Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2

Achintya Mohan Goswami *
Department of Physiology, Krishnagar Govt. College, Krishnagar, Nadia, West Bengal, India

A R T I C L E   I N F O
Article history:
Received 24 May 2015
Accepted 23 July 2015
Available online 8 August 2015

Keywords:
sSNP
3β-Hydroxyosteroid dehydrogenases/Δ4,5-isomerase type 2
In silico
Homology modeling

A B S T R A C T
Single-nucleotide polymorphisms (SNPs), a most common type of genetic mutations, result from single base pair alterations. Non-synonymous SNPs (nsSNP) occur in the coding regions of a gene and result in single amino acid substitution which might have the potential to affect the function as well as structure of the corresponding protein. In human the 3β-hydroxysteroid dehydrogenases/Δ4,5-isomerase type 2 (HSD3B2) is an important membrane-bound enzyme involved in the dehydrogenation and Δ4,5-isomerization of the Δ5-steroid precursors into their respective Δ4-ketosteroids in the biosynthesis of steroid hormones such as glucocorticoids, mineralocorticoids, progesterone, androgens, and estrogens in tissues such as adrenal gland, ovary, and testis. Most of the nsSNPs of HSD3B2 are still uncharacterized in terms of their disease causing potential. So, this study has been undertaken to explore and extend the knowledge related to the effect of nsSNPs on the stability and function of the HSD3B2. In this study sixteen nsSNP of HSD3B2 were subjected to in silico analysis using nine different algorithms: SIFT, PROVEAN, PolyPhen, MutPred, SNPeffect, PhD SNP, ssSNP, and I Mutant 2.0. The results obtained from the analysis revealed that the prioritization of diseases associated amino acid substitution as evident from possible alteration in structure–function relationship. Structural phylogenetic analysis using ConSurf revealed that the functional residues are highly conserved in human HSD3B2; and most of the disease associated nsSNPs are within these conserved residues. Structural theoretical models of HSD3B2 were created using HHPreed, Phyre2 and RaptorX server. The predicted models were evaluated to get the best one for structural understanding of amino acid substitutions in three dimensional spaces.

© 2015 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction
Single-nucleotide polymorphisms (SNPs) are the most common type of genetic mutations which alter single base pair in alleles either in or between individuals. There is an exponential expansion of SNPs in databases due to the development of new techniques for the large-scale identification of SNPs in the human genome (Wang et al., 1998). There are several publicly available databases for SNPs, such as dbSNP, GWAS Central, and SwissVar. By release of 135 hosting number of human SNPs reached more than 50 million, including 535,660 synonymous and 873,308 non-synonymous SNPs (Luu et al., 2012). The major goal of mining this database is to find the relevance of these genetic variations and genotypes; thus providing a basis for the mechanisms of and therapies for human diseases (Syvanen, 2001).

Non-synonymous single nucleotide polymorphisms (nsSNP) occur in the coding regions of a gene and result in single amino acid substitution that has the potential to alter the function of the corresponding protein, either directly or via disruption of structure leading to pathogenic phenotypes (Capriotti and Altman, 2011). Hence nsSNP are of particular interest as candidates for further assessment. There are many diseases associated with nsSNPs, such as the well-known sickle-cell anemia (Noguchi and Schechter, 1985); rheumatoid arthritis, which is caused by the dysfunction of the protein tyrosine phosphatase (Begovich et al., 2004); Li-Fraumeni syndrome (Ruijs et al., 2006) and congenital cataract-microcornea syndrome (Wang et al., 2011) and so on.

In human the 3β-hydroxysteroid dehydrogenases (HSD3B)/Δ4,5-isomerase is an important enzyme involved in the dehydrogenation and Δ4,5-isomerization of the Δ5-steroid precursors into their respective Δ4-ketosteroids (Mason et al., 1997). The enzyme uses dehydroepiandrosterone (DHEA) as substrate in a two-step reaction catalyzed by HSD3B. During this reaction, the reduction of NAD+ to NADH occurs by a rate-limiting activity of HSD3B followed by the NADH recruitment for the activation of isomerase activity on the same enzyme. This activity is crucial for the synthesis of hormonal steroids, which includes aldosterone, cortisol, and testosterone (Thomas et al., 2003).

* Corresponding author.
E-mail address: achintya02@gmail.com.
There are two HSD3B isoenzymes designated as type 1 and type 2. These isoenzymes are 93.5% homologous and two different genes located on chromosome 1p13.1 encode these two isozymes (Rheumae et al., 1991). The type 1 gene, called HSD3B1, is almost exclusively expressed in the placenta and peripheral tissues, including prostate, mammary gland, and skin. The type 2 gene, called HSD3B2, is predominantly expressed in adrenal gland, ovary, and testis (Simard et al., 1996).

HSD3B2 is present as membrane-bound form and it is involved in the biosynthesis of steroid hormones such as glucocorticoids, mineralocorticoids, progestrogens, androgens, and estrogens in tissues such as adrenal gland, ovary, and testis (Simard et al., 1996). In the adrenal, HSD3B2 is the key enzyme required for the synthesis of cortisol and aldosterone (Rainey et al., 2002). It is well known that these steroid hormones play an important role in various physiological processes (Labrie et al., 1992).

