Journal of Applied Biology & Biotechnology Vol. 6(05), pp. 1-11, September-October, 2018
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2018.60501

Comprehensive analysis of damage-associated single-nucleotide polymorphisms of sex hormone-binding globulin gene

Richa Bhatnager[1], Mehak Dangi[2], Amita Suneja Dang[1]*

[1]M.D. University, Rohtak, Haryana, India, [2]Centre for Bioinformatics, Maharshi Dayanand University, Rohtak, Haryana, India

ARTICLE INFO

Article history:
Received on: December 11, 2017
Accepted on: April 27, 2018
Available online: August 01, 2018

Key words:
Sex hormone-binding globulin,
Androgens,
Untranslated regions,
Splice site,
Non-synonymous single-nucleotide polymorphism.

ABSTRACT

Sex hormone-binding globulin (SHBG) controls the bioavailability of androgens, and its association has been found in a number of disorders such as hyperandrogenism, obesity, diabetes, and cancer. Many human single-nucleotide polymorphisms (SNPs) that are now recognized provide an opportunity to understand the association between genotype and phenotype. In our analysis, we found P185L (rs6258) substitution possess damaging effect on protein structure. ConSurf analysis predicted P185L is conserved and exposed in protein structure. Secondary and tertiary structure of mutated protein was predicted by PSIPRED and Swiss Modeller which were by superimposed using UCSF Chimera to predict their side-chain modification. FT site server predicted amino acid residues that are involved in ligand-binding site of SHBG protein and none of the substitution was involved in ligand-binding site. Six SNPs associated with untranslated region affect the miRNA seed region, thereby affect gene regulation. Ten SNPs associated with splice site were found to alter slicing signal by our study, hence affect the mRNA processing and resulted in faulty polypeptide. Alteration in SHBG polypeptide affects its affinity toward androgen binding and its physiological level as well. These SNPs are still uncharacterized; hence providing a baseline for validation of their association with the susceptibility of diseases and develop personalized therapeutics.

1. INTRODUCTION

Human sex hormone-binding globulin (SHBG) is a transport protein specific for dihydrotestosterone, testosterone, and estradiol [1]. It helps to regulate the free androgen availability in the body and has higher affinity toward testosterone than estrogen. Physiological levels of SHBG are higher in females than in males. Both in males and females, altered SHBG levels affect the testosterone availability in their body and can cause physiological and reproductive imbalance. Low SHBG levels are associated with a number of diseases in females such as breast cancer, endometrial cancer, and polycystic ovary syndrome. Similarly, physiological levels of SHBG were found to be altered in a number of male disorders such as erectile disorder, infertility, osteoporosis, and prostate cancer. SHBG is a homodimeric molecule, with a single steroid-binding site per dimer [2]. SHBG gene is located on the short arm of human chromosome 17 and is nearly 4 kb in size [3,4]. In humans, SHBG is synthesized mainly in liver under the control of various hormonal and metabolic regulators [5,6]. Its synthesis is stimulated by estrogens and thyroid hormones and is inhibited by androgens and insulin [7,8].

*Corresponding Author:
Amita Suneja Dang, M.D. University, Rohtak, Haryana, India. E-mail: suneja_a@yahoo.co.in

Single-nucleotide polymorphisms (SNPs) are the most common source of human genetic variation and are undoubtedly a valuable resource for investigating the genetic basis of diseases. SNPs, together with DNA copy number variations, have become one of the most active research areas of genomics in recent years [9] because of their association with diseases. In the present study, we aimed to find the effect of polymorphism present in non-coding region and non-synonymous SNPs on protein structure and function as these are highly abundant, stable, and distributed throughout the genome. This type of variation is associated with diversity in the population, individuality, and although the majority of these variations probably result in neutral phenotypic outcomes, certain polymorphisms can predispose individuals to disease, or influence its severity, progression, or individual response to medicine [10]. Viewed at the molecular level, these functional SNPs can affect the human phenotype. Non-coding SNPs may disrupt transcription factor binding (TFD) sites, splice sites, and other functional sites on the transcriptional level, whereas coding SNPs can cause an amino acid change and alter the functional or structural properties of the translated protein [11-14]. In this analysis, we employed different softwares to access the effect of SNPs on SHBG structure and function.

2. MATERIALS AND METHODS

2.1. SNP Retrieval

The data on human SHBG gene was collected from Entrez Gene on National Center for Biological Information (NCBI) website. The SNP information (protein accession number and SNP ID) of the SHBG gene was retrieved from the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/) and SwissProt databases (http://expasy.org/). Other databases such as Exome Aggregation Consortium, Genome Variation Server, and F-SNP were also searched to cross-check the non-synonymous SNP (nsSNP) data for SHBG gene [15].

2.2. Effect Prediction Tool used for nsSNP

2.2.1. Evaluation of the functional impact of coding nsSNPs using a sequence homology tool (sorting intolerant from tolerant [SIFT])

SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence [16]. It tells whether the substitution is tolerated at that position. It takes reference SNP ids to make its prediction.

2.2.2. PROVEAN

Protein variation effect analyzer predicts whether the particular substitution of amino acid or deletion, insertion is deleterious or tolerated. The threshold for a mutation to be deleterious is −2.5; if below threshold, prediction will be deleterious and will be neutral if it is above threshold [17]. The input data were those amino acid substitutions that are analyzed by SIFT program. Protein id used for PROVEAN analysis is NCBI ref sequence id for SHBG protein.

2.2.3. PolyPhen (http://genetics.bwh.harvard.edu/pph2/)

PolyPhen predicts the effect of amino acid substitution on the structure and function of proteins by following the specific empirical rules. Protein sequence, database ID/accession number, amino acid position, and amino acid variant details are the input options for Polyphen [18] and calculate the score difference between variants and wild-type amino acid.

2.2.4. nsSNPAnalyzer (http://snpanalyzer.uthsc.edu/)

It uses a machine learning method called Random Forest to classify the nsSNPs. It predicts the effect of nsSNPs on protein structure in terms of the environmental polarity, solvent accessibility, and secondary structure [19]. Its input includes FASTA sequence of protein and substitutions as predicted by SIFT server.

2.4.5. PhD-SNP (http://snps.biofold.org/phd-snp/phd-snp.html)

It is support vector machine (SVM)-based software which supports the local sequence environment and output of multiple sequence alignment to predict the nature of a particular mutation. Input format includes FASTA sequence of protein along with substituted residue [20]. The output is based on reliability score which predicts whether the substitution is disease-causing or neutral.

2.4.6. Mutation Assessor (http://mutationassessor.org/r3/)

It predicts the effect of amino acid substitutions on the function of proteins using "combinatorial entropy optimization" globally and identifies key functional residues and finally assigns a conservation score to them. Its output contains functional impact (high, medium, and neutral) and FI score (functional impact score).

2.4.7. Consensus generation

A consensus of all the predictions was generated to prioritize the deleterious substitution predicted by various softwares used. It was done by manual method. Results of all the software were analyzed and substitution was selected which are found to be deleterious in all the predictions.

2.4.8. Mutpred (http://mutpred.mutdb.org/)

MutPred is a web tool developed to predict whether the amino acid substitution is disease-associated or neutral along with the molecular cause of that substitution [21-23]. Input format required FASTA sequence of protein along with substitutions. The output is in the form of probability score (g), that is, P (deleterious) the probability that the amino acid substitution is deleterious or disease-associated, and top five characteristic scores (p), where p is the P-value that certain functional and structural characteristics of the protein are impacted.

