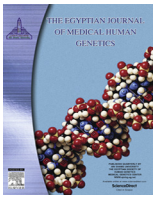
The Egyptian Journal of Medical Human Genetics 19 (2018) 367–378

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Medical Human Genetics

journal homepage: www.sciencedirect.com



Original article

In silico analysis of the functional non-synonymous single nucleotide polymorphisms in the human CYP27B1 gene

Solomon O. Rotimi *, Olamide Peter, Oluwayomi Oguntade, Oluwakemi A. Rotimi

Department of Biochemistry, Covenant University, Ota, Nigeria

ARTICLE INFO

Article history:
Received 19 February 2018
Accepted 7 March 2018
Available online 17 March 2018

Keywords:
Polymorphisms
CYP27B1
Vitamin D
Mutation
Cancer
Diabetes

ABSTRACT

Background: CYP27B1 gene codes for 25-hydroxyvitamin D_3 1-α-hydroxylase, an enzyme that catalyses the activation of vitamin D to the 1-α, 25 dihydroxyvitamin D_3. The activity of this enzyme is altered by non-synonymous single nucleotide polymorphisms (nsSNPs) located within its gene. Such alterations consequently affect the synthesis of the active form of the hormone, 1-α, 25 dihydroxyvitamin D_3, resulting in vitamin D deficiency or insufficiency.

Objective: We aimed to investigate the impact of nsSNPs in the CYP27B1 gene on the structure and/or function of 25-hydroxyvitamin D_3 1-α-hydroxylase.

Methods: The pathogenic nsSNPs in the human CYP27B1 obtained from National Centre for Biotechnology Information (NCBI) were analysed for their structural and functional consequence using mutation analysis algorithms like Consurf, I-Mutant, and MutPred. The effects of the mutation on tertiary structure of the human CYP27B1 protein was predicted using SWISS-MODEL while STRING was used to investigate its protein–protein interaction.

Results: Out of 938 SNPs in the human CYP27B1 gene, 455 that are responsible for missense mutations in the protein were subjected to various prediction algorithms to identify the pathogenic variants. Out of 24 consensus pathogenic nsSNPs, our Consurf analysis showed that mutations at conserved positions T321, R389 and G125 will significantly alter the structure of human CYP27B1 protein. These mutations also alter the metal binding and result in intrinsic structural disorder. These consequently, alter the 3D structure of the protein and could impact its ability to interact with other proteins like Cytochrome P450, family 2, subfamily R, polypeptide 1; Cytochrome P450, family 24, subfamily A, polypeptide 1 and Vitamin D receptor, that are involved in vitamin D pathway, as revealed by STRING.

Conclusion: These nsSNPs could contribute to vitamin D deficiency and its associated pathological conditions.

1. Introduction

Single nucleotide polymorphisms (SNPs) are regarded as one of the most common genetic variations in the human genome. They represent a single base change in a DNA sequence, with a common substitute of two possible nucleotides at a given position. The non-synonymous SNPs (nsSNPs) trigger genomic disparities within protein-coding areas resulting in mutations that could alter the structure and/or function of the protein. Nonetheless, such structural and/or functional modifications due to nsSNPs are not damaging or deleterious in most cases [1]. Non-synonymous SNPs could either alter the structural properties or the function of the protein. Such alterations may result in disease phenotypes. However, SNPs in genes involved in regulating key enzymes of nutrient metabolism may have significant phenotypic outcomes.

One of such enzymes is 25-hydroxyvitamin D_3 1-α-hydroxylase and its gene, cytochrome p450 27B1 (CYP27B1) in humans is sited on 12q14.1 (Fig. 1). This is the long (q) arm of chromosome 12 at position 14.1 that also encodes enzymes which belong to the cytochrome P450 superfamily. Generally, cytochrome P450 proteins are monooxygenases with catalytic activities including xenobiotic metabolism and synthesis of steroids and other lipids [2]. The protein encoded by human CYP27B1 is called 25-hydroxyvitamin D_3 1-α-hydroxylase and it hydroxylates 25-hydroxyvitamin D_3 at the 1-α position. It is localized in the inner mitochondrial membrane where it converts 25-hydroxyvitamin D_3 to the active form 1-α, 25 dihydroxyvitamin D_3 [3,4]. The active form of vitamin D_3 then binds to the vitamin D receptor to elicit its biological

Peer review under responsibility of Ain Shams University.

* Corresponding author.
E-mail address: Ola.rotimi@covenantuniversity.edu.ng (S.O. Rotimi).

https://doi.org/10.1016/j.ejmhg.2018.03.001
1110-8630/

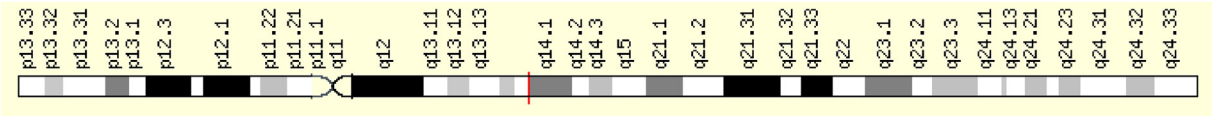


Fig. 1. CYP27B1 Gene on chromosome 12: bands according to Ensembl, locations according to GeneLoc [38].

functions [3,4]. Consequently, this gene is very important to the activation and biological activity of the hormone form of vitamin D, which involves calcium metabolism, cell proliferation and differentiation, and immunomodulatory functions, as well as calcidiol 1-monooxygenase activity. However, mutations in the CYP27B1 gene are one of the genomic factors that influence the levels of active vitamin D in circulation. Vitamin D insufficiency and/or deficiency is linked to rickets and a plethora of several chronic disorders, including Addison's disease, diabetes, cardiac health, and cancer [5,6].

In view of the major effects of circulating vitamin D, there is a consensus that the amount of active vitamin D available to bind vitamin D receptor is reduced by certain SNPs present in the CYP27B1 gene.

1.1. Aim

The aim of this study was to elucidate the impact of nsSNPs on the structural and functional properties of human CYP27B1.

2. Materials and methods

2.1. Sequence recovery

The nucleotide sequence data on human CYP27B1 gene was extracted from the National Centre for Biotechnology Information (NCBI) while the amino acid sequence of human CYP27B1 protein in FASTA format was retrieved from UniProt (http://www.uniprot.org/) [1].

2.2. Retrieval, mining and mapping of SNPs in human CYP27B1

Several databases were used in collecting the polymorphism data for the human CYP27B1 gene. The databases include: UniProt database (http://www.uniprot.org), NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/SNP/), and Ensembl genome browser (http://www.ensembl.org/index.html) [7]. The SNPs recovered from these databases were grouped into several functional classes, as described by Dakal et al. [1].

2.3. Analysis of nsSNP in human CYP27B1

A number of *in silico* algorithms were used in predicting functional effects of the nsSNPs in human CYP27B1 [8]. These included: PolyPhen-2 (http://genetics.bwh.harvard.edu/), SIFT (http://sift.jcvi.org/) and PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org). nsSNPAnalyzer (http://snpanalyzer.uthsc.edu) was used to predict nsSNPs that confer pathogenic effects and were considered as high-risk nsSNPs for further studies [9].

2.4. Evolutionary phylogenetic analysis of human CYP27B1 protein

The amino acid evolutionary conservation in CYP27B1 protein was predicted using the ConSurf web server (consurf.tau.ac.il/) [10]. ConSurf server uses an empirical Bayesian method for the determination of evolutionary conservation plus identification of putative structural and functional residues. The conservation score of 1–4 is considered as variable, 5–6 as intermediate, and 7–9 as conserved amino acid position [11,12].

