Received: 14 November 2020 | Revised: 8 January 2021 | Accepted: 21 January 2021

DOI: 10.1002/prot.26051

RESEARCH ARTICLE

PROTEINS WILEY

In silico screening and molecular dynamics simulation of deleterious PAH mutations responsible for phenylketonuria genetic disorder

Andrea Lopez | Brandon Havranek | George A. Papadantonakis | Shahidul M. Islam

Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois

Correspondence
Shahidul M. Islam, Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607.
Email: mshahidi@uic.edu

Abstract

Phenylketonuria (PKU) is a genetic disorder that if left untreated can lead to behavioral problems, epilepsy, and even mental retardation. PKU results from mutations within the phenylalanine-4-hydroxylase (PAH) gene that encodes for the PAH protein. The study of all PAH causing mutations is improbable using experimental techniques. In this study, a collection of in silico resources, sorting intolerant from tolerant, Polyphen-2, PhD-SNP, and MutPred were used to identify possible pathogenetic and deleterious PAH non-synonymous single nucleotide polymorphisms (nsSNPs). We identified two variants of PAH, I65N and L311P, to be the most deleterious and disease causing nsSNPs. Molecular dynamics (MD) simulations were carried out to characterize these point mutations on the atomic level. MD simulations revealed increased flexibility and a decrease in the hydrogen bond network for both mutants compared to the native protein. Free energy calculations using the MM/GBSA approach found that BH_4, a drug-based therapy for PKU patients, had a higher binding affinity for I65N and L311P mutants compared to the wildtype protein. We also identify important residues in the BH_4 binding pocket that may be of interest for the rational drug design of other PAH drug-based therapies. Lastly, free energy calculations also determined that the I65N mutation may impair the dimerization of the N-terminal regulatory domain of PAH.

KEYWORDS
free energy calculations, in silico genetic screening, molecular dynamics, PKU variants, tetrahydrobiopterin

1 | INTRODUCTION

Phenylketonuria (PKU) is an autosomal recessive genetic disorder,[1] with an estimated occurrence of one in every 20 000 births in the United States.[2-4] PKU occurs mostly in infants and can continue into childhood. Common symptoms of PKU include odor, eczema, reduced hair, skin and iris pigmentation. If left untreated, symptoms can be as severe as mental retardation, behavioral problems, epilepsy, microcephaly, and reduced growth.[1,4] PKU results from mutations in the phenylalanine hydroxylase gene[1,2] that encodes for the phenylalanine-4-hydroxylase (PAH) protein. PAH is an enzyme that catalyzes the hydroxylation of the aromatic side-chain of L-phenylalanine to produce L-tyrosine. The rate-limiting step of the reaction is also dependent on a cofactor, tetrahydrobiopterin (BH_4) and a non-heme iron. This reaction is very important as it prevents the buildup of phenylalanine, which if present in excessive amounts can be neurotoxic. Only a synthetically formulated BH_4 has shown to benefit some patients who suffer from PKU suggesting that the supplemental BH_4 may again activate the reaction so that more phenylalanine can be converted to tyrosine.

The PAH protein structure, found in humans and other mammals, has a tetrameric[5] conformation where the tetrameric complex is formed by a 2-fold symmetry of two dimers (Figure 1A). Crystal structures of both unbound (apo) and BH_4-bound (holo) states (Figure 1B) as well as modeling studies[6] revealed that BH_4 interacts with the serine-23 residue via hydrogen bonds with both O1′ and O2′ in the dihydroxypropyl side-chain (Figure 1C). Serine-251 and Glycine-247 also form hydrogen bonds with BH_4, while Leucine-249, Leucine-248 and Phenylalanine-254 help to stabilize BH_4 within the protein core via van der Waals, pi-sigma, and pi-pi stacking interactions. Moreover, it is been suggested by Patel et al[7] that phenylalanine binding at the N-terminal regulatory domain (PAH-RD) results in dimerization of the regulatory domain to reveal a homodimer. Disease associated PAH mutants are thought to impair the dimer equilibrium of PAH-RD.

Some studies[1,8,9] have shown that missense mutations may account for the majority of patients that exhibit increased amounts of phenylalanine. The PAH gene has approximately 548 mutations, with 50% of these mutations being missense mutations resulting from non-synonymous single nucleotide polymorphisms (nsSNPs).[2] NsSNPs occur when a single nucleotide change in the DNA sequence results in the formation of a different amino acid in the encoded protein sequence.[10] PAH mutations have also been shown to be different in various populations. A study[9] on 33 Korean patients have identified two novel missense mutations, A202T and G344D, which were not identified in the previous studies.[11] The study showed that the A202T mutation may weaken hydrophobic interactions among Ala202, Tyr198, Tyr206, and Leu348 residues in the PAH protein, while the G344D mutations may alter intermolecular interactions with some residues in the β4 strand and the α10 helical domain as the Gly344 is located between the β4 strand and the α10 helix in the catalytic domains of the PAH protein. Song et al[12] found Y154H, R157K, S349A, A395D, E280G, and D282G mutations in PAH from investigating 185 patients in northern China. These investigations also reveal that some mutations are very different from those found in Europe.[13] Patients in the United States have shown the largest mutational heterogeneity in PAH with the main mutations in amino acids R408W, Y414C, F39L, I65T, L348V, E280K and P281L.[11] Due to the possibility of various PAH variants, it is important to identify the variants of PAH that are most likely to be harmful in the human body.

Many studies[14-19] have shown that nsSNPs can alter the structure of proteins making the proteins more flexible in 3-dimensional space, therefore affecting the native protein function. In particular, mutations in the PAH gene have been closely associated with the misfolding and instability of the PAH protein.[20,21] Reblova et al[22] used molecular dynamics simulations to characterize R408Q, R408W, L348V, and V388M mutations in PAH and how they cause structural defects in the

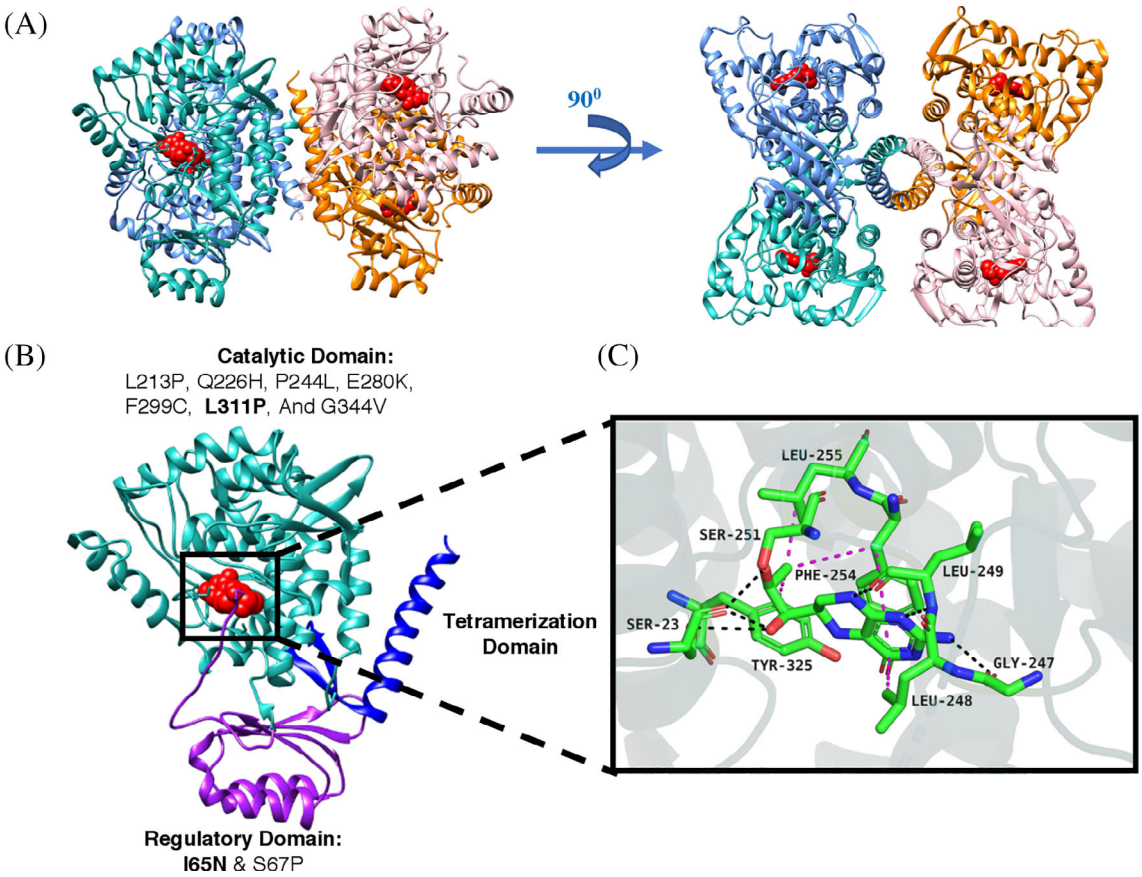

FIGURE 1 A, Crystal structure of tetrameric, a dimer of a dimer, full-length human phenylalanine hydroxylase (PAH) in complex with cofactor and negative regulator, tetrahydrobiopterin (Red), PDB: 6HYC; B, One of the chains of PAH which shows three subdomains (blue, purple and yellow), tetrahydrobiopterin (red) and nine important mutations; C, Interactions of tetrahydrobiopterin with binding domain of PAH

protein. Currently there are only two forms of treatment for PKU, a low phenylalanine diet and sapropterin dihydrochloride.[23,24] Sapropterin dihydrochloride is a synthetic form of the naturally occurring BH_4.[23,24] The FDA approved medication works by flooding the mutant PAH structure with excess BH_4, which may help suppress their low binding affinity, and allows phenylalanine to be converted to tyrosine in order to prevent PKU.[25,26] As of now, this is the only effective treatment that can suppress the mutants low binding affinity to BH_4, and help convert phenylalanine to tyrosine.[5,26]

