Published in final edited form as:
Proteins. 2012 January ; 80(1): 61–70. doi:10.1002/prot.23159.

Protein Stability and *in Vivo* Concentration of Missense Mutations in Phenylalanine Hydroxylase

Zhen Shi[1,2], **Jenn Sellers**[4], and **John Moult**[1,3,*]

[1]Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850

[2]Molecular and Cell Biology Graduate Program, University of Maryland, College Park, MD 20742

[3]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742

[4]Food and Drug Administration (FDA), 10903 New Hampshire Ave., Silver Spring, MD 20993

Abstract

A previous computational analysis of missense mutations linked to monogenic disease found a high proportion of missense mutations affect protein stability, rather than other aspects of protein structure and function. The purpose of the present study is to relate the presence of such stability damaging missense mutations to the levels of a particular protein present under '*in vivo*' like conditions, and to test the reliability of the computational methods. Experimental data on a set of missense mutations of the enzyme phenylalanine hydroxylase (PAH) associated with the monogenic disease phenylketonuria (PKU) have been compared with the expected *in vivo* impact on protein function, obtained using SNPs3D, an *in silico* analysis package. A high proportion of the PAH mutations are predicted to be destabilizing. The overall agreement between predicted stability impact and experimental evidence for lower protein levels is in accordance with the estimated error rates of the methods. For these mutations, destabilization of protein three dimensional structure is the major molecular mechanism leading to PKU, and results in a substantial reduction of *in vivo* PAH protein concentration. Although of limited scale, the results support the view that destabilization is the most common mechanism by which missense mutations cause monogenic disease. In turn, this conclusion suggests the general therapeutic strategy of developing drugs targeted at restoring wild type stability.

Keywords

missense mutation; machine learning; support vector machine; protein structure; protein stability; phenylketonuria (PKU)

[*]Corresponding author: jmoult@umd.edu, Phone: (240) 314-6241, FAX: (240) 314-6255.

Introduction

In a previous study of monogenic disease-associated missense mutations, we found that many are predicted to reduce protein stability.[1,2] These results are surprising since the associated reduction in stability of approximately 1∼3 Kcal/mol would not be expected to affect protein function significantly *in vitro*, given a typical free energy of stabilization of the folded state of the order of 10 Kcal/mol.[3] There are few experimental data on the relationship between disease missense mutations and protein properties under *in vivo* like conditions. One excellent source of information is for missense mutations found in the human hepatic enzyme phenylalanine hydroxylase (PAH) (EC 1.14.16.1), associated with phenylketonuria (PKU, OMIM 261600).

PKU is an autosomal recessive inherited disorder and the most common inborn error of amino acid metabolism, with average birth incidence of about 1 in 10,000 among European descent and Asian populations.[4-7] The conversion of dietary L-Phe to L-Tyr is catalyzed by PAH. The enzyme is the major means of degrading dietary L-Phe and the rate-limiting step controlling the catabolism of L-Phe.[8] Deficiency in PAH enzyme activity results in elevated phenylalanine concentration in the body and abnormally high levels of metabolites from phenylalanine by other metabolic pathways. L-Tyr is the substrate for the biosynthesis of the thyroid hormone thyroxine, the neurotransmitter dopamine, the adrenal hormones, and the pigment molecule melanin.[9] Lack of L-Tyr and excess of L-Phe, which acts as an antagonist to L-Tyr, leads to various clinical manifestations such as mental retardation and decreased pigmentation. Clinically, patients are assigned to one of four phenotype categories based on a continuum of blood phenylalanine level and dietary phenylalanine tolerance. The most severe is “classic PKU”, followed by “moderate PKU”, “mild PKU”, and the least severe, “mild hyperphenylalaninemia”(MHP) (summaried in Guldberg et al. [10]). More than 500 naturally occurring DNA mutations which affect the function of human PAH *in vivo* have been identified and archived in the PAH Mutation Analysis Consortium database (PAHdb [11], www.pahdb.mcgill.ca). About sixty percent of these are missense mutations arising from single base changes.[12] Homozygous or compound heterozygous genotypes of these missense mutations generally result in PKU.

The effects of a subset of PKU-associated PAH missense mutations have been studied in cultured cells and cell lysate extract, representing *in vivo* like conditions. Data on these are available through the PAHdb.[11] In these experiments, the mutant and wild type PAH cDNA constructs were transiently transfected and expressed in the host cells. The total enzyme activity, the PAH immune-reactive protein level, and sometimes the mRNA level were measured. These data provide a basis for testing the relationship between destabilization of protein structure and protein *in vivo* activity.

Crystal structures of PAH have shown that the human enzyme is a homo-tetramer.[13] Each chain has an N-terminal regulatory domain (residues 1-110), a catalytic domain containing an iron atom (residues 111-410) and a tetramerization domain (residues 411-452) (Fig. 1). The substrate L-Phe and cofactor tetrahydrobiopterin (BH4) both have binding sites in the catalytic domain. The availability of the crystal structures of PAH makes it possible to model missense mutations and their effects on protein structure and molecular function (see

Methods). An extensive review of the location of disease-associated missense mutations in the structure has been published.[14] Here we focus on relating predictions of lower protein stability to protein characteristics under *in vivo* conditions, and testing the computational assignments against the experimental data.

A number of computational methods have been developed to identify which missense base substitutions have a high impact on *in vi*vo protein function. These methods are based sequence conservation patterns,[15,16] features of protein three dimensional structure,[1,17] or a combination of both.[18-21] A variety of machine learning [1,16,18-20] and statistical [15,21] approaches are employed together with appropriate training data to utilize the sequence and structure features. We have developed a method that identifies substantial changes in the thermodynamic stability of a protein structure, based on the detailed structural environment of a mutation.[1] The method uses a Support Vector Machine (SVM[22]), trained on data for mutations that are considered to cause monogenic disease, taken from the Human Gene Mutation database (HGMD [23], www.hgmd.cf.ac.uk, as of 02/09/2002, (later versions of the database include many non-causative mutations)) and a control set of amino acid differences between corresponding mouse and human orthologs. Each mutation is characterized by 15 features, including perturbation of electrostatic factors, packing efficiency, steric clashes, breakage of disulfide bonds, polypeptide backbone strain, and the relative extent of local structural rigidity. Full details of the method and its benchmarking have been previously published.[1] In jack-knifed testing the SVM assigns 74% of the HGMD monogenic disease mutations as destabilizing. Comparison with experimental data for a set of site directed mutations in bacterial and phage proteins established that destabilizing monogenic disease mutations typically reduce the free energy difference between the folded and unfolded state of a protein by 2 to 3 Kcal/mol.[1] . We have also developed a support vector machine utilizing sequence conservation features to detect those mutations that have a high impact on any aspect of the protein function, not just destabilization.[16] These two support vector machines are implemented in a web interface and database infrastructure, SNPs3D (www.SNPs3D.org), which contains an analysis of human SNPs using the two methods. Blind testing against experimental data on a small set of common non-synonymous SNPs produced a high level of agreement between predicted destabilization and lower melting temperature for the variant containing proteins.[24] The experimental properties of mutations in monogenic disease proteins such as PAH provide the most direct test of the earlier finding that destabilization of protein structure plays a major role in monogenic disease.