It was observed that a number of mutations in HSD3B2 gene has been found to cause congenital adrenal hyperplasia (CAH; OMIM # +201,810), an autosomal recessive disease that impairs steroidogenesis in both the adrenals and gonads (Rheumae et al., 1992; Simard et al., 1993; Pang, 2001; Simard et al., 2002). The clinical manifestation of HSD3B2 deficiency ranges from salt-losing to non-salt-losing forms in both sexes. In newborns, HSD3B2 deficiency results in ambiguity of the external genitalia in genetic males, while affected females exhibit normal sexual differentiation or partial virilization. During adolescence, HSD3B2 deficiency results in variable degrees of hypogonadism in boys and hyperandrogenism (premature pubarche and hirsutism) in girls. Again a study has provided insight regarding the structure–function relationship of HSD3B2 mutants at codon 222 using molecular homology modeling. Their findings suggest an important catalytic role of amino acid at codon 222 (Lusa et al., 2010).

Most of the nsSNPs of HSD3B2 are still uncharacterized in terms of their disease causing potential. So, this study has been undertaken to explore and extend the knowledge related to the effect of nsSNPs on the stability and function of the HSD3B2. Here we have used an effective set of computational techniques to prioritize the most deleterious nsSNPs reported in the HSD3B2 gene. The future of SNP analysis greatly lies in the development of personalized medicines that can facilitate the treatment of genomic variation induced disorders at a higher extent (Sherry et al., 2001).

2. Materials and methods

2.1. Datasets

HSD3B2 gene SNPs and their protein sequences in the FASTA format were retrieved from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) for computational analysis in this study (Bhagwat, 2010).

2.2. Sequence homology-based prediction of deleterious nsSNPs by using SIFT

The Sorting Intolerant from Tolerant (SIFT) server available at (http://sift.jcvi.org) was used to predict the deleterious non-synonymous SNPs (Ng and Henikoff, 2003). The SIFT program utilizes amino acid sequence homology and the physical properties of the proteins in combination with naturally occurring nsSNPs by aligning paralogous and orthologous protein sequences for the prediction of functional consequences of nsSNPs. The threshold for the intolerance index is ≥0.05.

2.3. Predicting the functional effect of nsSNPs by PROVEAN

PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org) is a sequence based predictor that estimates the effect of protein sequence variation on protein function (Choi et al., 2012). It is based on a clustering method where BLAST hits with more than 75% global sequence identity clustered together and top 30 of such clusters from a supporting sequence are averaged within and across clusters to generate the final PROVEAN score. A protein variant is predicted to be “Deleterious” if the final score is below a certain threshold (default is −2.5), and is predicted to be “Neutral” if the score is above the threshold.

2.4. Structural homology-based prediction of functional consequences of coding nsSNPs by using PolyPhen

The PolyPhen server (http://genetics.bwh.harvard.edu/pph/) was used to study the functional consequences of nsSNPs (Ramsden et al., 2002; Adzhubei et al., 2010). Input option used here for the tool is protein sequence and sequence position with amino acid variants. PolyPhen classifies the SNPs as “benign,” “possibly damaging” or “probably damaging” based on site-specific sequence conservation among mammals, as well as their location in the three-dimensional structure of the protein molecule. PolyPhen then calculated PSIC scores for each of the two variants based on three parameters such as (i) sequence-based characterization of the substitution site, (ii) profile analysis of homologous sequences and (iii) mapping of a substitution site to a known three dimensional protein structure. The PSIC score difference between the two variants elucidates the amount of functional consequences that the nsSNP exerts. The PSIC score difference is regarded to be directly proportional to the impact of a particular amino acid substitution. PolyPhen identifies homologs of the input sequences via a BLAST and calculates position-specific independent count (PSIC) scores for every variant and estimates the difference between the variant scores, the difference of >0.339 is detrimental.

2.5. Prediction of disease related amino acid substitution by MutPred

MutPred (http://mutpred.mutdb.org/) is a web based tool to predict the molecular cause of disease related amino acid substitution (Li, 2009). It utilizes several attributes related to protein structure, function, and evolution. It uses SIFT (Kumar et al., 2009), PSI-BLAST (Altschul, 1997), and Pfam profiles (Punta, 2012), along with some structural disorder prediction algorithms, including TMHMM (Krogh et al., 2001), MARCOIL (Dolorenzi and Speed, 2002), and DisProt (Sickmeier, 2007). Functional analysis includes the prediction of DNA-binding site, catalytic domains, calmodulin-binding targets (Radivojac, 2006), and post-translational modification sites (Thisberg et al., 2001; Iakoucheva, 2004; Radiwojic, 2010). Thus by combining the scores of all three servers, the accuracy of prediction rises to a greater extent.

2.6. Molecular phenotypic characterization of nsSNPs by SNPeffect

The SNPeffect 4.0 (http://snpeffect.switchlab.org/) server provides sequence and structure-based bioinformatics tools to predict the effect of protein-coding SNVs on the structural phenotype of proteins (De Baets et al., 2012). SNPeffect integrates aggregation prediction (TANGO) (Fernandez-Escamilla et al., 2004), amyloid prediction (WALTZ) (Maurer-Stroh et al., 2010), chaperone-binding prediction (LIMBO) (Van Durme et al., 2009) and protein stability analysis (FoldX) (Schymkowitz et al., 2005) for structural phenotyping.

2.7. Disease associated SNP prediction by nsSNP analyzer and PhD SNP

nsSNP Analyzer (http://snpanalyzer.uthsc.edu) is a tool to predict the phenotypic effect (disease-associated vs. neutral) of a nsSNP by using a machine learning method called RandomForest, and extracting structural and evolutionary information from a query nsSNP (Bao et al., 2005).

PhD-SNP (http://snps.biofold.org/phd-snp/phd-snp.html) is also an SVM based classifier, trained over the million amino acid polymorphism
datasets using supervised training algorithm, to predict if the given nsSNP has any pathological effect (Capriotti et al., 2006).