2.5. Extrapolation of amino acid changes and disease phenotypes

The online server used for predicting the molecular basis of the disease linked amino acid replacement in a mutant protein is MutPred (http://mutpred.mutdb.org/)[13]. It uses several attributes associated with protein structure, function, and evolution. Its accuracy of prediction is increased by combining three servers, SIFT [9], PSI-BLAST [14], and Pfam, profiles alongside TMHMM, MARCOIL which are structural disorder prediction algorithms and DisProt.

2.6. Stability analysis of human CYP27B1 protein

I-Mutant adaptation 2.0, which is an internet support vector dependent upon ProTherm databases, helped to evaluate nsSNP prompted adjustments in protein dependability [15]. I-Mutant estimates the free energy changes value (DDG) as a difference between the unfolding Gibbs free energy value (DG) for the wild type protein and that of the mutant protein (DDG or DDG = DG mutant – DG wild type). Potential (surge or reduction) in free energy change rate (DDG) is also predicted, along with a reliability index (RI) for the results, where the lowest and highest reliability are 0 and 10, respectively [16].

2.7. Prediction of post translational modification sites for CYB27B1

The post translational modifications sites within the amino acid sequence of human CYP27B1 protein was identified using a web based tool called ModPred (www.modpred.org/) [17].

2.8. Prediction of structural effect of point mutation on human CYB27B1 protein

Project HOPE (www.cmbi.ru.nl/hope/) was used to identify the structural effect certain point mutations in the human CYP27B1 protein sequence. A BLAST was initially performed by the algorithm against UniProt and PDB to retrieve tertiary structure information and generate a homology model. This was followed by a prediction of the protein features using the Distributed Annotation System server [18].

2.9. Prediction of secondary structure and 3D structure of CYP27B1

Protein secondary structures was predicted based on their specific matrices using the PSIPRED tool (http://bioinf.cs.ucl.ac.uk/psipred/), with matrices developed by PSI-BLAST [14]. Since there was no crystal structure of human CYP27B1 protein available in protein databank, we modelled its 3D structure using SWISS-MODEL (https://swissmodel.expasy.org/) [19]. SWISS-MODEL is a fully automated server for predicting 3D structure of proteins using the crystal structure of similar protein as template. For this purpose, we used rat mitochondrial 1,25-dihydroxyvitamin D_3 24-hydroxylase (pdb ID: 3k9v.2.A) as template.

2.10. Protein-protein interactions prediction

Protein-protein interactions are very vital in the assessment of all protein functional interactions present within the cell. The search tool for the retrieval of interacting genes/proteins (STRING) (https://string-db.org/) was used to predict the interaction of human CYP27B1 protein with other proteins. The STRING generates protein-protein interaction through, either direct or indirect, associations between a known protein and other proteins by utilizing its database of 5,214,234 proteins of 1113 organisms [20]. For this study, CYP27B1 and *Homo sapiens* were used as the input.

3. Results and discussion

3.1. SNP dataset

In this study, we used extensive computational techniques to screen for functional genetic variants within the coding region of human CYP27B1. In total, 938 SNPs were recorded in human CYP27B1 gene sequence. All the reported SNPs of CYP27B1 gene were recovered from NCBI dbSNP, Ensembl genome browser, and the UniProt. In the Ensemble dbSNP database, 456 rsIDs were mapped that were connected with SNPs in the coding area, but the present investigation concentrates on SNPs that confer single amino acid changes into human CYP27B1 protein. The SNPs were arranged based on their functional classes and are reported in Table 1.

However, to ascertain whether an assumed missense mutation affected CYP27B1 function, 441 nsSNPs were subjected to a multiplicity of *in silico* SNP prediction algorithms. A summary of the outcomes is presented in Table 2.

Table 1
Functional consequence types of SNPs present in the human CYP27B1 gene.

Consequence type	Count
Coding sequence variant	203
Frameshift variant	27
Inframe deletion	1
Inframe insertion	1
Missense variant	441
Missense variant ~ splice region variant	14
Splice region variant ~ coding sequence variant	6
Splice region variant ~ synonymous variant	7
Start lost	3
Stop gained	6
Synonymous variant	229
Grand Total	938

Data derived from Ensembl [38].

Table 2
Summary of prediction results for nsSNPs in the CYP27B1 protein.

Prediction	Number of nsSNPs (%)		
	Poly-Phen 2	SIFT	nsSNPAnalyzer
Benign	222 (49%)		
Possibly damaging	82 (18%)		
Probably damaging	134 (29%)		
Neutral	17 (4%)		
Deleterious		197 (43%)	
Deleterious – low confidence		11 (2.5%)	
Tolerated		236 (52%)	
Tolerated – low confidence		11 (2.5%)	
Likely pathogenic			4 (0.9%)
Pathogenic			24 (5.3%)
Uncertain significance			16 (3.5%)
Blank			411 (90.3%)

Data derived from Poly-Phen 2, SIFT and nsSNPAnalyzer [9].

In general, nsSNPs are known to modify protein conformation, interaction, and function. Previous studies have shown that the SNP of CYP27B1 is one of the factors responsible for low level of circulating 1-α, 25-dihydroxyvitamin D_3 and with correlative relationship to increased risk of cancer, Addison's disease and diabetes [21–23]. Hence, it is important to investigate the potential of each nsSNP in human CYP27B1 in distorting the structure and function of 25-hydroxyvitamin D_3 1-α-hydroxylase.

3.2. Analysis of nsSNP in human CYP27B1

We progressed further to utilizing three *in silico* SNP prediction algorithms: PolyPhen-2, SIFT, nsSNPs Analyzer. With regards to the PolyPhen-2 results, 134 nsSNPs (29%) exist as 'probably damaging' to CYP27B1 protein function, while 222 nsSNPs (49%) are benign. Also, 82 nsSNPs (18%) have 'possibly damaging' potential as predicted by PolyPhen-2 (Table 2). Our SIFT analysis projected that 197 nsSNPs (43%) had deleterious effects to CYP27B1protein function and 236 nsSNPs (52%) are tolerated. Subsequently, SIFT analy-

Table 3
CYP27B1 nsSNPs projected to be functionally noteworthy by SNPs estimation algorithms.

nsSNP ID	Mutation	SNP Type	# Del. Pred.
Rs1057520815	G73E G102E	Non-synonymous	2
Rs118204007	T86R T321R	Non-synonymous	2
Rs118204008	T409I	Non-synonymous	1
Rs118204010	R154C R389C	Non-synonymous	2
	R154G R389G		2
Rs118204011	L108F L343F	Non-synonymous	2
Rs118204012	E160G E189G	Non-synonymous	2
Rs28934605	G96E G125E	Non-synonymous	2
Rs28934606	R100P R335P	Non-synonymous	2
	R100Q R335Q	Non-synonymous	2
Rs28934607	P147S P382S	Non-synonymous	2
Rs568165894	R429H	Non-synonymous	1
	R429P	Non-synonymous	1
Rs759208930	R453H	Non-synonymous	1

Data derived from PROVEAN [1].

Table 4
PolyPhen-2 analysis of Pathogenic coding nsSNPs in human CYP27B1.