Destabilization of protein structure presumably reduces *in vivo* protein abundance, either through unsuccessful protein folding, or increased chaperone scavenging of transiently unfolded molecules. Destabilization alone is not expected to alter enzyme specific activity, but a destabilizing mutation may additionally impact molecular function, in ways that may be identified from the structural context. For example, the mutation lies in the ligand binding site. Other mutations may only impact molecular function, and not stability. On this basis, there are five categories of prediction from the computational methods that may be tested against the experimental data:

Category 1

Where a mutation is assigned as destabilizing, and is not directly involved in molecular function, we expect low *in vivo* protein abundance, and wild type specific activity.

Category 2

Where a mutation is assigned as destabilizing, and there is structural evidence of an impact on molecular function as well, we expect low *in vivo* protein abundance, and low specific activity.

Category 3

Where a mutation is not assigned as destabilizing, but is assigned as affecting molecular function, we expect wild type protein abundance, low specific activity, and evidence of involvement in function from the structure.

Category 4

Where a mutation is assigned as not destabilizing and as not affecting any aspect of function, we expect wild type protein abundance, wild type specific activity and a mild disease classification. Below, we consider each of these prediction categories and the extent to which these expectations are met.

Results

Impact analysis of missense mutations on PAH function and protein stability

46 distinct human PAH missense mutations with suitable experimental data were selected from the PAHdb database (version of January 2010). These had all been expressed in a mammalian COS or A293 cell expression system, with total enzyme activity and protein level measured for mutants and wild type under the same conditions. In most cases, mRNA levels are also available. There are multiple experimental results available for 16 of the mutations, providing an indication of experimental precision. Fig. 1 shows the distribution of the mutants in the PAH monomer structure. The two retrained SNPs3D methods were used to analyze each mutant. The results, together with the experimental data, are shown in Table 1. Of the 46 mutations, 35 (76%) are assigned as high impact on protein stability. The sequence conservation method assigns 42 as high impact from all causes. Only two mutations show no impact by either method. Eleven of the mutations are within 6.5 Å of substrate, cofactor, or Fe++ ion, and so expected to affect molecular function through altered ligand binding or catalysis.

Fig. 2 shows the distribution of experimental total protein activity (vertical axis) and protein level (horizontal axis). There are multiple experimental protein levels for three mutations, R158Q, R261Q, and Y414C, are inconsistent. 30 of the 35 mutants categorized as destabilizing by the SVM are consistent with aspects of the experimental data in accordance with expectations for category 1 and 2 mutations. Conversely, 8 of the 11 assigned as not destabilizing, show close to normal protein level (more than 50% of normal level), also consistent with expectations. Below, we examine the results in detail.

Category 1: 28 missense mutations are expected to affect stability only

28 of the 35 mutations with destabilization assignments are remote from any known ligand binding or the catalytic site, and so are expected to have a low experimental protein level, and wild type specific activity. 16 of the 28 (F39L, G46S, L48S, I65T, A104D, P122Q, R157N, F161S, R243Q, R252G, R252Q, R252W, A259T, A259V, L311P, R408W) have protein levels less than 50% wild type, as expected. Of these, all but two have wild type specific activity. The two exceptions, F39L and L48S, have approximately three fold higher specific activities than the wild type. These mutations lie in the regulatory domain, suggesting a possible explanation for the high activity level. The 16 mutants are classified into clinical categories of mild PKU (A104D), moderate PKU (F39L, L48S, I65T), and classic PKU (G46S, R243Q, R252G/Q/W, A259V, L311P, R408W).

Nine of remaining mutations expected to affect stability only (L41F, R68G, R68S, E76G, G218V, P244L, A309V, A403V, R408Q) have reported experimental protein levels greater than 50% of wild type (all 100%, except one of the R408Q experiments with 70%), inconsistent with the computational assignment. For five of these mutations, there is other experimental evidence supporting an impact on stability. R68S, P244L, A309V, and R408Q all exhibit BH4 responsiveness, that is, the disease phenotype is relieved by oral administration of BH4.[26] Additionally, in *in vitro* experiments, A309V (moderate or classic PKU) and R68S (mild PKU) have been shown to have longer protein half lives in the presence of BH4 than in its absence,.[27,28] and cellular studies of R408Q (MHP or mild PKU) show protein aggregation.[33] It has been suggested that BH4 acts as a chemical chaperone, facilitating correct folding.[28] The standard experimental BH4 (or analog) concentration[29] is 10 times higher than that of physiological conditions[30,31] and this difference has been demonstrated to result in significant variation in experimental results.[32] Thus, for these four mutations, the observed high experimental protein levels are consistent with masking of destabilization effects by the presence of excess BH4. For a fifth mutation, G218V (classic PKU), a large fraction of aggregates have also been reported.[33] A sixth mutation, R68G, appears from the structural context to be destabilizing, but no disease classification or additional experimental evidence is available in this case. The three remaining inconsistent mutations in this category are likely false positives of the computational method. Two (E76G (no disease classification) and A403V (mild PKU)) have low impact (i.e. inconsistent) assignments from the sequence conservation method. Visual inspection of the third, L41F (no disease classification), suggests it may not affect stability.

The final three mutations expected to impact stability only, R158Q, R261Q, Y414C, (classic, moderate, and mild PKU or MPH respectively) have inconsistent experimental results. At least one experiment is consistent with that assignment for each mutant, with less 50% of wild type protein level. Two of these R261Q and Y414C have short *in vitro* half lives, and clinical symptoms can be alleviated by BH4 supplement.[28]

Overall, the computational category assignment is consistent with at least some of the experimental evidence for 24 (16 with low protein level, five with high protein level but other experimental evidence for destabilization, and three agreeing with at least one experimental low protein level result) of the 28 in this category. One more, R68G, is likely

destabilizing, but requires additional experimental evidence.. The remaining three (G68G, E76G, and A403V) are likely false positive assignments of destabilization.

Category 2: Seven missense mutations are expected to affect both stability and molecular function

There are seven mutations (G247V, L255S, R270S, E280K, S349L, S349P and Y277D) with atomic contacts of 6.5 Å or less to the phenylalanine substrate, the BH4 cofactor or the Fe^{++} ion, and that are assigned as destabilizing by the structure SVM. These mutant proteins are therefore expected to exhibit a combination of lower specific activity and a lower total protein level. Six of the seven (G247V, L255S, R270S, E280K, S349L, S349P) have protein levels less than half or in one case close to half (G247V, 56%) that of wild type, and very low protein activity, consistent with expectations. Clinical categories are available for E280K, S349L, and S349P, and are all “classic PKU”, consistent with the results and with experiment.