2.8. Predicting the effect of nsSNP on protein structure-function by stSNP

Structure SNP (stSNP) (http://stsynlab.org/stSNP/) is a server which provides the ability to analyze and compare human non-synonymous SNPs in protein structures, protein complexes, protein–protein interfaces and metabolic networks (Uzun et al., 2007).

2.9. Prediction of change in stability due to mutation

A support vector machine based tool I Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) was used to predict the change in the stability of the protein upon mutation. This tool automatically predicts protein stability changes upon single point mutations. I-Mutant 2.0 can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related change in Gibbs-free energy (∆G) (Capriotti et al., 2005). Prediction has been performed using the protein sequence of HSD3B2.

2.10. Structural conformation and conservation analysis of HSD3B2

ConSurf (http://consurf.tau.ac.il/) was used for high-throughput characterization of the functional regions in the protein (Ashkenazy et al., 2010). The degree of conservation of the amino-acid sites among 50 homologs with similar sequences was estimated. The conservation grades were then projected onto the molecular surface of the human HSD3B2 to reveal the patches with highly conserved residues that are often important for biological function.

2.11. Secondary structure prediction of HSD3B2 by PSIPRED

The PSIPRED program (http://bioinf.cs.ucl.ac.uk/psipred/) incorporates PSIPRED, GenTHREADER, and MEMSAT2 methods for protein structure prediction. This prediction method employs two feed-forward neural networks, which perform an analysis on the output obtained from PSI-BLAST (Buchan et al., 2013).

2.12. Homology modeling of HSD3B2

Homology modeling of HSD3B2 was carried out to predict its three dimensional (3D) structure as there was no crystal structure of HSD3B2 available. So the 3D structure of HSD3B2 has been modeled using the available protein sequence for homology based modeling.

### Table 2

| Substitution | dbsNP ID | Prediction | Score | Median info | Number of seqs at position |
|--------------|----------|------------|-------|-------------|---------------------------|
| G2V          | rs116449508 | Damaging   | 0.00  | 2.38        | 45                        |
| A10E         | rs28934880  | Damaging   | 0.00  | 1.88        | 88                        |
| V19A         | rs115344376 | Damaging   | 0.03  | 1.88        | 88                        |
| E44D         | rs11133222  | Tolerated  | 0.62  | 1.96        | 70                        |
| D74N         | rs4896059   | Damaging   | 0.02  | 1.89        | 92                        |
| E94Q         | rs86211     | Tolerated  | 0.28  | 1.89        | 92                        |
| E142K        | rs80358219  | Damaging   | 0.00  | 1.89        | 93                        |
| A164V        | rs34562248  | Tolerated  | 0.51  | 1.89        | 92                        |
| A167V        | rs35486059  | Tolerated  | 0.14  | 1.89        | 92                        |
| P222T        | rs80358220  | Tolerated  | 0.39  | 1.89        | 93                        |
| E236S        | rs89878237  | Damaging   | 0.17  | 1.87        | 92                        |
| T295M        | rs80358221  | Damaging   | 0.00  | 1.89        | 91                        |
| E270T        | rs75428981  | Tolerated  | 0.44  | 1.91        | 87                        |
| R316C        | rs114032180 | Damaging   | 0.02  | 1.86        | 90                        |
| P341L        | rs121964897 | Tolerated  | 0.00  | 1.89        | 90                        |
| T366N        | rs45609334  | Tolerated  | 0.31  | 2.38        | 45                        |

Web based servers like HHpred (Biegert et al., 2006), Phyre2 (Kelley and Sternberg, 2009) and RaptorX (Källberg et al., 2012) are used for homology modeling of HSD3B2. Structure refinement of the predicted models was carried out using ModRefiner (Xu and Zhang, 2011).

2.12.1. Homology modeling using HHpred server

HHpred (http://toolkit.tuebingen.mpg.de/hhpred) is a free protein function and protein structure prediction server based on the HHsearch method. HHpred profiles are calculated from a multiple sequence alignment of related sequences which are typically collected using the PSI-BLAST program. The template structures for HSD3B2 modeling were obtained initially by online submission of the amino acid sequences to HHpred (Biegert et al., 2006). The alignment suggested by HHpred was used for modeling. Atomic coordinates built on these target-template alignments were generated using MODELLER v. 9.2 (Fiser and Sali, 2003), choosing the best out of 20 models based on objective function score and visual inspection.

2.12.2. Homology modeling using phyre2

Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) uses the Hidden Markov Method to generate alignments of a submitted protein sequence against proteins with published structures (Kelley and Sternberg, 2009). The resulting alignments are then used to produce homology-based models of the query sequence to predict its three-dimensional structure. In addition, Phyre2 uses an ab-inito folding simulation called Poing to model regions of a query with no detectable similarities to known structures (Jefferys et al., 2010). Poing combines multiple proteins using a high-throughput search algorithm to find the best match for each residue in the query sequence.
templates of known structures to produce the final model of the query sequence. The 3D structure models for HSD3B2 were developed using Phyre2 for predicting the protein structure by homology modeling under ‘intensive’ mode.

2.12.3. Homology modeling using RaptorX

RaptorX (http://raptorx.uchicago.edu/StructurePrediction/predict/) uses a non-linear scoring function to combine homologous information with structural information for a given template-sequence alignment. It uses NEFF to adjust relative importance of homology and structural information. RaptorX uses a combination of RaptorX-Boost and RaptorX-MSA to build 3D models for a target-template alignment. In the absence of good quality templates RaptorX models the alignment using an in-house free modeling program (Källberg et al., 2012) to generate 5 models. Unaligned portions of the template are also folded by free modeling.

