Original

In-silico Evaluation of Rare Codons and their Positions in the Structure of ATP8b1 Gene

Zarenezhad M.[1,2*], Dehghani S. M.[3], Ejtehadi F.[3], Fattahi M. R.[3], Mortazavi M.[4], Tabei S. M. B.[5]

[1]MD, PhD, Gastroenterohepatology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[2]MD, PhD, legal medicine research center, legal medicine organization, Tehran , iran
[3]MD, Gastroenterohepatology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[4]PhD, Department of Biotechnology, Institute of Science and High Technology and Environmental Science, Graduate University of Advanced Technology, Kerman, Iran
[5]MD, Genetic Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

ABSTRACT

Background: Progressive familial intrahepatic cholestases (PFIC) are a spectrum of autosomal progressive liver diseases developing to end-stage liver disease. ATP8B1 deficiency caused by mutations in ATP8B1 gene encoding a P-type ATPase leads to PFIC1. The gene for PFIC1 has been mapped on a 19-cM region of 18q21-q22, and a gene defect in ATP8B1 can cause deregulations in bile salt transporters through decreased expression and/or activity of FXR. Point mutations are the most common, with the majority being missense or nonsense mutations. In addition, approximately 15% of disease-causing ATP8B1 mutations are annotated as splicing disrupting alteration given that they are located at exon-intron borders.

Objective: Here, we describe the hidden layer of computational biology information of rare codons in ATP8B1, which can help us for drug design.

Methods: Some rare codons in different locations of ATP8b1 gene were identified using several web servers and by in-silico modelling of ATP8b1 in Phyre2 and I-TASSER server, some rare codons were evaluated.

Results: Some of these rare codons were located at special positions which seem to have a critical role in proper folding of ATP8b1 protein. Structural analysis showed that some of rare codons are related to mutations in ATP8B1 that are responsible for PFIC1 disease, which may have a critical role in ensuring the correct folding.

Conclusion: Investigation of such hidden information can enhance our understanding of ATP8b1 folding. Moreover, studies of these rare codons help us to clarify their role in rational design of new and effective drugs.

Keywords

Progressive Familial Intrahepatic Cholestasis, Bioinformatics Analysis, ATP8b1, Rare Codon

Introduction

Cholestatic disorders are among the most severe liver diseases in infancy and childhood [1]. Cholestasis is defined as an impairment of normal bile flow and is divided into extra-hepatic cholestasis and intra-hepatic cholestasis, the latter can be of hepatocanalicular or ductal origin [2]. Progressive familial intrahepatic cholestases (PFICs) are a spectrum of autosomal liver disorders [3]. The three types of PFIC have distinctive clinical, biochemical and histological features [4]. PFIC1 or Byler disease [1] and PFIC2 or bile salt export pump (BSEP) disease [5] are associated with a low or normal serum gamma-glutamyltranspeptidase (GGT) activity, whereas PFIC3 or multidrug resistance

*Corresponding author:
M. Zarenezhad
MD, PhD, Gastroenterohepatology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
E-mail: zarenezhad@hotmail.com

Received: 27 July 2016
Accepted: 18 August 2016

protein 3 (MDR3) disease is associated with a high serum GGT activity [1]. All mentioned genes encode hepatocanalicular transporters. ATP8B1 encodes an amino-phospholipid flippase translocating phospholipids from the outer to the inner leaflet of the plasma membrane; ABCB11 encodes the bile salt export pump, a liver-specific adenosine triphosphate (ATP)-binding cassette transporter; ABCB4 encodes the multidrug resistance protein 3 functioning as a phospholipid floppase translocating phosphatidylcholine from the inner to the outer leaflet of the membrane [6, 7]. There are several studies on molecular evaluation of three different types of PFIC diseases including sequencing [8], evaluation of mutations of the genes [1, 9], locus mapping [10] and exon characterization [11]. On the other hand, recent studies show that rare codons have a critical role in protein folding and activity [12]. However, there is no study about computational biology and bioinformatics evaluation of PFIC. Furthermore, some reports indicate that ribosomal pausing occurs with decrease of tRNAs concentration in rare codons until the rare activated tRNA brings the next amino acid to ensure the independent folding of some regions of polypeptide chains [13, 14]. Rare codons studies can provide insights into the diseases background and help in problem solving of drug design [14]. So, the aim of this study is to evaluate the computational biology of PFIC1 with regard to rare codons and mutation leading to disease. These findings may help in operational development of this technology and in elucidating the ATP8B1 folding mechanism, as well as rational design of new and effective drugs.

Material and Methods

We studied, for the first time, rare codons in AT8B1_HUMAN and identified the location of these rare codons in the structure of ATP8B1. Detection of rare codons were performed using the ATGme (http://atgme.org/), Rare codon calculator (RaCC) (http://nihserver.mbi.ucla.edu/RACC/), LaTcOm (http://structure.biol.ucy.ac.cy/latcom.html), and Sherlocc program (http://bcb.med.usherbrooke.ca/sherlocc.php) [15]. By these analyses, some rare codons were identified and by molecular modeling in the I-TASSER and [16], the situation of these rare codons were studied using PyMOL Molecular Graphics System (21) and Swiss PDB Viewer software [17]. In the following, for study of PFIC1 and their relationships with rare codon, the position of some mutation was evaluated in comparison with rare codon.

Detection of Rare Codon in Gene and Protein Structure of ATP8b1

Rare codon detection in ATGme was performed in four steps: (i) Input of ATP8b1 sequence; (ii) Input of the codon usage table of Homo sapiens [gbpri]: 93487 CDS's (40662582 codons) that was obtained from the Codon Usage Database (http://www.kazusa .or.jp/codon/); (iii) Detection of rare codons. LaTcOm is a new web tool designed for detecting and visualizing rare codon clusters (RCC) [18]. In this tool, three core RCC detection algorithms are implemented: i) % minimax algorithm, ii) sliding window approach, and iii) a linear-time algorithm named MSS. RCC was used with the following parameters: MSS, Scale: Dong table codon usage [19], cluster length: 21 and transformation: linear + sigmoid. Then RCC positions were visualized within the submitted sequences.

Study of Rare Codons in Structure of ATP8b1

To investigate the position of mutations and rare codon in the structure of ATP8b1, the 3D structure of this enzyme was modelled by the submission of ATP8b1sequence in the I-TASSER [16] and Phyre2 web servers. I-TASSER web server was used to generate a total of five most suitable models of target protein. In this web server, 3D models are built based on multiple-threading alignments by LOMETS (Local Meta-Threading-Server)

[20] and interactive template fragment assembly simulations. The models with the best "Confidence Score" and Z-score were chosen by I-TASSER server. Phyre2 web server applies the alignment of hidden Markov models via HHsearch to improve the accuracy of alignment and detection rates. The model with the best confidence and Z-score was selected and visualized using Swiss PDB viewer [21] and PyMOL molecular graphics system [22]. Hydrogen bonds were also detected by WHAT IF web server [23] and PIC web server [24].

Results

Detection of Rare Codon Clusters

Rare and highly rare codons are highlighted in orange and red, respectively. The protein family (Pfam) accession number of ATP8B1 was identified using UniProt database (http://www.uniprot.org/). Pfam is a comprehensive collection of protein domains and families represented as multiple sequence alignments and as profile hidden Markov models [25]. Analysis of this Pfam in Sherlocc program showed that this gene does not have any rare codon cluster. RaCC introduced codons for arginine (AGG, AGA and, CGA), leucine (CTA), isoleucine (ATA), and proline (CCC) with probable problem. Analysis of this gene in this server show some rare codons throughout the gene sequence. The Pfam accession numbers of AT8B1_HUMAN was identified as PF00122 (E1-E2_ATPase. 1 hit) PF16212 (PhoLip_ATPase_N. 1 hit) and PF16209 (PhoLip_ATPase_N. 1 hit) in the UniProt database (http://www.uniprot.org/). The results of Sherlocc program [15] show that this program did not identify any rare codon cluster in AT8B1 (Table 1).

Next, the nucleotide sequence of AT8B1_HUMAN was analyzed in ATGme server. This server identifies rare codons and gives several options for codon usage optimization. By the use of codon usage table of Homo sapiens [gbpri]: 93487 CDS's (40662582 codons) (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606), this gene was analyzed and the rare and highly rare codons were shown and highlighted in orange and red, respectively (Figure 1). Moreover, GC and AT contents of this gene were GC%: 45.42, AT%:54.58, calculated by this server.

As these findings demonstrate, AT8B1_HUMAN gene has some rare and highly rare codons in which for results refinement, RaCC server was used to introduce the problematic residue codons as Arg, Leu, Ile and Pro. The results show that AT8B1_HUMAN gene has 35 rare codons for Arg, eleven rare codons for Ile, 9 single rare codons for Leu, and 11 rare codons for Pro (Table 2). This analysis also revealed AT8B1_HUMAN gene does not have any tandem double or triple repeats of rare Arg codon.