Replacement	Effect	Score	Sensitivity	Specificity
G73E	Probably Damaging	1.000	0.00	1.00
G102E	Probably Damaging	0.989	0.72	0.97
T321R	Probably Damaging	1.000	0.00	1.00
T409I	Benign	0.004	0.97	0.59
R154C	Probably Damaging	1.000	0.00	1.00
L343F	Probably Damaging	1.000	0.00	1.00
E189G	Probably Damaging	0.984	0.74	0.96
G125E	Probably Damaging	1.000	0.00	1.00
R335P	Probably Damaging	0.994	0.69	0.97
R335Q	Possibly Damaging	0.610	0.87	0.91
P147S	Probably Damaging	0.994	0.69	0.97
P382S	Probably Damaging	0.999	0.14	0.99
R429H	Benign	0.065	0.94	0.84
R429P	Probably Damaging	0.976	0.76	0.96
R453H	Probably Damaging	1.000	0.00	1.00

Data derived from Poly-Phen 2 [9].

sis also provided sub-data's predicting 'deleterious-low confidence' of 11 nsSNPs (3%) and 'tolerated-low confidence' of 11 nsSNPs (2%). The nsSNP Analyzer, which was used to predict the clinical significance of SNPs from human CYP27B1 predicted that 24 nsSNPs (55%) are pathogenic or disease causing, 4 nsSNPs (9%) are 'likely pathogenic', and 16 nsSNPs (36%) having uncertain significance. nsSNPs without any clinical significance were left blank.

Table 5
Conservation profile of amino acids in CYP27B1.

nsSNP ID	Amino Acid	CS	ConSurf Prediction
rs1057520815	G73	9	Buried
	G102	5	Exposed
rs118204007	T321	9	Highly conserved and exposed (f)
rs118204008	T409	8	Buried
rs118204010	R389	9	Highly conserved and exposed (f)
rs118204011	L343	8	Buried
rs118204012	E189	5	Exposed
rs28934605	G125	9	Highly conserved and buried (s)
rs28934606	R335	5	Exposed
rs28934607	P382	9	Highly conserved and exposed (f)
	P147	8	Highly conserved and exposed (f)
rs568165894	R429	1	Exposed
rs759208930	R453	9	Highly conserved and exposed (f)

Data obtained from ConSurf [10]/

Different *in silico* techniques used different algorithms, resulting in often considerable difference in results. Thusberg and Vihinen [24] reviewed these tools and suggested that SIFT and PolyPhen-2 have better performance in identifying deleterious nsSNPs. This observation was validated by Hicks et al. [25] and this makes these tools useful for our study. However, nsSNPs with greater confidence were expected to be truly deleterious. At this point, high-risk nsSNPs were classified based on their predicted pathogenicity or disease related prediction by the algorithms. Twenty-four nsSNPs in human CYP27B1 met these conditions and were carefully selected for extra examination as shown in Table 3. This group was then analysed further with PolyPhen-2 as well as PROVEN, ConSurf and I-Mutant. The PROVEN analysis identified 11 to be responsible for deleterious amino acid substitution.

In particular, PolyPhen-2 results showed that nine amino acid switches (G73E, G102E, T321R, R154C, L343F, E189G, G125E, R335P, P147S, P382S, R429P and R453H) were recorded as probably damaging (score > 0.96), 1 amino acid substitution (R335Q) was noted as possibly damaging (score >0.2 and <0.96), and 2 amino acid substitutions (T409I and R429H) was scored as benign (score < 0.2) (Table 4).

3.3. High-risk non-synonymous SNPs conservation profile

Several significant biological processes constantly involve the presence of amino acids; either being present in catalytic sites or

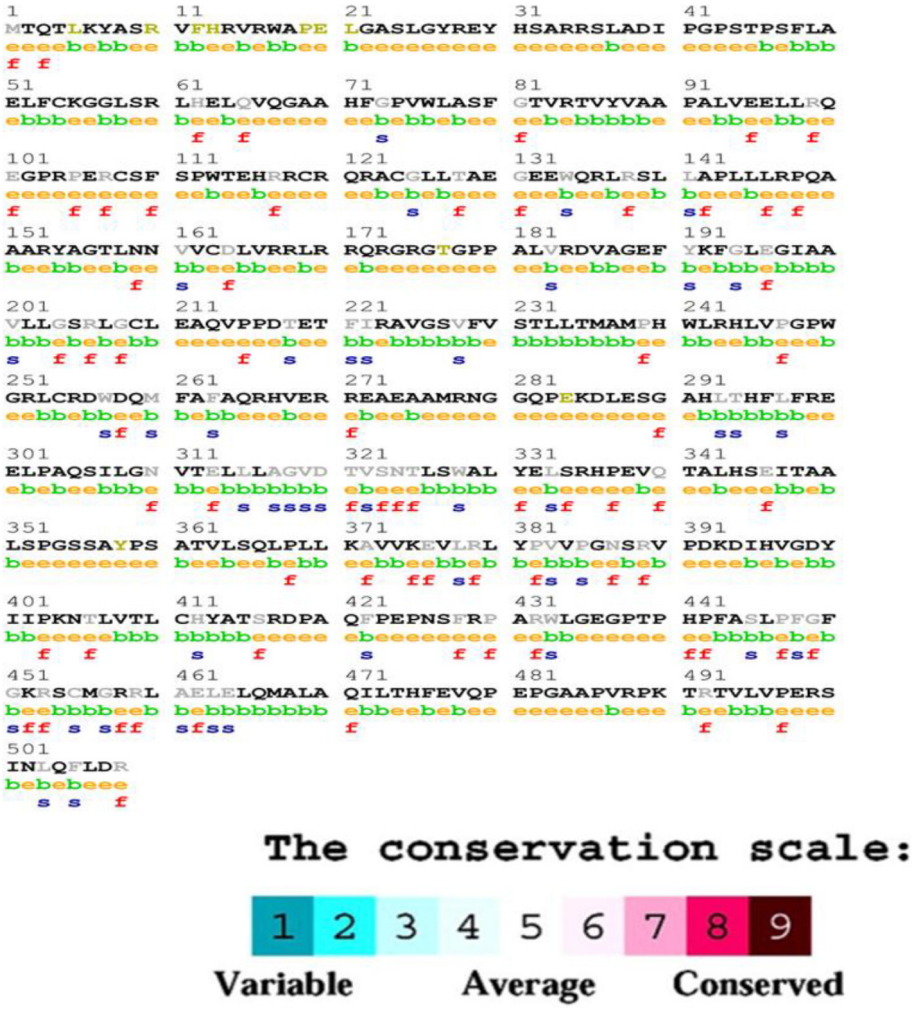
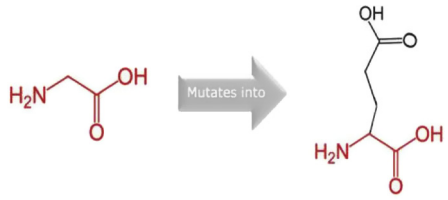
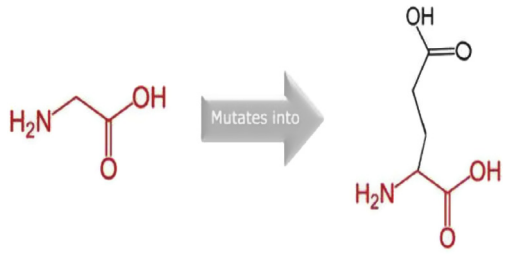
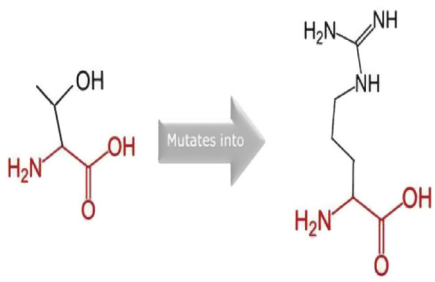
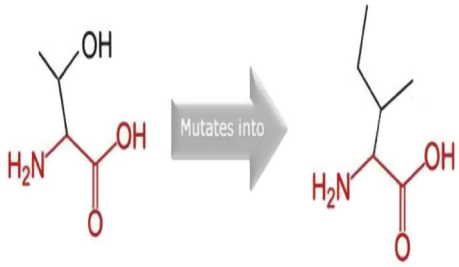
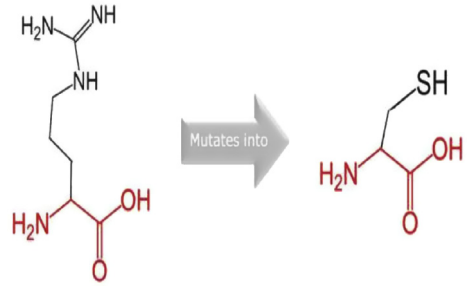
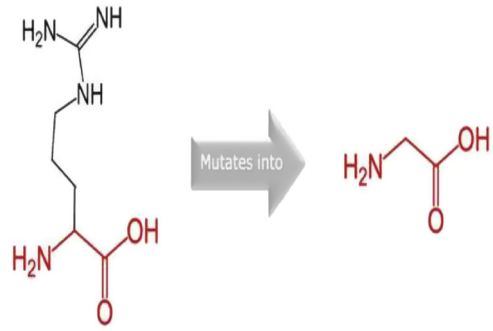


Fig. 2. Results from the Analysis of human CYP27B1 protein by ConSurf [10].