The deleterious and disease causing nsSNPs can affect the structure, stability, and residual fluctuation when compared to the wildtype. Therefore, identifying and predicting these diseases causing nsSNPs are very important as they can be possible targets for the development of effective therapeutic drugs. In order to characterize the most deleterious and disease causing nsSNPs in the PAH enzyme, in silico approaches, such as the sorting intolerant from tolerant (SIFT),[27,28] Polyphen-2,[29] PhD-SNP,[30] and MutPred,[31] were used to study 50 nsSNPs in the PAH gene. Many studies have used these methods to predict the effects of nsSNPs on protein function with great success.[32-34] Of the 50 PAH nsSNPs, the most deleterious and disease-causing mutations were further studied using molecular Dynamics (MD) simulations in order to understand the most deleterious mutation's effects on protein stability, structure, and dynamics. This study utilized the crystal structure of a full-length human phenylalanine hydroxylase in complex with the cofactor and negative regulator, tetrahydrobiopterin, from the PDB[35] with PDB ID 6HYC.[5] Furthermore, MM/GBSA calculations were performed between BH_4 and the wildtype and mutant PAH structures to analyze how pathogenic mutations in the PAH enzyme may impact the binding affinity of BH_4. This study may help lead to more novel approaches in the treatment of PKU. Additional free energy calculations were performed to analyze the affect a deleterious mutation in the N-terminal regulatory domain (RD) may have on the binding ability of PAH-RD, which may be consequential in its propensity to form a homodimer.

2 | METHODS

2.1 | Predicting deleterious and pathogenic nsSNPs

The Universal Protein Resource database (UniProt)[36] was used to obtain potential mutations that are likely to occur in the PAH gene. The mutations were then subjected to the algorithm of the SIFT server[27,28] (https://sift.bii.a-star.edu.sg/). SIFT is a prediction tool that can help determine if the nsSNPs affects protein function.[27] SIFT functions by using sequence homology and assuming evolutionary conserved regions tend to be less tolerant of mutations, indicating that missense mutations affect protein function.[37] Furthermore, SIFT prediction is based upon the degree of conservation of amino acid residues in sequence alignments that are derived from related structures and are collected using Position-Specific Iterated BLAST (PSI-BLAST).[38] The SIFT score ranges from 0 to 1, and a score < 0.05 is predicted to affect protein function.[14,37] After SIFT, the missense mutations were subjected to Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/). PolyPhen-2 is able to predict the possible impact of missense mutation on the stability and function of human proteins by assuming evolutionary conservations.[29] The server is recognized for analyzing the homology of SNPs, with over a 68% accuracy in predicting whether nsSNPs are disease related.[30] PolyPhen-2 analyzes the damaging probability of nsSNPs and determines the functional impact using the following factors: ligand binding, single peptide, transmembrane, and protein interactions,[14] the predictions are essential for interpreting large sets of rare genetic variants.[29] Therefore, the scores obtained from PolyPhen-2 can help determine which mutations are highly probable to disrupt protein function. Like SIFT, PolyPhen-2 aligns multiple protein sequences; however, PolyPhen-2 is not as dependent on alignment as SIFT.[29,39] Therefore, SIFT and PolyPhen-2 can be used together to best predict how the missense mutations will affect protein function. PolyPhen-2 gives a confidence score that ranges from 0 to 1. Scores that are <0.15 are predicted to be benign, 0.15 to 1 are possibly damaging; however, scores between 0.85 and 1 indicates that the mutation has a higher probability of being damaging.

The mutations agreed to be highly deleterious, with a SIFT tolerance score of 0 and a PolyPhen-2 score of 1, were further subjected to PhD-SNP and MutPred. PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html) is a support vector machine that can predict if the nsSNP will result in a new structure of the protein studied and if it can be related to a genetic disease in humans.[30] PhD-SNP solely depends on sequence based features.[40] More specifically, PhD-SNP can classify the disease-causing point mutation.[14] PhD-SNP has been reported to have more than 74% accuracy in predicting whether a singular nsSNP can be disease related.[30] MutPred (http://mutpred.mutdb.org/) is another widely used and developed web server based upon protein sequence.[31,40,41] From the protein sequence, MutPred models the changes of structural features and functional sites between the wildtype and the nsSNP.[31] MutPred has been developed to improve classification accuracy with respect to human disease mutations.[31] MutPred prioritizes the amino acid substitutions that are most likely to be involved with disease, and assesses the probabilities that a mutation affects structural and functional components of the protein.[41] MutPred gives two scores: a general score (g-score) that determines the probability that the nsSNP is related to disease, and the property value (p-score), which indicates whether there are certain structural or functional properties that are disrupted. The following indicates the interpretations of the two scores: $g > 0.5$ and $P < .05$ is considered to have an actionable hypothesis, $g > 0.75$ and $P < .05$ is referred to as confident hypotheses, and $g > 0.75$ and $P < .01$ will yield a very confident hypothesis that the nsSNP is disease related and affects structural properties of the protein.[14]

2.2 | MD simulations

MD Simulations were performed using Nanoscale Molecular Dynamics (NAMD).[42] The crystal structure of full-length human

phenylalanine hydroxylase is available in the protein data bank (PDB)[35] with PDB ID: (6HYC).[5] Since all four chains in 6HYC are identical, chain A is used for the molecular dynamic's simulations. The wildtype PAH and I65N and L311P mutant protein systems were prepared using the CHARMM-gui solution builder[43,44] with CHARMM36m additive forcefield for proteins.[45] The protein systems were fit in a rectangular water box filled with TIP3P water molecules with a radius of 10 Å from the surface of the complex. In order to mimic physiological conditions, 0.15 M of KCl ions were added using the Monte-Carlo ion placing method. All simulations were carried out under periodic boundary condition using the NPT ensemble where the pressure is kept at 1 atm and temperature is kept at 303.15 K. The Particle Mesh Ewald algorithm[46] was used for long-range electrostatic forces between atoms and the cut-off for non-bonded atom interactions was set to 12 Å. The temperature was controlled using Langevin dynamics[47] and pressure was controlled using the Nose-Hoover Langevin piston method.[47] The systems were minimized for 2 ns and then subjected to 30 ns production MD simulations. The integration time step of dynamics was set to 2-fs. All bonds involving hydrogen were constrained using the SHAKE algorithm[48] and water was restrained using the SETTLE algorithm.[49]

2.3 | Free energy calculations

The wildtype and two mutant systems (I65N, and L311P) bound with the ligand BH_4 were prepared for binding free energy calculations using the CHARMM-GUI solution builder using the CHARMM36m additive forcefield for proteins[45] and CHARMM General Force Filed (CGenFF) for ligands.[50] The binding free energy simulations were conducted using the Amber 18.[51] The systems were minimized with a steepest decent energy minimization for 5000 cycles and a conjugate gradient algorithm for 10 000 cycles. Equilibration was carried out for 1 ns under NVT conditions with a positional restraint of 1 kcal/mol and 1 fs time step. The temperature was set to 303.15 K and maintained using Langevin dynamics. Production MD simulations were conducted for 10 ns with a 2 fs time step under NPT conditions using a friction coefficient (γ) of 1.0 ps^{-1} for the Langevin thermostat. The pressure was held constant using the Monte Carlo barostat with constant pressure dynamics performed using isotropic scaling. Integration was performed with a leap-frog algorithm and all bonds involving hydrogen were restrained using SHAKE.[48,49] The Ambertools20[52] ante-MMPBSA python script[53] was used to create the complex, receptor, and ligand parameter files from the solvated parameter files used in the MD simulations. The binding free energies between BH_4 and the wildtype and two mutant systems were performed using the Molecular Mechanics-Generalized Born Surface Area (MM/GBSA) method with the MMPBSA python script[53] inside of Ambertools20. The binding free energy was calculated from 50 frames throughout the 10 ns simulation. The Generalized born method was set to igb = 5, along with salt concentration set to 0.15 M. The dielectric constant of solvent was set to 78.5 and the dielectric constant of solute was set to 1.0. Both of these dielectric constants are Amber default and recommended values. Wang et al.[54] investigated 21 protein-protein complexes to access the performance of MM/GBSA in different conditions and suggested using a solute dielectric constant of 1.0. Additionally, the solvent-accessible surface area (SASA) was calculated using a γ = 0.0072 kcal/mol/Å2 and β = 0.0 kcal/mol, respectively.