Afterwards, rare codon clusters (RCC) were detected and visualized in LaTcOm web tool [18]. In LaTcOm, three algorithm of % MINMAX, sliding window and MSS were employed. Codon usage table from CUTG database [19] was used as a reference in these three algorithms, and the situation of rare codon clusters were identified in AT8B1_HUMAN gene using MSS, minmax and sliding window algorithms (Figure 2; A, B and C).

These results indicated different features of these three algorithms. As shown, MSS detected 2 clusters, Minmax detected 10 and sliding_window detected 16 clusters. It is important to note that the cluster length selected for

Table 1: The result of PF07969 ID analysis in Sherlocc program.

PFAM ID	PFAM Name	No. of rare codon clusters	Rare codon frequency threshold	Size of largest cluster	Number of sequences	Number of organisms
Your query gave 0 match.						

ATG AGT ACA GAA AGA GAC TCA GAA **ACG** ACA TTT GAC GAG GAT TCT CAG CCT AAT GAC GAA GTG GTT
CCC TAC AGT GAT GAT GAA ACA GAA GAT GAA CTT GAT GAC CAG GGG TCT GCT GTT GAA CCA GAA CAA
AAC **CGA** GTC AAC AGG GAA GCA GAG GAG AAC CGG GAG CCA TTC AGA AAA GAA TGT ACA TGG CAA GTC
AAA GCA AAC GAT CGC AAG TAC CAC GAA CAA CCT CAC TTT ATG AAC ACA AAA TTC TTG TGT ATT AAG
GAG AGT AAA TAT **GCG** AAT AAT GCA ATT AAA ACA TAC AAG TAC AAC GCA TTT ACC TTT **ATA** CCA ATG
AAT CTG TTT GAG CAG TTT AAG AGA GCA GCC AAT **TTA** TAT TTC CTG GCT CTT CTT ATC **TTA** CAG GCA GTT
CCT CAA ATC TCT ACC CTG GCT TGG TAC ACC ACA **CTA** GTG CCC CTG CTT GTG GTG CTG GGC GTC ACT
GCA ATC AAA GAC CTG GTG GAC GAT GTG GCT CGC CAT AAA ATG GAT AAG GAA ATC AAC AAT AGG **ACG**
TGT GAA GTC ATT AAG GAT GGC AGG TTC AAA GTT GCT AAG TGG AAA GAA ATT CAA GTT GGA GAC GTC
ATT **CGT** CTG AAA AAA AAT GAT TTT GTT CCA GCT GAC ATT CTC CTG CTG TCT AGC TCT GAG CCT AAC
AGC CTC TGC TAT GTG GAA ACA GCA GAA CTG GAT GGA GAA ACC AAT **TTA** AAA TTT AAG ATG TCA CTT
GAA ATC ACA GAC CAG TAC CTC CAA AGA GAA GAT ACA TTG GCT ACA TTT GAT GGT TTT ATT GAA TGT
GAA GAA CCC AAT AAC AGA **CTA** GAT AAG TTT ACA GGA ACA **CTA** TTT TGG AGA AAC ACA AGT TTT CCT
TTG GAT GCT GAT AAA ATT TTG **TTA** **CGT** GGC TGT **GTA** ATT AGG AAC ACC GAT TTC TGC CAC GGC **TTA**
GTC ATT TTT GCA GGT GCT GAC ACT AAA **ATA** ATG AAG AAT AGT GGG AAA ACC AGA TTT AAA AGA ACT
AAA ATT GAT TAC TTG ATG AAC TAC ATG GTT TAC **ACG** ATC TTT GTT GTT CTT ATT CTG CTT TCT GCT GGT
CTT GCC ATC GGC CAT GCT TAT TGG GAA GCA CAG GTG GGC AAT TCC TCT TGG TAC CTC TAT GAT GGA
GAA GAC GAT ACA CCC TCC TAC **CGT** GGA TTC CTC ATT TTC TGG GGC TAT ATC ATT GTT CTC AAC ACC
ATG **GTA** CCC ATC TCT CTC TAT GTC AGC GTG GAA GTG ATT **CGT** CTT GGA CAG AGT CAC TTC ATC AAC
TGG GAC CTG CAA ATG TAC TAT GCT GAG AAG GAC ACA CCC GCA AAA GCT AGA ACC ACC ACA CTC AAT
GAA CAG CTC GGG CAG ATC CAT TAT ATC TTC TCT GAT AAG **ACG** GGG ACA CTC ACA CAA AAT ATC ATG
ACC TTT AAA AAG TGC TGT ATC AAC GGG CAG **ATA** TAT GGG GAC CAT CGG GAT GCC TCT CAA CAC AAC
CAC AAC AAA **ATA** GAG CAA GTT GAT TTT AGC TGG AAT ACA TAT GCT GAT GGG AAG CTT GCA TTT TAT
GAC CAC TAT CTT ATT GAG CAA ATC CAG TCA GGG AAA GAG CCA GAA **GTA** **CGA** CAG TTC TTC TTC TTG
CTC GCA GTT TGC CAC ACA GTC ATG GTG GAT AGG ACT GAT GGT CAG CTC AAC TAC CAG GCA GCC TCT
CCC GAT GAA GGT GCC CTG **GTA** AAC GCT GCC AGG AAC TTT GGC TTT GCC TTC CTC GCC AGG ACC CAG
AAC ACC ATC ACC ATC AGT GAA CTG GGC ACT GAA AGG ACT TAC AAT GTT CTT GCC ATT TTG GAC TTC
AAC AGT GAC CGG AAG **CGA** ATG TCT ATC ATT **GTA** AGA ACC CCA GAA GGC AAT ATC AAG CTT TAC TGT
AAA GGT GCT GAC ACT GTT ATT TAT GAA CGG **TTA** CAT **CGA** ATG AAT CCT ACT AAG CAA GAA ACA CAG
GAT GCC CTG GAT ATC TTT GCA AAT GAA ACT CTT AGA ACC **CTA** TGC CTT TGC TAC AAG GAA ATT GAA
GAA AAA GAA TTT ACA GAA TGG AAT AAA AAG TTT ATG GCT GCC AGT GTG GCC TCC ACC AAC CGG GAC
GAA GCT CTG GAT AAA **GTA** TAT GAG GAG ATT GAA AAA GAC **TTA** ATT CTC CTG GGA GCT ACA GCT ATT
GAA GAC AAG **CTA** CAG GAT GGA GTT CCA GAA ACC ATT TCA AAA CTT GCA AAA GCT GAC ATT AAG ATC
TGG GTG CTT ACT GGA GAC AAA AAG GAA ACT GCT GAA AAT **ATA** GGA TTT GCT TGT GAA CTT CTG ACT
GAA GAC ACC ACC ATC TGC TAT GGG GAG GAT ATT AAT TCT CTT CTT CAT GCA AGG ATG GAA AAC CAG
AGG AAT AGA GGT GGC GTC TAC GCA AAG TTT GCA CCT CCT GTG CAG GAA TCT TTT TTT CCA CCC GGT
GGA AAC **CGT** GCC **TTA** ATC ATC ACT GGT TCT TGG TTG AAT GAA ATT CTT CTC GAG AAA AAG ACC AAG
AGA AAT AAG ATT CTG AAG CTG AAG TTC CCA AGA ACA GAA GAA GAA AGA CGG ATG CGG ACC CAA AGT
AAA AGG AGG **CTA** GAA GCT AAG AAA GAG CAG CGG CAG AAA AAC TTT GTG GAC CTG GCC TGC GAG TGC
AGC GCA GTC ATC TGC TGC CGC GTC ACC CCC AAG CAG AAG GCC ATG GTG GTG GAC CTG GTG AAG AGG
TAC AAG AAA GCC ATC **ACG** CTG GCC ATC GGA GAT GGG GCC AAT GAC GTG AAC ATG ATC AAA ACT GCC
CAC ATT GGC GTT GGA **ATA** AGT GGA CAA GAA GGA ATG CAA GCT GTC ATG **TCG** AGT GAC TAT TCC TTT
GCT CAG TTC **CGA** TAT CTG CAG AGG **CTA** CTG CTG GTG CAT GGC **CGA** TGG TCT TAC **ATA** AGG ATG TGC
AAG TTC **CTA** **CGA** TAC TTC TTT TAC AAA AAC TTT GCC TTT ACT TTG GTT CAT TTC TGG TAC TCC TTC TTC
AAT GGC TAC TCT **GCG** CAG ACT GCA TAC GAG GAT TGG TTC ATC ACC CTC TAC AAC GTG CTG TAC ACC
AGC CTG CCC GTG CTC CTC ATG GGG CTG CTC GAC CAG GAT GTG AGT GAC AAA CTG AGC CTC **CGA** TTC
CCT GGG **TTA** TAC **ATA** GTG GGA CAA AGA GAC **TTA** **CTA** TTC AAC TAT AAG AGA TTC TTT **GTA** AGC TTG
TTG CAT GGG GTC **CTA** ACA **TCG** ATG ATC CTC TTC TTC **ATA** CCT CTT GGA GCT TAT CTG CAA ACC **GTA**
GGG CAG GAT GGA GAG GCA CCT TCC GAC TAC CAG TCT TTT GCC GTC ACC ATT GCC TCT GCT CTT **GTA**
ATA ACA GTC AAT TTC CAG ATT GGC TTG GAT ACT TCT TAT TGG ACT TTT GTG AAT GCT TTT TCA ATT TTT
GGA AGC ATT GCA CTT TAT TTT GGC ATC ATG TTT GAC TTT CAT AGT GCT GGA **ATA** CAT GTT CTC TTT CCA
TCT GCA TTT CAA TTT ACA GGC ACA GCT TCA AAC GCT CTG AGA CAG CCA TAC ATT TGG **TTA** ACT ATC
ATC CTG GCT GTT GCT GTG TGC **TTA** **CTA** CCC GTC GTT GCC ATT **CGA** TTC CTG TCA ATG ACC ATC TGG CCA
TCA GAA AGT GAT AAG ATC CAG AAG CAT CGC AAG CGG TTG AAG **GCG** GAG GAG CAG TGG CAG **CGA** CGG
CAG CAG GTG TTC CGC CGG GGC GTG TCA **ACG** CGG CGC **TCG** GCC TAC GCC TTC **TCG** CAC CAG CGG GGC
TAC **GCG** GAC CTC ATC TCC TCC GGG CGC AGC ATC CGC AAG AAG CGC **TCG** **CCG** CTT GAT GCC ATC GTG
GCG GAT GGC ACC **GCG** GAG TAC AGG CGC ACC GGG GAC AGC **TGA**