Table 6
Schematic structures of the original (left) and mutant (right) amino acid for each Mutation.

Residue	Structure	Properties
G73E		- the mutant residue is bigger than the wild-type residue. - the wild-type residue is charged neutral; the mutant residue is charged negative. - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is much conserved; the mutant residue is located near a highly conserved position.
G102E		- the mutant residue is bigger than the wild-type residue - the wild-type residue is charged neutral, the mutant residue is charged negative - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is much conserved, the mutant residue is located near a highly conserved position
T321R		- the mutant residue is bigger than the wild-type residue - the wild-type residue is charged neutral, the mutant residue is charged positive - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is much conserved, the mutant residue is located near a highly conserved position
T409I		- the mutant residue is bigger than the wild-type residue - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is not conserved, the mutant residue is among the observed residue types at this position.
R389C		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is very conserved, the mutant residue was not observed at this position.
R389G		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is very conserved, the mutant residue was not observed at this position.

(*continued on next page*)

Table 6 (*continued*)

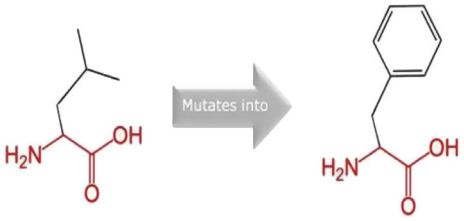
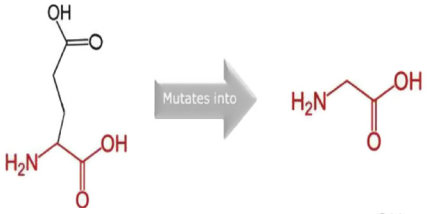
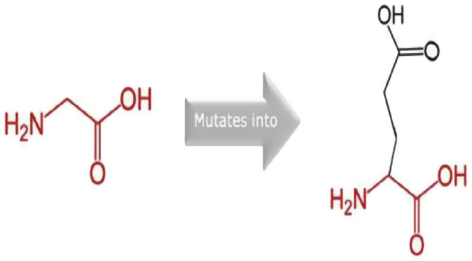
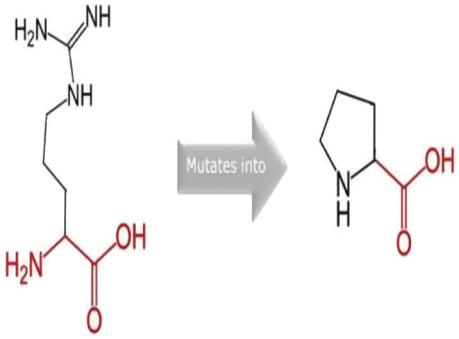
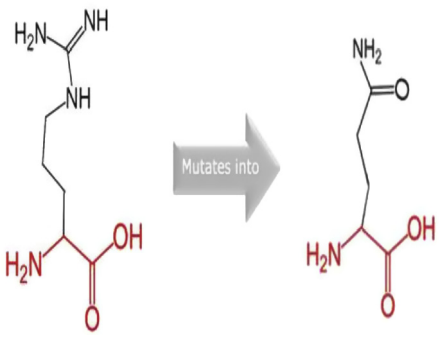
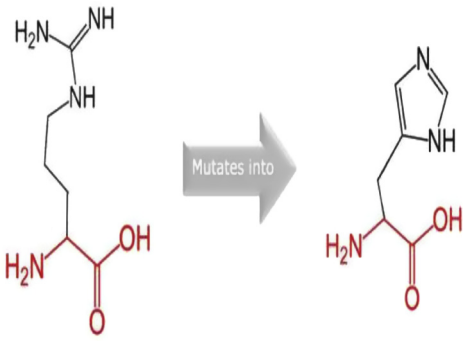
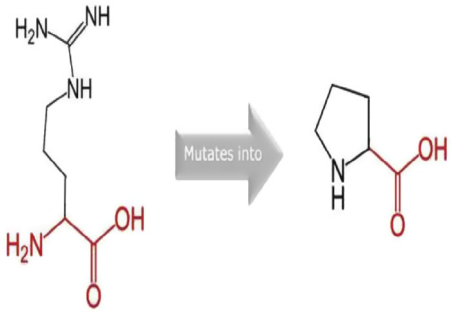
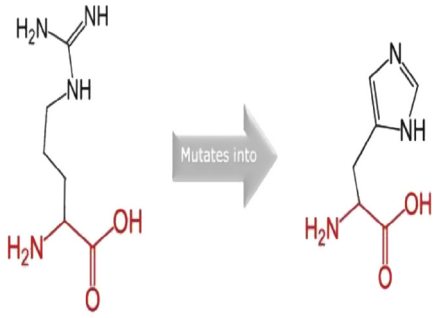
Residue	Structure	Properties
L343F		- the mutant residue is bigger than the wild-type residue - the wild type residue is very conserved, the mutant residue was not observed at this position.
E189G		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged negative, the mutant residue is charged neutral - the wild-type reside is more hydrophobic than the mutant residue
G125E		- the mutant residue is bigger than the wild-type residue - the wild-type residue is charged neutral, the mutant residue is charged negative - the wild-type reside is more hydrophobic than the mutant residue. - the mutant and wild-type residue are not very similar. based on this conservation information, this mutation is probably damaging to the protein.
R335P		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild-type reside is more hydrophobic than the mutant residue. - the wild type residue is observed alongside other residues, the mutant residue was not observed at this position.
R335Q		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild type residue is observed alongside other residues; the mutant residue was also observed amongst other residues at this position. This means the mutation can occur at this position and is probably not damaging to the protein.
R429H		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild type residue is not conserved at this region, - the mutant residue is located near a highly conserved position.

Table 6 (*continued*)

Residue	Structure	Properties
R429P		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild-type reside is more hydrophobic than the mutant residue - the wild type residue is not conserved at this region, - the mutant residue is located near a highly conserved position.
R453H		- the mutant residue is smaller than the wild-type residue - the wild-type residue is charged positive, the mutant residue is charged neutral. - the wild type residue is very conserved at this region, but a few other residue types have been observed at this position too, - the mutant residue was not observed at this position or any position. this mutation is probably damaging to the protein.

Data obtained from project HOPE [18].

Table 7
MutPred analysis of amino acid substitutions related to diseases and phenotypes.