The remaining mutant in this category, Y277D, has an experimental activity of zero, and is classified as mild or classic PKU, consistent with the profile SVM assignments. But the measured protein level is reported as 99% of wild type, inconsistent with a modest confidence stability assignment. This may be a computational false positive with respect to stability.

Category 3: Nine mutations are expected to impact molecular function only

A total of nine mutations are classified as high impact by the sequence conservation method, classified as not destabilizing by the stability method, and so are expected to impact molecular function but not stability, implying wild type protein levels and lower activity.

Four of these, L255V, P281L, A322G, and L348V have atomic contacts of 6.5 Å or less to a ligand. Experimental data for two, A322G and L348V, are consistent with expectations, with low activity and normal protein levels. The remaining two, L255V and P281L, have low activity, but also low protein level. Both are in direct contact with the BH4 cofactor, and would disrupt binding substantially. Experimental measurements for P281L show <1% or non-detectable for both total enzyme activity and protein level [34,35], and the mutant is classified as classic PKU. For L255V, two independent experimental measurements give <3% and 13% of wild type activity, and 8% and 18% of total protein [36,37]. For both mutants, it is probable that the low protein level is consequence of reduced protein stability, arising from reduced ability to bind BH4, rather than direct destabilization of the protein structure.

The other five mutations in this category, K42I, D59Y, D143G, V388M and R413P, are not near to any known ligand binding or catalytic site. Four are located on the protein surface. Two of these, D143G, and V388M, exhibit low total protein activity, and have near 100% wild type protein levels, consistent with the computational assignments. D143 is a conserved residue located on the dynamic loop (residue 136∼151) at the entrance to the active site, and is believed to play a role in the access of substrate and BH4 to the active site [13,14] and so the mutant likely affects catalytic efficacy. Although two independent reports give V388M wild type protein level, it has been demonstrated to affect tetramer formation, and co-expression

with additional GroESL chaperone partly overcomes that effect.[38] Also, it has a shorter *in vitro* half life, and patients respond to BH4 supplement,[27,28] all suggesting a destabilization effect. There are no inter-domain or inter-subunit contacts. It is possible that the larger exposed hydrophobic mutant side chain is responsible for a greater tendency to aggregate.

K42I has atomic contacts with a neighboring subunit, suggesting interference with tetramer formation, although this was not detected by the computational analysis, and the structural context does not appear destabilizing. D59Y has 92% of wild type activity, and is a MHP class mutant. The sequence conservation confidence score is low (-0.21), all suggesting this is a false positive. R413P has a low protein level and low activity, and inspection of the structure suggests it is likely destabilizing. Overall, for this category, there are two likely false negatives for stability impact (K42I and R413P), and one marginal false positive general high impact assignment (D59Y), and one unclear (V388M). The computational assignments for the other five are consistent with the experimental evidence.

Category 4: Two mutations are assigned low impact by both the sequence conservation and stability methods

Two mutations, T92I, and P211T, are assigned low impact by both computational methods. Both sets of experimental results show close to normal activity and protein levels, consistent with the analysis results. Also reasonably consistent, T92I is assigned to the mild MHP category of disease, suggesting a subtle effect on protein function. Inconsistent with both experiment and computational analysis, P211T is assigned to the “classic PKU” category, based on a single functionally hemizygous patient genotype.

Discussion

We have used the two computational methods to categorize the expected impact of a set of 46 PKU related mutations on the structural stability and molecular function of PAH, and compared the results to experimental data on *in vivo* like activity and protein levels, as well as the severity of disease. As in the general study of monogenic disease mutations,[1] about ¾ of the PAH mutations are assigned a high impact on protein stability. A primary objective was to test whether this computational assignment of a high fraction of mutations affecting protein stability is accurate Of these 35 mutations assigned as destabilizing, the experimental data for 30 support a role for destabilization, a true positive rate of 86%, and consistent with the benchmark 17% false positive rate. Results for an impact on molecular function are also reasonable, with three possible false predictions out of 16 (categories 2 and 3). Finally, the two cases where no functional impact is assigned by the computational methods agree with experiment. Even in the best of circumstances, comparisons between computational and experimental results are never entirely straightforward. In this study, because of variability in the experimental results and conditions, as well as the small number of cases considered, these accuracy rates are quite approximate. Variability in experimental results may arise from a number of factors, including the use of non-natural cofactors, or high cofactor concentrations, and can result in misleading disagreements with some computational assignments. Never-the-less, the results broadly confirm the primary conclusion from the

earlier computational work that destabilization plays a major role in monogenic disease, at least for this protein, and also support the estimated false positive rates.

The general tendency for PKU mutants to be associated with normal mRNA levels together with reduced protein levels and reduced total activity has been noted before.[11] The computational methods have allowed us to firmly link these observations to reduced stability of the protein three dimensional structure. Earlier comparison of the stability method results with experimental data for a set of site directed mutations in bacterial and phage proteins established that destabilizing monogenic disease mutations typically reduce the free energy difference between the folded and unfolded state of a protein by 2 to 3 Kcal/mol,[1] and the PAH mutations are likely similar in this regard. Given the typical range of free energy difference between the folded and unfolded state of a protein of 5 to 15 Kcal/mol, this level of destabilization would not have a measurable effect on *in vitro* activity, but evidently is usually critical *in vivo*.

There are two possible mechanisms. One mechanism is that these mutations sufficiently slow folding that a much smaller number of mature protein molecules are produced. Quality control mechanisms in eukaryotic organisms that remove ‘mis-folded’ proteins in the ER have been known for some time.[39] Available data on bacterial proteins show that about 40% of destabilizing mutations affect folding rate, but there are no extensive data for human proteins. The second possible mechanism is that the large increase in the concentration of unfolded protein (typically approximately 100 fold) produced by a 2-3 kcal/mol destabilization results in a high scavenging rate by molecular chaperones such as HSP90, which recognize unfolded protein molecules and target them to the mediated protein degradation system with the aid of proteins such as CHIP [40]. In either case, the result is a much lower *in vivo* concentration of proteins carrying destabilizing mutations.

Particularly where the mechanism involves scavenging of the unfolded species, these findings suggest that a general therapeutic strategy of restoring stability through binding of suitable small molecules may be suitable for a substantial fraction of PKU mutants. The effectiveness of excess cofactor treatment for some PKU mutants also supports the viability of such a strategy. The validity of this approach is also supported by cases of disease treatment restoration of stability for some other monogenic disease genes, for example transthyretin.[50]

Conclusions

There are three primary conclusions from this analysis of monogenic disease causing mutations in phenylalanine hydroxylase. First, the results support the conclusion of an earlier computational study that the large majority of missense mutations that cause monogenic disease involve destabilization of the protein structure. Second, the results confirm the link between destabilization and low *in vivo* protein levels. Third, although the numbers are small, the results also support the previous benchmark accuracy levels of the computational methods.