In MM/GBSA calculations, the binding free energy (ΔG_{bind}) between the ligand and PAH is calculated as

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM,gas} + \Delta G_{sol} - T\Delta S \quad (1)$$

where $\Delta E_{MM,gas}$, ΔG_{sol} and $-T\Delta S$ are the changes of the gas phase molecular mechanics energy, the solvation free energy, and the conformational entropy upon binding, respectively. The $\Delta E_{MM,gas}$ is the sum of the internal energy ($\Delta E_{internal}$) arising from bond, angle and dihedral interactions, non-bonded electrostatic energy (ΔE_{ELE}) and Van der Waals energy (ΔE_{VDW}).

$$\Delta E_{MM,gas} = \Delta E_{int} + \Delta E_{ELE} + \Delta E_{VDW} \quad (2)$$

In MM/GBSA calculations, the ΔE_{int} is canceled since the complex, receptor, and ligand parameter files are created from the same trajectory. The solvation free energy (ΔG_{sol}) is calculated from the sum of the polar solvation energy (ΔG_{GB}) using the Generalized Born model, and the non-polar energy is calculated based on the solvent-accessible surface area (ΔG_{surf}) according to the LCPO algorithm.[55]

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{surf} \quad (3)$$

The conformational entropy change $-T\Delta S$ is usually computed by normal-mode analysis on a set of conformational snapshots taken from MD simulations. In this case, contribution from the conformational entropy is neglected as inclusion of the entropy term may not improve ranking the binding affinities of BH_4 with the wild type and the two mutant systems.

MM/GBSA calculations were also performed according to the same procedures discussed above using PDB 5FII[7,35] which contains the homodimer of the PAH N-terminal regulatory domain. Due to smaller system size, long MD simulations of 100 ns were conducted for both the wildtype and I65N mutant structures. The free energy calculations were performed using 800 frames between 20 and 100 ns when the systems were properly equilibrated.

3 | RESULTS AND DISCUSSIONS

3.1 | PAH mutations that are damaging and affect protein function

Identifying pathogenic mutations in the PAH protein is useful as it may help provide insight into the structural consequences of deleterious and pathogenic mutations on the PAH protein that can be used in the future treatment of PKU. Of the potential mutations, 50 total

mutations that have been characterized in the literature were selected from the Universal Protein Resource database (UniProt)[36] in this study. The sorting intolerant from tolerant (SIFT) algorithm was used to best predict which of the 50 nsSNPs in the PAH protein were the most deleterious. Thirty-five of the 50 mutations were predicted to affect protein function (APF) according to SIFT as they have scores below a tolerance index of 0.05 (Table 1). Of the 35 mutations that APF, 9 of the missense mutations, I65N, S67P, L213P, Q226H,

TABLE 1 Scores from SIFT and PolyPhen-2 along with their prediction on how it impacts protein structure and function

Missense mutation	SIFT tolerance score	SIFT prediction	PolyPhen-2 score	PolyPhen-2 prediction
S16P	0.08	Tolerated	0	Benign
L41F	0.05	Tolerated	0.948	Probably damaging
G46S	0.46	Tolerated	0.994	Probably damaging
E56D	0.19	Tolerated	0.003	Benign
I65N	**0**	**APF**	**1**	**Probably damaging**
S67P	**0**	**APF**	**1**	**Probably damaging**
T92I	0.39	Tolerated	0.006	Benign
D129Y	0	APF	0.999	Probably damaging
G148S	0	APF	0.998	Probably damaging
R157S	0	APF	0.997	Probably damaging
Q160P	0.16	Tolerated	0.245	Benign
R169H	0	APF	0.083	Benign
G171R	0	APF	0.999	Probably damaging
P175A	0.05	Tolerated	0.999	Probably damaging
V190A	0	APF	0.999	Probably damaging
H201R	0.24	Tolerated	0.982	Probably damaging
N207D	0.08	Tolerated	0.918	Probably damaging
L213P	**0**	**APF**	**1**	**Probably damaging**
D222V	0	APF	0.645	Possibly damaging
Q226H	**0**	**APF**	**1**	**Probably damaging**
T238P	0.01	APF	1	Probably damaging
R241H	0.12	Tolerated	0.903	Probably damaging
P244L	**0**	**APF**	**1**	**Probably damaging**
A246D	0.02	APF	1	Probably damaging
R252G	0.01	APF	1	Probably damaging
L255V	0.01	APF	0.998	Probably damaging
R261P	0	APF	0.96	Probably damaging
C265G	0.03	APF	0.342	Benign
S273F	0	APF	0.996	Probably damaging
Y277C	0.01	APF	1	Probably damaging
E280K	**0**	**APF**	**1**	**Probably damaging**
I283N	0.04	APF	0.999	Probably damaging
F299C	**0**	**APF**	**1**	**Probably damaging**
Q304R	0	APF	0.999	Probably damaging
L311P	**0**	**APF**	**1**	**Probably damaging**
A322T	0.04	APF	0.99	Probably damaging
C334S	0.01	APF	1	Probably damaging
K341T	0.01	APF	0.996	Probably damaging
G344V	**0**	**APF**	**1**	**Probably damaging**
L348V	0.01	APF	0.964	Probably damaging
C357G	0.13	Tolerated	0.977	Probably damaging

(Continues)

TABLE 1 (Continued)

Missense mutation	SIFT tolerance score	SIFT prediction	PolyPhen-2 score	PolyPhen-2 prediction
P366H	0.03	APF	0.981	Probably damaging
Y387H	0	APF	0.987	Probably damaging
D394A	0.25	Tolerated	0.451	Possibly damaging
A403V	0.32	Tolerated	0.953	Probably damaging
P407L	0.29	Tolerated	0.987	Probably damaging
F410S	0.01	APF	1	Probably damaging
D415N	1	Tolerated	0.003	Benign
L430P	0.01	APF	0.978	Probably damaging
A447D	0.01	APF	0.976	Probably damaging

Note: Mutations in bold show the most confidence in affecting protein function (APF) and are further studied with PhD-SNP and MutPred.

P244L, E280K, F299C, L311P, and G344V, were considered to be highly deleterious with SIFT scores of 0. Next, PolyPhen-2 was used to verify the results gathered from SIFT. The PolyPhen-2 score (PSIC) <0.15 indicates the mutation is benign, between 0.15 and 1 is possibly damaging, and between 0.85 and 1 to be higher probability of damaging. Of the 50 nsSNPs, PolyPhen-2 predicted 7 of the mutations, S16P, E56D, T921, Q160P, R169H, C265G, and D451N, to be benign, 2 of the mutations, D222V and D394A, to be possibly damaging and 41 of the mutations, as shown in Table 1, to be probably damaging.

The SIFT server and Polyphen-2 server work together to provide the most confident predictions on the impact of specific nsSNPs. Of the 50 mutations, only 9, I65N, S67P, L213P, Q226H, P244L, E280K, F299C, L311P, G344V, were selected to undergo analysis with the PhD-SNP and MutPred servers, as these 9 mutations had the maximum deleterious scores with a SIFT tolerance score of 0 and a PSIC (PolyPhen-2) score of 1. With this criterion, we ensured that only the most confident deleterious nsSNPs moved on for further examination. A recommendation by the American College of Medical Genetics and Genomics suggest using multiple servers to predict the effects of mutations on protein function,[56] thus the agreeable deleterious mutations from SIFT and PolyPhen-2 were further studied with PhD-SNP and MutPred.

3.2 | PAH mutations that are pathogenic and affect protein structure

PhD-SNP and MutPred are used in conjunction to further identify if the 9 missense mutations, identified with SIFT and PolyPhen-2, may be pathogenic and affect protein structure. PhD-SNP was used to determine the pathogenic effect of the mutations. The 9 mutations, I65N, S67P, L213P, Q226H, P244L, E280K, F299C, L311P, G344V, which were found to have a high confidence level of being deleterious to the function of the PAH protein, were also found to be disease associated according to PhD-SNP (Table 2).

To further analyze the results from PhD-SNP, all 9 nsSNPs, listed in Table 1, were also studied with MutPred. Of the 9 mutations, MutPred found only two of the mutations, I65N and L311P, to have very confident reliability of structural disruption, by predicting a gain of disorder. Mutant I65N had g and *P* scores of 0.961 and .0063 respectively, which suggests a gain of disorder with I65N mutation. Furthermore, MutPred elicited a loss of catalytic residue with a confident hypothesis (g-score = 0.961 and p-score = .0359), and a loss of sheet (g-score = 0.961 and *P* = .0457). The L311P mutant also had a very confident hypothesis of gain of disorder due to the mutation, as the g and p-scores are found to be 0.918 and 0.0067, respectively. L311P was also predicted to have loss of catalytic residue (*P* = .0295) and loss of stability (*P* = .0441). MutPred predicted both the I65N and L311P mutations to lead to a loss of catalytic residue and a gain of disorder with confident and very confident hypotheses, respectively. Loss of catalytic residues is known to compromise the amino acids located in the active site, which can disrupt enzyme-catalyzed reactions.[57] Therefore, a loss in the catalytic residue may impact the metabolic pathway responsible for converting phenylalanine to tyrosine.