SNPs	Actionable/Confident Hypothesis	MutPred2 score	p-value
G73E	Altered DNA binding	0.914	0.03
	Altered Transmembrane protein		0.03
G102E	Gain of Intrinsic disorder	0.702	0.03
	Gain of B-factor		0.02
T321R	Altered Metal binding	0.954	1.40E−03
	Gain of Helix		0.02
	Loss of Loop		0.01
	Altered Transmembrane protein		3.90E−03
	Gain of Allosteric site at T321		0.04
	Loss of Catalytic site at D320		0.04
	Loss of GPI-anchor amidation at N324		0.03
T409I	Altered Ordered interface	0.558	7.40E−03
	Altered Metal binding		0.05
R389G	Loss of Allosteric site at R389	0.958	4.90E−06
	Altered Ordered interface		5.70E−04
	Altered Metal binding		0.01
	Loss of Catalytic site at R389		7.00E−03
	Gain of Ubiquitylation at K393		0.02
	Altered Transmembrane protein		0.02
	Altered Stability		0.12
	Gain of GPI-anchor amidation at N387		0.02
L343F	None	0.487	–
E189G	Loss of Acetylation at K192	0.775	0.02
	Gain of Intrinsic disorder		0.04
	Loss of Loop	0.91	0.02
	Gain of Strand		0.03
R335P	Altered Metal binding	0.727	5.90E−03
R335Q	None	0.19	–
R429H	None	0.064	–
R429P	None	0.47	
R453H	Altered Metal binding	0.893	0.01
	Loss of Catalytic site at R453		2.00E−03
	Loss of Allosteric site at R453		4.90E−03
	Loss of Helix		0.04
	Loss of Acetylation at K452		0.01
	Altered DNA binding		0.05

Data obtained from MutPred [13].

Table 8
nsSNPs analysis by ModPred.

Residue	Modification	Score	Confidence	Remarks
T409I	Proteolytic Cleavage	0.54	Low	Novel Prediction
R389C	ADP-Ribosylation	0.67	Medium	Novel Prediction
R389G	Proteolytic Cleavage	0.58	Low	Novel Prediction
E189G	Proteolytic Cleavage	0.58	Low	Novel Prediction
G125E	Proteolytic Cleavage	0.58	Low	Novel Prediction
R335P	ADP-Ribosylation	0.67	Medium	Novel Prediction
R335Q	ADP-Ribosylation	0.67	Medium	Novel Prediction
R429H	ADP-Ribosylation	0.60	Low	Novel Prediction
R429P	ADP-Ribosylation	0.60	Low	Novel Prediction
R453H	Proteolytic Cleavage	0.78	Medium	Novel Prediction

Data obtained from ModPred [39].

Table 9
I-Mutant results for carefully chosen nsSNPs in the CYP27B1 gene.

nsSNP ID	Mutation	DDG	Sign of DDG
rs1057520815	G73E	−0.25	Increase (3)
	G102E	−0.30	Increase (4)
rs118204007	T86R	−1.46	Decrease (9)
	T321R	−0.37	Decrease (2)
rs118204008	T409I	−0.44	Decrease (6)
rs118204010	R154C	−1.05	Decrease (3)
	R389G	−1.50	Decrease (8)
rs118204011	L108F	0.03	Increase (1)
	L343F	−0.87	Decrease (5)
rs118204012	E160G	−1.27	Decrease (9)
	E189G	−1.46	Decrease (9)
rs28934605	G125E	−0.30	Decrease (2)
rs28934606	R335P	−0.34	Decrease (4)
	R100Q	−0.40	Decrease (2)
	R335Q	−0.66	Decrease (7)
rs28934607	P147S	−0.82	Decrease (8)
	P382S	−1.55	Decrease (8)
rs568165894	R429H	−1.11	Decrease (8)
	R429P	−0.59	Decrease (5)
rs759208930	R453H	−1.30	Decrease (9)

Data obtained from I-Mutant [15].

other sites required for protein-protein interactions. Some of these amino acids lean towards being more conserved than other residues. NsSNPs situated at extremely conserved amino acid locations are more deleterious than nsSNPs at non-conversed locations. For advance examination of the possible effects of the high-risk nsSNPs, the degree of evolutionary conservation was calculated using the ConSurf web server for amino acids human CYP27B1 protein [10]. For this study, there was more concentration on amino acid sites that match locations where the high-risk nsSNPs are present; however, ConSurf also acknowledges some other residues that may have purposeful significance.

As presented in Table 5 and Fig. 2, ConSurf analysis revealed that residues G73, T321, T409, R389, L343, G125, P382, R453 are highly conserved (Conservation Score of 7–9) and predicted less conserved residues E189, R429 which were exposed (Conservation Score of 5). Furthermore, ConSurf prediction gave G125 as highly conserved and buried, which makes it an important structural residue, and that T321, R389, P382, R453 were functionally significant residues (highly conserved and exposed). The location of the highly conserved residues to the protein surface or protein core helps in predicting their structural or functional base. Interestingly, five of the high-risk nsSNPs that were predicted to be deleterious by the SNP prediction algorithms (P382, R453, R389, T321 and G125) were also recognized as significant structural or functional residues by ConSurf. Our data strongly backs up the claim that these nsSNPs are deleterious to CYP27B1 structure and/or function.

Furthermore, the project HOPE was used to investigate the structural effects of these amino acid substitutions. Its results showed that T321, R389 and G125 are highly conserved and their substitutions are probably damaging to the protein structure. It is noteworthy that the germline mutation at these positions have been reported to be responsible for abolishing enzymatic activity of CYP27B1 leading to Vitamin D-dependent hereditary rickets type 1 [5,26–28]. For example, the R389G mutation is a change to smaller amino acid while T321R and G125E, were changes to amino acid bigger than the residue in wild type. All the three substitution resulted in change in the net charge of CYP27B1 protein (Table 6). It is known that protein charge and mass affect spatio-temporal dynamics of protein-protein interaction [29,30]; hence, these changes could alter the ability of CYP27B1 to interact the other proteins.

3.4. Identification of disease phenotype related with nsSNPs in human CYP27B1

Phenotypic analysis of the amino acid replacements accounted as ‘pathogenic’ was done using MutPred. The MutPred is a web-tool that predicts diseased phenotype and also identifies the molecular mechanisms that result from amino acid replacement instigated by nsSNPs. Subsequently, valuable data such as addition or loss of solvent accessibility is also provided by MutPred. The g-value and p-value scores of the amino acid substitution are shown in Table 7. A g-value score greater than 0.5 and p-value scores less than 0.05 are referred to as actionable hypotheses, whereas the scores with g-value greater than 0.75 and p-value less than 0.05 are referred to as confident hypotheses. Of interest were the mutations of T321R, R389G and G125E, which were associated with altered metal binding, loss of allosteric site and gain of intrinsic disorder, respectively.

3.5. Prediction of post translational modification sites for human CYP27B1 protein

ModPred tool was used to analyse the effect of nsSNPs on post translational modification process in human CYP27B1 protein. ModPred predicted sites for proteolytic cleavage and ADP-Ribosylation as presented in Table 8.

The mutations at these conserved locations affect the posttranslational modification of the protein; particularly, R389C and G125E could result in the loss of ADP-ribosylation and proteolytic cleavage respectively. The loss of posttranslational modification sites has deleterious clinical implications. For example, ADP-ribosylation is involved in cellular processes like cell signalling and apoptosis, and its alteration has been implicated in carcinogenesis [31]. It is therefore possible that these mutations could account for alteration on vitamin D signalling that has been reported in some cancer [32–34].