Figure 1: Schematic representation of the codon usage and position of rare and highly rare codons in AT8B1_HUMAN gene (displayed in orange and red, respectively)

Table 2: Schematic representation the position of Arg, Leu, Ile, and Pro in the AT8B1_HUMAN gene. These residues have rare codons, displayed in red, blue, green, orange, and red, respectively.

Amino Acid	Codon	Codon position
Arg	AGA	5-59-118-252-271-282-327-330-437-608-652-775-817-827-832-1024-1032-1141
	CGA	46-631-930-941-952-1164-1193
	AGG	176-185-301-541-563-572-586-768-773-1246-
Leu	CTA	145-272-279-654-710-842-935-951-1027
Ile	ATA	108-319-475-490-742-910-945-1020-1050-1082-1122
Pro	CCC	23-147-268-381-401-433-553-793-870-996-1159

all of algorithms was 21. The characteristics and positions of these RCCs in the AT8B1_HUMAN gene are reported in Table 3.

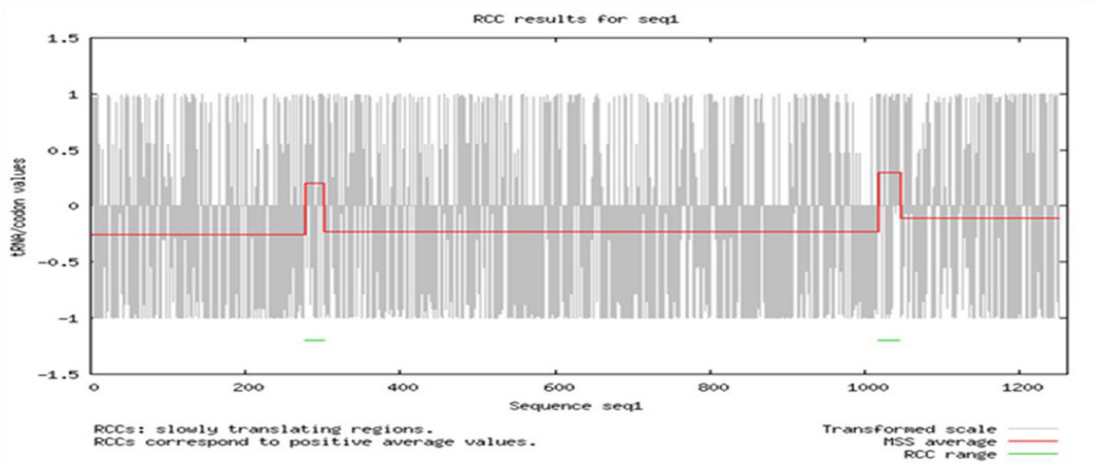
Later, for results better understanding, we focused on the rare codons in relation to mutations in ATP8B1 responsible for PFIC1 disease. After their comparison, three clusters located at codon sequence in the ATP8B1 structure were selected and studied precisely. These regions were identified in most of the findings and outputs obtained from various web servers.

Studying Rare Codons in ATP8B1 Structure

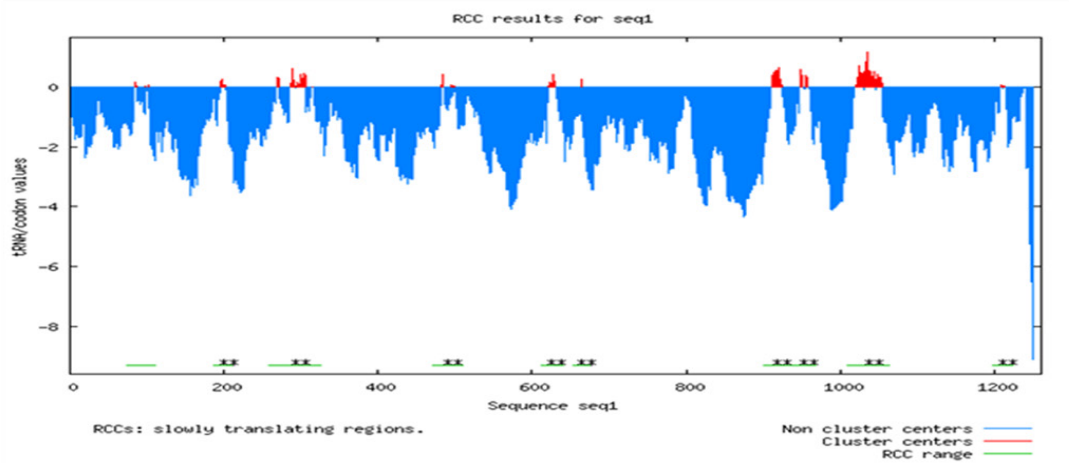
Knowledge of 3D structure is a useful prerequisite for understanding the proteins functions and studying rare codons in 3D protein structure as a cornerstone in many aspects of modern biology. One possible role of rare codon [14] is to play a regulating role in folding catalytically important domains, in protein structure and indirectly folding [15, 26]. As mentioned, 6 rare codon clusters were identified in AT8B1_HUMAN gene. Specific studies show that crystal structures of ATP8B1 protein have not been reported. For precise studying of the location and role of rare codons, it is necessary to gain 3D models from these sequences. In the beginning, to obtain the relative assessment from this protein, the sequence of ATP8B1 protein was analyzed in Predictprotein server (Figure 3).

Consequently, by submitting sequences of ATP8B1 in I-TSSAR and Phyre2 Web Servers, 3D models of these proteins were obtained. The crystal structures of selected proteins as templates in these servers approximately have <200 amino acids in comparison with ATP8B1 protein. So, these extra segments were predicted as disordered regions (I-TSSAR results) or were not included in the final structure of model (Phyre2 results). I-TSSAR Web Server generated five models and best model showed -1.82 value of overall C-score, 0.49±0.15 value of TM-Score and Exp. RMSD was 14.1±3.9 (Figure 4A). Phyre2 Web Servers used the crystal structure of the sodium-potassium pump in the e2.2k+.pi2 state as top template and the structure contents were as disordered (23%), Alpha helix (47%), Beta strand (15%) and TM helix (19%) (Figure 4B).

