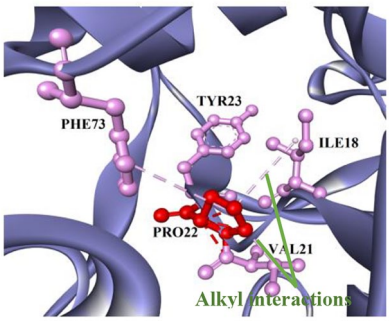
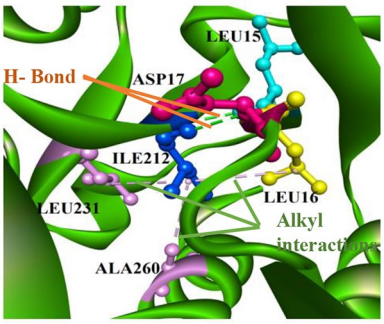
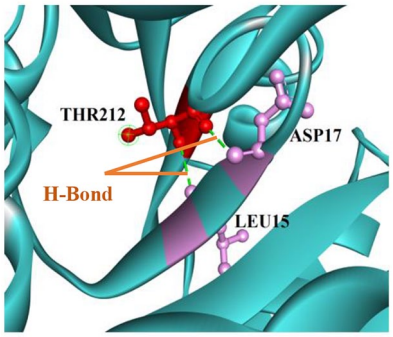
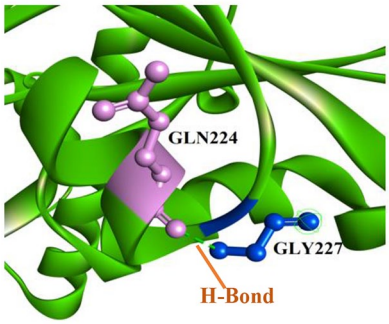
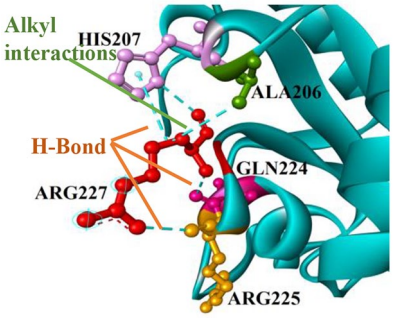
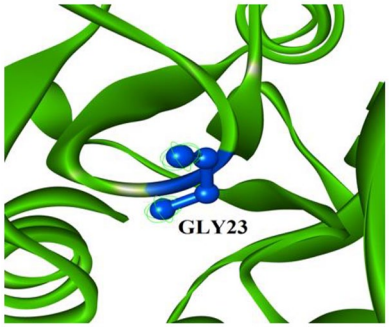
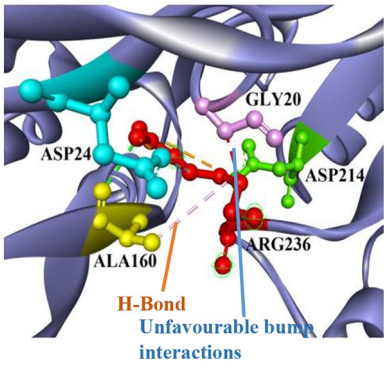
The NMSim-based RMSD analysis of wild-type LHPP protein with the 4 destabilizing revealed *L22P* as the most conformationally deviated mutant compared to the native model. To confirm this 20 ns MD, simulations of these 2 proteins were performed, which indicate higher fluctuations in RMSD values of *L22P* indicating it an unstable structure (Figure 5). Therefore, the substitution of leucine with proline at 22 positions of LHPP proteins, is likely to have a significant effect on the conformation of this protein, making this mutant an important pathogenic factor for HCC.

Discussion

Hepatocellular carcinoma is among the most pervasive forms of liver cancer that initiates gradually after the long-term inflammations. The correlation between genetic polymorphisms in multiple cancer-mediated genes and HCC initiation needs to be established.[46] *LHPP* gene encodes histidine phosphatase protein, which is a tumor-suppressor protein but the unregulated histidine-phosphorylation of *LHPP* expected to have oncogenic outcomes.[6,9] Currently, various computational tools are available to investigate the mutations that might induce alterations in structure, folding, conformation, or stability of the proteins and hence contribute to numerous genetic diseases.[47] This study employs multiple *in silico* tools based on different algorithms for the screening of potentially deleterious missense SNVs from the conserved regions of LHPP protein. The predicted variants are likely to associate with the incidence of HCC in various populations.