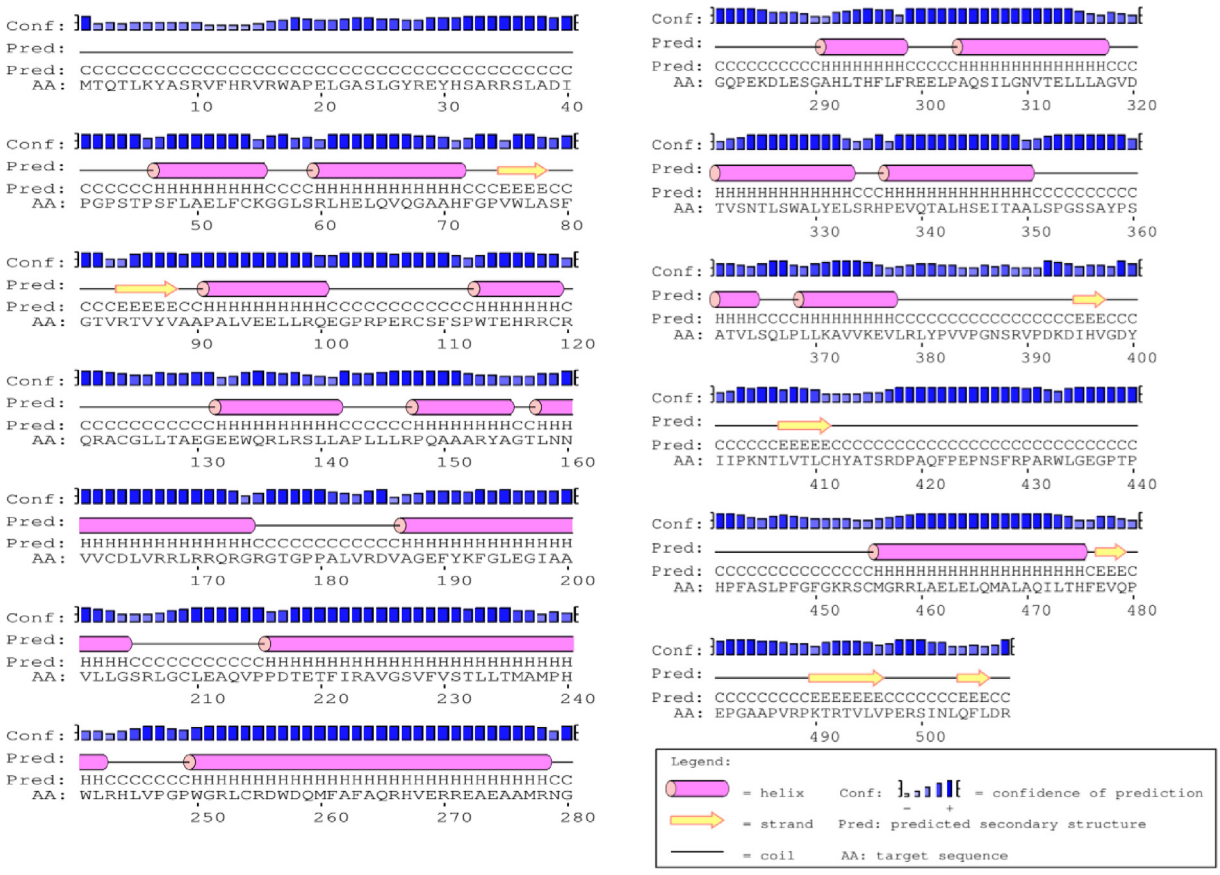


Fig. 3. Secondary structure of CYP27B1 showing beta helix, strand and coil. The results highlight a mix distribution of coil, alpha helix and strands. The coil was observed to be the major secondary structural motif among various structural elements (51%), followed by helix (35%) and strand (7%), as generated by PSIPRED [14].

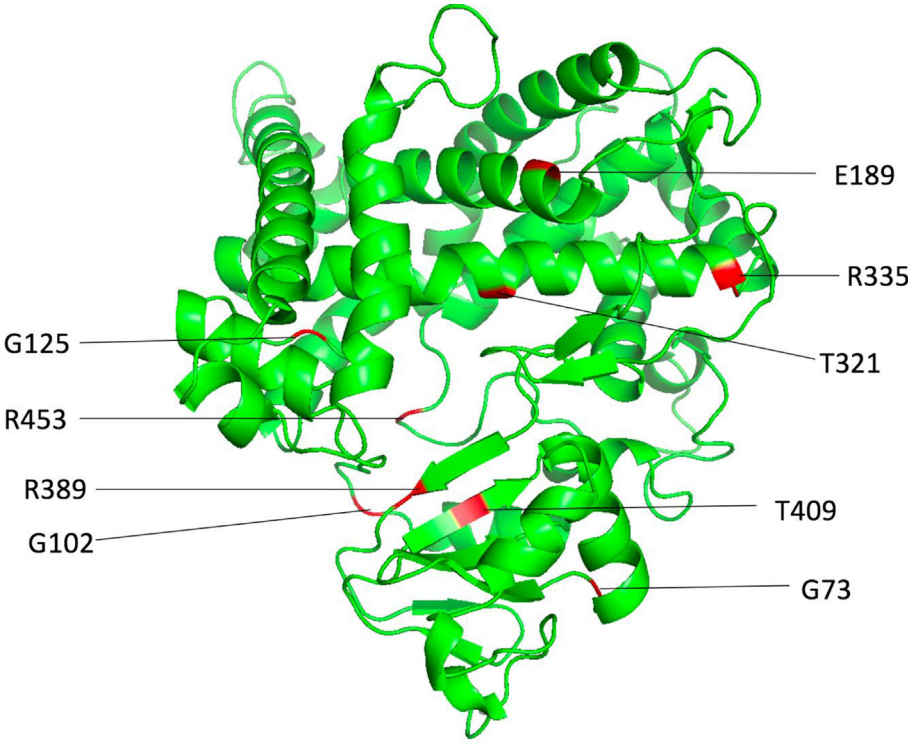


Fig. 4. Modelled 3D structure of CYP25B1. The positions of mutation with significant disease-related phenotypes are indicated in red, as generated from SWISS-MODEL [19].

3.6. Effects of mutation on protein stability and protein-protein interaction of human CYP27B1

There is a consensus that most diseases associated with nsSNPs affect the protein stability. The effect of mutation on the human CYP27B1 protein stability was investigated using the I-Mutant server. I-Mutant utilizes the structure based analysis of the mutant protein to deliver an estimate of the free energy change in mutant protein with amino acid replacement at a single location (Table 9). The I-Mutant results indicated that almost all the amino acid substitution studied will decrease the degree of free energy change value of the protein.

The predicted tertiary structure of CYP27B1 protein showed that all the amino acid substitutions investigated resulted in varying degrees of disruption to the tertiary structure of CYP27B1 (Figs. 3 and 4 and Table 10). For example, T321R and R389G would decrease the solvation and increase torsion of CYP28B1. Protein solvation and torsion are known to affect the energetically favourable conformation and chain elasticity of proteins [35] and they are used to predict the stability of mutant proteins [36]. These physical changes to the protein structure consequently affects the molecular interaction ability of its side chains [36].

Optimum protein-protein interaction is important to allow for metabolite interaction and maintenance of normal cellular function [29]. However, it is worthy of note that protein-protein interaction occurs at the level of protein tertiary structure and this makes it important to predict the impact of nsSNPs on tertiary structure of CYP27B1.

Therefore, these pathogenic mutations in the human CYP27B1 could result in altered stability and metal binding, as well as loss

Table 10
Comparative effects of disease-related mutations on properties of 3D structure of CYP27B1.

SNPs	QMEAN	Cβ	All Atom	Solvation	Torsion
WILD	−2.46	−1.45	−0.94	0.72	−2.28
G73E	−2.80	−1.58	−0.89	0.74	−2.61
G102E	−2.04	−2.35	−0.53	0.70	−1.77
T321R	−2.14	−2.27	−0.48	0.66	−1.89
T409I	−2.11	−2.45	−0.62	0.51	−1.79
R389G	−2.09	−2.30	−0.58	0.65	−1.82
E189G	−2.09	−2.34	−0.61	0.57	−1.80
G125E	−2.02	−1.37	−0.49	0.80	−1.90
R335P	−2.44	−1.42	−1.09	0.58	−2.24
R453H	−2.55	−1.25	−0.86	0.59	−2.40

Data generated from SWISS-MODEL [40].