Bosco et al[58] studied nine patients that suffer from PKU by isolating their DNA from peripheral blood leukocytes and found that the I65N mutation is prevalent in the impairment of PAH's enzymatic activity. The I65N mutation was found to be one of the most pathogenic mutations that leads to PKU[58]. The patient who had the I65N mutation demonstrated adverse effects of PKU due to a lack of early treatment which resulted in a mild form of mental retardation.[58] The L311P mutation has also been studied experimentally. Lichter-Konecki et.al[59] found that the missense mutation, L311P, that changes the leucine amino acid to proline resulted in an unstable PAH protein, as well as a high probability that the mutation leads to PKU. These experimental results agree very well with the results from both PhD-SNP and MutPred.

3.3 | MD analysis

In order to further understand the structure, dynamics and function of the possible deleterious and disease-causing mutations, I65N and L311P, molecular dynamics simulations were carried out and the results were further compared to the wildtype PAH protein. First, the root-mean-square deviation (RMSD) was calculated, with respect to

TABLE 2 Agreeable mutations from SIFT and PolyPhen-2 were studied with PhD-SNP and MutPred to predict possible pathogenic mutations and whether they may affect PAH protein structure

Missense mutation	PhD-SNP	g Score	*P* Score	Mechanism disrupted	Confidence
I65N	**Disease**	**0.961**	**.0063**	**Gain of disorder**	**Very confident**
I65N	**Disease**	**0.961**	**.0359**	**Loss of catalytic residue**	**Confident**
I65N	**Disease**	**0.961**	**.0457**	**Loss of sheet**	**Confident**
S67P	Disease	0.918	.0508	Loss of phosphorylation	Low confidence
L213P	Disease	0.927	.0406	Loss of stability	Confident
Q226H	Disease	0.957	.2137	Loss of disorder	Low Confidence
P244L	Disease	0.925	.0109	Gain of catalytic residue	Confident
E280K	Disease	0.965	.0351	Gain of ubiquitination	Confident
E280K	Disease	0.965	.0066	Gain of methylation	Very Confident
F299C	Disease	0.970	.0559	Gain of catalytic residue	Low Confidence
L311P	**Disease**	**0.918**	**.0067**	**Gain of disorder**	**Very Confident**
L311P	**Disease**	**0.918**	**.0295**	**Loss of catalytic residue**	**Confident**
L311P	**Disease**	**0.918**	**.0441**	**Loss of stability**	**Confident**
G344V	Disease	0.993	.0464	Loss of ubiquitination	**Confident**

Note: Mutations in bold are studied with MD simulations.

the Cα atoms, from the initial structure, to compare how the mutations impact the dynamic behavior of the PAH protein (Figure 2A).

The RMSD of the mutant proteins, I65N and L311P, deviate from their initial structures to a much greater extent than the wildtype protein. The RMSD of the wildtype protein stays relatively stable throughout the simulation, with an average RMSD of 2.28 Å. In comparison, the L311P and I65N mutants deviate far more from their initial crystal structures with an average RMSD of 3.30 and 4.95 Å, respectively. The RMSD of the I65N mutant increases drastically from its initial crystal structure until about 7 Å before the RMSD begins to stabilize (Figure 2A). This is further verified by overlapping the last frame of the wildtype and I65N and L311P mutant PAH structures (Figure 3).

There is a noticeable difference between the C-terminal region in the mutant structures compared to the wildtype. Furthermore, structural differences are observed in a few of the alpha helical regions including in the regulatory domain, especially in the I65N mutant, which helps to explain why the RMSD was much higher than the L311P and native structures. This may further help to show that mutations in the PAH gene are closely associated with the misfolding and instability of the PAH protein.[60]

Next, the root-mean-square-fluctuation (RMSF) was calculated for the Cα atoms of the PAH protein residues, to provide insight into the dynamic behavior of the mutant structure residues compared to the wildtype (Figure 2B). Both mutants, I65N and L311P, had larger residual fluctuation compared to the wildtype (Figure 2B). The wildtype protein had an average RMSF of 1.17 Å, while the I65N and L311P mutants had an average RMSF of 1.35 and 1.27 Å, respectively, demonstrating that there is greater flexibility in the mutant structures as compared to the wildtype. In particular, the I65N mutant has the largest fluctuation over the course of the simulation, in accordance with the RMSD calculation. Residues 125 to 135 are in the loop region and quite flexible, therefore, the RMSF for the wild type as well as the mutants are quite large between resides 125 and 135.

Figure 2C represents the radius of gyration, Rg, of the wild type PAH protein and its mutants, I65N and L311P. The Rg provide additional insights into the overall dynamics of these protein structures. Based on the radius of gyration calculation, the wildtype was found to be more stable compared to the I65N and L311P mutants, indicating the wildtype is more compact and tightly folded. The radius of gyration of I65N change more than the wildtype, while the L311P mutant had the largest radius of gyration, which agrees with the predicted MutPred results indicating a loss of stability. This may suggest that the L311P mutant is undergoing a significant structural transformation.

Hydrogen bonds play an important role in the stability of protein structure. Thus, the total number of intramolecular hydrogen bonds, with respect to time, were calculated for the mutant and wildtype PAH structures (Figure 2D). Both mutant structures had a disruption in their intramolecular hydrogen bond network with less hydrogen bonds observed than the wildtype. In particular, the L311P mutant had the fewest number of intramolecular hydrogen bonds, which likely contributed to the loss of stability that is also observed in the radius of gyration calculation. The theoretical hydrogen bond calculation agrees with the experimental results that also found the L311P mutation to lead to a disruption in the hydrogen bond network of PAH.[61]

To further assess the stability of the PAH mutant proteins compared to the wildtype, a principal component analysis of the wildtype and L311P and I65N mutant proteins, along their first two eigenvectors, were calculated to provide insight into the internal motion of the proteins during the simulation (Figure 4).

Both mutant proteins covered a larger area along the first two principal eigenvectors as compared to the wildtype. In particular, the

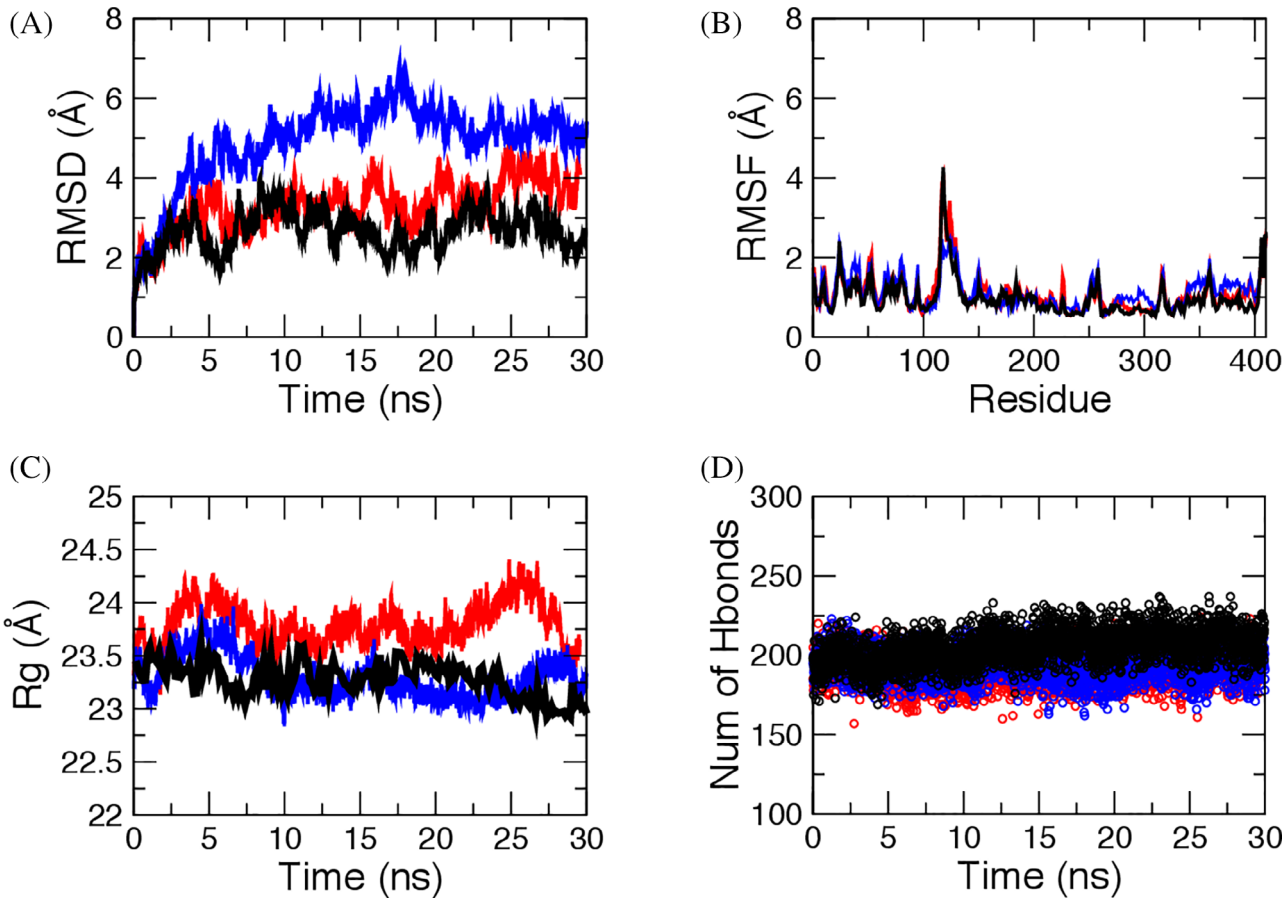

FIGURE 2 A, rootmean-square deviation (RMSD); B, root-mean-square-fluctuation (RMSF); C, Radius of gyration (Rg); and D, Total number of intramolecular hydrogen bonds. RMSD, RMSF, and Rg is calculated with respect to the Cα atoms. The values for wildtype, L311P mutant, and I65N mutant are shown in black, red and blue, respectively [Color figure can be viewed at wileyonlinelibrary.com]