2.4.9. Prediction of stability change by I-Mutant (http://folding.biofold.org/i-mutant/i-mutant2.0.html)

A SVM-based tool I-Mutant 2.0 predicts the change in the stability of the protein by a particular mutation. I-Mutant 2.0 provides output in the form of sign of energy change and relative energy change at standard temperature and pH [20]. Its prediction comes in two forms, i.e., change in DDG and ΔG. Positive G value leads to increased stability, whereas negative G values correspond to decreased stability

2.3. Identification of cis-Regulatory Elements (CRE)

2.3.1. PROSCAN version 1.7 Web Promoter Scan Service (http://bimas.dcrt.nih.gov/molbio/proscan/)

PROSCAN version 1.7 Web Promoter Scan Service predicts promoter regions based on homologies with putative eukaryotic Pol II promoter sequences. The site is serviced and maintained by Dr. Dan Prestridge at the Advanced Biosciences Computing Center, University of Minnesota.

2.3.2. Promoter 2.0 Prediction Server (http://www.cbs.dtu.dk/services/Promoter/)

Promoter 2.0 predicts the initiation site of transcription at promoters in DNA sequences. It is also a machine learning program whose size is growing day by day. The site is serviced and maintained by Steen Knudsen at The Center for Biological Sequence Analysis at the Technical University of Denmark.

2.3.3. TSSG (http://www.softberry.com/)

TSSG predicts the CRE of mammals. It predicts the potential initiation site for transcription by employing linear discriminant function and combined characteristics which described the oligonucleotide composition and functional motifs of these sites. It uses promoter. dat file with selected factor binding sites (TFD) developed by Dan Prestridge to calculate the density of functional sites.

2.4. Prediction of the Effect of SNPs Associated with 3' Untranslated Regions (UTRs) by PolymiRTS Database (http://compbio.uthsc.edu/miRSNP/)

It is an integrated platform for analyzing the functional impact of genetic polymorphisms (SNPs and Indels) within microRNAs-binding sites [24]. Single variants within 3' UTR were selected from total variants and submitted to PolymiRTS server, to check if these variants could disrupt or create new miRNA-binding sites or have no impact at all.

2.5. Effect of 3'/5' Splice Sites SNPs/Indels by HSF Tool (http://www.umd.be/HSF3/index.html)

Human Splicing Finder (HSF) predicts the effects of substitutions on splicing signals and also identify splicing motifs in the human

genome [25]. In this study, HSF was used to detect the functional SNPs and Indels within 3'/5' splice sites. Input data were nucleotide sequence containing the single substitution as SNP or insertion/deletion as Indel.

2.6. Prediction of Amino Acid Conservation by ConSurf (ConSurf.tau.ac.il)

It calculates the evolutionary conservation of amino acid within a protein sequence using empirical Bayesian inference. It gives conservation score along with color scheme. Score 9 was given to most conserved amino acid, whereas 1 is given to variable amino acid. FASTA sequence of SHBG protein was submitted for ConSurf analysis.

2.7. Secondary Structure Prediction by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)

PSIPRED (PSIBLAST-based secondary structure prediction) predicted secondary structure of protein based on related sequences and position-specific scoring matrix. It predicted whether the residues were form strand, helix, and coils. Input format was the FASTA sequence of SHBG protein.

2.8. Three-Dimensional (3D) Structure Prediction by Swiss Model (https://swissmodel.expasy.org/)

Prediction of 3D structure was done by Swiss Modeller which allow to model the amino acid on the basis of structure homology. It allows modeling using manual template selection or by automated selection mode. It identifies the template, align the sequence, generate model then assess the model quality in terms of QMEAN value. FASTA sequence (mutation incorporated) was modeled against PDB structure of SHBG protein. Swiss Pdb viewer tool was used to visualize and analyze generated model.

2.9. Ligand-Binding Site Prediction by FTSite Server (http://FTSite.bu.edu/)

FTSite server predicted the ligand-binding site in 3D structure of protein. Its prediction is energy based and identifies the binding site accurately over 94% of the apoproteins listed today. It requires PDB data to predict the ligand-binding hotspot.

3. RESULTS

3.1. SNP Retrieval

SNP for SHBG gene were retrieved from dbSNP database. Search queries used were SHBG and SHBG. A total of 4630 SNP of SHBG gene were reported in database (build150). Search was then further restricted to get various SNP of *Homo sapiens*, i.e., nsSNP (missense, non-sense, and stop gained), SNP located in 3' and 5' UTRs, splice sites, frameshift, and intronic SNPs. 2687 SNPs were found in intron region, 166 were coding synonymous, 15 in splice site, 303 in UTRs, and 328 were non-synonymous SNPs [Figure 1].

3.2. Prediction of Deleterious nsSNP

A total of 328 nsSNP were screened to find their effect on protein structure and function. The first step is to predict the nsSNP carried out amino acid substitution. SIFT predicts the effect of nsSNP on protein structure and tells whether the induced amino acid is tolerable at that position or not. SIFT predicted eight nsSNP caused amino acids substitution out of total 328nsSNPs [Table 1]. Out of these 8

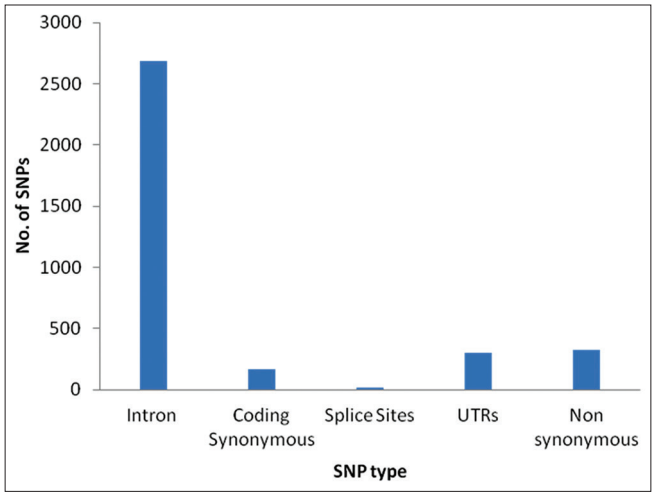


Figure 1: Different single-nucleotide polymorphism types of sex hormone-binding globulin gene

substitutions, 4 substitutions (P185L, R22H, R126C, Y59H) were found to damaging at the substituted position, while remaining 4 (D356N, R25H, Q371H, A179H) were found to be tolerated. These eight substitutions were further used as input for other tools to predict their effect. PROTEIN variation effect analyzer (PROVEAN) predicts whether a protein sequence variation affects protein function. It predicts the effect the amino acid substitutions as well as InDels on protein function. Out of eight substitutions predicted by SIFT, only one P185L was found to have deleterious effect on protein function by PROVEAN. Polyphen (polymorphism phenotyping) predicts the effect of nsSNP on protein structure and function. Its output predicts the effect in the form of three categories, i.e., probably damaging, possibly damaging, and benign with a specificity and sensitivity score in human var and hum Div forms. It predicted 2 substitutions (P185L, Q371H) as probably damaging, 2 substitutions (A179H, R126C) as possibly damaging, and remaining 4 (D356N, R25H, R22H, Y59H) as benign in nature. nsSNP Analyzer predicts whether a nsSNP has a phenotypic effect with some additional information, e.g., structural environment to facilitate the interpretation of results. It predicted only one (R25H) result in diseased phenotype [Table 1]. PhD-SNP is a SVM-based classifier, predicts the phenotypic effect of a non-synonymous SNP. Eight substitution given by SIFT server were used as input for PhD-SNP and it predicted 2 substitutions (P185L, Y59H) as disease-causing and remaining 6 as neutral [Table 1]. Mutation Accessor predicts the functional impact of non-synonymous mutations on proteins. Its prediction is in the form of impact scale, i.e., high, medium, low, and neutral. Out of 8 substitutions submitted for analysis, 2 (P185L, Q371H) have medium impact, 4 substitutions (D356N, A179H, R126C and Y59H) possess low impact, and remaining 2 (R25H, R22H) were neutral in nature [Table 2]. Effect of these nsSNPs on protein stability was predicted by I-Mutant which is a SVM-based neural network. All the eight substitutions were found to decrease the stability [Table 1]. Further effect of these substitutions on protein structure and function was accessed by Mutpred server as shown in Table 2.