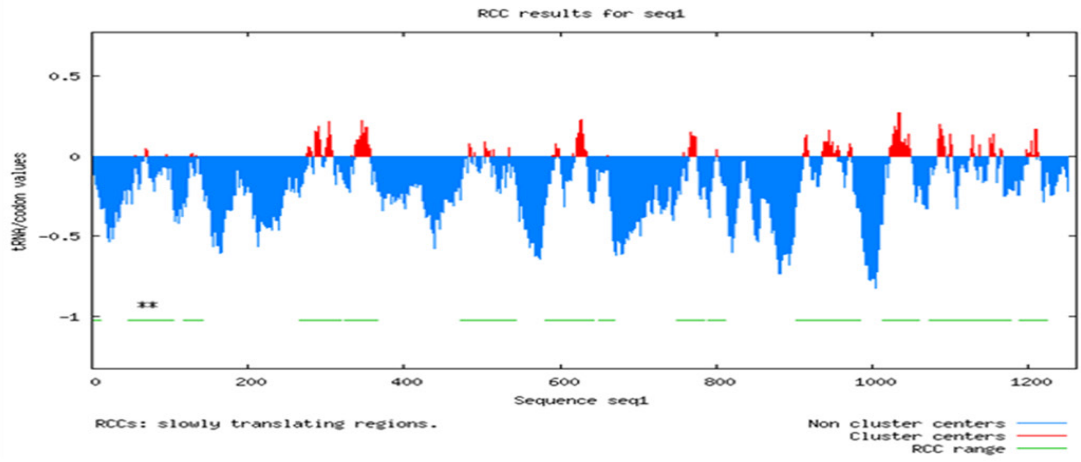
Since the selected templates have fewer amino acids, these extra segments were predicted as disorder loops in I-TSSAR Web Servers (A) and were not included in the final structure in Phyre2 Web Servers. Thus, the best model of Phyre2 Web Server was used in the study of rare codon and mutation in the structure



A



B



C

Figure 2: The position of rare codon clusters in AT8B1_HUMAN gene. Detection of RCC using MSS algorithm (A), minmax algorithm (B), and sliding window method (C)

Table 3: The rare codon clusters characteristics in AT8B1_HUMAN gene retrieved from LaTcOm web tool

RCC identification Algorithms	Cluster length	Position of clusters	Score (per position)	Expected value
MSS	21	276-301	0.205	0.996
		1018-1044	0.300	-0.099
Min- max	21	74-111		0.057
		186-211	0.061	0.139
		259-325	0.152	0.115
		472-509	0.128	0.261
		611-640	0.289	0.391
		654-674	0.432	0.610
		901-935	0.673	0.427
		938-968	0.475	0.550
		1010-1064	0.609	0.417
		1198-1224		0.927
Sliding window	21	1-11		-0.274
		46-82		-0.305
		85-105	0.042	-0.834
		117-142	-0.080	-0.827
		265-319	-0.132	-0.604
		323-366	-0.147	-0.907
		473-544	-0.066	-0.779
		580-643	-0.094	-1.081
		650-670	-0.081	-3.486
		748-784	-0.144	-2.166
		789-811	-0.437	-3.693
		901-983	-0.306	-1.139
		1012-1059	-0.474	-0.203
		1073-1112	-0.147	-0.237
		1115-1176		-0.189
		1187-1223		-0.415

of ATP8B1. Following, the physiochemical properties of ATP8B1 protein was calculated in ProtParam tool (Table 4).

ATP8B1 has 1251 residues, in which rare codons are distributed throughout the protein sequence. Based on modelling results, these rare codons are located in different regions of ATP8B1 structure. Furthermore, analyzing the 3D model of ATP8B1 structure in PIC server showed the interaction of these residues with other residues (Figure 5). A spectrum of mutations in ATP8B1 responsible for PFIC1 disease was presented previously. Next, the situation of some important mutations in relation to rare codons were studied in the ATP8B1 structure. Analyzing the 3D model of ATP8B1 demonstrated that Arg^{600} residue (detected as rare codon in our analysis) forms a hydrogen bond

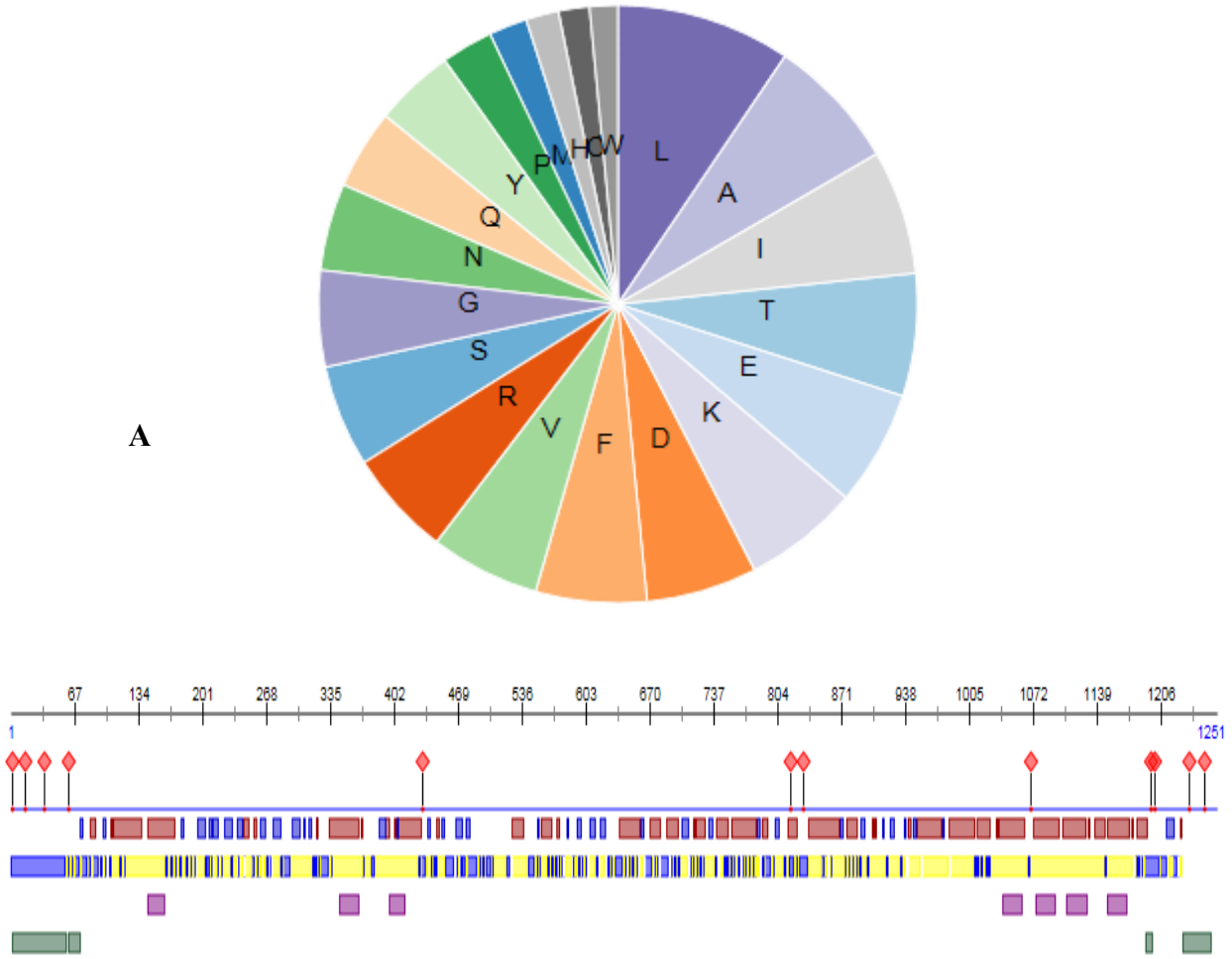


Figure 3: Schematic representation properties of ATP8B1 sequence in Predictprotein server. A) The amino acid composition. B) Red diamond in line 2 shows the position of binding site. At line 3, the red and blue color rectangular show the α–helix and β–sheet, respectively. In line 4, blue, yellow and white color rectangular show the buried, exposed and intermediate, region respectively. Helical transmembrane region shows with Purple Square at line 5. The disordered regions show with green rectangular in the last line.

with ASN^{597} (Figure 5). But, with mutation of this residue to Trp or Gln, this hydrogen bond was disrupted. The significance of this change is that the mutation caused the BRIC disorder in patients with this mutation in the ATP8B1 gene (Figure 5).

The non-covalent interactions was calculated by WHAT IF [23] and PIC [24] web servers, and the results are shown in Table 5.

Another missense mutation (2197 G>A) was detected in the nucleotide sequence of AT8B1_HUMAN gene that resulted in PFIC (Figure 6A). In this mutation, the codon sequence of Gly^{733} (GGA) changed to Arg^{733} (AGA) diagnosed as rare codon. This substituted residue, Arg^{733} residue constitutes hydrogen bonds with Asp^{232}, Leu^{231} and Arg^{867} as shown in Figure 6B.