Materials and Methods

Data source

The experimental data for a set of 46 PKU-causing missense mutations of PAH and the results of 66 transient expression experiments in mammalian cell hosts was taken from PAHdb [11]. All the wild type and mutant cDNAs in this set have been expressed in monkey COS or human A293 cells. The expression experiments had enzyme activity, immuno-reactive protein level and sometimes mRNA level measured and reported as a percentage of wild type.

Templates selected for modeling missense mutations

Under physiological conditions, human PAH is a homo-tetramer, with each subunit composed of three domains. From N terminal to C terminal these are the regulatory, catalytic and tetramerization domains. To date, no experimentally determined structure of the complete human molecule is available., Three PDB structures were selected to model specific mutations in different domains based on crystal structure resolution, structure quality, and coverage: 1j8u, human PAH structure containing mainly the catalytic domain (resolution 1.50Å, R-free 0.203, in monomeric form); 2pah, human PAH structure covering the catalytic and tetramerization domains (resolution 3.10Å; R-free 0.326, in a tetrameric complex); and 1phz, rat PAH structure covering the regulatory and catalytic domains (resolution 2.20Å; R-free 0.297, in a dimeric complex).

The high resolution human 1j8u structure was used to model catalytic domain mutations. Regulatory domain mutations were modeled using a homology model of the human domain, based on the rat 1phz structure, as were three catalytic domain mutations, R261Q, R413P, and Y414C, that are in contact with the regulatory domain across a subunit interface. Rat PAH protein has 93% sequence identity with human PAH. There are no insertions or deletions in sequence between the two proteins. Main chain coordinates were taken directly from the rat structure. Side chains conformations were optimized using SCRWL.[25] Catalytic domain mutations R408W and R408Q are in contact with the tetramerization domain of another subunit and were modeled using 2pah.

Structure and sequence conservation methods for missense mutation impact analysis

The detailed methodology has been described previously.[1,16] The stability method optimizes the side chain conformations of a mutated residue and calculates 15 stability factors, including solvent accessible surface area, electrostatic interactions, steric clashes, buried hydrophobic area and local main chain flexibility. Based on these factors, a Support Vector Machine model (classifier) was trained using a set of monogenic disease causing mutations from the HGMD database and a non-disease control set of genomic variation between human and closely related mammals. For the present project, the model was retrained, excluding all PAH variants. False positive and false negative rates were assessed using the same bootstrap procedure as in Yue & Moult [16]. The false positive rate and false negative rates for the retrained model are 16.4% and 26.6% respectively, little different from the published model with 17% and 26%. Note that the high false negative rate is expected, since not all monogenic disease mutations include an effect on stability.

The sequence conservation method has also previously been published.[16] Five features are used to characterize the relative sequence conservation across the protein family at each residue position and the probability of accepting a specific substitution at that position. The same training data as for the stability method was used to train another SVM classifier. For the present application, the model was again retrained omitting all PAH variants, resulting in false positive and false negative rates of 9.5% and 20.1% respectively, the same as for the original model.

Acknowledgments

We are indebted to the authors of the PAHdb database for providing a compilation of the experimental data on PKU mutants. This work was supported in part by grant R01 LM07174 from the National Library of Medicine.

References

1. Yue P, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol. 2005; 353(2):459–473. [PubMed: 16169011]
2. Wang Z, Moult J. SNPs, protein structure, and disease. Hum Mutat. 2001; 17(4):263–270. [PubMed: 11295823]
3. Privalov PL. Stability of proteins: small globular proteins. Adv Protein Chem. 1979; 33:167–241. [PubMed: 44431]
4. Guthrie R, Susi A. A simple Phenylalanine Method for Detecting Phenylketonuria in Lare Populations of New Born Infants. Pediatrics. 1963; 32:338–343. [PubMed: 14063511]
5. DiLella AG, Kwok SC, Ledley FD, Marvit J, Woo SL. Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. Biochemistry. 1986; 25(4):743–749. [PubMed: 3008810]
6. Pitt D, Connelly J, Francis I, Wilcken B, Brown DA, Robertson E, Hill G, Masters P, Raby J, McFarlane J, Bowling F, Hancock J. Genetic screening of newborn in Australia. Results for 1981. Med J Aust. 1983; 1(7):333–335. [PubMed: 6835136]
7. Liu SR, Zuo QH. Newborn screening for phenylketonuria in eleven districts. Chin Med J (Engl). 1986; 99(2):113–118. [PubMed: 3093157]
8. S, CR.; K, S.; E, RC.; W, SLC. The Hyperphenylalaninemias. In: V, D., editor. The Metabolic and Molecular Bases of Inherited Disease. 7th. New York: McGraw-Hill; 1995. p. 1015-1075.
9. Fitzpatrick PF. Mechanism of aromatic amino acid hydroxylation. Biochemistry. 2003; 42(48): 14083–14091. [PubMed: 14640675]
10. Guldberg P, Rey F, Zschocke J, Romano V, Francois B, Michiels L, Ullrich K, Hoffmann GF, Burgard P, Schmidt H, Meli C, Riva E, Dianzani I, Ponzone A, Rey J, Guttler F. A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. Am J Hum Genet. 1998; 63(1):71–79. [PubMed: 9634518]
11. Scriver CR, Waters PJ, Sarkissian C, Ryan S, Prevost L, Cote D, Novak J, Teebi S, Nowacki PM. PAHdb: a locus-specific knowledgebase. Hum Mutat. 2000; 15(1):99–104. [PubMed: 10612829]
12. Scriver CR. The PAH gene, phenylketonuria, and a paradigm shift. Hum Mutat. 2007; 28(9):831–845. [PubMed: 17443661]
13. Fusetti F, Erlandsen H, Flatmark T, Stevens RC. Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria. J Biol Chem. 1998; 273(27):16962–16967. [PubMed: 9642259]
14. Erlandsen H, Stevens RC. The structural basis of phenylketonuria. Mol Genet Metab. 1999; 68(2): 103–125. [PubMed: 10527663]
15. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001; 11(5): 863–874. [PubMed: 11337480]

16. Yue P, Moult J. Identification and analysis of deleterious human SNPs. J Mol Biol. 2006; 356(5): 1263–1274. [PubMed: 16412461]

17. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet. 2001; 10(6):591–597. [PubMed: 11230178]

18. Bao L, Cui Y. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. Bioinformatics. 2005; 21(10):2185–2190. [PubMed: 15746281]

19. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 2007; 35(11):3823–3835. [PubMed: 17526529]

20. Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A. LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. Bioinformatics. 2005; 21(12):2814–2820. [PubMed: 15827081]

21. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248–249. [PubMed: 20354512]

22. Cortes C, Vapnik V. Support-Vector Networks. Machine Leaming. 1995; 20:273–297.

23. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat. 2003; 21(6): 577–581. [PubMed: 12754702]

24. Allali-Hassani A, Wasney GA, Chau I, Hong BS, Senisterra G, Loppnau P, Shi Z, Moult J, Edwards AM, Arrowsmith CH, Park HW, Schapira M, Vedadi M. A survey of proteins encoded by non-synonymous single nucleotide polymorphisms reveals a significant fraction with altered stability and activity. Biochem J. 2009; 424(1):15–26. [PubMed: 19702579]

25. Canutescu AA, Shelenkov AA, Dunbrack RL Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. 2003; 12(9):2001–2014. [PubMed: 12930999]

26. Erlandsen H, Pey AL, Gamez A, Perez B, Desviat LR, Aguado C, Koch R, Surendran S, Tyring S, Matalon R, Scriver CR, Ugarte M, Martinez A, Stevens RC. Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations. Proc Natl Acad Sci U S A. 2004; 101(48):16903–16908. [PubMed: 15557004]

27. Perez B, Desviat LR, Gomez-Puertas P, Martinez A, Stevens RC, Ugarte M. Kinetic and stability analysis of PKU mutations identified in BH4-responsive patients. Mol Genet Metab. 2005; 86(1):S11–16. [PubMed: 16091306]

28. Pey AL, Perez B, Desviat LR, Martinez MA, Aguado C, Erlandsen H, Gamez A, Stevens RC, Thorolfsson M, Ugarte M, Martinez A. Mechanisms underlying responsiveness to tetrahydrobiopterin in mild phenylketonuria mutations. Hum Mutat. 2004; 24(5):388–399. [PubMed: 15459954]

29. Ledley FD, Grenett HE, DiLella AG, Kwok SC, Woo SL. Gene transfer and expression of human phenylalanine hydroxylase. Science. 1985; 228(4695):77–79. [PubMed: 3856322]

30. Harding CO, Neff M, Wild K, Jones K, Elzaouk L, Thony B, Milstien S. The fate of intravenously administered tetrahydrobiopterin and its implications for heterologous gene therapy of phenylketonuria. Mol Genet Metab. 2004; 81(1):52–57. [PubMed: 14728991]

31. Kaufman S, Holtzman NA, Milstien S, Butler LJ, Krumholz A. Phenylketonuria due to a deficiency of dihydropteridine reductase. N Engl J Med. 1975; 293(16):785–790. [PubMed: 1160969]

32. Knappskog PM, Eiken HG, Martinez A, Bruland O, Apold J, Flatmark T. PKU mutation (D143G) associated with an apparent high residual enzyme activity: expression of a kinetic variant form of phenylalanine hydroxylase in three different systems. Hum Mutat. 1996; 8(3):236–246. [PubMed: 8889583]

33. Pey AL, Desviat LR, Gamez A, Ugarte M, Perez B. Phenylketonuria: genotype-phenotype correlations based on expression analysis of structural and functional mutations in PAH. Hum Mutat. 2003; 21(4):370–378. [PubMed: 12655546]

34. Okano Y, Wang T, Eisensmith RC, Longhi R, Riva E, Giovannini M, Cerone R, Romano C, Woo SL. Phenylketonuria missense mutations in the Mediterranean. Genomics. 1991; 9(1):96–103. [PubMed: 1672294]

35. Dworniczak B, Grudda K, Stumper J, Bartholome K, Aulehla-Scholz C, Horst J. Phenylalanine hydroxylase gene: novel missense mutation in exon 7 causing severe phenylketonuria. Genomics. 1991; 9(1):193–199. [PubMed: 1672290]
36. Li J, Eisensmith RC, Wang T, Lo WH, Huang SZ, Zeng YT, Yuan LF, Liu SR, Woo SL. Identification of three novel missense PKU mutations among Chinese. Genomics. 1992; 13(3): 894–895. [PubMed: 1639423]
37. Bjorgo E, Knappskog PM, Martinez A, Stevens RC, Flatmark T. Partial characterization and three-dimensional-structural localization of eight mutations in exon 7 of the human phenylalanine hydroxylase gene associated with phenylketonuria. Eur J Biochem. 1998; 257(1):1–10. [PubMed: 9799096]
38. Gamez A, Perez B, Ugarte M, Desviat LR. Expression analysis of phenylketonuria mutations. Effect on folding and stability of the phenylalanine hydroxylase protein. J Biol Chem. 2000; 275(38):29737–29742. [PubMed: 10875932]
39. Plemper RK, Wolf DH. Retrograde protein translocation: ERADication of secretory proteins in health and disease. Trends Biochem Sci. 1999; 24(7):266–270. [PubMed: 10390615]
40. Hohfeld J, Cyr DM, Patterson C. From the cradle to the grave: molecular chaperones that may choose between folding and degradation. EMBO Rep. 2001; 2(10):885–890. [PubMed: 11600451]
41. Guttler F, Azen C, Guldberg P, Romstad A, Hanley WB, Levy HL, Matalon R, Rouse BM, Trefz F, de la Cruz F, Koch R. Relationship among genotype, biochemical phenotype, and cognitive performance in females with phenylalanine hydroxylase deficiency: report from the Maternal Phenylketonuria Collaborative Study. Pediatrics. 1999; 104(2 Pt 1):258–262. [PubMed: 10429004]
42. Eiken HG, Knappskog PM, Apold J, Flatmark T. PKU mutation G46S is associated with increased aggregation and degradation of the phenylalanine hydroxylase enzyme. Hum Mutat. 1996; 7(3): 228–238. [PubMed: 8829656]
43. Desviat LR, Perez B, Gamez A, Sanchez A, Garcia MJ, Martinez-Pardo M, Marchante C, Boveda D, Baldellou A, Arena J, Sanjurjo P, Fernandez A, Cabello ML, Ugarte M. Genetic and phenotypic aspects of phenylalanine hydroxylase deficiency in Spain: molecular survey by regions. Eur J Hum Genet. 1999; 7(3):386–392. [PubMed: 10234516]
44. Zekanowsk C, Perez B, Desviat LR, Wiszniewski W, Ugarte M. In vitro expression analysis of R68G and R68S mutations in phenylalanine hydroxylase gene. Acta Biochim Pol. 2000; 47(2): 365–369. [PubMed: 11051201]
45. Kayaalp E, Treacy E, Waters PJ, Byck S, Nowacki P, Scriver CR. Human phenylalanine hydroxylase mutations and hyperphenylalaninemia phenotypes: a metanalysis of genotype-phenotype correlations. Am J Hum Genet. 1997; 61(6):1309–1317. [PubMed: 9399896]
46. Mirisola MG, Cali F, Gloria A, Schinocca P, D'Amato M, Cassara G, Leo GD, Palillo L, Meli C, Romano V. PAH gene mutations in the Sicilian population: association with minihaplotypes and expression analysis. Mol Genet Metab. 2001; 74(3):353–361. [PubMed: 11708866]
47. Gjetting T, Petersen M, Guldberg P, Guttler F. In vitro expression of 34 naturally occurring mutant variants of phenylalanine hydroxylase: correlation with metabolic phenotypes and susceptibility toward protein aggregation. Mol Genet Metab. 2001; 72(2):132–143. [PubMed: 11161839]
48. Apold J, Eiken HG, Odland E, Fredriksen A, Bakken A, Lorens JB, Boman H. A termination mutant prevalent in Norwegian haplotype 7 phenylketonuria genes. Am J Hum Genet. 1990; 47(6):1002–1007. [PubMed: 1978553]
49. De Lucca M, Perez B, Desviat LR, Ugarte M. Molecular basis of phenylketonuria in Venezuela: presence of two novel null mutations. Hum Mutat. 1998; 11(5):354–359. [PubMed: 9600453]
50. Hammarstrom P, Wiseman RL, Powers ET, Kelly JW. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. Science. 2003; 299:713–6. [PubMed: 12560553]