3.3. Identification of CRE

CREs, regulate the transcription of nearby genes are regions of non-coding DNA. 3 programs, i.e., Proscan, Promoter2, and TSSG were used to analyze Cis-regulatory regions. Proscan predicted the 8 promoter region in SHBG gene, 4 were residing at forward, and 4

Table 1: Prediction of the effect of nsSNP by various tools.

rsIDS	AAS	SIFT	PROVEAN	Polyphen	nsSNP	PhD-SNP	Mutation assessor	I-Mutant (stability)
rs6258	P185L	Damaging	Deleterious	Probably Damaging	Neutral	Disease	Medium	↓se
rs6259	D356N	Tolerated	Neutral	Benign	Neutral	Neutral	Low	↓se
rs6260	R25H	Tolerated	Neutral	Benign	Disease	Neutral	Neutral	↓se
rs9282845	R22H	Damaging	Neutral	Benign	Neutral	Neutral	Neutral	↓se
rs113214318	Q371H	Tolerated	Neutral	Probably Damaging	Neutral	Neutral	Medium	↓se
rs1153367000	A179H	Tolerated	Neutral	Possibly Damaging	Neutral	Neutral	Low	↓se
rs13894	R126C	Damaging	Neutral	Possibly Damaging	Neutral	Neutral	Low	↓se
rs113136747	Y59H	Damaging	Neutral	Benign	Neutral	Disease	Low	↓se

nsSNP: Non-synonymous single-nucleotide polymorphism, SIFT: Sorting intolerant from tolerant.

Table 2: Effect of nsSNP on protein structure and function by Mutpred.

rsIDS	AAS	Top 5 affected features
rs6258	P185L	Loss of helix (*P*=0.2662) Loss of catalytic residue at L186 (*P*=0.2825) Loss of disorder (*P*=0.3542) Loss of loop (*P*=0.3664) Loss of methylation at R183 (*P*=0.385)
rs6259	D356N	Loss of loop (*P*=0.0804) Loss of solvent accessibility (*P*=0.1202) Gain of sheet (*P*=0.1208) Loss of relative solvent accessibility (*P*=0.1807) Loss of disorder (*P*=0.2304)
rs6260	R25H	Loss of MoRF binding (*P*=0.0131) Gain of helix (*P*=0.0854) Loss of disorder (*P*=0.1675) Loss of solvent accessibility (*P*=0.2034) Gain of methylation at R22 (*P*=0.2042)
rs9282845	R22H	Loss of MoRF binding (*P*=0.0045) Loss of stability (*P*=0.1215) Loss of helix (*P*=0.1299) Loss of disorder (*P*=0.1879) Loss of phosphorylation at T24 (*P*=0.1931)
rs113214318	Q371H	Gain of catalytic residue at Q369 (*P*=0.0859) Loss of sheet (*P*=0.1907) Gain of disorder (*P*=0.2388) Gain of loop (*P*=0.3485) Loss of stability (*P*=0.4433)
rs1153367000	A179H	Gain of loop (*P*=0.2045) Gain of solvent accessibility (*P*=0.2902) Gain of catalytic residue at A179 (*P*=0.2988) Loss of sheet (*P*=0.302) Loss of stability (*P*=0.3132)
rs13894	R126C	Loss of MoRF binding (*P*=0.0045) Loss of disorder (*P*=0.0349) Loss of sheet (*P*=0.0817) Loss of solvent accessibility (*P*=0.1077) Gain of loop (*P*=0.2045)
rs113136747	Y59H	Gain of disorder (*P*=0.0312) Loss of phosphorylation at Y59 (*P*=0.0855) Loss of sheet (*P*=0.1501) Gain of loop (*P*=0.2045) Loss of stability (*P*=0.2456)

nsSNP: Non-synonymous single-nucleotide polymorphism.

were resides at reverse stand as shown in Table 3. Promoter2 server predicted 16 promoter region, out of which 14 were at marginal prediction and two position (8300 and 17100 with a score 1.162 and 1.149, respectively) were at strong promoter prediction score. TSSG is highly advanced server predicted 7 promoter region at position 818, 1127, 407, 4760, 4029, 2614, and 14724, out of which 5 having TATA box as shown in Table 3.

3.4. Effect of SNP 3'UTRs Prediction by PolymiRTS Database

Polymorphism in microRNAs and their target site (PolymiRTS) is database of naturally occurring DNA variations in microRNA seed region and miRNA target sites. One hundred and eight SNPs in UTRs region given by dbSNP database were used as input of this server and it predicted the 6 SNPs (rs113214318, rs138612626, rs200408334, rs199525667, rs142693170, and rs144865079) were have an impact on miRNA target site and binding site. Results are given in Table 4. Two SNPs (rs113214318, rs138612626) allowed wobbling at their position while remaining four do not allow any wobbling. At rs113214318 wobbling of both A and G disrupt the conserved miRNA site while remaining SNPs creates a new miRNA site by replacement of ancestral allele by A, G, C, A and T, respectively.

3.5. Effect of 3'/5' Splice Sites SNPs/Indels (HSF tool)

HSF (HSF tool) predicts the effect of SNP on splicing signals and also identifies the slicing motifs in any human sequence. A total of 15 SNPs were found in 3' and 5' splice site of SHBG gene and submitted in HSF server. Ten SNPs of splice region were predicted by HSF tool to alter the splice signals. rs141042763 created new ESS site which cause potential alteration of splicing.rs763346578 and rs370605449 were found to create new exonic splicing enhancer (ESE) site and probably do not have any impact on splicing. Remaining 7 sites altered the donor and acceptor signals, thereby have the tendency to alter splicing signals [Table 5].

3.6. Conservation Analysis by ConSurf

Figure 2 represents the result predicted by the ConSurf tool which contains nine color codes based on conservation score, which indicates evolutionary relationships among their sequence homologs. It was found that P185L, Q371H, and R126C are exposed in protein structure and had a conservation score 7, 7, and 9, respectively, i.e., conserved in protein structure whereas D356N, R25H, and A1791 had score 1, 2, 1, i.e., variable in nature. R22H and Y59H had average score, i.e., 5.

Table 3: Identification of cis-regulatory elements by three different servers.