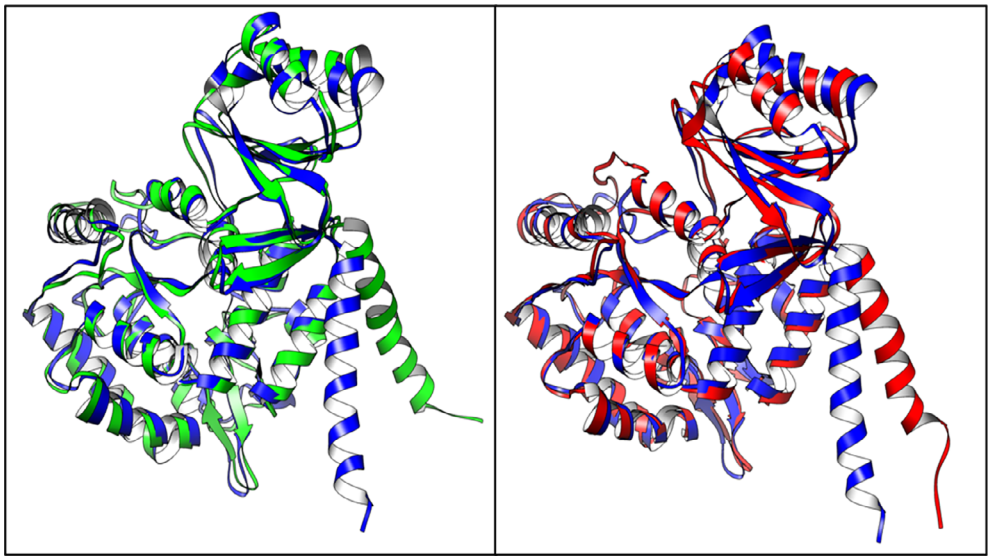

FIGURE 3 Overlap image of wildtype (blue), mutant I65N (green) and mutant L311P (red) compared at the last frame of the simulation [Color figure can be viewed at wileyonlinelibrary.com]

I65N mutant had the largest motion during the course of the simulation along the first eigenvector, while the L311P had a larger range of motion along both the first and second eigenvectors compared to the wildtype PAH protein.

To further understand the effects of the L311P mutation on the stability of the PAH protein, an intramolecular hydrogen bond analysis was conducted between Leu-311 and all the residues in the wildtype protein, as well as between the mutant, Pro-311, and all the residues

from the mutant structure, L311P. The amino acid, Leu, from the wildtype formed hydrogen bonds between ARG-408^{NH1}, LEU-308^{O}, LEU-308^{HA}, and GLY-307^{O} with frequencies of 98, 42, 40, and 32%, respectively. However, the mutated L311P protein demonstrates an overall decrease in both the number of hydrogen bonds as well as its occupancies (Table 3). Pro-311 in the mutated structure mainly formed hydrogen bonds with ARG-408 and LEU-308 with 97% and 25% occupancies respectively. This indicates that the L311P mutation is likely to lead to instability in the PAH protein, as represented in the RMSD, RMSF, radius of gyration analysis and principal component analysis.

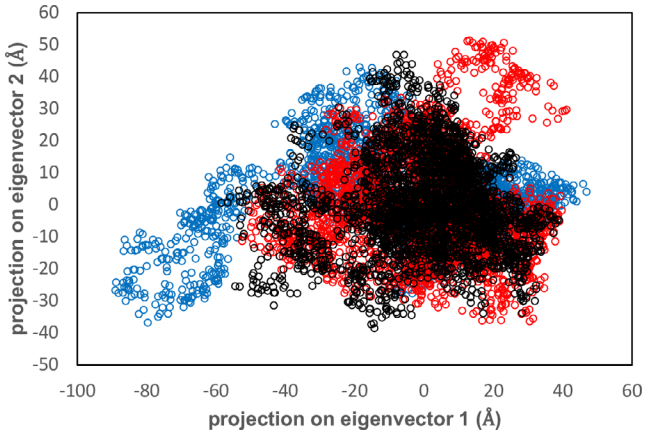

FIGURE 4 Projection of motion along first two principal eigenvectors for Cα atoms of phenylalanine hydroxylase (PAH) proteins. Wildtype shown in black, I65N mutant shown in blue, and L311P mutant shown in red [Color figure can be viewed at wileyonlinelibrary.com]

On the contrary, the 65th residue of the wildtype protein, ILE, forms hydrogen bonds with LEU-65^{O} and PHE-80^{O} at frequencies of 49% and 24%, respectively. After the mutation to ASN, more hydrogen bonds are formed with residues such THR-81^{HG1} and HSD-82^{HN} at frequencies of 53% and 30%, respectively (Table 3).

The intramolecular bonds were then further analyzed by creating a distribution of the relative frequency of the number of hydrogen bonds between all the residues and wildtype/mutant residues (Figure 5). The native structure, before the mutation, had a greater relative frequency of intramolecular hydrogen bonds between the Ile-65 and the rest of the residues in the protein, compared to after the mutation to Asn-65. Ile-65 is found to form more one or two hydrogen bonds than Asn-65 with the other resides of protein (Figure 5A). Furthermore, the number of hydrogen bonds for Leu-311 in wildtype PAH is compared with those of Pro-311 in L311P. The Leu-311 forms more hydrogen compared to those of Pro-311 with the rest of the residues in PAH, as shown in Figure 5B. The decrease in hydrogen bonds in L311P indicates a loss of protein stability when leucine is replaced with proline in the 311th residue of PAH.

In addition to the intramolecular hydrogen bond analysis, an intermolecular hydrogen bond analysis between the Leu-311 wildtype and water molecules, as well as between the mutated amino acid, Pro-311, and the solvent was conducted. The intermolecular hydrogen bond analysis was performed to further understand the interactions of the solvent molecules with the wildtype and L311P mutant (Table 4). The wildtype amino acid, Leu-311, formed only two hydrogen bonds with the solvent molecules. In contrast, the mutated amino acid, Pro-311, formed six hydrogen bonds between PRO-311^{CA}/WAT-11774^{OH2}, PRO-311^{CA}/WAT-2643^{OH2}, PRO-311^{CA}/WAT-19666^{OH2},

TABLE 3 Average lengths (r), angles (θ), and occupancy (f) of intramolecular hydrogen bonds between all residues in wildtype PAH and Leu-311 and Ile-65, as well as between all residues in L311P and I65N mutant and their respective substitution from the MD simulation

Wildtype with 311th residue				Wildtype with 65th residue			
Donor/acceptor pair	***f* (%)**	**r (Å)**	**θ (deg)**	**Donor/acceptor pair**	***f* (%)**	**r (Å)**	**θ (deg)**
ARG-408^{NH1}/LEU-311^{O}	98%	2.81	159.76	ILE-65^{HN}/LEU-65^{O}	49%	3.23	157.65
LEU-311^{N}/LEU-308^{O}	42%	3.31	151.29	ILE-65^{HA}/PHE-80^{O}	24%	3.39	146.62
LEU-311^{CG}/LEU-308^{HA}	40%	3.26	146.35				
LEU-311^{CG}/GLY-307^{O}	32%	3.39	149.00				
LEU-308^{CA}/LEU-311^{HG}	31%	3.22	141.07				
LEU-311^{CD2}/PHE-121^{HB2}	21%	3.20	151.80				
LEU-311^{N}/GLY-307^{O}	17%	3.31	144.95				
PHE-121^{CB}/LEU-311^{HD21}	15%	3.21	150.55				
L311P Mutant				**I65N Mutant**			
Donor/acceptor pair	***f* (%)**	**r (Å)**	**θ (deg)**	**Donor/acceptor pair**	***f* (%)**	**r (Å)**	**θ (deg)**
ARG-408^{NH1}/PRO-311^{O}	97%	2.88	161.22	THR-81^{HG1}/ASN-65^{OD1}	53%	2.75	162.39
PRO-311^{CD}/LEU-308^{O}	25%	3.40	150.32	HSD-82^{HN}/ASN-65^{OD1}	30%	3.00	154.40
VAL-412^{CG2}/PRO-311^{HB2}	17%	3.23	149.98	THR-81^{HA}/ ASN-65^{HA}	18%	3.35	150.71
PRO-311^{CA}/PHE-121^{HB1}	16%	3.35	150.32	THR-81^{HA}/ ASN-65^{OD1}	17%	3.31	143.63
VAL-412^{CG2}/PRO-311^{HB2}	16%	3.24	150.44				

Note: Only occupancies (f) up to 15% are reported. 3.5 Å hydrogen bond distance is given with respect to the donor heavy atoms.