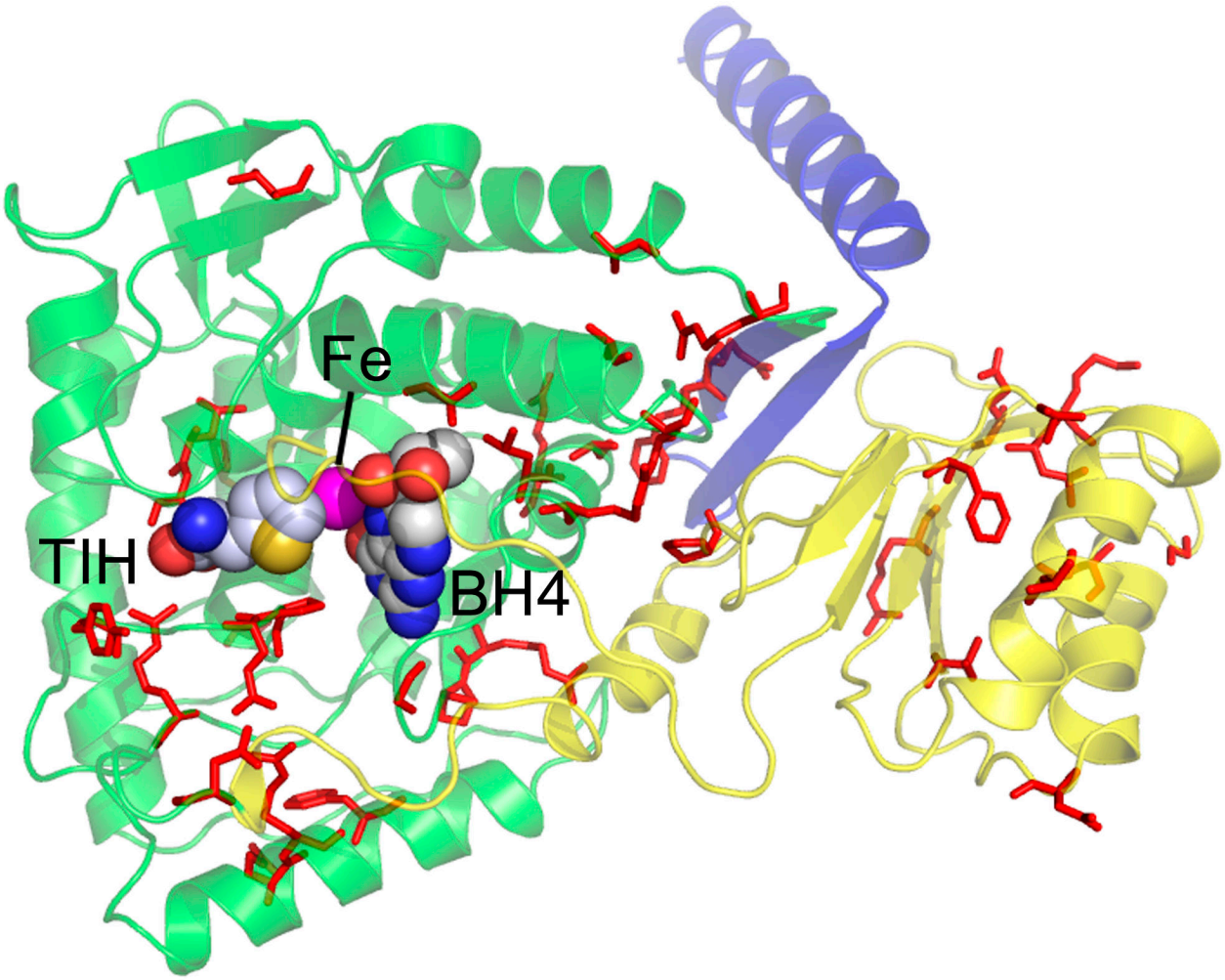

Figure 1. Structure model of phenylalanine hydroxylase used for mutation analysis

Domains are: regulatory (yellow); catalytic (green); tetramerization (blue). The 39 residues with mutations discussed in this study are in red. The Fe (++) ion is magenta, and a substrate analog, Beta(2-thienyl) alanine (TIH) and cofactor Tetrahydrobiopterin (BH4) are shown space filled. This composite model is build from PDB structures 1j8u, 2pah, and 1phz.