PROSCAN			Promoter2			TSSG		
Strand	**Position**	**TATA BOX**	**Strand**	**Score**	**Prediction**	**Position**	**Score**	**TATA box**
Forward	449–699		1700	0.522	Marginal prediction	818	14.63	
Forward	839–1089		2500	0.638	Marginal prediction	1127	9.25	
Forward	7897–8147	8101	4100	0.727	Marginal prediction	407	8.43	390
Forward	14414–14164	14198	5600	0.636	Marginal prediction	4760	7.17	4729
Reverse	12875–12625	12644	6400	0.616	Marginal prediction	4029	5.57	3998
Reverse	11404–11154	11204	7800	0.566	Marginal prediction	2614	4.46	2553
Reverse	1144–894		8300	1.162	Highly likely prediction	14724	4.43	14694
Reverse	720–470		9000	0.545	Marginal prediction			
			10600	0.547	Marginal prediction			
			12300	0.649	Marginal prediction			
			13100	0.603	Marginal prediction			
			13700	0.578	Marginal prediction			
			14200	0.564	Marginal prediction			
			15600	0.570	Marginal prediction			
			17100	1.149	Highly likely prediction			
			18700	0.664	Marginal prediction			

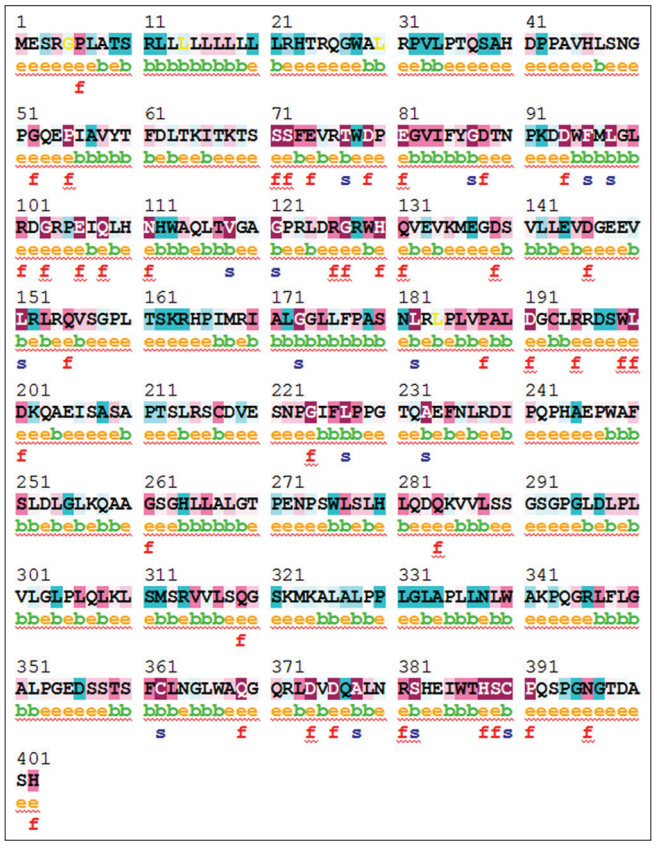
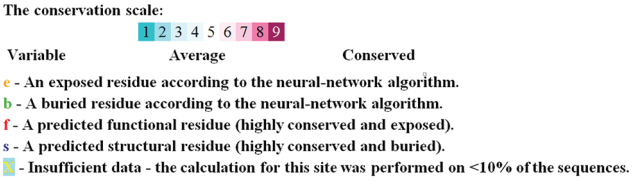


Figure 2: Evolutionary conservation analysis by ConSurf.



3.7. Secondary Structure Prediction by PSIPRED

Secondary structure of SHBG was predicted by PSIPRED which showed the distribution of alpha helix, beta sheet, and coils. By analysis, it was found that in native structure coils contribute major portion in protein structure (53.4%) followed by β-strand (43.1%) and alpha helix (3.44%) [Figure 3]. D165N substitution resulted in loss of strand at 149–150, 190–200, and 290–300th residues and gain of strand at 220–230 and 230–240th residues, thereby distort the native secondary structure [Figure 4]. On insertion of all the substitutions predicted by SIFT server, major distortion was loss of strand and gain of helix at residues 250–330th residues [Figure 5].

3.8. 3D Structure Prediction by Swiss-Modeler

Nine models were generated by Swiss modeler for SHBG protein. Models with the Z-score between the ranges of 0–1 are considered as good models. Wild-type 3D Model had a Z-score of 0.259 whereas P185L mutant had a Z-score of 0.45. Both the native and mutated models were further visualized and analyzed by UCSF Chimera [Figures 6 and 7]. 3D structure of SHBG protein was of 373 amino acid residues where former 29 amino acid residues are not included in the protein structure. The residue P185L hence lie on the 156th position in 3D structure. This substituted position is labeled in both the models, i.e., wild and mutant, respectively. Energy minimization of P185L substituted model was done using Swiss PDB viewer, and it was found to be reduced from −6459.697 KJ/mol to −8834.880 KJ/mol. Energy minimization value of mutated model was found to be greater than wild-type model. Superimposed structure of native and mutated SHBG protein presented in ribbon form are presented in Figure 8 where brown color side chain is of Proline in native structure and blue side chain is of leucine in mutated structure.

3.9. Ligand-Binding Site Prediction by FTSite

Sites for ligand binding were predicted by FTSite algorithms and visualized and further analyzed using Pymol. By this tool, 3 ligand-binding sites were identified in human SHBG protein [Figure 9]. Site 1

Table 4: Predicted results of functional 3'UTRs SNPs/Indels.

dbSNP ID	Variant type	Ancestral allele	Allele	miR ID	Conservation	miRSite	Function class	Context+Score change
rs113214318	SNP	A	A	hsamiR301a5p	2	aagGTCAGAGgct	D	−0.135
				hsamiR4256	2	aaGGTCAGAggct	D	−0.194
				hsamiR47503p	2	aaGGTCAGAggct	D	−0.141
				hsamiR67593p	2	AAGGTCAgaggct	D	−0.09
				hsamiR68193p	4	aaggtcAGAGGCT	D	−0.146
				hsamiR68773p	4	aaggtcAGAGGCT	D	−0.184
rs138612626	SNP	G	G	hsamiR71523p	13	gatgtGGACCAGg	D	−0.225
			A	hsa-miR-130a-5p	2	gATGTGAAccagg	D	−0.018
				hsa-miR-23a-3p	2	gATGTGAAccagg		−0.018
				hsa-miR-23b-3p	2	gATGTGAAccagg		−0.018
				hsa-miR-23c	2	gATGTGAAccagg		−0.022
rs200408334	SNP	C	C	hsa-miR-26b-3p	3	cctGAACAGAagc	D	−0.056
				hsa-miR-3182	3	cctgaACAGAAGc	D	−0.079
				hsa-miR-4652-3p	3	cctgAACAGAAgc	D	0.023
			G	hsa-miR-1236-3p	3	cctGAAGAGAagc	C	−0.044
				hsa-miR-4753-3p	3	cctgAAGAGAAgc	C	−0.021
				hsa-miR-6515-3p	3	ccTGAAGAGAagc	C	−0.178
				hsa-miR-6809-3p	3	cctgAAGAGAAgc	C	0.003
				hsa-miR-942-5p	3	cctgaAGAGAAGc	C	−0.114
rs199525667	SNP	A	A	hsa-miR-135a-5p	2	gAAGCCATgagat	D	−0.129
				hsa-miR-135b-5p	2	gAAGCCATgagat	D	−0.129
				hsa-miR-6815-3p	9	GAAGCCAtgagat	D	−0.184
			C	hsa-miR-1266-5p	2	gaagCCCTGAGAt	C	−0.379
				hsa-miR-3127-5p	2	gaaGCCCTGAgat	C	−0.227
				hsa-miR-3158-3p	2	gAAGCCCTgagat	C	−0.162
				hsa-miR-3664-3p	2	gaagcCCTGAGAt	C	−0.109
				hsa-miR-3918	2	gaaGCCCTGAgat	C	−0.156
				hsa-miR-4518	2	gaagCCCTGAGAt	C	−0.348
rs142693170	SNP	C	C	hsa-miR-148a-3p	5	atgGCACTGAcgc	D	−0.124
				hsa-miR-148b-3p	5	atgGCACTGAcgc	D	−0.124
				hsa-miR-152-3p	5	atgGCACTGAcgc	D	−0.124
				hsa-miR-4697-3p	5	atggCACTGACgc	D	−0.151
			A	hsa-miR-4317	5	atGGCAATGAcgc	C	−0.354
				hsa-miR-4659a-5p	6	ATGGCAAtgacgc	C	−0.091
				hsa-miR-4659b-5p	6	ATGGCAAtgacgc	C	−0.091
				hsa-miR-4776-3p	6	ATGGCAAtgacgc	C	−0.144
				hsa-miR-5590-5p	6	ATGGCAAtgacgc	C	−0.106
				hsa-miR-616-3p	5	atggCAATGACgc	C	−0.069
rs144865079	SNP	C	C	hsa-miR-4697-3p	5	CACTGACgcttcc	D	−0.151
			T	hsa-miR-4711-5p	2	caCTGATGCttcc	C	−0.191
				hsa-miR-5094	5	CACTGATgcttcc	C	−0.124
				hsa-miR-6828-5p	2	cactgaTGCTTCC	C	−0.2