PRO-311^{CA}/WAT-2643^{H2}, PRO-311^{CA}/WAT-11774^{H2}, and PRO-311^{CA}/WAT-11774^{H1} with frequencies of 8, 6, 5, 3, 3, and 2%, respectively (Table 4). The mutant Pro-311 may demonstrate greater structural flexibility as the number of hydrogen bonds increased with the solvent and decrease with the other residues in protein, in comparison to the wildtype, Leu-311.

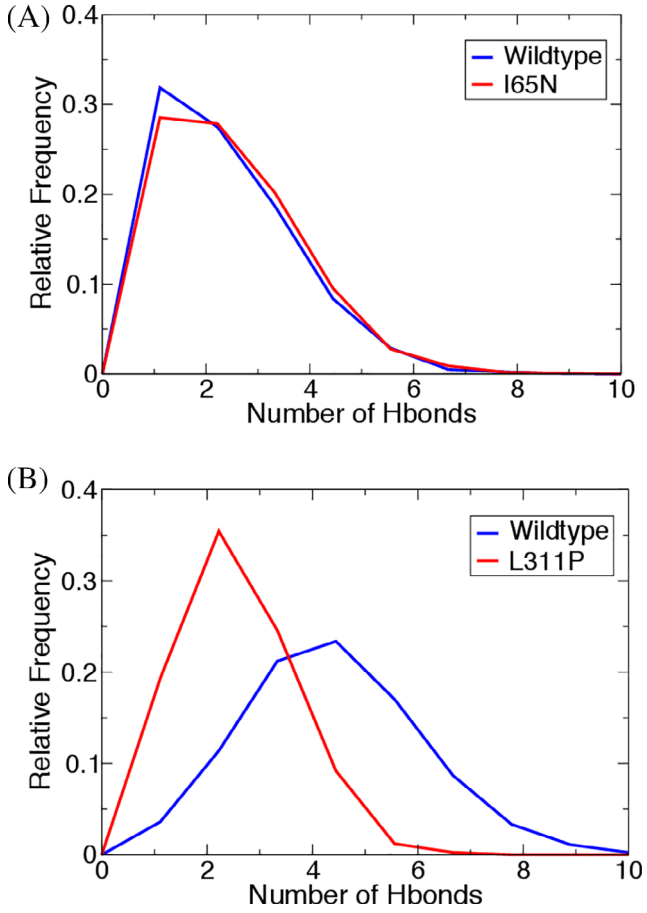

FIGURE 5 Frequency of the number of hydrogen bonds, A, between Ile-65 and all residues in wildtype phenylalanine hydroxylase (PAH) and between Asn-65 and all residues in I65N as well as, B, between Leu-311 and all residues in wildtype PAH and between Pro-311 and all residues in L311P [Color figure can be viewed at wileyonlinelibrary.com]

An intermolecular hydrogen bond analysis was also performed between the Ile-65 residue and water molecules, as well as between the mutant, Asn-65, and solvent. The wildtype amino acid, Ile-65, formed only two hydrogen bonds at an occupancy of 1%. In comparison, the mutant residue, Asn-65, formed six hydrogen bonds between, WAT-4095^{OH2}/ASN-65^{OD1}, WAT-16618^{OH2}/ASN-65^{OD1}, WAT-6806^{OH2}/ASN-65^{OD1}, WAT-8483^{OH2}/ASN-65^{OD1}, WAT-18494^{OH2}/ASN-65^{OD1}, and WAT-6806^{OH2}/ASN-65^{OD1} with occupancies of 10, 5, 2, 2, 2, and 2% respectively (Table 4). Both mutants displayed greater structural flexibility compared to the wildtype PAH protein.

3.4 | Free energy calculations

A synthetic form of BH_4, Kuvan, is a pharmacological agent that can lower phenylalanine levels in the blood and is currently the only drug-based therapy for a subset of phenylketonuria patients.[5] The synthetic formulation of BH_4 works by partially acting as a pharmacological chaperone to help misfolded mutant proteins fold correctly.[5] Here, free energy calculations using the MM/GBSA approach were conducted between BH_4 and native PAH as well as BH_4 and the two mutants, I65N and L311P, in order to understand how the binding affinity may differ between BH_4 with native PAH and the mutants. Interestingly, the calculated binding affinity between BH_4 and the wildtype was −6.45 kcal/mol, which is less than the binding affinity for the I65N and L311P mutants at −10.03 and − 8.70 kcal/mol, respectively (Table 5).

For all systems, the contribution of Van der Waals forces (ΔE_{VDW}), electrostatic energies (ΔE_{ELE}), and non-polar solvation energy (ΔG_{surf}) was favorable, while the polar solvation energy (ΔG_{GB}) calculated from the Generalized born model was unfavorable (Table 5). The increase in binding affinity for the mutants is mostly

TABLE 4 Average lengths (r), angles (θ), and occupancy (f) of intermolecular hydrogen bonds of solvent water molecules with wildtype of the 65th and 311th residue, with each mutant from the MD simulation

Wildtype with 311th residue				Wildtype with 65th residue			
Donor/acceptor pair	*f* (%)	r (Å)	θ (deg)	Donor/acceptor pair	*f* (%)	r (Å)	θ (deg)
LEU-311^{CA}/WAT-13814^{OH2}	2%	3.41	148.72	WAT-6362^{OH2}/ILE-65^{O}	1%	2.98	152.87
LEU-311^{CD2}/WAT-13814^{H1}	2%	3.31	149.78	WAT-811^{OH2}/ILE-65^{O}	1%	3.00	153.81
L311P Mutant				**I65N Mutant**			
Donor/acceptor pair	***f* (%)**	**r (Å)**	**θ (deg)**	**Donor/acceptor pair**	***f* (%)**	**r (Å)**	**θ (deg)**
PRO-311^{CA}/WAT-11774^{OH2}	8%	3.40	149.09	WAT-4095^{OH2}/ASN-65^{OD1}	10%	2.96	153.44
PRO-311^{CA}/WAT-2643^{OH2}	6%	3.42	148.65	WAT-16618^{OH2}/ASN-65^{OD1}	5%	2.94	153.50
PRO-311^{CA}/WAT-19666^{OH2}	5%	3.40	149.28	WAT-6806^{OH2}/ASN-65^{OD1}	2%	2.91	155.15
PRO-311^{CA}/WAT-2643^{H2}	3%	3.38	154.78	WAT-8483^{OH2}/ASN-65^{OD1}	2%	2.93	151.24
PRO-311^{CA}/WAT-11774^{H2}	3%	3.36	156.18	WAT-18494^{OH2}/ASN-65^{OD1}	2%	2.97	151.23
PRO-311^{CA}/WAT-11774^{H1}	2%	3.33	157.55	WAT-6806^{OH2}/ASN-65^{OD1}	2%	2.93	153.57

Note: Only occupancies (*f*) up to 2% are reported for mutant. 3.5 Å hydrogen bond distance is given with respect to the donor heavy atoms.

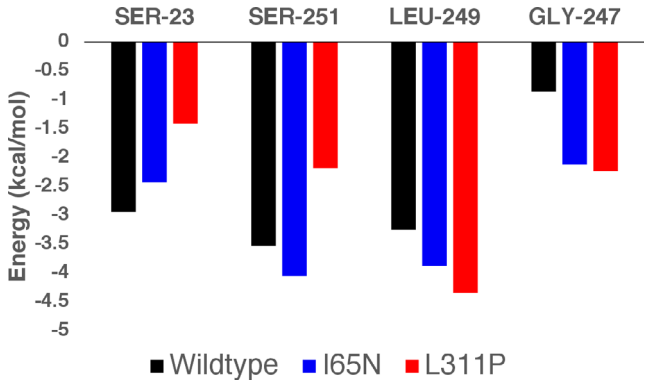

FIGURE 6 Pair-wise energy decomposition analysis of residues in the binding site that make hydrogen bonds with BH_4 ligand [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Binding free energies of BH_4 bound to wildtype, I65N, and L311P mutant systems

System	ΔE_{VDW}	ΔE_{ELE}	ΔG_{GB}	ΔG_{surf}	ΔG bind (kcal/mol)
BH_4-wildtype	−23.52	−5.18	25.91	−3.66	−6.45 ± 3.68
BH_4- I65N mutant	−26.19	−12.03	32.58	−4.39	−10.03 ± 2.97
BH_4- L311P mutant	−22.20	−9.62	26.81	−3.69	−8.70 ± 3.62

Note: All units are in kcal/mol.

due to the increase in ΔE_{ELE}, and ΔG_{surf}, as well as an increase in the ΔE_{VDW} for the I65N mutant.