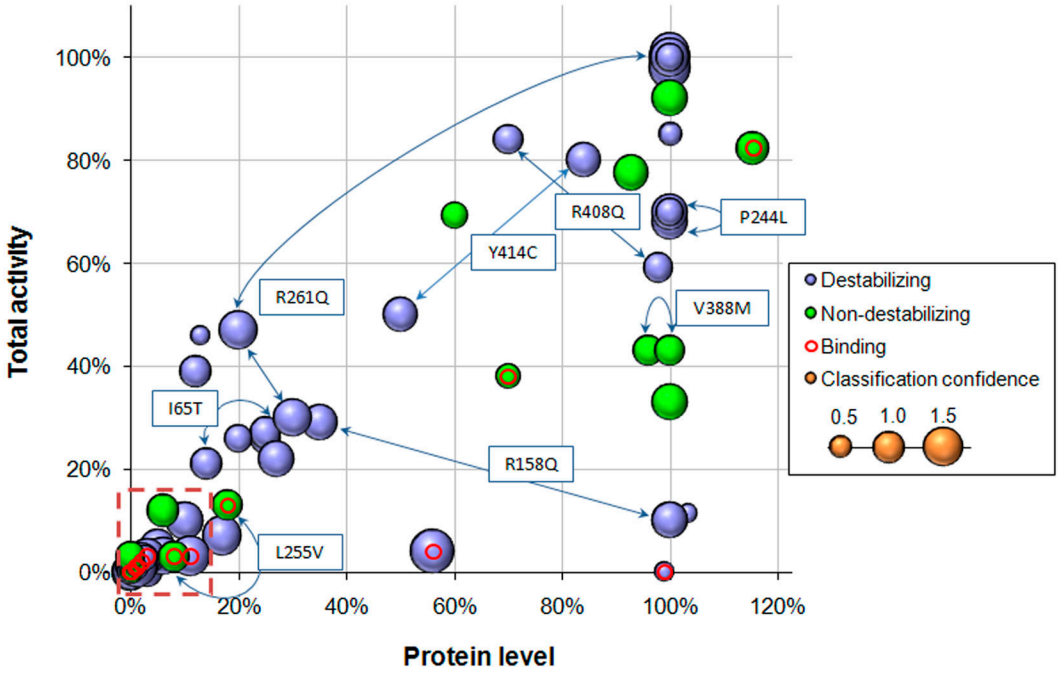

Figure 2. The relationship between stability impact and experimental measurements of mutant enzyme activity and protein level

66 experiment measurements for 46 mutations are plotted by percentage of wild type protein level (X axis) and enzyme activity (Y axis). Both axes are normalized by the mRNA level compared to wildtype, where available. Experimental results for differing independent measurements of the same mutation are connected by double headed arrows. Each point is colored according to the predicted mechanism of action, with blue for an assignment of destabilization, green for an assignment of normal stability, and red circles indicating an involvement in ligand binding. The size of each point is proportional to the confidence of the computational assignment. Near the origin there is a cluster of 29 points for 20 mutations (red dashed box). 16 of these mutations (24 experimental results) are blue, and 4 mutations (5 experiments) are green. Most predicted destabilizing mutations show significantly reduced protein level (<50%), while most of the mutations with no stability assignment have close to wild type protein level. All but one of the mutations close to an active site (red open circles) show low total activity, consistent with the experimental assignments.

Table 1

Functional and stability impact analysis results of 46 PAH missense mutations together with experimental measurements of activity and protein level, and clinical PKU classification

A negative score for the stability method indicates an expected impact in protein stability. A negative score for the profile method indicates an expected impact on protein function in vivo from any cause, including stability. The absolute value of a score shows the confidence for a particular assignment. Benchmarking (Yue, Moult 2006) has shown that the higher confidence assignments (|*Score*| ≥*0.5*) are more reliable.

mutation	template	contact with	vector	host	total activity	protein level	specific activity	mRNA level	Profile method	Stability method	stability impact #	clinical category
p.F39L	1phz_A		pcDNA3	A293	46%	13%	3.54	100%	−0.01	−0.39	hydrophobic interaction decreased	classic-moderate-mild [10]; moderate [41]
p.L41F	1phz_A		pCDNA3	COS-7	10%	91%	0.11	88%	−0.57	−0.35	NA(overpacking 2.68 Å; gain of hydrophobic interaction)	NA
p.K42I	1phz_A		pcDNA3	A293	12%	6%	2.00	100%	−0.53	1.06	on surface;saltbridge lost	NA
p.G46S	1phz_A		pcDNA3	A293	0.1%	3%	0.03	Not stated	−0.05	−0.90	backbone strain	classic [42]; mild [10]
p.L48S	1phz_A		pcDNA3	A293	39%	12%	3.25	100%	−2.52	−1.00	hydrophobic interaction decreased	classic-moderate-mild [10]
p.D59Y	1phz_A		pRc/CMV	COS	92%	100%	0.92	Not stated	−0.21	1.27	on surface	MHP [33]
p.I65T	1phz_A		pcDNA3	A293	21%	14%	1.50	100%	−1.23	−0.97	buried polar;hydrophobic interaction decreased	mild [33,43]; moderate [41]
p.I65T	1phz_A		p91023(B)	COS	26%	25%	1.04	100%				
p.I65T	1phz_A		pRc/CMV	COS	27%	25%	1.08	Not stated				
p.R68G	1phz_A		pRc/CMV	COS	100%	100%	1.00	Not stated	−2.52	−1.56	hydrogen bond lost;saltbridge lost;hydrophobic interaction decreased;	suggested mild when patients are combined with ‘classic’ mutations. [44]
p.R68S	1phz_A		pRc/CMV	COS	98%	100%	0.98	Not stated	−1.50	−1.72	hydrogen bond lost;saltbridge lost;hydrophobic interaction decreased	mild [41,43]
p.E76G	1phz_A		pRc/CMV	COS	85%	100%	0.85	Not stated	0.95	−0.55	saltbridge lost	NA
p.T92I	1phz_A		pCDNA3	COS-7	76%	91%	0.84	98%	2.32	1.21	on surface; loss of polar-polar interaction	MHP [45,46]
p.A104D	1phz_A		pcDNA3	A293	26%	20%	1.30	100%	1.01	−0.75	hydrophobic interaction decreased	mild [41,43,45,47]
p.P122Q	1j8u_A		pRc/CMV	COS	22%	27%	0.81	Not stated	−2.23	−1.28	NA (gain of polar-polar interaction; minor decrease of hydrophobic interaction)	NA
p.D143G	1j8u_A		pcDNA3	A293	33%	100%	0.33	Not stated	−1.12	1.30	on surface	classic [48]
p.R157N	1j8u_A		pcDNA3	A293	5%	5%	1.00	100%	−0.67	−1.18	hydrogen bond lost; saltbridge lost	NA
p.R158Q	1j8u_A		pRc/CMV	COS	29%	35%	0.83	Not stated	−0.95	−1.21	hydrogen bond lost;saltbridge lost	classic-moderate [10,45] mild [33]
p.R158Q	1j8u_A		p91023(B)	COS	10%	100%	0.10	100%				
p.R158Q	1j8u_A		pCDNA1	COS	10%	100%	0.10	Not stated				