SNP: Single-nucleotide polymorphism, UTR: Untranslated region.

consisted of 28 residues; site 2 and site 3 consisted of 9 residues. None of the eight substituted positions predicted by SIFT server lie in the predicted ligand-binding sites [Table 6].

4. DISCUSSION

SHBG is a transport protein specific for androgens [1]. It is located on the short arm of chromosome 17 and made up of eight exons [3]. SHBG is mainly secreted by the liver into the bloodstream although it has been found to be secreted by testis germ cells and brain tissue also. It also recognizes a specific binding site located on membranes of sex steroid target (Rosner *et al.* 1984; Plymate *et al.* 1988). Its synthesis is stimulated by estrogens and thyroid hormones and is inhibited by androgens and insulin [7,8]. Abnormally low serum SHBG values are frequently found in women with polycystic ovary syndrome and contribute to hyperandrogenic symptoms such as hirsutism and acne [26]. Serum SHBG values are also reduced in patients with an increased accumulation of abdominal fat [27-29], hyperinsulinemia, glucose intolerance, insulin resistance, and in people at risk of coronary heart disease as well as type 2 diabetes [30,31]. Reduced SHBG levels are also associated with an adverse blood lipid profile, including reduced HDL-cholesterol [32,33].

Effect of nsSNP on the protein structure and function was assessed by software having different prediction algorithm. The 328 nsSNPs listed in human SHBG gene were first submitted to SIFT server to predict those SNP which cause amino acid substitutions. SIFT server analyzed them and predicted 8 substitutions, out of which 4 substitutions (P185L, R22H, R126C, Y59H) were damaging. These

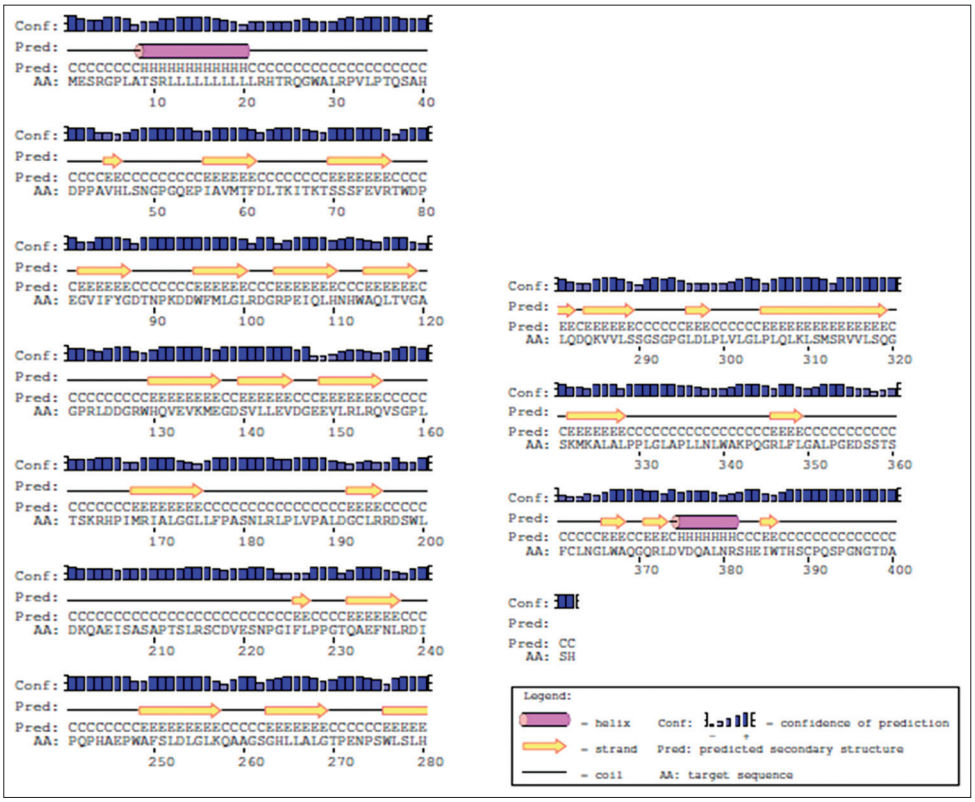


Figure 3: Secondary structure analysis of native sex hormone-binding globulin by PSIPRED.

Table 5: Effect of splice sites associated SNP predicted by HSF.

rsids	Substitution	Predicted signal	Interpretation	Exon position
rs141042763	c. 447G>A	New ESS Site	Creation of an exonic ESS site Potential alteration of splicing	4
		ESE Site Broken	Alteration of an exonic ESE site Potential alteration of splicing	
rs370605449	c. 853-25C>A	New ESE Site	Creation of an intronic ESE site Probably no impact on splicing	7
rs750069375	c. 852+2_852+3insA	Broken WT Donor Site	Alteration of the WT donor site, most probably affecting splicing	6
rs754083080	c. 112-1G>C	Broken WT Acceptor Site	Alteration of the WT acceptor site, most probably affecting splicing	2
rs754804182	c. 394-2A>C	Broken WT Acceptor Site	Alteration of the WT acceptor site, most probably affecting splicing	4
rs755177340	c. 609G>C	ESE Site Broken	Alteration of an exonic ESE site Potential alteration of splicing	5
rs758832324	c. 716-1G>T	Broken WT Acceptor Site	Alteration of the WT acceptor site, most probably affecting splicing	6
rs763346578	c. 853-24G>A	New ESE Site	Creation of an intronic ESE site Probably no impact on splicing	7
rs764464601	c. 112-2A>G	Broken WT Acceptor Site	Alteration of the WT acceptor site, most probably affecting splicing	2
rs777398160	c. 394-1G>A	Broken WT Acceptor Site	Alteration of the WT acceptor site, most probably affecting splicing	4
		New Acceptor Site	Activation of an intronic cryptic acceptor site Potential alteration of splicing	
rs780112120	c.-61-3712C>T	-		
rs900657041	c.-61-3191C>T	-		

SNP: Single-nucleotide polymorphism, HSF: Human Splicing Finder, ESE: Exonic splicing enhance.