2.13. Energy minimization of modeled structures of HSD3B2

Structure refinement of the predicted models was carried out using ModRefiner (Xu and Zhang, 2011). Then to improve the quality of predicted model of HSD3B2, energy minimization was performed with the GROMOS 96 force field (Van Gunsteren et al., 1996) implementation of DeepView v4.04 (spdb viewer) tool (Guex and Peitsch, 1997). The PDB model of HSD3B2 generated from RaptorX server was used for this purpose. Swiss-PDB Viewer allows browsing through a rotamer library to change amino acids. A mutation tool was used for this purpose. Swiss-PDB Viewer allows browsing through a rotamer library to change amino acids.

2.15. Modeling amino acid substitution and energy minimization

Swiss-PDBViewer (v.4.04) was used to generate the mutated models of HSD3B2 for the corresponding amino acid substitutions (Guex and Peitsch, 1997). The PDB model of HSD3B2 generated from RaptorX server was used for this purpose. Swiss-PDB Viewer allows browsing through a rotamer library to change amino acids. A “mutation tool” was used to replace the native amino acid with a new one. The mutation tool facilitates the replacement of the native amino acid by the “best” rotamer of the new amino acid. The “.pdb” files were saved for all the models. Then to improve the quality of predicted model energy minimization was performed with the GROMOS 96 force field (Van Gunsteren et al., 1996) implementation of DeepView v4.04 (spdb viewer) tool.

3. Results and discussions

To determine the deleterious nsSNPs, involved in inducing disease associated phenomena, is now among the most important field of computational genomic research. The advanced methods in computational biology has now enabled us to determine the disease associated or deleterious nsSNPs in the target candidate genes. The computational methods were applied to study the protein structural and functional effect on point mutation at molecular level. In this investigation we have implemented multiple computational methods and tools to identify the most likely pathogenic mutations in HSD3B2 gene along with structural modeling.

3.1. Collecting nsSNP information from dbSNP

The dbSNP database contains both validated and non-validated polymorphisms. In spite of this data has been collected from the dbSNP because allelic frequency of most of nsSNPs of HSD3B2 has been recorded there and it is the most extensive SNP database. Table 1 shows the list of nsSNPs present in the coding region of HSD3B2 gene which were used for this present study.

3.2. Deleterious nsSNPs predicted by SIFT

The sequence homology-based tool SIFT was used to calculate the level of conservation of a particular amino acid position in a protein.
### Table 5

| Mutation | Probability of deleterious mutation | Top 5 features |
|----------|------------------------------------|----------------|
| G2V      | 0.586                              | Loss of disorder (P = 0.004) |
|          |                                    | Gain of sheet (P = 0.0827)   |
|          |                                    | Gain of MoRF binding (P = 0.1923) |
|          |                                    | Gain of catalytic residue at (P = 0.2325) |
|          |                                    | Gain of solvent accessibility (P = 0.4137) |
|          |                                    | Loss of MoRF binding (P = 0.0643) |
|          |                                    | Gain of disorder (P = 0.1322)   |
| A10E     | 0.942                              | Loss of helix (P = 0.2662)    |
|          |                                    | Gain of loop (P = 0.2754)     |
|          |                                    | Loss of sheet (P = 0.3635)    |
|          |                                    | Loss of stability (P = 0.077)  |
|          |                                    | Gain of disorder (P = 0.1143)  |
| V19A     | 0.649                              | Gain of MoRF binding (P = 0.1788) |
|          |                                    | Gain of methylation at R20 (P = 0.5498) |
|          |                                    | Gain of catalytic residue at 118 (P = 0.6328) |
|          |                                    | Loss of helix (P = 0.1299)    |
|          |                                    | Gain of ubiquitination at K48 (P = 0.1317) |
| E44D     | 0.444                              | Gain of MoRF binding (P = 0.1702) |
|          |                                    | Loss of disorder (P = 0.1919)  |
|          |                                    | Loss of methylation at K48 (P = 0.2057) |
|          |                                    | Gain of methylation at K69 (P = 0.1364) |
|          |                                    | Gain of ubiquitination at K69 (P = 0.2584) |
| D74N     | 0.255                              | Gain of catalytic residue at D74 (P = 0.307) |
|          |                                    | Gain of loop (P = 0.3485)     |
|          |                                    | Loss of sheet (P = 0.3635)    |
|          |                                    | Loss of sheet (P = 0.1158)    |
|          |                                    | Gain of catalytic residue at N98 (P = 0.1576) |
| E94Q     | 0.122                              | Gain of helix (P = 0.2684)    |
|          |                                    | Gain of MoRF binding (P = 0.3065) |
|          |                                    | Gain of disorder (P = 0.4148)  |
|          |                                    | Gain of methylation (P = 0.0075) |
|          |                                    | Gain of ubiquitination (P = 0.0117) |
| E142K    | 0.906                              | Gain of glycosylation (P = 0.0224) |
|          |                                    | Gain of solvent accessibility (P = 0.0314) |
|          |                                    | Gain of relative solvent accessibility (P = 0.0479) |
|          |                                    | Loss of ubiquitination at K163 (P = 0.0755) |
|          |                                    | Loss of disorder (P = 0.0859)  |
| A164V    | 0.572                              | Gain of methylation at K163 (P = 0.1682) |
|          |                                    | Gain of MoRF binding (P = 0.1847) |
|          |                                    | Loss of glycosylation at K163 (P = 0.3349) |
|          |                                    | Loss of ubiquitination at K163 (P = 0.0755) |
|          |                                    | Loss of disorder (P = 0.1388)  |
| A167V    | 0.617                              | Gain of MoRF binding (P = 0.2181) |
|          |                                    | Loss of glycosylation at K163 (P = 0.3737) |
|          |                                    | Gain of catalytic residue at A167 (P = 0.4112) |
|          |                                    | Gain of helix (P = 0.2099)     |
|          |                                    | Gain of relative solvent accessibility (P = 0.2363) |
| P222T    | 0.918                              | Gain of MoRF binding (P = 0.2525) |
|          |                                    | Loss of glycosylation at T219 (P = 0.2882) |
|          |                                    | Loss of catalytic residue at P222 (P = 0.3449) |
|          |                                    | Gain of glycosylation at L236 (P = 0.012) |
|          |                                    | Loss of stability (P = 0.0143)  |
| L236S    | 0.903                              | Gain of disorder (P = 0.0182)  |
|          |                                    | Gain of MoRF binding (P = 0.0824) |
|          |                                    | Gain of methylation at R240 (P = 0.207) |
|          |                                    | Loss of phosphorylation at T209 (P = 0.0342) |
|          |                                    | Loss of disorder (P = 0.1162)  |
| T259M    | 0.888                              | Gain of helix (P = 0.1736)    |
|          |                                    | Loss of sheet (P = 0.3635)    |
|          |                                    | Gain of solvent accessibility (P = 0.3956) |
|          |                                    | Gain of disorder (P = 0.0039)  |
|          |                                    | Gain of disorder (P = 0.0233)  |
| I270T    | 0.587                              | Gain of ubiquitination at K271 (P = 0.0562) |
|          |                                    | Loss of catalytic residue at I270 (P = 0.2609) |
|          |                                    | Loss of phosphorylation at Y269 (P = 0.1396) |
|          |                                    | Loss of disorder (P = 0.0514)  |
|          |                                    | Gain of catalytic residue at T118 (P = 0.0761) |
| R316C    | 0.571                              | Loss of glycosylation at P313 (P = 0.1732) |
|          |                                    | Loss of solvent accessibility (P = 0.2668) |
|          |                                    | Loss of stability (P = 0.284)   |
|          |                                    | Gain of MoRF binding (P = 0.0516) |
| P341L    | 0.924                              | Loss of methylation at K434 (P = 0.0533) |