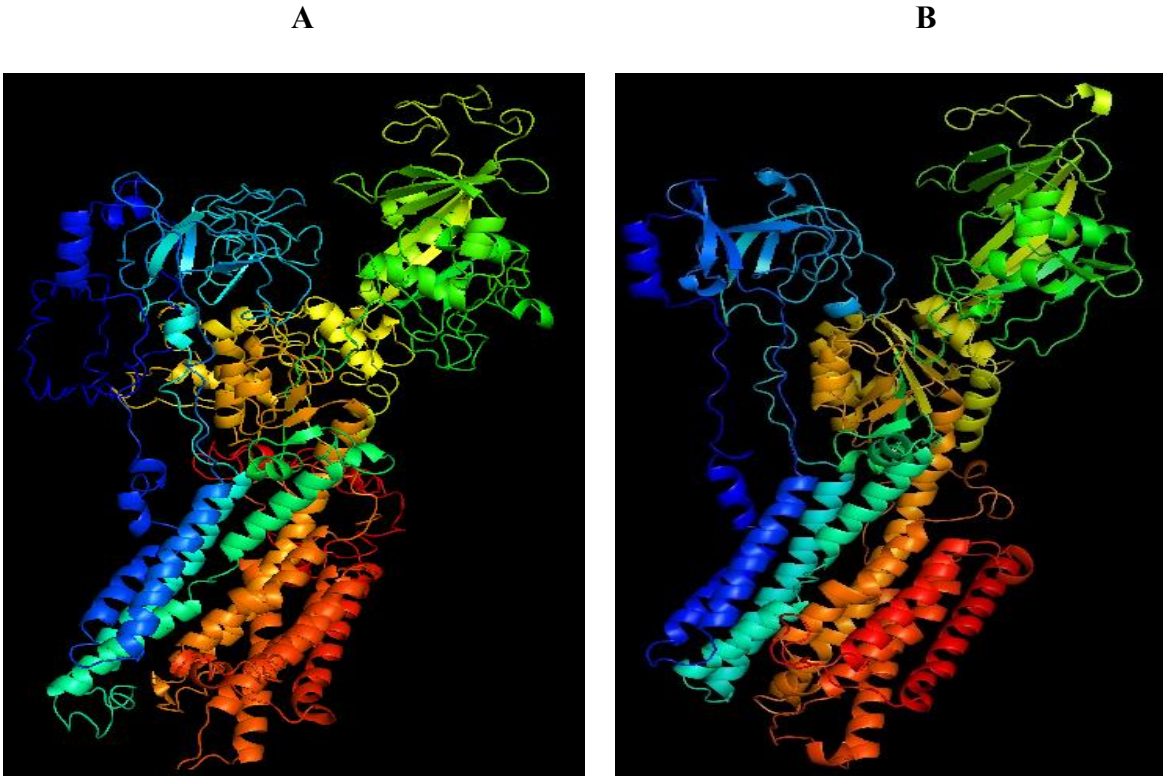


Figure 4: The ribbon diagram of ATP8B1 modeled in I-TSSAR (A) and Phyre2 Web Servers (B).

Table 4: In silico physico-chemical properties of ATP8B1 protein obtained from ProtParam tool. * First value is based on the assumption both cysteine residues form cystine and the second assumes that both cysteine residues are reduced.

Parameters	ATP8b1
Theoretical pI	6.77
Molecular weight	143695.4
Sequence length	1251
Extinction coefficients (M-1 cm-1at 260 nm)*	186210-184960
Asp + Glu	153
Arg + Lys	150
Instability index	40.60
Grand average of hydropathicity	-0.241
Aliphatic index	87.32

The non-covalent interactions were calculated by PIC [24] web servers, and the results are shown in Table 6.

Similarly, in PFIC patients, the missense mutation (2674 G>A) was detected and analyzed (Figure 7A). In this mutant, the codon sequence of Gly^{892} (GGA) changed to rare codon of Arg^{892} (AGA). This Arg^{8923} residue constitutes hydrogen bonds with Phe^{452}, Leu^{731}, Asp^{897} and Glu^{914} as shown in Figure 7B.

The results of this analysis are shown in Table 7.

Also, Gly^{1040} forms hydrogen bond with His^{356}, His^{357} as the missense mutation (2674 G>A) was detected and analyzed (Figure 8A). At missense mutation in PICF patient, the codon sequence of Gly^{1040} (GGG) changes to rare codon of Arg^{1040} (AGG). This Arg^{892} residue constitutes hydrogen bonds with Phe^{452}, Leu^{731}, Asp^{897} and Glu^{914} shown in Figure 8B.

The results of this analysis are shown in Ta-

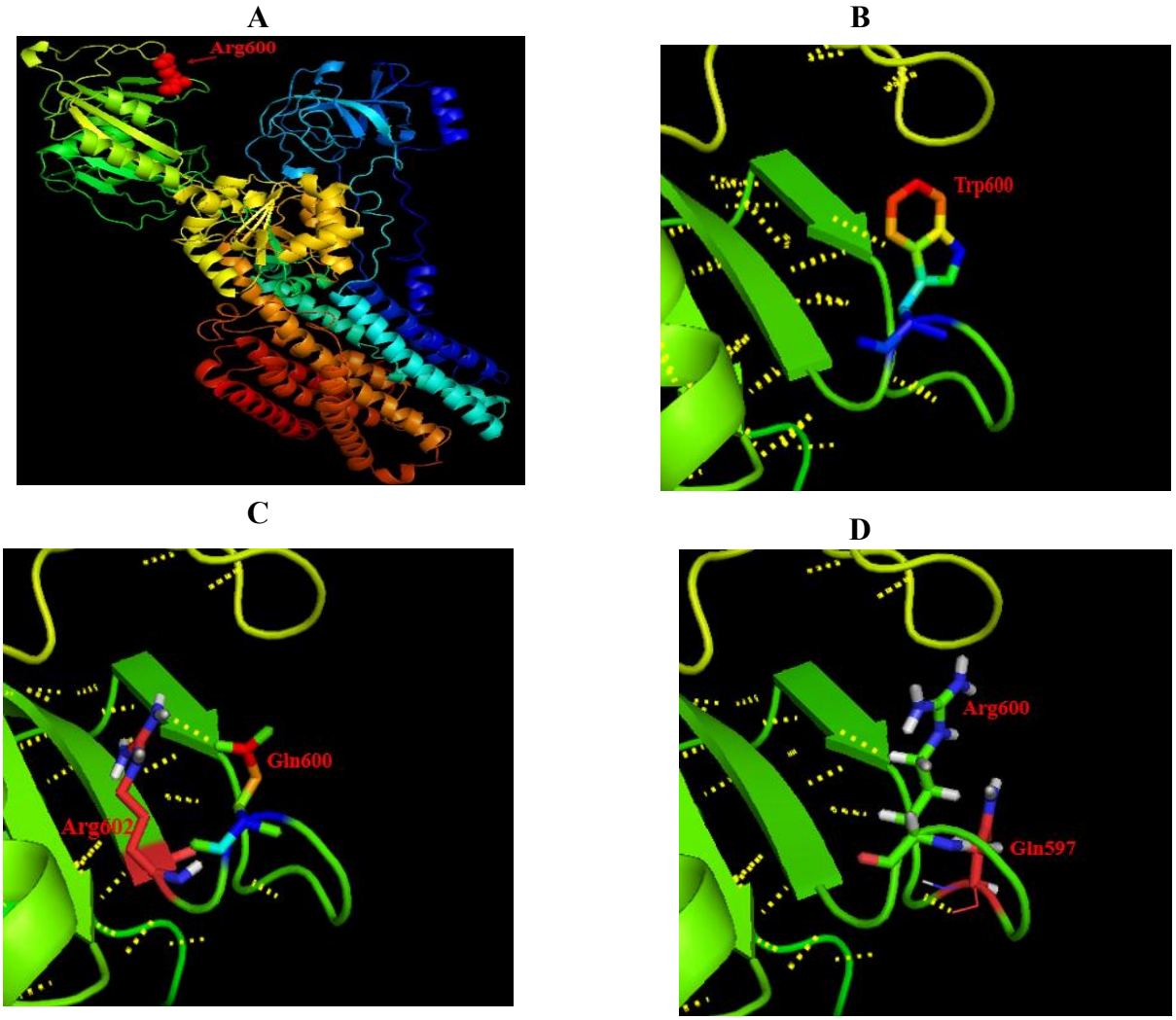


Figure 5: A) The ribbon diagram of ATP8B1, with location of Arg^{600} (rare codon residues) in blue color. The Arg^{600} residue form the hydrogen bond with Gln^{597} (B), mutation of Arg^{600} to Gln^{600} (C) and Trp^{600} (D) are shown.

ble 8.