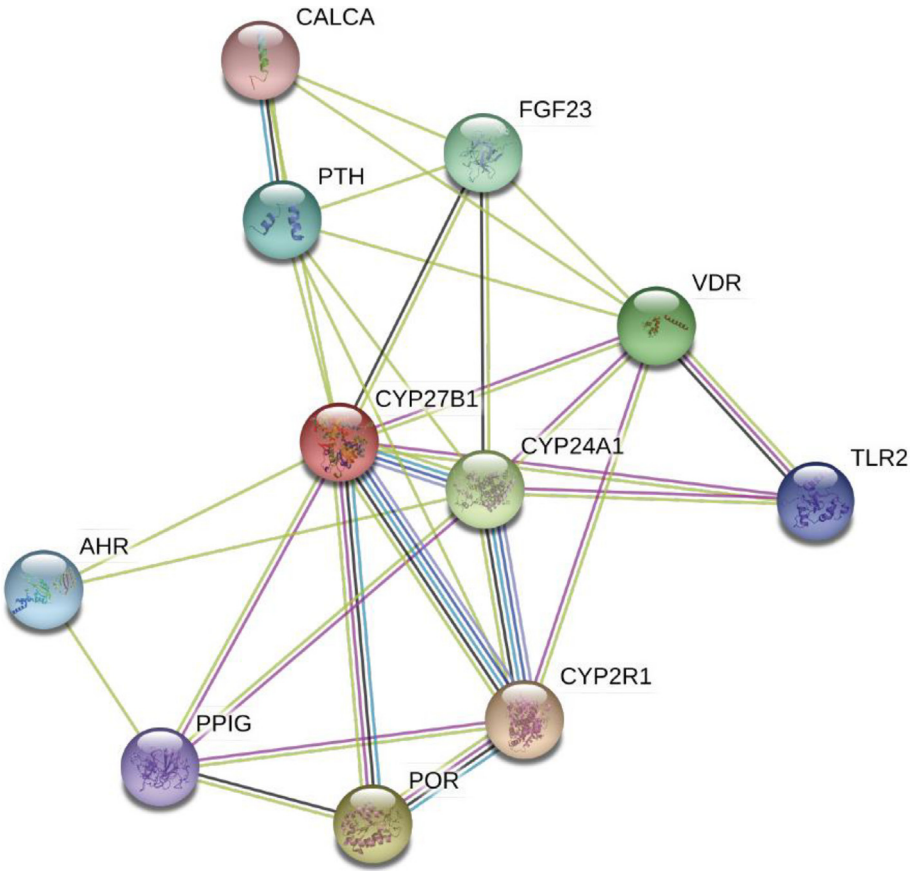


Fig. 5. Protein–protein interaction network of CYP27B1 protein shown by STRING. STRING database is used to predict functional interaction between the proteins in the cell, using STRING [20].

of allosteric and catalytic residue with a concomitant loss of enzyme activity. Furthermore, we used the STRING maps to investigate the protein–protein interaction of CYP27B1 (Fig. 5). This is useful in understanding the genotype-phenotype consequence of the mutation [37].

STRING database was used to predict functional interaction between the proteins in the cell. STRING results predicted the functional association partner of CYP27B1 protein with CYP24A1 (Cytochrome P450, family 24, subfamily A, polypeptide 1), CALCA (Calcitonin-related polypeptide alpha), FGF23 (Fibroblast growth factor 23), PTH (Parathyroid hormone), VDR (Vitamin D (1, 25-dihydroxyvitamin D3) receptor; Nuclear hormone receptor), TLR2 (Toll-like receptor 2), CYP2R1 (Cytochrome P450, family 2, subfamily R, polypeptide 1), POR (P450 (cytochrome) oxidoreductase), PPIG (Peptidylprolyl isomerase G (cyclophilin G)) and AHR (Aryl hydrocarbon receptor; Ligand-activated transcriptional activator). Out of these proteins, CYP24A1, CYP2R1 and VDR directly affect the level of circulating synthesis and actions of 1-α, 25 dihydroxyvitamin D_3. This shows that the consequence of CYP27B1 mutation could affect other proteins involved in vitamin D metabolism and function.

4. Conclusion

Out of the 411 nsSNPs found in human CYP27B1, 11 have significant disease related phenotype. Out of these, three-rs118204007, rs118204010 and rs28934605 affect conserved positions. The amino acid substitutions R389C and G125E affect posttranslational modification sites. These mutations can consequently affect the function of CYP27B1 resulting in vitamin D deficiency or insufficiency; thereby, leading to diverse pathological conditions like rickets, diabetes and various cancer.

References

[1] Dakal TC, Kala D, Dhiman G, Yadav V, Krokhotin A, Dokholyan NV. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. Sci Rep 2017;7(1):6525. doi: https://doi.org/10.1038/s41598-017-06575-4. PubMed PMID: 28747718; PubMed Central PMCID: PMCPMC5529537.

[2] Sawada N, Sakaki T, Kitanaka S, Takeyama K, Kato S, Inouye K. Enzymatic properties of human 25-hydroxyvitamin D3 1alpha-hydroxylase coexpression with adrenodoxin and NADPH-adrenodoxin reductase in Escherichia coli. Eur J Biochem 1999;265(3):950–6. PubMed PMID: 10518789.

[3] Holick MF. Vitamin D status: measurement, interpretation, and clinical application. Ann Epidemiol 2009;19(2):73–8. doi: https://doi.org/10.1016/j.annepidem.2007.12.001. PubMed PMID: 18329892; PubMed Central PMCID: PMCPMC2665033.

[4] Holick MF. Vitamin D: the other steroid hormone for muscle function and strength. Menopause 2009;16(6):1077–8. doi: https://doi.org/10.1097/gme.0b013e3181bd9804. PubMed PMID: 19801955.

[5] Kato S, Yanagisawa J, Murayama A, Kitanaka S, Takeyama K. The importance of 25-hydroxyvitamin D3 1 alpha-hydroxylase gene in vitamin D-dependent rickets. Curr Opin Nephrol Hypertens 1998;7(4):377–83. PubMed PMID: 9690035.

[6] Ahmad S, Chowdhury TA, Boucher BJ. Diabetes and cancer: could vitamin D provide the link? J Diabetes Complications 2013;27(2):184–90. doi: https://doi.org/10.1016/j.jdiacomp.2012.10.005. PubMed PMID: 23164631.

[7] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucl Acids Res 2001;29(1):308–11. PubMed PMID: 11125122; PubMed Central PMCID: PMCPMC29783.

[8] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Meth 2010;7(4):248–9. doi: https://doi.org/10.1038/nmeth0410-248. PubMed PMID: 20354512; PubMed Central PMCID: PMCPMC2855889.

[9] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073–81. doi: https://doi.org/10.1038/nprot.2009.86. PubMed PMID: 19561590.

[10] Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N, et al. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl Acids Res 2010;38(Web Server issue). pp. W529-33, doi: 10.1093/nar/gkq399. PubMed PMID: 20478830; PubMed Central PMCID: PMCPMC2896094.

[11] Mayrose I, Graur D, Ben-Tal N, Pupko T. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. Mol Biol Evol 2004;21(9):1781–91. doi: https://doi.org/10.1093/molbev/msh194. PubMed PMID: 15201400.

[12] Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 2002;18 (Suppl 1):S71–7. PubMed PMID: 12169533.