The study starts with the selection of 238 missense SNVs, reported for the *LHPP* in dbSNP. The selected missense SNVs were sequentially analyzed using multiple computational algorithms to predict their pathogenic impacts (Figure 1). To increase the prediction accuracy, more than one computational tools from each algorithm were considered. Among these algorithms, the first was a sequence homology–based approach, which employs sequence homology information to mark the effect of a particular substitution on the protein function.[30,48] The missense mutations marked as deleterious by the sequence homology algorithm that were further filtered by using the machine learning algorithm, which marked the pathogenic substitutions based on statistical values. The next approach was the sequence to the structure-based approach, which evaluated the protein's sequence and structural features while classifying the missense mutations as neutral or deleterious.[36,49] The further evaluation of the missense SNVs via the consensus-based approach offered an accurate and robust functional assessment as compared to the individual prediction methods.[50] The application of these multiple algorithms helped to eliminate the false positive hits from the analysis, with the selection of 5 potentially pathogenic missense SNVs (Table 1). The evolutionary conservation analysis through ConSurf marked 4 missense

Table 2. The structural comparison of native and mutated *LHPP* models. The molecular interactions of native residues are shown in blue while the red color designate the mutated residue.

MUTATIONS	WILD-TYPE MODELS	MUTATED MODELS
L22P		
I212T		
G227R		
G236R		

substitutions, that is, *L22P, I212T, G227R*, and *G236R*, to be part of the evolutionarily conserved regions of LHPP protein. These evolutionarily conserved mutations might increase the risk of cancer by deregulating the histidine phosphatase (Table 2). Fold-X marked all these 4 mutations as destabilizing mutations for the LHPP protein. As protein stability is