Discussion

Progressive familial intrahepatic cholestasis (PFIC1) refers to autosomal- recessive liver disorders of childhood in which cholestasis of hepatocellular origin often presents in the neonatal period or first year of life and leads to liver failure and death [1, 27]. Three types of progressive familial intrahepatic cholestasis including type 1 (PFIC1), type 2 (PFIC2) and type 3 (PFIC3) are related to mutations in hepatocellular transport-system genes involved in bile formation [28].

ATP8B1 deficiency is an autosomal recessive liver disease caused by mutations in ATP8B1, encoding a P-type ATPase [29]. Deficiency of ATP8B1 in the hepatocyte leads to loss of asymmetric distribution of phospholipids in the canalicular membrane, decreasing both membrane stability and function of transmembrane transporters such as ABCB11, the bile salt export pump, resulting in intrahepatic cholestasis (IC) [30]. PFIC1 also known as Byler disease, is characterized by cholestasis often arising in the neonatal period leading to

Table 5: The characteristics of non-covalent interactions of Arg[600] with Gln[597] (Dd-a= Distance Between Donor and Acceptor, Dh-a= Distance Between Hydrogen and Acceptor, A(d-H-N)= Angle Between Donor-H-N , A(a-O=C) = Angle Between Acceptor-O=C, MO=Multiple Occupancy).

DONOR			ACCEPTOR			PARAMETERS				
POS	RES	ATOM	POS	RES	ATOM	MO	Dd-a	Dh-a	A(d-HN)	A(aO=C)
600	ARG	N	597	ASN	O	1	3.45	3.03	107.09	100.51

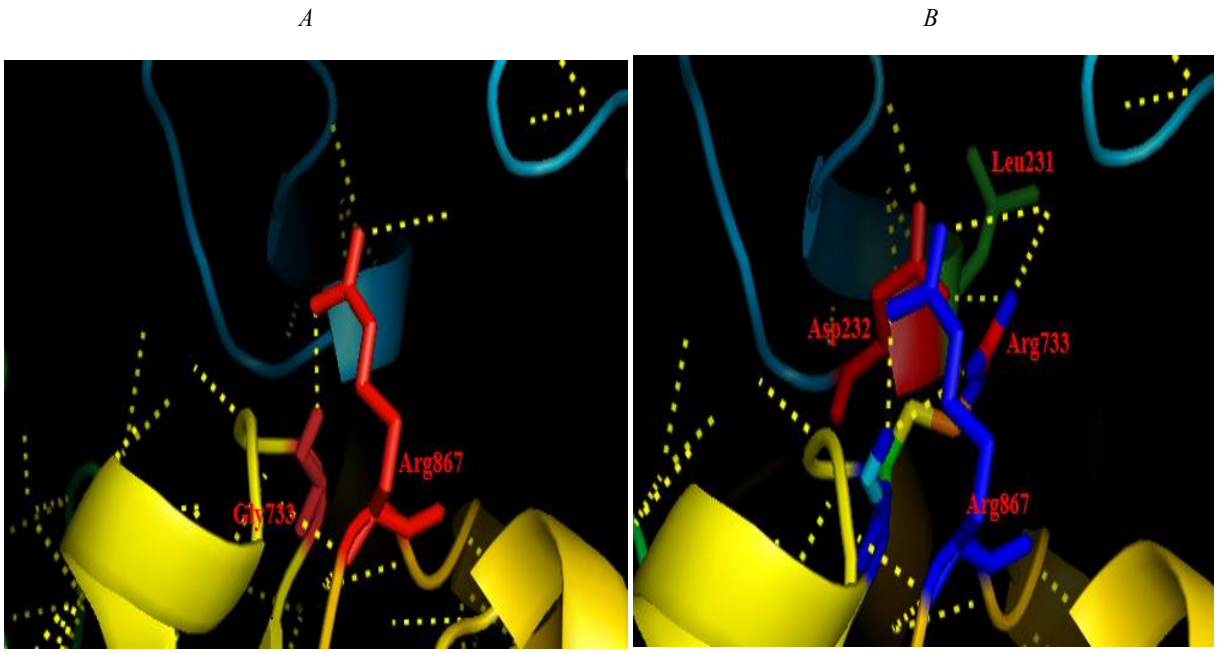


Figure 6: The ribbon diagram of ATP8B1, with location of Gly[733] (A) and mutation to rare codon of Arg[733] (B) in blue color. The hydrogen interaction of these residues with Asp[232], Leu[231] and Arg[867] are shown in yellow color.

Table 6: The characteristics of non-covalent interactions of Gly[733] and Arg[733] with other residues.

DONOR			ACCEPTOR			PARAMETERS				
POS	RES	ATOM	POS	RES	ATOM	MO	Dd-a	Dh-a	A(d-HN)	A(aO=C)
867	ARG	NH1	733	Gly	O	1	2.93	2.16	128.60	150.57
867	ARG	NH1	733	Gly	O	2	2.93	3.45	52.14	150.57
733	ARG	NE	231	LEU	O	-	3.22	2.52	127.01	121.45
867	ARG	NH1	733	ARG	O	1	2.93	2.16	128.60	150.57
733	ARG	NH1	232	ASP	OD1	1	2.97	3.34	60.82	999.99
733	ARG	NH2	289	ASP	OD1	2	3.40	2.40	163.31	999.99

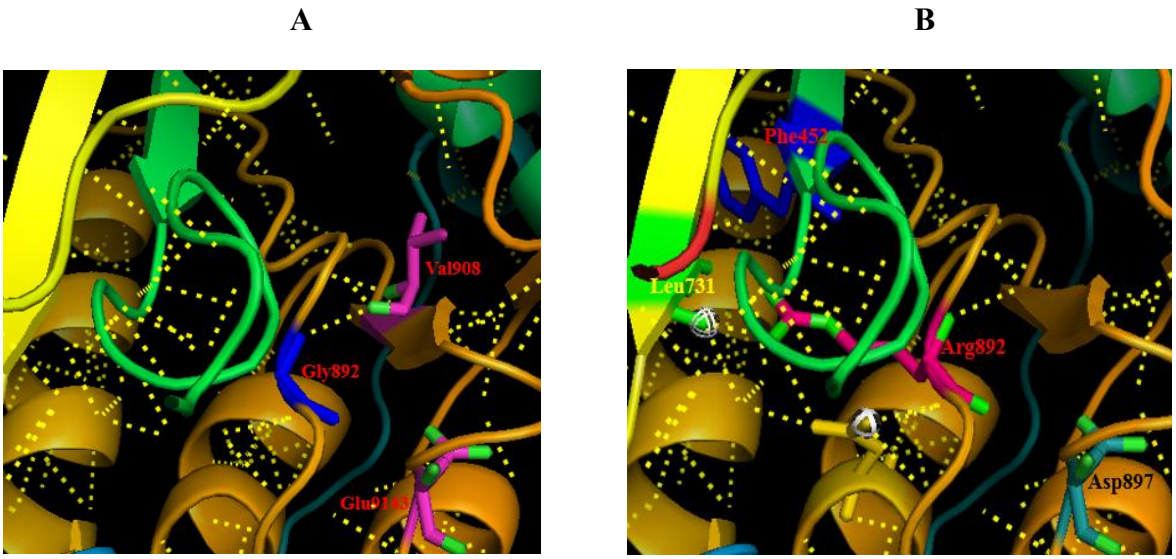


Figure 7: The ribbon diagram of ATP8B1, with location of Gly[892] (A) and mutation to rare codon of Arg[892] (B) (blue color). The hydrogen interaction of these residues are shown in yellow color.

Table 7: The characteristics of non-covalent interactions of Gly[892] (A) and Arg[892] (B) with other residues.

DONOR			ACCEPTOR				PARAMETERS			
POS	RES	ATOM	POS	RES	ATOM	MO	Dd-a	Dh-a	A(d-HN)	A(aO=C)
892	GLY	N	908	VAL	O		3.08	2.31	134.82	156.35
914	GLU	OE2	892	GLY	O	1	3.41	3.96	51.99	168.78
892	ARG	NH2	452	PHE	O	1	3.46	2.79	122.31	81.75
892	ARG	NH1	731	LEU	O	2	2.53	1.53	162.16	149.29
914	GLU	OE2	892	ARG	O	1	3.41	3.96	51.99	168.78
892	ARG	NH1	897	ASP	OD1	1	3.21	2.37	136.91	999.99

death due to liver failure [31].