- **All the 16 nsSNPs were submitted to the SIFT server page to calculate the tolerance index (TI).** The functional impact of the amino acid substitution is inversely proportional to the tolerance index. *Table 2* summarizes the results obtained from SIFT server. It was observed that out of 16 nsSNPs, 8 were predicted as ‘DAMAGING’ and had a TI of ≤0.05. The corresponding amino acid substitutions of rs116449508, rs28934880, rs80358219, rs80358221, and rs121964897 had a TI score of 0.00. The TI score was 0.02 for rs4986954 and rs114032180 and 0.03 for rs115344376. SIFT has been tested on many human SNP databases and was found able to distinguish the disease associated SNPs from a neutral one with only a 20% false positive error. Seventy four percent (74%) of nsSNPs identified by the SNP Consortium, were sufficiently similar to homologs in protein sequence databases for SIFT prediction *(Ng and Henikoff, 2003)*. Furthermore, the SIFT algorithm works mainly sequence for prediction as the crystal structure of HSD3B2 is not known.

### 3.3. Deleterious nsSNPs predicted by PROVEAN

The results obtained from the sequence based predictor PROVEAN *(Table 3)* sorted out the HSD3B2 variants that were predicted to be “Deleterious” when the final score is below the threshold (default is −2.5), and that were predicted to be “Neutral” when the score was above the threshold. It was observed that out of 16 nsSNPs, 8 nsSNPs were predicted as “Deleterious” and had a PROVEAN score of below −2.5. The amino acid substitutions G2V, P222T, T259M, R316C and P341L have PROVEAN score of below −5.

### 3.4. Damaging nsSNPs predicted by PolyPhen

PolyPhen predicts the fate of the structure and function of a protein due to an amino acid change through specific empirical rules on the sequence. For sequence-based characterization of the substitution site PolyPhen uses the TMHMM algorithm, Coils2 program and SignalP program to predict transmembrane, coiled coil and signal peptide regions of the protein sequences. There are certain empirical rules applied on the sequences and the accuracy of that is approximately 82% with a chance of 8% false positive prediction *(Ramensky et al., 2002)*. *Table 4* summarized the results obtained from the PolyPhen server. A position-specific independent count (PSIC) score difference was assigned using the categories ‘probably damaging’, ‘possibly damaging’, ‘potentially damaging’, ‘borderline’ and ‘benign’. It was observed that the variants G2V, A10E, V19A, E142K, P222T, T259M, and P341L are probably damaging whereas variants D74N, A167V and R316C are potentially damaging and variants E44D, E94Q, A164V, L236S, I270T and T366N are benign.

### 3.5. Disease related amino acid substitution prediction by MutPred

It uses SIFT, PSI-BLAST, and Pfam profiles, along with some structural disorder prediction algorithms, including TMHMM, MARCOIL, B-factor prediction, and DisProt. *Table 5* summarized the result obtained from
I270T decreases the amyloid propensity of the protein as predicted by the dTANGO score. It was observed that variant G2V whereas variant L236S decreases the aggregation tendency. D74N, A164V, A167V and P222T increase the aggregation tendency. This implied that some nsSNPs may account for potential structural and functional alterations of HSD3B2.