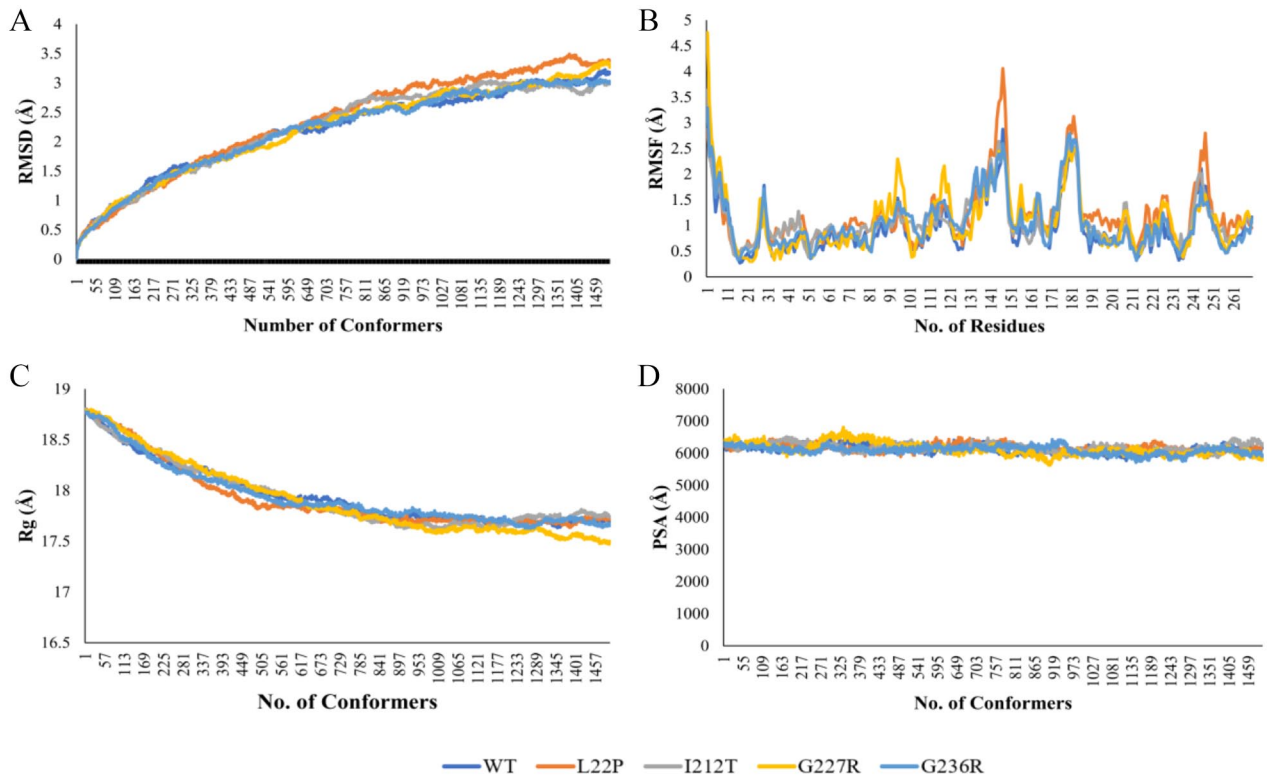


Figure 4. Trajectory analysis for native *LHPP* and its mutants: (A) RMSD, (B) RMSF, (C) Rg, and (D) PSA of the variants are identified by different colored trajectories showed that the *L22P* variant displayed higher variations from its wild-type structure. RMSD indicates root mean square deviation; RMSF, root mean square fluctuation; Rg, radius of gyration; PSA, polar surface area.

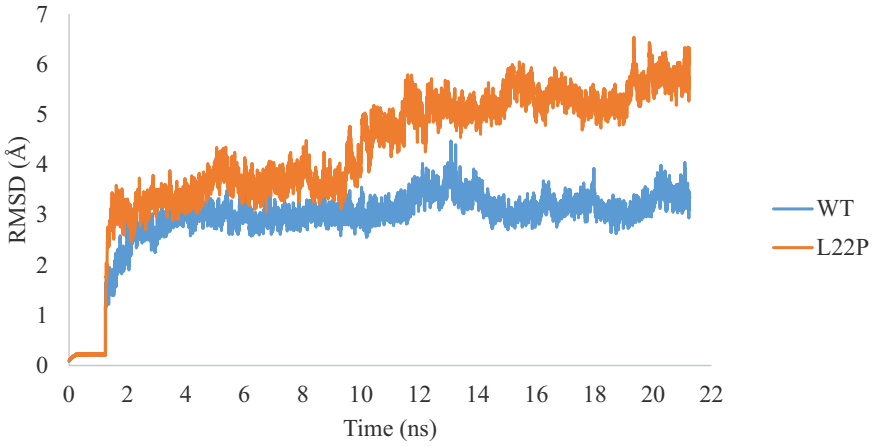


Figure 5. The MD simulation analysis for native LHPP protein and *L22P* mutant. The *L22P* mutant showed significant variations in RMSD values compared to the native protein. MD indicates molecular dynamics; RMSD, root mean square deviation.

essential for maintaining the conformation and functionality of a protein, and variations in the protein stability can cause misfolding of the proteins, resulting in the structural and functional disruption.[51-53] Furthermore, the active site analysis showed that selected mutations occur in the active site of protein and therefore may have significant contributions in altering the protein's function (Figure 3). Moreover, the comparative analysis of the native and mutated protein revealed an altered pattern of molecular interactions, with the neighboring residues inducing detrimental effects on the protein.[43]

The mutant *I212T* retained H-bonding with the same residues, but after mutation, the alkyl interactions with leucine and alanine became distorted. The resulting unbound residues in the mutated model might attract other amino acids causing conformational distortions (Table 2). In *L22P* mutant, the substitution of aliphatic leucine with aromatic proline imparts rigidity to the polypeptide chain by imposing certain torsion angels. Moreover, the formation of tertiary amide and unfavorable bumps interactions might break the alpha helices and beta sheets that can affect the

protein-protein or protein-ligand interactions.[54,55] In *G227R* and *G236R* mutants, nonpolar amino acid glycine was substituted by a polar and charged amino acid; arginine resulted in the development of hydrogen bonding and various additional interactions in the mutated models (Table 2). These deformed or distorted interactions of all mutated models could cause structural destabilization in one way or the other and would ultimately disturb the enzymatic function of the LHPP protein.[56] Geometric simulation analysis determines the stability and functionality of any protein by generating conformational trajectories. The results are represented in terms of RMSD, RMSF, Rg, and PSA, among which RMSD and RMSF are the main parameters for determining the protein's stability.[57] The RMSD of the 4 mutants marked that *L22P* is a significantly destabilizing mutant with the highest deviations in the RMSD values. The RMSD fluctuations in the other mutants were as follows, *L22P* > *I212T* > *G227R* > *G236R* (Figure 4A). Likewise, high RMSF was observed in mutant *G227R* (Figure 4B). The other 2 structural parameters used to evaluate *LHPP* stability including Rg and PSA. Furthermore, the computed Rg and PSA values of the mutants were not significantly divergent from the native LHPP protein. The mutant *G236R* displayed the least divergence behavior from the wild-type LHPP protein. The other 3 variants, *L22P*, *I212T*, and *G227R*, were suggested to have more pronounced structural and functional effects (Figure 4). Among the 4 mutants, *L22P* (*LHPP*, rs754022892) presented considerable deviation and rigidity as compared to the native LHPP protein. This divergence might be because of the distortion in the secondary conformation and folding of the protein.[15,44] The same results were also confirmed by the MD simulation analysis (Figure 5). Hence, these findings suggest a significant decrease in the stability of the LHPP protein by *L22P* mutant, which might contribute to the pathogenesis of HCC by inducing tumorigeneses. Thus, more research based on static model analysis and MD simulations is required to unveil the contribution of *LHPP* mutant in the pathogenesis of HCC.