[13] Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009;25(21):2744–50. doi: https://doi.org/10.1093/bioinformatics/btp528. PubMed PMID: 19734154; PubMed Central PMCID: PMCPMC3140805.

[14] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25(17):3389–402. PubMed PMID: 9254694; PubMed Central PMCID: PMCPMC146917.

[15] Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22 (22):2729–34. doi: https://doi.org/10.1093/bioinformatics/btl423. PubMed PMID: 16895930.

[16] Mavroconstanti T, Johansson S, Winge I, Knappskog PM, Haavik J. Functional properties of rare missense variants of human CDH13 found in adult attention deficit/hyperactivity disorder (ADHD) patients. PLoS One 2013;8(8):e71445. doi: https://doi.org/10.1371/journal.pone.0071445. PubMed PMID: 23936508; PubMed Central PMCID: PMCPMC3731280.

[17] Pejaver RK, Nisarga R, Gowda B. Temperature monitoring in newborns using thermospot. Indian J Pediatr 2004;71(9):795–6. PubMed PMID: 15448385.

[18] Venselaar H, Te BeekA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinf 2010;11:548. doi: https://doi.org/10.1186/1471-2105-11-548. PubMed PMID: 21059217; PubMed Central PMCID: PMCPMC2992548.

[19] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucl Acids Res 2014;42(Web Server issue). pp. W252-8, doi: 10.1093/nar/gku340. PubMed PMID: 24782522; PubMed Central PMCID: PMCPMC4086089.

[20] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucl Acids Res 2011;39(Database issue). pp. D561-8, doi: 10.1093/nar/gkq973. PubMed PMID: 21045058; PubMed Central PMCID: PMCPMC3013807.

[21] Jacobs ET, Van Pelt C, Forster RE, Zaidi W, Hibler EA, Galligan MA, et al. CYP24A1 and CYP27B1 polymorphisms modulate vitamin D metabolism in colon cancer cells. Cancer Res 2013;73(8):2563–73. doi: https://doi.org/10.1158/0008-5472.CAN-12-4134. PubMed PMID: 23423976; PubMed Central PMCID: PMCPMC3630267.

[22] Lopez ER, Regulla K, Pani MA, Krause M, Usadel KH, Badenhoop K. CYP27B1 polymorphisms variants are associated with type 1 diabetes mellitus in Germans. J Steroid Biochem Mol Biol 2004;89–90(1–5):155–7. doi: https://doi.org/10.1016/j.jsbmb.2004.03.095. PubMed PMID: 15225764.

[23] Agnello L, Scazzone C, Lo Sasso B, Bellia C, Bivona G, Realmuto S, et al. VDBP, CYP27B1, and 25-Hydroxyvitamin D Gene Polymorphism Analyses in a Group of Sicilian Multiple Sclerosis Patients. Biochem Genet 2017;55(2):183–92. doi: https://doi.org/10.1007/s10528-016-9783-4. PubMed PMID: 27904983.

[24] Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum Mutat 2009;30(5):703–14. doi: https://doi.org/10.1002/humu.20938. PubMed PMID: 19267389.

[25] 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):661–8. doi: https://doi.org/10.1002/humu.21490. PubMed PMID: 21480434; PubMed Central PMCID: PMCPMC4154965.

[26] Kato S. Genetic mutation in the human 25-hydroxyvitamin D3 1alpha-hydroxylase gene causes vitamin D-dependent rickets type I. Mol Cell Endocrinol 1999;156(1–2):7–12. PubMed PMID: 10612418.

[27] Kitanaka S, Murayama A, Sakaki T, Inouye K, Seino Y, Fukumoto S, et al. No enzyme activity of 25-hydroxyvitamin D3 1alpha-hydroxylase gene product in pseudovitamin D deficiency rickets, including that with mild clinical manifestation. J Clin Endocrinol Metab 1999;84(11):4111–7. doi: https://doi.org/10.1210/jcem.84.11.6131. PubMed PMID: 10566658.

[28] Kitanaka S, Takeyama K, Murayama A, Sato T, Okumura K, Nogami M, et al. Inactivating mutations in the 25-hydroxyvitamin D3 1alpha-hydroxylase gene in patients with pseudovitamin D-deficiency rickets. N Engl J Med 1998;338 (10):653–61. doi: https://doi.org/10.1056/NEJM199803053381004. PubMed PMID: 9486994.

[29] Xu Y, Wang H, Nussinov R, Ma B. Protein charge and mass contribute to the spatio-temporal dynamics of protein-protein interactions in a minimal proteome. Proteomics 2013;13(8):1339–51. doi: https://doi.org/10.1002/pmic.201100540. PubMed PMID: 23420643; PubMed Central PMCID: PMCPMC3762602.

[30] Peleg O, Choi JM, Shakhnovich EI. Evolution of specificity in protein-protein interactions. Biophys J 2014;107(7):1686–96. doi: https://doi.org/10.1016/j.

bpj.2014.08.004. PubMed PMID: 25296322; PubMed Central PMCID: PMCPMC4190652.
[31] Scarpa ES, Fabrizio G, Di Girolamo M. A role of intracellular mono-ADP-ribosylation in cancer biology. FEBS J 2013;280(15):3551–62. doi: https://doi.org/10.1111/febs.12290. PubMed PMID: 23590234.
[32] Chen TC. 25-Hydroxyvitamin D-1 alpha-hydroxylase (CYP27B1) is a new class of tumor suppressor in the prostate. Anticancer Res 2008;28(4A):2015–7. PubMed PMID: 18649741.
[33] Kang W, Lee S, Jeon E, Yun YR, Kim KH, Jang JH. Emerging role of vitamin D in colorectal cancer. World J Gastrointest Oncol 2011;3(8):123–7. doi: https://doi.org/10.4251/wjgo.v3.i8.123. PubMed PMID: 22007275; PubMed Central PMCID: PMCPMC3192221.
[34] Gupta D, Vashi PG, Trukova K, Lis CG, Lammersfeld CA. Prevalence of serum vitamin D deficiency and insufficiency in cancer: review of the epidemiological literature. Exp Ther Med 2011;2(2):181–93. doi: https://doi.org/10.3892/etm.2011.205. PubMed PMID: 22977487; PubMed Central PMCID: PMCPMC3440651.
[35] Sorokina I, Mushegian A. The role of the backbone torsion in protein folding. Biol Direct 2016;11(1):64. doi: https://doi.org/10.1186/s13062-016-0166-x. PubMed PMID: 27906033; PubMed Central PMCID: PMCPMC5133741.
[36] Parthiban V, Gromiha MM, Hoppe C, Schomburg D. Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility. Proteins 2007;66(1):41–52. doi: https://doi.org/10.1002/prot.21115. PubMed PMID: 17068801.
[37] Mohamoud HS, Hussain MR, El-Harouni AA, Shaik NA, Qasmi ZU, Merican AF, et al. First comprehensive in silico analysis of the functional and structural consequences of SNPs in human GalNAc-T1 gene. Comput Math Methods Med. 2014;2014:904052. doi: https://doi.org/10.1155/2014/904052. PubMed PMID: 24723968; PubMed Central PMCID: PMCPMC3960557.
[38] Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucl Acids Res 2018;46(D1). doi: https://doi.org/10.1093/nar/gkx1098. PubMed PMID: 29155950; PubMed Central PMCID: PMCPMC5753206. D754-D61.
[39] Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, Radivojac P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci 2014;23(8):1077–93. doi: https://doi.org/10.1002/pro.2494. PubMed PMID: 24888500; PubMed Central PMCID: PMCPMC4116656.
[40] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22(2):195–201. doi: https://doi.org/10.1093/bioinformatics/bti770. PubMed PMID: 16301204.