### 3.6. Molecular phenotypic characterization of nsSNP by SNPeffector

This method uses sequence- and structure-based bioinformatics tools to predict the effect of protein-coding SNVs on the structural phenotype of proteins. The results obtained from the SNPeffector server were summarized in Table 6. It was observed that variants G2V, D74N, A164V, A167V and P222T increase the aggregation tendency of HSD3B2 whereas variant L236S decreases the aggregation tendency as predicted by the dTANGO score. It was observed that variant I270T decreases the amyloid propensity of the protein as predicted by WALTZ. It was observed that no one variant affects the chaperone binding tendency of HSD3B2 as predicted by LIMBO. No protein stability change due to nsSNP has been predicted by FoldX because it couldn’t find reliable structural information.

### 3.7. Disease associated SNP prediction by nsSNP Analyzer and PhD SNP

The results obtained from nsSNP Analyzer and PhD SNP server were summarized in Table 7. It was observed that out of 16 nsSNPs, 8 were predicted to be “Diseases associated” (viz. G2V, A10E, V19A, E44D, E142K, T259M, R316C, P341L substitutions) by nsSNP Analyzer server. On the other hand results obtained from PhD SNP server showed that 11 nsSNPs are “Diseases associated”.

### 3.8. Predicting the effect of nsSNP on HSD3B2 structure-function by stSNP server

Out of 16 nsSNPs, stSNP server was able to predict the effect of seven nsSNPs viz. rs ID 6211, 4986954, 28934880, 34562248, 35486059, 35887327 and 72631744. Table 8 showed the results obtained from the server. Increase in volume difference has been observed for amino acid substitutions at A10E, A164V, and A167V and decrease in volume difference has been observed at L236S.

### 3.9. Prediction of change in stability due to mutation

The I-Mutant 2.0 server was developed and tested with the data extracted from ProTherm, the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability due to mutation. It was observed that there was a large decrease of stability for variants V19A, E94Q, A167V, P222T, L236S, I270T, R316C and T366N. Other mutants exhibited decreased stability. These results were summarized in Table 9. Hence, I-Mutant

---

**Table 6**

| rsID of SNP | Variant | TANGO | WALTZ | LIMBO | FoldX |
|------------|---------|-------|-------|-------|-------|
| rs116449508 | G2V     | Increased aggregation tendency | No effect | No effect | Not available |
| rs28934880 | A10E    | No effect | No effect | No effect | Not available |
| rs115344376 | V19A   | No effect | No effect | No effect | Not available |
| rs111333222 | E44D   | No effect | No effect | No effect | Not available |
| rs4986954 | D74N    | Increased aggregation tendency | No effect | No effect | Not available |
| rs6211    | E94Q    | No effect | No effect | No effect | Not available |
| rs80358219 | E142K  | No effect | No effect | No effect | Not available |
| rs34562248 | A164V  | Increased aggregation tendency | No effect | No effect | Not available |
| rs53486059 | A167V  | Increased aggregation tendency | No effect | No effect | Not available |
| rs80358220 | P222T  | Increased aggregation tendency | No effect | No effect | Not available |
| rs35887327 | L236S  | Decreased aggregation tendency | No effect | No effect | Not available |
| rs80358221 | T259M  | No effect | No effect | Decreased amyloid propensity | Not available |
| rs75429891 | I270T  | No effect | No effect | No effect | Not available |
| rs114032180 | R316C | No effect | No effect | No effect | Not available |
| rs121964897 | P341L | No effect | No effect | No effect | Not available |
| rs45609334 | T366N | No effect | No effect | No effect | Not available |

**Table 7**

| SNP rsID | Amino acid change | Phenotypic prediction by nsSNP Analyzer | Sequence and profile based PhD SNP Prediction |
|----------|-------------------|----------------------------------------|--------------------------------------------|
| rs116449508 | G2V | Disease | Disease |
| rs28934880 | A10E | Disease | Disease |
| rs115344376 | V19A | Disease | Disease |
| rs111333222 | E44D | Disease | Neutral |
| rs4986954 | D74N | Neutral | Disease |
| rs6211 | E94Q | Neutral | Disease |
| rs80358219 | E142K | Disease | Disease |
| rs34562248 | A164V | Neutral | Neutral |
| rs35486059 | A167V | Neutral | Neutral |
| rs80358220 | P222T | Neutral | Disease |
| rs35887327 | L236S | Neutral | Disease |
| rs80358221 | T259M | Neutral | Disease |
| rs75429891 | I270T | Neutral | Disease |
| rs114032180 | R316C | No effect | Disease |
| rs121964897 | P341L | No effect | Disease |
| rs45609334 | T366N | No effect | Disease |

**Table 8**

| SNP rsID | Variation | Volume Difference |
|----------|-----------|-------------------|
| 6211     | E94Q      | 5.40 (increased)  |
| 4986954  | D74N      | 3.00 (increased)  |
| 28934880 | A10E      | 49.80 (increased) |
| 34562248 | A164V     | 51.40 (increased) |
| 35486059 | A167V     | 51.40 (increased) |
| 35887327 | L236S     | −77.70 (decreased) |
| 72631744 | T366N     | −2.00 (decreased) |
2.0 efficiently predicted mutations that affected the stability of the HSD3B2. Although the predictions were 80% or 70% accurate depending upon the usage of structural or sequence information, respectively.