8 substitutions were further analyzed to predict their effect on the structure and function of protein. PROVEAN predicted P185L as deleterious. Polyphen predicted 2 substitutions (P185L, Q371H) were probably damaging, 2 (A179H, and R126C) were possibly damaging, and remaining 4 were benign. nsSNP Analyzer and PhD-SNP predicted R25H and P185L, Y59H were disease associated, respectively. Mutation Assessor predicted all eight substituted positions would impact protein structure and function. Stability of

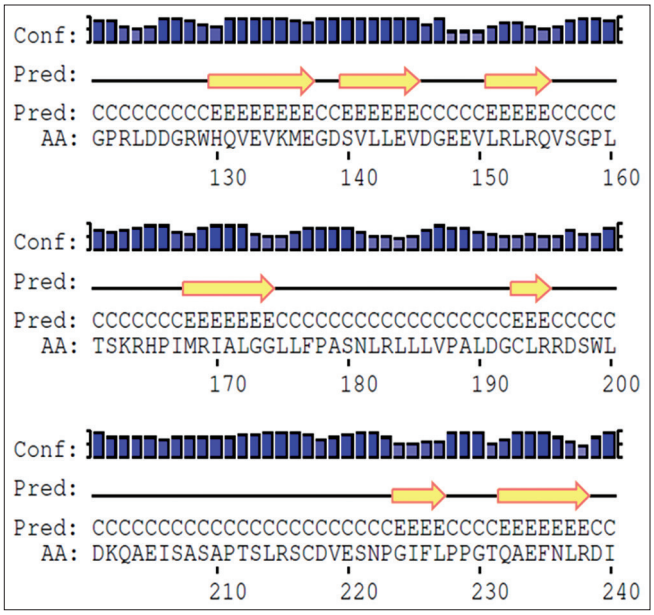


Figure 4: Secondary structure distortion.

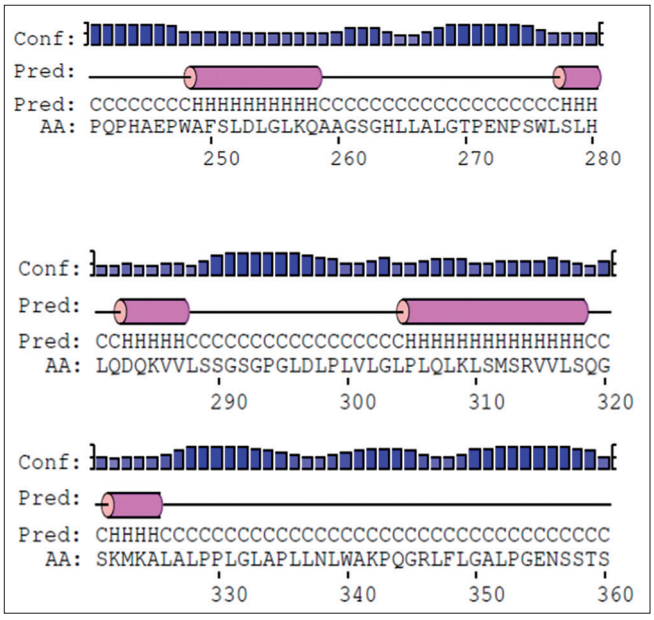


Figure 5: Secondary structure distortion after D165N substitution after all the substitutions given by sorting intolerant from tolerant.

protein was found to be decreased in all the substituted positions. Four positions (R22H, R25H, P185L, and D356N) are natural variant of SHBG gene, but till date, only rs6259 (D356N) was the only site validated to find its association with diseases and it was found that this polymorphism is not pathogenic rather it provides protection against hyperandrogenism development in females. Our *in silico* results were also consistence with this finding as rs6259 polymorphism is found to be neutral by all the softwares listed in our study. This is the only nsSNP site validated in laboratory except (TAAAA) n promoter polymorphism. In our study, we found P185L substitution to be highly pathogenic as predicted damaging by most of the softwares listed in the study. By Mutpred analysis, it was found that P185L substitution resulted in loss of helix, catalytic

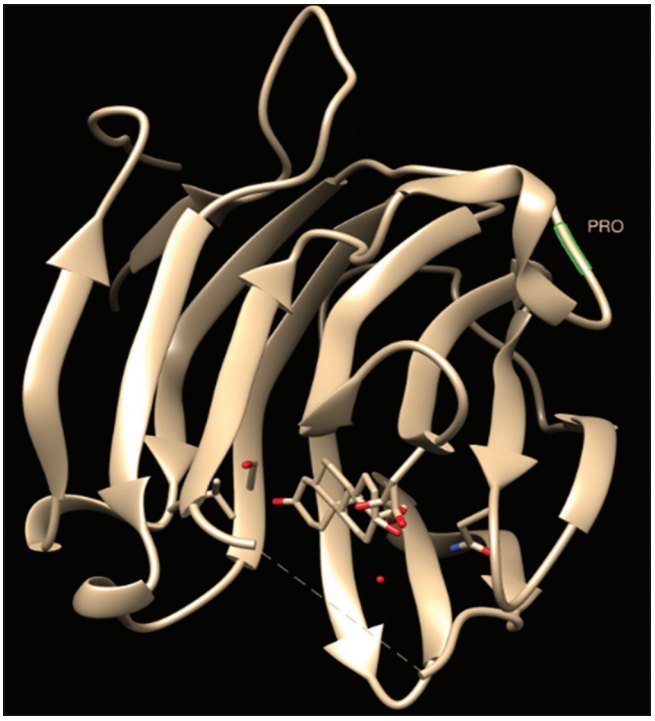


Figure 6: Native sex hormone-binding globulin protein (373amino acids) 3D structure with Proline marked at 156th position.

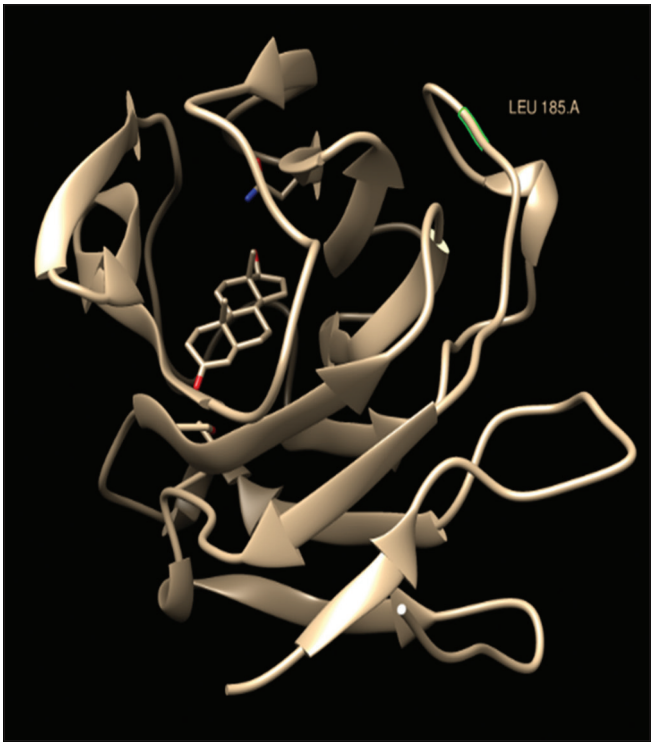


Figure 7: P185L substituted sex hormone-binding globulin protein (402 amino acid residues) 3D structure.