Conclusion

The computational pipeline employed in this study identifies 4 potentially deleterious missense mutations from the evolutionarily conserved regions of phospholysine phosphohistidine inorganic pyrophosphate phosphatase. The comparison of these mutants with the native protein model reveals significant alterations in the interaction profile after mutation. These mutations might, therefore, lead to a protein instability and contributes to HCC. All these mutants need further experimental validation via wet-laboratory practices.

Acknowledgements

The authors are thankful for the online available open-access databases which were used in this research.

Author Contributions

M.S.M., M.A. and H.B. carried out the experimental analysis. A.S. and T.A. wrote the manuscript. N.M.A. and M.S. supervised the study and finalized the manuscript. All the authors have read and revised the manuscript.

Data Availability Statement

The data that support the findings of this study are available in the Supplementary Material of this article.

Ethical Statement

This study is a computational work. No humans or animal were used during the study. Therefore, ethical approval was not made necessary.

Informed Consent

Being an *in silico* study, there is no need of informed consent.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog*. 2017;16:1.
2. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019;16:589-604.
3. El-Serag HB. Epidemiology of hepatocellular carcinoma. *Liver Biol Pathobiol*. 2020:758-772.
4. Kim YJ, Jang H, Lee K, et al. PAIP 2019: liver cancer segmentation challenge. *Med Image Anal*. 2020;67:101854.
5. Niu Z-S, Niu X-J, Wang W-H. Genetic alterations in hepatocellular carcinoma: an update. *World J Gastroenterol*. 2016;22:9069.
6. Hindupur SK, Colombi M, Fuhs SR, et al. The protein histidine phosphatase LHPP is a tumour suppressor. *Nature*. 2018;555:678-682.
7. Li Y, Zhang X, Zhou X, Zhang X. LHPP suppresses bladder cancer cell proliferation and growth via inactivating AKT/p65 signaling pathway. *Biosci Rep*. 2019;39:BSR20182270.
8. Zheng J, Dai X, Chen H, Fang C, Chen J, Sun L. Down-regulation of LHPP in cervical cancer influences cell proliferation, metastasis and apoptosis by modulating AKT. *Biochem Biophys Res Commun*. 2018;503:1108-1114.
9. Liao L, Duan D, Liu Y, Chen L. LHPP inhibits hepatocellular carcinoma cell growth and metastasis. *Cell Cycle*. 2020;19:1846-1854.
10. Lesseur C, Diergaarde B, Olshan AF, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet*. 2016;48:1544-1550.
11. Cui L, Gong X, Tang Y, et al. Relationship between the LHPP gene polymorphism and resting-state brain activity in major depressive disorder. *Neural Plast*. 2016;2016:9162590.
12. Polimanti R, Wang Q, Meda SA, et al. The interplay between risky sexual behaviors and alcohol dependence: genome-wide association and neuroimaging support for LHPP as a risk gene. *Neuropsychopharmacology*. 2017;42:598-605.
13. Arshad M, Bhatti A, John P. Identification and *in silico* analysis of functional SNPs of human TAGAP protein: a comprehensive study. *PLoS ONE*. 2018;13:e0188143.
14. Doss CGP, Sudandiradoss C, Rajasekaran R, et al. Applications of computational algorithm tools to identify functional SNPs. *Func Integr Genom*. 2008;8:309-316.
15. Mahmood MS, Irshad S, Butt TA, Batool H, Batool S, Ashraf NM. In-silico analysis of deleterious missense SNPs of human TYR gene associated with oculocutaneous albinism type 1 (OCA1). *Meta Gene*. 2020;24:100674.
16. Benson D, Lipman DJ, Ostell J. GenBank. *Nucleic Acids Res*. 1993;21:2963-2965.
17. Fang LT, Afshar PT, Chhibber A, et al. An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biol*. 2015;16:197.

18. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4: 1073-1081.
19. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31:2745-2747.
20. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13: 2129-2141.
21. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*. 2012;7:e46688.
22. Dabhi B, Mistry KN. *In silico* analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. *Meta Gene*. 2014;2:586-595.
23. Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res*. 2017;45:W247-W252.
24. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genom*. 2013;14:S6.
25. Yates CM, Filippis I, Kelley LA, Sternberg MJ. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol*. 2014;426:2692-2701.
26. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*. 2017;2017:134981.
27. Hurst JM, McMillan LE, Porter CT, Allen J, Fakorede A, Martin AC. The SAAPdb web resource: a large-scale structural analysis of mutant proteins. *Hum Mutat*. 2009;30:616-624.
28. Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. *Bioinformatics*. 2008;24:2397-2398.
29. López-Ferrando V, Gazzo A, De La Cruz X, Orozco M, Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res*. 2017;45:W222-W228.
30. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013; Chapter7:Unit7.20.
31. Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res*. 2017;45:W229-W235.
32. Buß O, Rudat J, Ochsenreither K. FoldX as protein engineering tool: better than random based approaches. *Comput Struct Biotechnol J*. 2018;16:25-33.
33. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*. 2011;12:1-12.
34. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*. 2011;88:440-449.
35. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genom*. 2013;14:S2.
36. Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS ONE*. 2015;10:e0117380.
37. Bendl J, Stourac J, Salanda O, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*. 2014;10:e1003440.
38. Goldenberg O, Erez E, Nimrod G, Ben-Tal N. The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res*. 2009;37:D323-D327.
39. Sussman JL, Lin D, Jiang J, et al. Protein data bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr*. 1998;54:1078-1084.
40. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33:W382-W388.
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283-291.
42. Tian W, Chen C, Lei X, Zhao J, Liang J.CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46:W363-W367.
43. Bouafi H, Bencheikh S, Mehdi Krami AL, et al. Prediction and structural comparison of deleterious coding nonsynonymous single nucleotide polymorphisms (nsSNPs) in human LEP gene associated with obesity. *Biomed Res Int*. 2019;2019:1832084.
44. Lopus M, Paul DM, Rajasekaran R. Unraveling the deleterious effects of cancer-driven stk11 mutants through conformational sampling approach. *Cancer Inform*. 2016;15:35-44.
45. Krüger DM, Ahmed A, Gohlke H. NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res*. 2012;40:W310-W316.
46. Dondeti MF, El-Maadawy EA, Talaat RM. Hepatitis-related hepatocellular carcinoma: insights into cytokine gene polymorphisms. *World J Gastroentero*. 2016;22:6800.
47. Broom A, Jacobi Z, Trainor K, Meiering EM. Computational tools help improve protein stability but with a solubility tradeoff. *J Biol Chem*. 2017;292:14349-14361.
48. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812-3814.
49. Gorlov IP, Moore JH, Peng B, Jin JL, Gorlova OY, Amos CI. SNP characteristics predict replication success in association studies. *Hum Genet*. 2014;133:1477-1486.
50. Hassan MS, Shaalan AA, Dessouky MI, Abdelnaiem AE, ElHefnawi M. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics*. 2019;111:869-882.
51. Deller MC, Kong L, Rupp B. Protein stability: a crystallographer's perspective. *Acta Crystallogr F Struct Biol Commun*. 2016;72:72-95.
52. Witham S, Takano K, Schwartz C, Alexov E. A missense mutation in CLIC2 associated with intellectual disability is predicted by *in silico* modeling to affect protein stability and dynamics. *Proteins*. 2011;79:2444-2454.
53. Hossain MS, Roy AS, Islam MS. *In silico* analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. *Sci Rep*. 2020;10:14542.
54. Bach TM, Takagi H. Properties, metabolisms, and applications of L-proline analogues. *Appl Microbiol Biotechnol*. 2013;97:6623-6634.
55. Prajapati RS, Das M, Sreeramulu S, et al. Thermodynamic effects of proline introduction on protein stability. *Proteins*. 2007;66:480-491.
56. Samsel A, Seneff S. Glyphosate pathways to modern diseases V: amino acid analogue of glycine in diverse proteins. *J Biol Phys Chem*. 2016;16:9-46.
57. Senthilkumar B, Meshachpaul D, Sethumadhavan R, Rajasekaran R. Selection of effective and highly thermostable Bacillus subtilis lipase A template as an industrial biocatalyst-A modern computational approach. *Frontiers in Biology*. 2015;10:508-519.