3.10. Structural conformation and conservation analysis of HSD3B2

The results generated by the ConSurf tool consist of a structural representation of the protein (Fig. 1) which contain a colorimetric conservation score. Evolutionary information is of fundamental importance for detecting mutations that affect human health (Ramensky et al., 2002). ConSurf identifies functional regions in proteins, taking into account the evolutionary relationships among their sequence homologs (Glaser et al., 2003). The ConSurf conservation analysis was performed by evolutionarily related conservation scores of the residues for functional region identification from proteins of known three dimensional structures (Jimenez-Lopez et al., 2010). The ConSurf analysis also revealed, as expected, that the functional regions of the protein are highly conserved. It was observed

| Position | ΔΔG | Stability |
|----------|-----|-----------|
| G2V      | −0.16 | Decrease Stability |
| A10E     | −0.42 | Decrease Stability |
| V19A     | −1.71 | Large Decrease of Stability |
| E44D     | −0.37 | Decrease Stability |
| D74N     | −0.15 | Decrease Stability |
| E94Q     | −1.45 | Large Decrease of Stability |
| E142K    | −0.18 | Decrease Stability |
| A164V    | 0.17  | Increase Stability |
| A167V    | −0.26 | Decrease Stability |
| P222T    | −2.13 | Large Decrease of Stability |
| L236S    | −2.40 | Large Decrease of Stability |
| T259M    | −0.23 | Decrease Stability |
| I270I    | −3.39 | Large Decrease of Stability |
| K316C    | −1.99 | Large Decrease of Stability |
| P341L    | −0.10 | Decrease Stability |
| T366N    | −1.31 | Large Decrease of Stability |

Table 9
Stability prediction of HSD3B2 protein upon amino acid substitution by I Mutant. Stability was predicted as ΔΔG Value = ΔG (NewProtein) - ΔG (WildType) in Kcal/mol. Stability change was calculated at pH 7 and 25 °C.

The conservation scale:

Nature of amino acids:
- e - Exposed residue
- b - Buried residue
- f - Predicted functional residue
- s - Predicted structural residue

Fig. 1. Analysis of evolutionary conserved amino acid residues of HSD3B2 by ConSurf. The color coding bar shows the coloring scheme representation of conservation score.
that variants A10E, E142K, T259M, and P341L have a conservation scale of 9; G2V and D74N have a conservation scale of 8; P222T has a conservation scale of 7; L236S and R316C have a conservation scale of 6 and V19A has a conservation scale of 5.

3.11. Secondary structure prediction by PSIPRED

The results revealed a clear distribution of alpha helix, beta sheet and coil (Fig. 2). Coils dominated among secondary structure elements (51.32%) followed by alpha helix (37.75%) and beta sheet (10.93%).

3.12. Homology modeling of HSD3B2

The ability of the protein to interact with other molecules or to have different functions depends upon its tertiary structure (Hasan et al., 2011; Alshatwi et al., 2011). Therefore, analysis of damaged coding nsSNPs at the structural level is necessary to understand the activity of the protein. There is no crystal structure of HSD3B2 available in the protein data bank. So the 3D structure of HSD3B2 has been modeled using the available protein sequence for homology based modeling. Web based servers like HHPred, Phyre 2 and Raptor X are used for homology modeling of HSD3B2. Fig. 3A, B and C were cartoon representation of the protein structure obtained from HHPred server, Phyre2 server and RaptorX server respectively. HHpred profiles are calculated from a multiple sequence alignment of related sequences which are typically collected using the PSI-BLAST program. Phyre2 uses the Hidden Markov Method to generate alignments of a submitted protein sequence against proteins with published structures. RaptorX uses a non-linear scoring function to combine homologous information with structural information for a given template-sequence alignment.

3.13. Energy minimization of modeled structures of HSD3B2

Structure refinement and energy minimization of the predicted models of HSD3B2 was carried out using ModRefiner and DeepView v4.04 tools respectively. Fig. 4A, B and C showed the image of HSD3B2 structure, generated by HHPred, Phyre2 and RaptorX respectively, after refinement and energy minimization. The energy minimization repaired as well as overhauled distorted geometries of HSD3B2.

3.14. Model validation for HSD3B2

Validation of the model is very important in protein structural prediction since ultimately the modeled protein structure is used to design further experiments and understand the protein’s biological function. A Ramachandran plot was obtained from PROCHECK for each of the generated pdb structures of HSD3B2. Fig. 5A, B and C showed the phi/psi Ramachandran plot of energy minimized HSD3B2 structures obtained from HHPred, Phyre2 and RaptorX server respectively. Total number of non-glycine and non-proline residues in the HSD3B2 structure is 329. Table 10 showed the analysis of Ramachandran Plot of modeled structures of HSD3B2. Protein backbone conformations were evaluated by inspection of the Ramachandran Plot which is an x–y plot of phi/psi dihedral angles between N-Cα and Cα-Cα planar peptide bonds in the protein’s backbone. Both these angles are able to rotate freely in proteins (−180 to +180). Any combination of these angles is theoretically possible but in actual biological conditions many combinations are rarely or never seen due to steric clashes in the proteins’ backbone structure.

The QMEAN6 server stands for Qualitative Model Energy ANalysis along with its clustering method QMEANclust (Benkert et al., 2009a, Benkert et al., 2009b). Six structural descriptors are employed to assess global quality of models. They are (a) torsion energy potential based on three consecutive amino acids used to measure local geometry, (b) two distance dependent potentials to assess long range interactions based on Cβ atoms and all atoms, (c) solvent potential (d) solvent accessibility. The QMEAN Z-score provides an estimate of the absolute quality of a model by comparing it to same sized reference structures present in the PDB and solved by experimental techniques (Srivastava et al., 2012). QMEAN Z-score was used to estimate the ‘degree of nativeness’ of the predicted structure. Table 11 showed the results from QMEAN Server using energy minimized structures of HSD3B2. The total QMEAN score
Fig. 3. Homology based prediction of HSD3B2 structure using HHpred (A), Phyre (B) and RaptorX (C) web-based server. The visual images of the structures were generated in PyMol.