residue, loop, and methylation thus predicted as damaging. As reported in Claes Ohlsson (2011) study, rs6258 (P185L) results in the production of an *SHBG* variant with reduced affinity for testosterone provides an explanation for the association between rs6258 and low

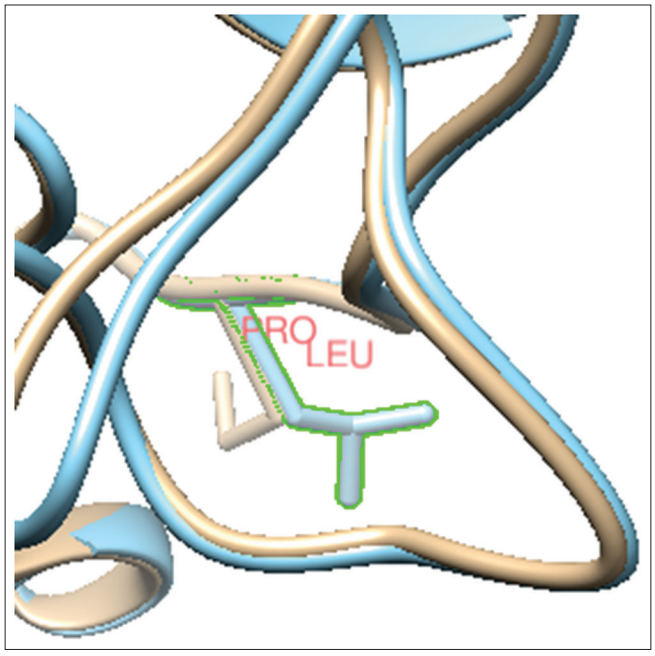


Figure 8: Superimposed 3D structure of native and P185L substituted sex hormone-binding globulin protein with their side chain where brown chain in for Proline and blue is for Leucine.

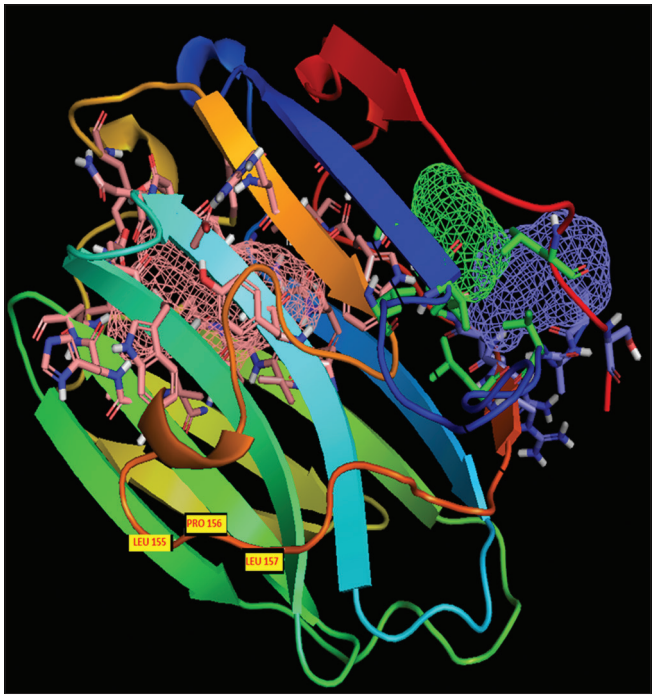


Figure 9: Ligand-binding site prediction by FT site. 3 meshes were predicted as ligand-binding site. P156 is not lie in any binding site.

serum testosterone concentrations [34]. This study was not based on polymorphic analysis, but on the binding of wild and recombinant SHBG, hence need to be validated genotypically.

Position of cis-regulatory sequences was also identified by three different softwares. Proscan identified 8 different promoter sites, 4 on forward and 4 on reverse strand but only 4 TATA-binding sites were located on forward and reverse strand, respectively. TSSG tool predicted 7 promoter sites out, of which 4 were same as predicted by Proscan program.

Table 6: Residues at ligand-binding sites of SHBG protein.

Site 1	Site2	Site 3
THR A 40	VAL A 16	PRO A 14
SER A 41	ILE A 27	ALA A 15
SER A 42	ALA A 28	VAL A 16
PHE A 56	MET A 30	LEU A 165
TYR A 57	LEU A 143	ARG A 166
GLY A 58	LEU A 165	ARG A 167
ASP A 59	ASP A 168	ASP A 168
THR A 60	TRP A 170	LEU A 185
ASN A 61	LEU A 185	SER A 187
ASP A 65		
TRP A 66		
PHE A 67		
LEU A 80		
HIS A 81		
ASN A 82		
VAL A 105		
LYS A 106		
MET A 107		
VAL A 112		
LEU A 113		
SER A 128		
LEU A 131		
LYS A 134		
MET A 139		
ARG A 140		
ILE A 141		
TRP A 170		
LEU A 171		

SHBG: Sex hormone-binding globulin.

Both the sides of mRNA are flanked by UTR. Information contained in DNA is transferred in mRNA and then converted into the functional effector, i.e., protein. However, all the regions of the mRNA are usually not translated into protein, i.e., 5' and 3' UTRs. These regions as the name implies generally not contribute to the protein coding region of the gene but some 5' UTR element (UoRFs) get translated into protein. Termination codons are flanked by 3' UTR plays a crucial role in termination of translation as well as post-transcriptional gene expression. 108 SNPs in UTRs region given by dbSNP database were analyzed by PolymiRTS server. It predicted 6 SNPs (rs113214318, rs138612626, rs200408334, rs199525667, rs142693170, and rs144865079) have an impact on miRNA target site and binding site. Two SNPs (rs113214318, rs138612626) allowed wobbling at their position. At rs113214318 wobbling of both A and G disrupt, the conserved miRNA site while remaining SNPs creates a new miRNA site by replacement of ancestral allele. These 6 UTR-associated SNPs possess great impact on SHBG structure and are still uncharacterized yet.

Splice sites are the consensus sequences that make the exon recognition by marking the termini location of introns. The deletion of the splicing site results in one or more introns remaining in mature mRNA and may lead to the production of abnormal proteins. Mutation is splice site affect the proper removal of intron, thereby leaving introns segments between the exon and resulted in non-functional protein. Effect of SNPs associated with splice sites was predicted by HSF tool. It predicted 10 SNPs would alter splicing signal. rs370605449 and rs763346578 would create new ESE site and probably do not affect splicing. rs141042763 and rs755177340 would result in broken

ESE, thereby affecting the splicing signals. Five sites (rs754083080, rs754804182, rs758832324, rs764464601, and rs777398160) were predicted to break acceptor site and affect splicing. rs777398160 broke wild-type acceptor signal and found to create new acceptor site which activates an intronic cryptic acceptor site resulting in faulty splicing.

Residues involved in the function of protein are generally found to be conserved in protein structure. Evolutionary conservation of functional regions was predicted by ConSurf. Score was assigned on the basis of their relatedness among proteins and their homologs. P185L substitution is predicted as damaging by SIFT server. This position is exposed and has a conservation score 7 indicating the mutation is least evolving. R126C substitution is highly conserved hence does not allow evolutionary changes.

The primary structure of SHBG composed of 402 amino acids, i.e., signal sequence of 29 amino acid residues followed by residues contributing to 3D structure (373 amino acids). The secondary structure of SHBG gene was predicted by PSIPRED which represented higher portion of strand than coil and helix in SHBG protein. 3D structure predicted by Swiss modeler on SHBG template of 373 amino acid residues (Z score 0.295).