Previously, the detection of rare codons was performed on the genome and proteins of cytosine deaminase and HCV (article in press). Furthermore, no similar analyses have been reported on ATP8b1. It is very important to recognize the cause of disease in genetic disorders of the liver such as "rare" codons infrequently used by cells. Besides, in drug research and designing, considering the structural situation, the hidden computational biology information and the roles of specific residues in catalytic function are critical. In spite of the large number of studies on PFIC1, there are a number of unresolved issues regarding the structure of ATP8b1. In this bioinformatic study, several web servers were used for detecting mutation and rare codons in the structure of ATP8b1.

The Sherlocc program identified no rare codon clusters in the ATP8b1 protein family with three Pfam IDs of PF00122, PF16212 and PF16209. Following, rare and highly rare codons were identified using ATGme web server. The results indicated that ATP8b1 had 67 rare

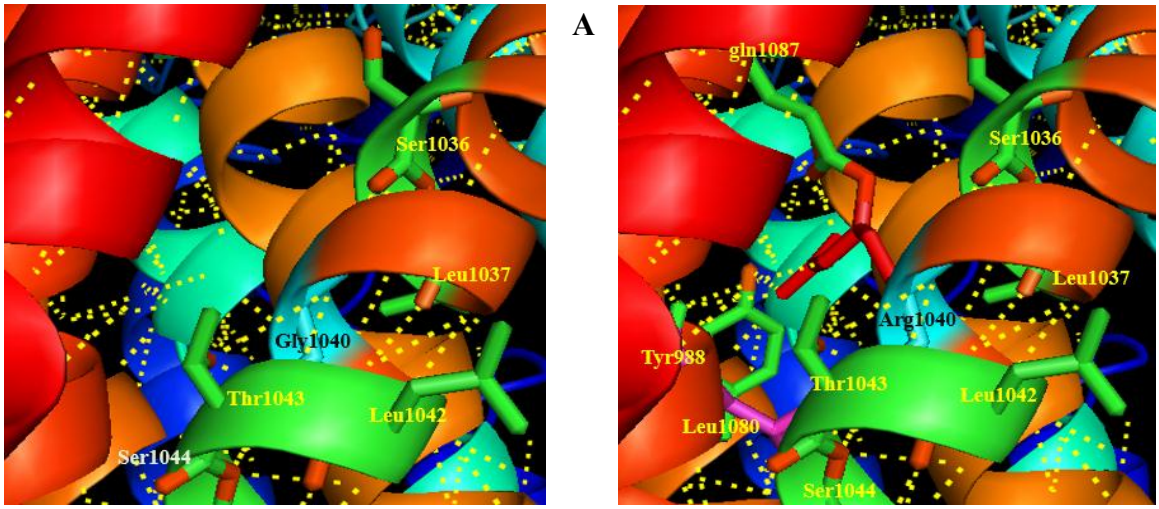


Figure 8: The ribbon diagram of ATP8B1, with location of Gly^{1040} (A) and Arg^{1040} rare codon residue (B) in blue color. The residues that form hydrogen interaction are shown in red color.

Table 8: The characteristics of non-covalent interactions of Gly^{1040} and Arg^{1040} with other residues. The Arg^{1040} has a similar interaction with Ser^{1036}, leu^{1037}, Leu^{1042}, Thr^{1043} and Ser^{1044}

DONOR			ACCEPTOR					PARAMETERS	
POS	RES	ATOM	POS	RES	ATOM	Dd-a	Dh-a	A(d-HN)	A(aO=C)
1040	GLY	N	1036	SER	O	3.01	2.13	147.85	158.16
1040	GLY	N	1037	LEU	O	3.31	2.74	117.05	98.24
1042	LEU	N	1040	GLY	O	3.37	3.44	77.23	75.69
1043	THR	OG1	1040	GLY	O	2.41	9.99	999.99	116.60
1044	SER	OG1	1040	GLY	O	3.14	9.99	999.99	137.57
1040	ARG	NH2	1080	LEU	O	3.40	2.62	131.29	127.28
1040	ARG	NE	1087	GLN	OE1	3.21	4.02	31.64	999.99
1040	ARG	NH2	988	TYR	OH1	3.22	3.95	40.29	999.99

codons and 11 very rare codons. These rare and very rare codons can play a critical role in folding protein chain. Moreover, AT8B1_HUMAN gene was analyzed in RaCC server which focused on Arg, Leu, Ile and Pro. Results showed that AT8B1_HUMAN gene had 35 rare codons of Arg, 11 single rare codons for Ile, 9 rare codons for Leu and 11 rare codons for amino acid Pro. Later, rare codon clusters of AT8B1_HUMAN gene were also detected using LaTcOm web tool (17).

Results of this study showed that 2 rare codon clusters were identified via MSS, 10 rare codon clusters via minmax and 16 rare codon clusters via sliding_window algorithm. The difference of outputs is because these algorithms have different primary databases. The results also showed the high frequency of rare codons that make susceptible this gene in mutation and disrupt the proper folding of the protein. An initial review of location of these Arg residues and the large number of formed

hydrogen bonds demonstrate that these residues have a critical role in proper folding of ATP8b1. Because the large number of rare codons are difficult to consider, we focus on some rare codons related to mutation causing PFIC1. Results summarization led to choose three rare codons for precisely studying the structure of ATP8b1, 3D structure of ATP8b1 which was modelled by Phyre2 and I-TSSAR Web Server.

The initial analysis showed that some regions of this protein have disorder structure. These regions were modelled as disordered in I-TSSAR or not included in final structure of model as shown in Phyre2 results. Disordered regions are dynamically flexible and are distinct from irregular loop secondary structures which are static in solution. Phyre2 prediction has been made by the knowledge-based Disopred method. The superimposition of these models from Phyre2 and I-TSSAR show high degree of similarity. We used Phyre2 results in structure study because some rare codons were difficult to study in model from I-TSSAR which has the Disordered regions, for which results were not reported.

Structure analyses revealed that these rare codons and mutations were scattered across different regions of the ATP8b1 structure. Results of 3D modelling indicated that Arg600 residue forms some hydrogen bonds with other residues, with which mutation to the Gln600 and Trp600 in PFIC1 patients, these hydrogen bonds changed, and this affects the rate of folding that may affect the proper folding and catalytic activity of ATP8b1 (Figure 4). It seems that this residue has a critical role in the process of protein folding, in such positions to grantee the proper folding was necessary to slow down the rate of the folding. These effects may result in inefficiency of ATP8b1 and subsequently PFIC1. However, other hypothesis can be considered for the pathogenicity causes of this mutation.

Previously, three mutations that caused the PFIC1 were identified in Gly location at Gly^{733}, Gly^{892} and Gly^{1040}. All of these Glys were mutated to Arg in PFIC1 patients. Arg has six codons including AGG, AGA, CGT, CGC, CGA and CGG. In PFIC1 patients, the Gly codons, in these position, were mutated to Arg (AGA^{733}, AGA^{892} and AGG^{1040}) with low frequency (identified as rare codons). These mutations reduced the rate of protein folding in these residues and may be interfering with proper rate of folding affecting the final structure and catalytic activity. Furthermore, these Args can form some hydrogen bonds involving different parts of the protein and may disrupt the ATP8b1 folding. Besides, Gly contains no side chain and has the ability to fit within the structure, conveniently. In comparison with Gly, the Arg has a large and bulky side chain. With mutation of Gly to Arg, these may create the structural repulsion that interfere with folding and functional activity of PFIC1. All these hypotheses have a negative effect on the correct activity of ATP8b1 resulting in PFIC1 disease. Meanwhile, experimental evidence as introducing another mutation in these position is needed for our theoretical studies confirmation; other mutations and rare codons should undergo further study.

Next, by the use of molecular docking as a prerequisite of performing structure-based virtual screening (SBVS) [32], the residues involved in the binding site and enzyme's activity of ATP8b1 will be analyzed. This has profound applications in drug discovery. Therefore, we tried to introduce zinc ion in the substrate-docking region to determine the proper cytosine binding site in further studies.

We have previously modeled the structure of a number of proteins and have a good experience in homology modeling technique [33-36]. In this regard, the RCCs properties in the protein and genome of ATP8b1 was evaluated.

Conclusion

Our study identified nearly some of these regions that might involve in the substrate binding site or proper folding. Our data showed

that rare codon positions might have an essential role in folding and activity of ATP8b1. This study may also provide new insights into drug design for the treatment of PFIC1 in the future.