mutation	template	contact with	vector	host	total activity	protein level	specific activity	mRNA level	Profile method	Stability method	stability impact #	clinical category
p.F161S	1j8u_A		pRc/CMV	COS	7%	17%	0.41	Not stated	−2.02	−1.48	hydrophobic interaction decreased	NA
p.P211T	1j8u_A		pCDNA3	COS-7	72%	63%	1.15	104%	0.62	0.73	on surface	classic [45]
p.G218V	1j8u_A		pRc/CMV	COS	101%	100%	1.01	Not stated	−0.14	−1.38	overpacking; backbone strain	classic [10]
p.R243Q	1j8u_A		pCDNA1	COS	10%	10%	1.00	100%	−0.90	−1.34	saltbridge lost	classic [10,41]; mild [33]
p.R243Q	1j8u_A		pRc/CMV	COS	10%	10%	1.00	Not stated				
p.P244L	1j8u_A		pRc/CMV	COS	68%	100%	0.68	Not stated	−2.01	−1.28	NA (overpacking 2.64 Å; gain of hydrophobic interaction)	NA
p.P244L	1j8u_A		pRc/CMV	COS	70%	100%	0.70	100%				
p.G247V	1j8u_A	BH4	pRc/CMV	COS	4%	56%	0.07	Not stated	−2.94	−1.82	backbone strain	NA
p.R252G	1j8u_A		pcDNA3	A293	3%	5%	0.60	100%	−1.29	−1.26	hydrogen bond lost;saltbridge lost	classic [10]
p.R252Q	1j8u_A		pcDNA3	A293	3%	3%	1.00	100%	−0.95	−1.12	hydrogen bond lost;saltbridge lost	classic [10,41,45]
p.R252W	1j8u_A		pRc/CMV	COS	0%	0%	NA	Not stated	−2.31	−0.92	hydrogen bond lost;overpackingsaltbridge lost	classic [10,33,41,43]
p.R252W	1j8u_A		pCDNA1	COS	1%	1%	1.00	100%				
p.L255S	1j8u_A	BH4	pcDNA3	A293	3%	11%	0.27	100%	−2.83	−1.55	hydrophobic interaction decreased	NA
p.L255V	1j8u_A	BH4	pcDNA3	A293	3%	8%	0.38	100%	−2.15	0.93	NA	NA
p.L255V	1j8u_A	BH4	pRc/CMV	COS	13%	18%	0.72	Not stated				
p.A259T	1j8u_A		pcDNA3	A293	3%	3%	1.00	100%	−2.57	−1.09	buried polar; overpacking 2.52 Å	NA
p.A259V	1j8u_A		pcDNA3	A293	3%	6%	0.50	100%	−2.91	−1.32	overpack 2.51 Å	classic [10]
p.R261Q	1phz_A		pRc/CMV	COS	47%	20%	2.35	Not stated	−1.48	−1.42	hydrogen bond lost	classic-moderate [10]
p.R261Q	1phz_A		pCDNA1	COS	30%	30%	1.00	Not stated				
p.R261Q	1phz_A		p91023(B)	COS	100%	100%	1.00	100%				
p.R270S	1j8u_A	TIH	pRc/CMV	COS	1%	1%	1.00	100%	−3.25	−1.24	hydrogen bond lost; saltbridge lost; hydrophobic interaction decreased	NA
p.R270S	1j8u_A	TIH	pcDNA3	A293	3%	3%	1.00	100%				
p.Y277D	1j8u_A	TIH	pRc/CMV	COS	0%	99%	0.00	Not stated	−2.77	−0.40	hydrophobic interaction decreased	mild [10,41]; classic [33,43]
p.E280K	1j8u_A	FE,TIH	pRc/CMV	COS	0%	0%	NA	Not stated	−2.83	−1.34	saltbridge lost	classic [10,41,43,45]; moderate [33]
p.E280K	1j8u_A	FE,TIH	pCDNA1	COS	2%	2%	1.00	Not stated				
p.P281L	1j8u_A	TIH,BH4	pSVL	COS	0%	0%	NA	100%	−3.85	0.47	on surface	classic [10,41,43,45]
p.P281L	1j8u_A	TIH,BH4	pCDNA1	COS	1%	1%	1.00	100%				
p.A309V	1j8u_A		pRc/CMV	COS	70%	100%	0.70	Not stated	−0.47	−0.76	NA (overpacking 2.68 Å; gain of hydrophobic interaction)	Moderate [33] Pey classic Desviat

mutation	template	contact with	vector	host	total activity	protein level	specific activity	mRNA level	Profile method	Stability method	stability impact #	clinical category
p.L311P	1j8u_A		pRc/CMV	COS	0%	0%	NA	Not stated	−3.21	−1.08	hydrogen bond lost;hydrophobic interaction decreased	classic-moderate [10]; classic [43]
p.L311P	1j8u_A		p91023(B)	COS	1%	1%	1.00	100%	−3.21	−1.08		
p.A322G	1j8u_A	BH4	pRc/CMV	COS	75%	105%	0.71	91%	−0.37	1.08	on surface	MHP [43]
p.L348V	1j8u_A	TIH	pRc/CMV	COS	38%	70%	0.54	Not stated	−1.12	0.59	cavity creation; hydrophobic interaction decreased	mild [43]; moderate [10,41]
p.S349L	1j8u_A	TIH	pRc/CMV	COS	0%	0%	NA	100%	−3.42	−1.22	hydrogen bond lost;overpacking	classic [10,49]
p.S349L	1j8u_A	TIH	pRc/CMV	COS	0%	0%	NA	Not stated				
p.S349P	1j8u_A	TIH	pRc/CMV	COS	1%	1%	1.00	90%	−3.08	−1.67	hydrogen bond lost;backbone strain	classic [10,41,43,45]
p.V388M	1j8u_A		pRc/CMV	COS	43%	96%	0.45	Not stated	−0.12	0.87	NA	mild [43]; moderate [10,41]
p.V388M	1j8u_A		pRc/CMV	COS	43%	100%	0.43	100%				
p.A403V	1j8u_A		pRc/CMV	COS	100%	100%	1.00	100%	0.46	−0.78	NA (overpacking 2.53 Å; on surface; gain of hydrophobic interaction)	MHP [41,43]; MHP-mild [45]
p.R408Q	2pah_A		pRc/CMV	COS	84%	70%	1.20	Not stated	−0.67	−0.89	hydrogen bond lost	MHP [45]; MHP-mild [33]; mild [10,41,43]
p.R408Q	2pah_A		pRc/CMV	COS	55%	91%	0.60	93%				
p.R408W	2pah_A		pRc/CMV	COS	0%	0%	NA	Not stated	−2.03	−0.63	hydrogen bond lost	classic [10,33,43,45]
p.R408W	2pah_A		p91023(B)	COS	1%	1%	1.00	100%				
p.R408W	2pah_A		pRc/CMV	COS	3%	3%	1.00	102%				
p.R413P	1phz_A		pCDNA1	COS	3%	0%	NA	100%	−0.93	0.80	hydrogen bond lost;backbone strain;saltbridge lost	NA
p.Y414C	1phz_A		pCDNA1	COS	50%	50%	1.00	100%	−2.48	−1.18	hydrophobic interaction decreased	MHP [33]; mild [41,43]; classic-moderate-mild-MHP [10]
p.Y414C	1phz_A		pRc/CMV	COS	80%	84%	0.95	Not stated				