Function of protein depends on the tertiary structure of protein. Using template SHBG from PDB, mutated (P185L) SHBG structure was derived which has a Z score of 0.45. Structure was further refined and analyzed by Swiss PDB viewer to map substituted position. Energy of mutated (P185L) was higher than native model. This indicated that this substitution decrease the stability of SHBG protein. Quality of mutated model was validated by RAMPAGE which predicted 97.1% residues in favored region and 2.9% in allowed region. In P185L substitution, both the amino acid residues different their side chain hence contribute to different energy state and interaction property, thereby making functional difference.

Effect of P185L substitution on protein function was further accessed by FTSite. FTSite server predicted 3 ligand-binding site made up of 29 and 9 residues, respectively. By this, we found that P185L substitutions do not involve in ligand-binding site and not even form the catalytic coordination sphere hence do not affect the binding affinity of the SHBG protein. But by other software and prediction method, it was found that this substitution affects the SHBG structure and decreases its stability.

5. CONCLUSION

SHBG, a transporter protein is a regulator of free androgen pool of human body. Although SHBG levels directly affect the androgen pool, thereby is an important factor for maintaining the proper reproductive physiology. By different software and Mutpred analysis, we can say that SNP filtered by our study have impact on the structure and function of SHBG. In addition to nsSNP, we also found six SNPs associated with UTRs and 10 SNPs associated with splice site also regulates the gene expression. Most of the SNPs predicted by present analysis are not characterized in their disease-causing potential, hence provide a new pavement for further clinical and experimental studies to validate the effect of these mutations for the development of new therapeutics models.

6. ACKNOWLEDGMENT

The authors acknowledge Centre for Medical Biotechnology, Maharshi Dayanand University, Rohtak, India, for providing research work platform. This work was supported by INSPIRE division of the Department of Science and Technology in the form of a research fellowship.

REFERENCES

1. Aydın B, Winters SJ. Sex hormone-binding globulin in children and adolescents. J Clin Res Pediatr Endocrinol 2016;8:1.
2. Petra PH, Namkung PC, Senear DF, McCrae DA, Rousslang KW, Teller DC, *et al.* Molecular characterization of the sex steroid binding protein (SBP) of plasma. Re-examination of rabbit SBP and comparison with the human, macaque and baboon proteins. J Steroid Biochem 1986;25:191-200.
3. Hammond GL, Underhill DA, Rykse HM, Smith CL. The human sex hormone-binding globulin gene contains exons for androgen-binding protein and two other testicular messenger RNAs. Mol Endocrinol 1989;3:1869-76.
4. Berube D, Seralini GE, Gagne R, Hammond GL. Localization of the human sex hormone-binding globulin gene (SHBG) to the short arm of chromosome 17 (17p12→ p13). Cytogenet Genome Res 1990;54:65-7.
5. Rosner W, Aden DP, Khan MS. Hormonal influences on the secretion of steroid-binding proteins by a human hepatoma-derived cell line. J Clin Endocrinol Metab 1984;59:806-8.
6. Plymate SR, Matej LA, Jones RE, Friedl KE. Inhibition of sex hormone-binding globulin production in the human hepatoma (Hep G2) cell line by insulin and prolactin. J Clin Endocrinol Metab 1988;67:460-4.
7. Edmunds SE, Stubbs AP, Santos AA, Wilkinson ML. Estrogen and androgen regulation of sex hormone binding globulin secretion by a human liver cell line. J Steroid Biochem Mol Biol 1990;37:733-9.
8. Cibula D, Skrha J, Hill M, Fanta M, Haaková L, VrbIková J, *et al.* Prediction of insulin sensitivity in nonobese women with polycystic ovary syndrome. J Clin Endocrinol Metab 2002;87:5821-5.
9. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931-45.
10. Dabhi B, Mistry KN. *In silico* analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. Meta Gene 2014;2:586-95.
11. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, *et al*. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 1999;402:880-3.
12. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. J Mol Biol 2001;307:683-706.
13. Lander ES. The new genomics: Global views of biology. Science 1996;274:536-9.
14. Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, *et al.* Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 1994;331:1056-61.
15. Bhagwat M. Searching NCBI's dbSNP database. Curr Protoc Bioinformatics 2010;Chapter 1:Unit 1.19.
16. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812-4.
17. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and Indels. PLoS One 2012;7:e46688.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, *et al.* A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.
19. Bao L, Zhou M, Cui Y. NsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms.

Nucleic Acids Res 2005;33:W480-2.
20. Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:W306-10.
21. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, *et al.* The pfam protein families database: Towards a more sustainable future. Nucleic Acids Res 2016;44:D279-85.
22. Delorenzi M, Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 2002;18:617-25.
23. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, *et al.* DisProt: The database of disordered proteins. Nucleic Acids Res 2007;35:D786-93.
24. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 2010;464:768-72.
25. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C, *et al.* Human splicing finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.
26. Anderson DC. Sex-hormone-binding globulin. Clin Endocrinol 1974;3:69-96.
27. Evans DJ, Hoffmann RG, Kalkhoff RK, Kissebah AH. Relationship of androgenic activity to body fat topography, fat cell morphology, and metabolic aberrations in premenopausal women. J Clin Endocrinol Metab 1983;57:304-10.
28. Raison J, Bonithon-Kopp C, Egloff M, Ducimetiere P, Guy-Grand B. Hormonal influences on the relationships between body fatness, body fat distribution, lipids, lipoproteins, glucose and blood pressure in French working women. Atherosclerosis 1990;85:185-92.
29. Haffner SM, Katz MS, Dunn JF. Increased upper body and overall adiposity is associated with decreased sex hormone binding globulin in postmenopausal women. Int J Obes 1991;15:471-8.
30. Lapidus L, Lindstedt G, Lundberg PA, Bengtsson C, Gredmark T. Concentrations of sex-hormone binding globulin and corticosteroid binding globulin in serum in relation to cardiovascular risk factors and to 12-year incidence of cardiovascular disease and overall mortality in postmenopausal women. Clin Chem 1986;32:146-52.
31. Lindstedt G, Lundberg PA, Lapidus L, Lundgren H, Bengtsson C, Björntorp P, *et al.* Low sex-hormone-binding globulin concentration as independent risk factor for development of NIDDM 12-yr follow-up of population study of women in Gothenburg, Sweden. Diabetes 1991;40:123-8.
32. Soler JT, Folsom AR, Kaye SA, Prineas RJ. Associations of abdominal adiposity, fasting insulin, sex hormone binding globulin, and estrone with lipids and lipoproteins in post-menopausal women. Atherosclerosis 1989;79:21-7.
33. Pugeat M, Moulin P, Cousin P, Fimbel S, Nicolas MH, Crave JC, *et al.* Interrelations between sex hormone-binding globulin (SHBG), plasma lipoproteins and cardiovascular risk. J Steroid Biochem Mol Biol 1995;53:567-72.
34. Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JR, Koster A, *et al.* Genetic determinants of serum testosterone concentrations in men. PLoS Genet 2011;7:e1002313.

How to cite this article:
Bhatnager R, Dangi M, Dang AS. Comprehensive analysis of damage associated single nucleotide polymorphisms of sex hormone binding globulin gene. J App Biol Biotech. 2018;6(05):1-11.
DOI: 10.7324/JABB.2018.60501