The largest increase in binding affinity was observed for the I65N mutant. Patients with a very similar mutation, I65T, have previously shown positive BH_4 response.[62] Since BH_4 binds with the I65N mutant at a higher affinity than the wildtype, this may help to explain why patients with the similar I65T mutant responded to treatment with BH_4. Moreover, this may also mean that patients with the I65N mutation might respond similarly to treatment using BH_4.

In the binding pocket of the PAH protein, Ser-23, Ser-251, Leu-249, and Gly-247 are all very important residues since they make hydrogen bonds with the BH_4 ligand (Figure 1C). To assess how the interaction energy between these residues and BH_4 changes after mutation, a pairwise decomposition analysis was performed from the MM/GBSA calculations (Figure 6).

In all systems, Ser-251 and Leu-249 contribute substantially to the total binding affinity (Figure 6). In particular, there is an increase in the interaction energy between Leu-249 and Gly-247 with BH_4 for both the I65N and L311P mutants. Therefore, these residues may be very important in the rational drug design for other PAH pharmacological agents. Only with Ser-23 was the interaction energy with BH_4 larger for the wildtype than the mutants. These results reveal that the increase in the binding affinity for the mutant PAH proteins is largely due to Leu-249 and Gly-247 residues. Moreover, Ser-251 also contributes greatly to the I65N mutant having the largest binding affinity with BH_4.

Free energy calculations, using MM/GBSA, were also carried out using the PDB 5FII which contains the homodimer structure of the

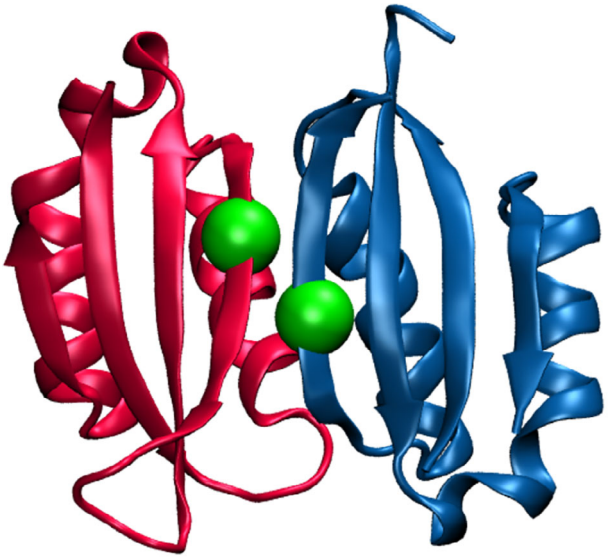

FIGURE 7 Phenylalanine hydroxylase (PAH) N-terminal regulatory domain homodimer showing the site for I65N mutation (green) [Color figure can be viewed at wileyonlinelibrary.com]

PAH N-terminal regulatory domains (RD). Wildtype PAH undergoes dimerization due to intermolecular force at the RD of two PAH proteins. A mutation in the RD domain may disrupt or weaken the dimerization process. The goal was to elicit how the I65N mutation that results from an amino acid change from isoleucine to asparagine may impact the ability of the N-terminal regulatory domain monomer to bind to the other identical monomer that leads to homodimer formation.

Disease causing mutations such as G46S, T63P, I65S, and E76A have previously shown to affect the structural stability of the N-terminal regulatory domain impacting dimerization.[7] Here, the free energy of binding between the two wildtype N-terminal regulatory domain monomers was calculated at −119.58 ± 13.18 kcal/mol (Figure 7). After the I65N mutation was introduced to both monomers, the binding free energy dropped substantially to −24.06 ± 10.65 kcal/mol (Figure 7). The lower free energy for the I65N mutant agrees with literature[7] that also finds that mutations in the N-terminal regulatory domain lead to instability that likely results in decreased dimerization ability. However, future work is still garnered to better understand how the dimerization of the PAH N-terminal domain activates enzyme activity and the mechanism in which mutations may impact dimerization that might lead to lower enzyme activity and subsequently PKU disease.

4 | CONCLUSION

Computational prediction of the effects of nsSNPs can be imperative to further understand how missense mutations may impact protein

function and stability. In this study, we prioritized the most deleterious and disease-causing mutations, I65N and L311P, out of 50 nsSNPs from the UniProt database. In order to examine the impact of the nsSNPs, we carried out an in-silico approach via molecular dynamics to further understand the consequences of these mutations on the structure and dynamics of these mutant proteins. It was observed, through RMSF, radius of gyration, eigenvector value calculations that there was a loss of stability in these proteins. This observation was further supported after calculating the number of hydrogen bonds which were less for the mutants compared to the wildtype and indicated that the mutants are more flexible. Moreover, MM/GBSA free energy calculations reveal that BH_4, a drug-based therapy for PKU patients, had a higher binding affinity for I65N and L311P mutants compared to the wildtype protein. In addition, important residues are identified in the BH_4 binding pocket that may be of interest for the rational drug design of other PAH drug-based therapies. Free energy calculations also revealed that the I65N mutation may impair the dimerization of the N-terminal regulatory domain. Our computational investigation provides great insight into two probable point mutations that can lead to PKU. This study may also be used to improve treatment and early detection methods in efforts to prevent children from developing severe symptoms of PKU.

ACKNOWLEDGMENTS

The authors acknowledge the Advanced Cyberinfrastructure for Education and Research (ACER) at The University of Illinois at Chicago that have contributed to the research results reported within this paper. We are also grateful for Latinos Gaining Access to Networks for Advancement in Science (Las GANAS) at the University of Illinois at Chicago for financial support.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/prot.26051.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Shahidul M. Islam https://orcid.org/0000-0001-5769-6844

REFERENCES

1. Mitchell JJ, Trakadis YJ, Scriver CR. Phenylalanine hydroxylase deficiency. *Genet Med*. 2011;13:697-707. https://doi.org/10.1097/GIM.0b013e3182141b48.
2. Blau N, Van Spronsen FJ, Levy HL. Phenylketonuria. *Lancet*. 2010;376(9750):1417-1427. https://doi.org/10.1016/S0140-6736(10)60961-0.
3. Sumaily K, Mujamammi A. Phenylketonuria: a new look at an old topic, advances in laboratory diagnosis, and therapeutic strategies. *Int J Health Sci*. 2017;11(5):63-70.
4. Strisciuglio P, Concolino D. New strategies for the treatment of phenylketonuria (PKU). *Metabolites*. 2014;4:1007-1017. https://doi.org/10.3390/metabo4041007.
5. Flydal MI, Alcorlo-Pagés M, Johannessen FG, et al. Structure of full-length human phenylalanine hydroxylase in complex with tetrahydrobiopterin. *Proc Natl Acad Sci U S A*. 2019;116:11229-11234. https://doi.org/10.1073/pnas.1902639116.
6. Seeliger D, De Groot BL. Conformational transitions upon ligand binding: Holo-structure prediction from apo conformations. *PLoS Comput Biol*. 2010;6(1):e1000634. https://doi.org/10.1371/journal.pcbi.1000634.
7. Patel D, Kopec J, Fitzpatrick F, Mccorvie TJ, Yue WW. Structural basis for ligand-dependent dimerization of phenylalanine hydroxylase regulatory domain. *Sci Rep*. 2016;6:1-10. https://doi.org/10.1038/srep23748.
8. McGee MM, Greengard O, Knox WE. Liver phenylalanine hydroxylase activity in relation to blood concentrations of tyrosine and phenylalanine in the rat. *Biochem J*. 1972;127:675-680. https://doi.org/10.1042/bj1270675.
9. Lee YW, Dong HL, Kim ND, et al. Mutation analysis of PAH gene and characterization of a recurrent deletion mutation in Korean patients with phenylketonuria. *Exp Mol Med*. 2008;40(5):533-540. https://doi.org/10.3858/emm.2008.40.5.533.
10. Yates CM, Sternberg MJE. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. *J Mol Biol*. 2013;425(21):3949-3963. https://doi.org/10.1016/j.jmb.2013.07.012.
11. Guldberg P, Levy HL, Hanley WB, et al. Phenylalanine hydroxylase gene mutations in the United States: report from the maternal PKU collaborative study. *Am J Hum Genet*. 1996;59(1):84-94.
12. Song F, Qu YJ, Zhang T, Jin YW, Wang H, Zheng XY. Phenylketonuria mutations in northern China. *Mol Genet Metab*. 2005;86:107-118. https://doi.org/10.1016/j.ymgme.2005.09.001.
13. Van Wegberg AMJ, MacDonald A, Ahring K, et al. The complete European guidelines on phenylketonuria: diagnosis and treatment. *Orphanet J Rare Dis*. 2017;12:162. https://doi.org/10.1186/s13023-017-0685-2.
14. Kumar A, Purohit R. Computational investigation of pathogenic nsSNPs in CEP63 protein. *Gene*. 2012;503(1):75-82. https://doi.org/10.1016/j.gene.2012.04.032.
15. Singh PK, Mistry KN. A computational approach to determine susceptibility to cancer by evaluating the deleterious effect of nsSNP in XRCC1 gene on binding interaction of XRCC1 protein with ligase III. *Gene*. 2016;580(2):141-149. https://doi.org/10.1016/j.gene.2015.09.084.
16. Al Mehdi K, Fouad B, Zouhair E, et al. Molecular Modelling and dynamics study of nsSNP in STXBP1 gene in early infantile epileptic encephalopathy disease. *Biomed Res Int*. 2019;2019(14):1-14. https://doi.org/10.1155/2019/4872101.
17. Loong CKP, Zhou HX, Bryant Chase P. Familial hypertrophic cardiomyopathy related E180G mutation increases flexibility of human cardiac α-tropomyosin. *FEBS Lett*. 2012;586:3503-3507. https://doi.org/10.1016/j.febslet.2012.08.005.
18. Verma D, Jacobs DJ, Livesay DR. Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. *PLoS Comput Biol*. 2012;8(3):e1002409. https://doi.org/10.1371/journal.pcbi.1002409.
19. Sharma M, Kumar R, Singh R, Kaur J. Thirty-degree shift in optimum temperature of a thermophilic lipase by a single-point mutation: effect of serine to threonine mutation on structural flexibility. *Mol Cell Biochem*. 2017;430:21-30. https://doi.org/10.1007/s11010-017-2950-z.
20. Flydal MI, Martinez A. Phenylalanine hydroxylase: function, structure, and regulation. *IUBMB Life*. 2013;65(4):341-349. https://doi.org/10.1002/iub.1150.
21. Waters PJ. How PAH gene mutations cause hyperphenylalaninemia and why mechanism matters: insights from in vitro expression. *Hum Mutat*. 2003;21:357-369. https://doi.org/10.1002/humu.10197.