Conflict of Interest

None

References

1. Jacquemin E. Progressive familial intrahepatic cholestasis. *Clinics and research in hepatology and gastroenterology*. 2012;**36**:S26-S35. doi.org/10.1016/S2210-7401(12)70018-9.
2. Sira AM, Sira MM. Progressive familial intrahepatic cholestasis: INTECH Open Access Publisher; 2013.
3. Srivastava A. Progressive familial intrahepatic cholestasis. *J Clin Exp Hepatol*. 2014;**4**:25-36. doi.org/10.1016/j.jceh.2013.10.005. PubMed PMID: 25755532. PubMed PMCID: 4017198.
4. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E. Progressive familial intrahepatic cholestasis. *Orphanet J Rare Dis*. 2009;**4**:1. doi.org/10.1186/1750-1172-4-1. PubMed PMID: 19133130. PubMed PMCID: 2647530.
5. Engelmann G, Wenning D, Herebian D, Sander O, Droge C, Kluge S, et al. Two Case Reports of Successful Treatment of Cholestasis With Steroids in Patients With PFIC-2. *Pediatrics*. 2015;**135**:e1326-32. doi.org/10.1542/peds.2014-2376. PubMed PMID: 25847799.
6. Groen A, Romero MR, Kunne C, Hoosdally SJ, Dixon PH, Wooding C, et al. Complementary functions of the flippase ATP8B1 and the floppase ABCB4 in maintaining canalicular membrane integrity. *Gastroenterology*. 2011;**141**(5):1927-37. e4. https://doi.org/10.1053/j.gastro.2011.07.042.
7. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E, editors. The spectrum of liver diseases related to ABCB4 gene mutations: pathophysiology and clinical aspects. Seminars in liver disease; 2010: © Thieme Medical Publishers.
8. Dröge C, Kluge S, Häussinger D, Kubitz R, Keitel V. Sequencing of ATP8B1, ABCB11 and ABCB4 revealed 135 genetic variants in 374 unrelated patients with suspected intrahepatic cholestasis. Zeitschrift für Gastroenterologie. 2015;53:A3_27.
9. Park JS, Ko JS, Seo JK, Moon JS, Park SS. Clinical and ABCB11 profiles in Korean infants with progressive familial intrahepatic cholestasis. *World J Gastroenterol*. 2016;**22**:4901-7. doi.org/10.3748/wjg.v22.i20.4901. PubMed PMID: 27239116. PubMed PMCID: 4873882.
10. Carlton VE, Knisely AS, Freimer NB. Mapping of a locus for progressive familial intrahepatic cholestasis (Byler disease) to 18q21-q22, the benign recurrent intrahepatic cholestasis region. *Hum Mol Genet*. 1995;**4**:1049-53. doi.org/10.1093/hmg/4.6.1049. PubMed PMID: 7655458.
11. Fathy M, Kamal M, Al-Sharkawy M, Al-Karaksy H, Hassan N. Molecular characterization of exons 6, 8 and 9 of ABCB4 gene in children with Progressive Familial Intrahepatic Cholestasis type 3. *Biomarkers*. 2016;**21**:573-7. doi.org/10.3109/1354750X.2016.1166264. PubMed PMID: 27075526.
12. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr Opin Biotechnol*. 1995;**6**:494-500. doi.org/10.1016/0958-1669(95)80082-4. PubMed PMID: 7579660.
13. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res*. 2000;**28**:292. doi.org/10.1093/nar/28.1.292. PubMed PMID: 10592250. PubMed PMCID: 102460.
14. Chartier M, Gaudreault F, Najmanovich R. Large-scale analysis of conserved rare codon clusters suggests an involvement in co-translational molecular recognition events. *Bioinformatics*. 2012;**28**:1438-45. doi.org/10.1093/bioinformatics/bts149. PubMed PMID: 22467916. PubMed PMCID: 3465090.
15. Thanaraj TA, Argos P. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci*. 1996;**5**:1973-83. doi.org/10.1002/pro.5560051003. PubMed PMID: 8897597. PubMed PMCID: 2143259.
16. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;**9**:40. doi.org/10.1186/1471-2105-9-40. PubMed PMID: 18215316. PubMed PMCID: 2245901.
17. Kaplan W, Littlejohn TG. Swiss-PDB Viewer (Deep View). *Brief Bioinform*. 2001;**2**:195-7. doi.org/10.1093/bib/2.2.195. PubMed PMID: 11465736.
18. Theodosiou A, Promponas VJ. LaTcOm: a web server for visualizing rare codon clusters in coding sequences. *Bioinformatics*. 2012;**28**:591-2. doi.org/10.1093/bioinformatics/btr706. PubMed PMID: 22199385.
19. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol*. 1996;**260**:649-63. doi.org/10.1006/jmbi.1996.0428. PubMed

PMID: 8709146.

20. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res*. 2007;**35**:3375-82. doi.org/10.1093/nar/gkm251. PubMed PMID: 17478507. PubMed PMCID: 1904280.
21. Guex N, Peitsch M. Swiss-PdbViewer: a fast and easy-to-use PDB viewer for Macintosh and PC. *Protein Data Bank Quaterly Newsletter*. 1996;**77**(7).
22. DeLano WL. The PyMOL molecular graphics system. 2002.
23. Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph*. 1990;**8**:52-6, 29. PubMed PMID: 2268628.
24. Tina KG, Bhadra R, Srinivasan N. PIC: Protein Interactions Calculator. *Nucleic Acids Res*. 2007;**35**:W473-6. doi.org/10.1093/nar/gkm423. PubMed PMID: 17584791. PubMed PMCID: 1933215.
25. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*. 1997;**28**:405-20. doi.org/10.1002/(SICI)1097-0134(199707)28:3<405::AID-PROT10>3.0.CO;2-L. PubMed PMID: 9223186.
26. Widmann M, Clairo M, Dippon J, Pleiss J. Analysis of the distribution of functionally relevant rare codons. *BMC Genomics*. 2008;**9**:207. doi.org/10.1186/1471-2164-9-207. PubMed PMID: 18457591. PubMed PMCID: 2391168.
27. Nicolaou M, Andress EJ, Zolnerciks JK, Dixon PH, Williamson C, Linton KJ. Canalicular ABC transporters and liver disease. *J Pathol*. 2012;**226**:300-15. doi.org/10.1002/path.3019. PubMed PMID: 21984474.
28. Erlinger S, Arias IM, Dhumeaux D. Inherited disorders of bilirubin transport and conjugation: new insights into molecular mechanisms and consequences. *Gastroenterology*. 2014;**146**:1625-38. doi.org/10.1053/j.gastro.2014.03.047. PubMed PMID: 24704527.
29. Bull LN, van Eijk MJ, Pawlikowska L, DeYoung JA, Juijn JA, Liao M, et al. A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. *Nat Genet*. 1998;**18**:219-24. doi.org/10.1038/ng0398-219. PubMed PMID: 9500542.
30. van der Woerd WL, van Mil SW, Stapelbroek JM, Klomp LW, van de Graaf SF, Houwen RH. Familial cholestasis: progressive familial intrahepatic cholestasis, benign recurrent intrahepatic cholestasis and intrahepatic cholestasis of pregnancy. *Best Pract Res Clin Gastroenterol*. 2010;**24**:541-53. doi.org/10.1016/j.bpg.2010.07.010. PubMed PMID: 20955958.
31. Morris AL, Bukauskas K, Sada RE, Shneider BL. Byler disease: early natural history. *J Pediatr Gastroenterol Nutr*. 2015;**60**:460-6. doi.org/10.1097/MPG.0000000000000650. PubMed PMID: 25825852.
32. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J*. 2012;**14**:133-41. doi.org/10.1208/s12248-012-9322-0. PubMed PMID: 22281989. PubMed PMCID: 3282008.
33. Mortazavi M, Zarenezhad M, Alavian SM, Gholamzadeh S, Malekpour A, Ghorbani M, et al. Bioinformatic Analysis of Codon Usage and Phylogenetic Relationships in Different Genotypes of the Hepatitis C Virus. *Hepatitis monthly*. 2016;**16**(10). doi: 10.5812/hepatmon.39196 PMCID: PMC5111459, PMID: 27882066
34. Mortazavi M, Zarenezhad M, Gholamzadeh S, Alavian SM, Ghorbani M, Dehghani R, et al. Bioinformatic Identification of Rare Codon Clusters (RCCs) in HBV Genome and Evaluation of RCCs in Proteins Structure of Hepatitis B Virus. *Hepatitis monthly.* 2016;**16**(10). doi: 10.5812/hepatmon.39909, PMCID: PMC5116127, PMID: 27882067
35. Fattahi M, Malekpour A, Mortazavi M, Safarpour A, Naseri N. The characteristics of rare codon clusters in the genome and proteins of hepatitis C virus; a bioinformatics look. *Middle East journal of digestive diseases*. 2014;**6**(4):214. PMCID: PMC4208930, PMID: 25349685
36. Mortazavi M, Hosseinkhani S. Design of thermostable luciferases through arginine saturation in solvent-exposed loops. *Protein Engineering, Design & Selection*. 2011;**24**(12):893-903. https://doi.org/10.1093/protein/gzr051