Fig. 4. Cartoon representation of predicted models of HSD3B2 after structure refinement and energy minimization. Homology models were developed by HHpred (A), Phyre2 (B) and RaptorX (C). Model images were generated in PyMol.

Fig. 5. Ramachandran plot of the predicted model of HSD3B2 by (A) HHpred, (B) Phyre2 and (C) RaptorX server respectively. The plots were generated in PROCHECK. The most favored regions are colored red, additional allowed, generously allowed and disallowed regions are indicated as yellow, light yellow and white fields, respectively.

Table 10
Analysis of Ramachandran plot of modeled structures of HSD3B2 by using PROCHECK server. Total number of non-glycine and non-proline residues in the structure is 329.

| Model             | Residues in most favored regions | Residues in additional allowed regions | Residues in generously allowed regions | Residues in disallowed regions |
|-------------------|----------------------------------|----------------------------------------|----------------------------------------|---------------------------------|
|                   | No. of residues | % of residues | No. of residues | % of residues | No. of residues | % of residues | No. of residues | % of residues |
| HSD3B2_hhpred     | 273             | 83.0%         | 50             | 15.2%         | 3               | 0.9%          | 3               | 0.9%          |
| HSD3B2_phyre2     | 268             | 81.5%         | 49             | 14.9%         | 9               | 2.7%          | 3               | 0.9%          |
| HSD3B2_RaptorX    | 282             | 85.7%         | 41             | 12.5%         | 5               | 1.5%          | 1               | 0.3%          |
for the model from HHpred, Phyre2 and RaptorX server was 0.497 (Z-score: −3.22), 0.523 (Z-score: −2.92), and 0.556 (Z-score: −2.53) respectively. It was observed that C\_beta interaction energy (−44.21; Z-score: −1.85), all-atom pairwise energy (−2957.90; Z score: −2.78), solvation energy (−11.89; Z-score: −2.27), torsion angle energy (−26.00; Z score: −3.11) and secondary structure agreement (77.3%; Z-score: −0.83) were better for the model of HSD3B2 obtained from RaptorX server than HHpred and Phyre2 server.

The predicted HSD3B2 energy minimized models were further evaluated using the ProSA web server. Z-score for modeled energy minimized PDB structure from HHpred, Phyre2 and RaptorX server were −5.46, −5.61, and −7.25 respectively (Table 12).

Thus it was observed that the model of HSD3B2 obtained from RaptorX server was better than the model generated from HHpred, Phyre2 server.

### 3.15. Modeling amino acid substitution and energy minimization

The amino acid substitutions which have potential impact on HSD3B2 structure-function were shown in the Fig. 6. Fig. 6A showed the wild type residues with their Van der Wall radius and Fig. 6B represented the mutant residues with their Van der Wall radius. Mapping of the variants on the predicted structure of HSD3B2 revealed that mutational hotspots were in close proximity to one another in three dimensional space.

### 4. Conclusions

Computational study has now got major importance to screen diseases specific SNP at molecular level. In this study in silico analysis has been performed to investigate the effect of nsSNPs on structure–function of HSD3B2. Out of 16 nsSNP, eight point mutations in the coding region may have significant effect in the HSD3B2 structure as well as in function. The computational analysis of free energy change due to mutation indicates that stability of HSD3B2 get decreased. The modeled structure of HSD3B2 provides the insight into the structural mapping of nsSNP in three dimensional space. For the first time this result provides a significant computational approach to detect the pathologically significant nsSNPs in HSD3B2. Furthermore, the predicted disease associated nsSNP can be studied for the further development in potent drug discovery.

### Conflict of interest

None.

### Acknowledgment

This project has been supported by grants from the University Grants Commission [F. No. PSW-092/14-15 (ERO)].

---

**Table 11**
Results obtained from the QMEAN Server using energy minimized modeled structures of HSD3B2.

| Model         | C\_beta interaction energy | All-atom pairwise energy | Solvation energy | Torsion angle energy | Secondary structure agreement | Solvent accessibility agreement | Total QMEAN-score |
|---------------|-----------------------------|--------------------------|------------------|----------------------|-------------------------------|---------------------------------|------------------|
| HSD3B2_hhpred | −18.01 (Z-score: −2.35)     | −1935.38 (Z-score: −3.19) | −5.05 (Z-score: −2.94) | −5.25 (Z-score: −4.33) | 76.5% (Z-score: −0.99)         | 72.0% (Z-score: −1.76)          | 0.497            |
| HSD3B2_phyre2 | −13.64 (Z-score: −2.46)     | −2499.39 (Z-score: −3.09) | −9.66 (Z-score: −2.49) | −9.01 (Z-score: −4.11) | 72.8% (Z-score: −1.73)         | 74.5% (Z-score: −1.26)          | 0.523            |
| HSD3B2_RaptorX| −44.21 (Z-score: −1.85)     | −2957.90 (Z-score: −2.27) | −11.89 (Z-score: −2.78) | −26.00 (Z score: −3.11)| 77.3% (Z-score: −0.83)         | 73.7% (Z-score: −1.42)          | 0.556            |

**Table 12**
Results obtained from ProSA server using energy minimized modeled structures of HSD3B2.

| Models         | Z-Score: |
|----------------|----------|
| HSD3B2_hhpred  | −5.46    |
| HSD3B2_phyre2  | −5.61    |
| HSD3B2_RaptorX | −7.25    |

**Fig. 6.** Mapping variants onto the predicted structure of HSD3B2. A- Wild type; B- Mutant. The mutational hotspots are revealed in the three dimensional structure.