22. Réblová K, Kulhánek P, Fajkusová L. Computational study of missense mutations in phenylalanine hydroxylase. *J Mol Model*. 2015;21:70. https://doi.org/10.1007/s00894-015-2620-6.
23. Burnett JR. Sapropterin dihydrochloride (Kuvan/Phenoptin), an orally active synthetic form of BH4 for the treatment of phenylketonuria. *IDrugs*. 2007;11:805-813.
24. Gordon P, Thomas JA, Suter R, Jurecki E. Evolving patient selection and clinical benefit criteria for sapropterin dihydrochloride (Kuvan®) treatment of PKU patients. *Mol Genet Metab*. 2012;105:672-676. https://doi.org/10.1016/j.ymgme.2011.12.023.
25. Hofto ME, Cross JN, Cafiero M. Interaction energies between tetrahydrobiopterin analogues and aromatic residues in tyrosine hydroxylase and phenylalanine hydroxylase. *J Phys Chem B*. 2007;111:9651-9654. https://doi.org/10.1021/jp072518w.
26. Erlandsen H, Stevens RC. A structural hypothesis for BH4 responsiveness in patients with mild forms of hyperphenylalaninaemia and phenylketonuria. *J Inherit Metab Dis*. 2001;24(2):213-230. https://doi.org/10.1023/A:1010371002631.
27. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. https://doi.org/10.1093/nar/gkg509.
28. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11:1-9. https://doi.org/10.1038/nprot.2015.123.
29. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;76(1):7.20.1-7.20.41. https://doi.org/10.1002/0471142905.hg0720s76.
30. Galehdari H, Saki N, Mohammadi-asl J, Rahim F. Meta-analysis diagnostic accuracy of SNP-based pathogenicity detection tools: a case of UTG1A1 gene mutations. *Int J Mol Epidemiol Genet*. 2013;44(2):77-85.
31. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009;25:2744-2750. https://doi.org/10.1093/bioinformatics/btp528.
32. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;48(8):1073-1082. https://doi.org/10.1038/nprot.2009.86.
33. Sheynkman GM, Shortreed MR, Frey BL, Scalf M, Smith LM. Large-scale mass spectrometric detection of variant peptides resulting from nonsynonymous nucleotide differences. *J Proteome Res*. 2014;13:228-240. https://doi.org/10.1021/pr4009207.
34. Venter M, Malan L, van Dyk E, Elson JL, van der Westhuizen FH. Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: the SABPA study. *J Genet Genomics*. 2017;44(3):139-149. https://doi.org/10.1016/j.jgg.2016.12.003.
35. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28(1):235-242. https://doi.org/10.1093/nar/28.1.235.
36. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(1):D158-D169. https://doi.org/10.1093/nar/gkw1099.
37. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(W1):W452-W457. https://doi.org/10.1093/nar/gks539.
38. Altschul SF, Koonin EV. Iterated profile searches with PSI-BLAST - a tool for discovery in protein databases. *Trends Biochem Sci*. 1998;23(11):444-447. https://doi.org/10.1016/S0968-0004(98)01298-5.
39. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat*. 2011;32(6):661-668. https://doi.org/10.1002/humu.21490.
40. Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res*. 2017;45(1):W247-W252. https://doi.org/10.1093/nar/gkx369.
41. Pejaver V, Urresti J, Lugo-Martinez J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*. 2020;11:1-13. https://doi.org/10.1038/s41467-020-19669-x.
42. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26:1781-1802. https://doi.org/10.1002/jcc.20289.
43. Lee J, Cheng X, Swails JM, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput*. 2016;12(1):405-413. https://doi.org/10.1021/acs.jctc.5b00935.
44. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical UserInterface for CHARMM. *J Comput Chem*. 2008;29:1859-1865.
45. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2016;14:71-73. https://doi.org/10.1038/nmeth.4067.
46. Wang H, Dommert F, Holm C. Optimizing working parameters of the smooth particle mesh Ewald algorithm in terms of accuracy and efficiency. *J Chem Phys*. 2010;133:1-13. https://doi.org/10.1063/1.3446812.
47. Wang X, Chen Y, Deng W. Lévy-walk-like Langevin dynamics. *New J Phys*. 2019;29:1-20. https://doi.org/10.1088/1367-2630/aaf764.
48. Elber R, Ruymgaart AP, Hess B. SHAKE parallelization. *Eur Phys J Spec Top*. 2011;200:211-223. https://doi.org/10.1140/epjst/e2011-01525-9.
49. Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*. 1992;13(8):952-962. https://doi.org/10.1002/jcc.540130805.
50. Kim S, Lee J, Jo S, Brooks CL, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem*. 2017;38(21):1879-1886. https://doi.org/10.1002/jcc.24829.
51. Le Grand S, Götz AW, Walker RCSPFP. Speed without compromise - a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun*. 2013;184:374-380. https://doi.org/10.1016/j.cpc.2012.09.022.
52. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T. E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T DMY and PA Kollman. AMBER 2020. University of California, San Fransisco 2020.
53. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.Py: an efficient program for end-state free energy calculations. *J Chem Theory Comput*. 2012;8(9):3314-3321. https://doi.org/10.1021/ct300418h.
54. Wang E, Weng G, Sun H, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein interactions. *Phys Chem Chem Phys*. 2019;21(35):18958-18969. https://doi.org/10.1039/c9cp04096j.
55. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*. 1999;20(2):217-230.
56. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. https://doi.org/10.1038/gim.2015.30.

57. Ribeiro AJM, Tyzack JD, Borkakoti N, Holliday GL, Thornton JM. A global analysis of function and conservation of catalytic residues in enzymes. *J Biol Chem*. 2020;295(2):314-324. https://doi.org/10.1074/jbc.REV119.006289.
58. Bosco P, Calì F, Meli C, et al. Eight new mutations of the phenylalanine hydroxylase gene in Italian patients with hyperphenylalaninemeia. *Hum Mutat*. 1998;11(3):240-243. https://doi.org/10.1002/(SICI)1098-1004(1998)11:3<240::AID-HUMU9>3.0.CO;2-L.
59. Lichter-konecki U, Konecki DS, DiLella AG, et al. Phenylalanine hydroxylase deficiency caused by a Single Base substitution in an exon of the human phenylalanine hydroxylase gene. *Biochemistry*. 1988;27:2881-2885. https://doi.org/10.1021/bi00408a032.
60. Shi Z, Sellers J, Moult J. Protein stability and in vivo concentration of missense mutations in phenylalanine hydroxylase. *Proteins Struct Funct Bioinformartics*. 2012;80(1):61-70. https://doi.org/10.1002/prot.23159.
61. Erlandsen H, Stevens RC. The structural basis of phenylketonuria. *Mol Genet Metab*. 1999;68(2):103-125. https://doi.org/10.1006/mgme.1999.2922.
62. Desviat LR, Pérez B, Bèlanger-Quintana A, et al. Tetrahydrobiopterin responsiveness: results of the BH 4 loading test in 31 Spanish PKU patients and correlation with their genotype. *Mol Genet Metab*. 2004;83:157-162. https://doi.org/10.1016/j.ymgme.2004.06.007.

How to cite this article: Lopez A, Havranek B, Papadantonakis GA, Islam SM. In silico screening and molecular dynamics simulation of deleterious PAH mutations responsible for phenylketonuria genetic disorder. *Proteins*. 2021;89:683–696. https://doi.org/10.1002/prot.26051