Molecular Modelling and Evaluation of Hidden Information in ABCB11 Gene Mutations

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

ABSTRACT

Background: Cholestatic disorders are divided in the extra and intra-hepatic that created due to the severe liver diseases. ABCB11 encodes the bile salt export pump and this gene is mutated in several forms of intrahepatic cholestasis. So far, some molecular features of this gene was studies.

Objective: Using a developed web server, we identified high number of rare codons in this gene, and four cases were related to BSEP-deficient patients. We also Methods: By in-silico modelling of ABCB11, some of rare codons in different locations of ATP8b1 gene were identified and evaluated. Using several web servers a number of mutations that converted non-rare codons to rare codon in these patients were identified.

Results: Some of these rare Codons were located at special positions by mutation of which, the new side chains do not seem suitable for protein structure and function. Furthermore, this mutation changed the protein folding rate that may have a critical role in proper folding. Thus, primary change of these codons contributes to BSEP deficiency.

Conclusion: This work is a comprehensive analysis of rare codons of ABCB11 and assessment of a number of these rare codon in protein levels. Rare codons evaluation can enhance our understanding of ABCB11 structural protein of ABCB11, and help us to develop mutation-specific therapies in design of new drugs.

Keywords
ABCB11, Bioinformatics Analysis, Rare Codon, Mutation

Introduction

The bile salt export pump (BSEP, protein product of the ABCB11 gene) is situated in the canalicular membrane of hepatocytes and is responsible for the translocation of bile salts [1]. BSEP belongs to ABC transporter superfamily and has 12-transmembrane span integral membrane proteins. Mutations in ABCB11 are related to a phenotypical spectrum of cholestatic liver diseases. Cholestasis is created due to severe liver diseases where bile cannot flow from the liver to the duodenum [2]. The causes of cholestasis are divided into two groups: those originating outside the liver and those originating within the liver [3]. Some reasons of the cholestasis within the liver include cirrhosis due to viral hepatitis B or C, drugs, acute hepatitis, alcoholic liver disease, primary biliary cirrhosis with inflammation and scarring.
of the bile ducts [4]. At present, specific gene defects have been identified for PFIC2 which are caused by the deficient in gene product required for bile formation and canalicular export [2, 5]. Previously, the molecular evaluation of these PFICs including exon characterization (6), locus mapping [7], sequencing [8] and gene mutations have been studied.

The rare codons that are introduced as genetic hidden information are shown to have a critical role in protein activity and folding, and can help in problem solving of diseases and drug design [9, 10]. Codon-usage analysis can also contribute to understanding the interaction between RNA viruses and the immune response of the hosts [10]. Several mutations of ABCB11 associated with BSEP deficiency have previously been introduced [1, 11]. However, there is no study about rare codons of ABCB11 gene, and in this study we evaluated the situation of these rare codons and variations in the structure of ABCB11. For this, by submission of PFIC3 gene in the I-TASSER server, a three dimensional model of BSEP protein was created [12]. In addition, rare codons of ABCB11 gene were detected using the Sherloc program [13], LaTcOm (http://structure.biol.ucy.ac.cy/latcom.html) [14], ATGme [15] and RaCC server (http://nihserver.mbi.ucla.edu/RACC/). By PyMOL [20] and SPD-BV software [21], the characteristics of these rare codons were identified in the nucleotide sequences of ABCB11 gene.

Material and Methods

Rare Codons Analysis
For bioinformatic analysis of ABCB11 gene, the nucleotide sequences and features of this gene were retrieved from http://www.ncbi.nlm.nih.gov/genome/. By use of nucleotide sequence of ABCB11 gene, rare codons of this gene were detected using the following servers. Rare codon calculator (RaCC) (http://nihserver.mbi.ucla.edu/RACC/) detected problematic residues as arginine (AGG, AGA, CGA), leucine (CTA), isoleucine (ATA) and proline (CCC). ATGme (15) detected rare codon in three steps: (i) Input of the ATP8b1 sequence; (ii) Input of the codon usage table of Homo sapiens [gbpri]: 93487 CDS’s (iii) Detection of rare codons. LaTcOm [14] reported three algorithms are implemented for the detection of rare codon clusters: i) % minimax algorithm, ii) sliding window approach and iii) a linear-time algorithm named MSS. Then, the RCC positions were visualized within the submitted sequences. Sherloc’s program [13] detected rare codon clusters by retrieving the nucleotide sequence of proteins in each Pfam protein family alignments. By these servers, some rare codons were identified in the nucleotide sequences of ABCB11 gene.

Study of Rare Codons in the Structure of ABCB11

To evaluate the position of these new identified mutations and rare codon in the structure of BSEP protein, a 3D structure of BSEP was created in the I-TASSER web server [12] based on multiple-threading alignments by LOMETS [16]. Models with the best “Confidence Score” and Z-score were chosen by I-TASSER server. The best model with suitable Z-score and confidence was visualized using PyMOL [17] and Swiss PDB viewer [18]. With Expasy’s ProtParam (http://us.expasy.org/tools/protparam.html) server, the total number of positive and negative residues, physico-chemical parameters, molecular weight and other features of
Rare Codon Evaluation of ABCB11 Gene

Detection of Rare Codon Clusters
Using UniProt database (http://www.uniprot.org/), the Pfam accession number of BSEP protein was identified as PF00664 (ABC_

Table 1: In silico physico-chemical features of BSEP protein obtained from ProtParam tool.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>ATP8b1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical pl</td>
<td>6.17</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>146407.18</td>
</tr>
<tr>
<td>Sequence length</td>
<td>1321</td>
</tr>
<tr>
<td>Extinction coefficients (M-1 cm-1at 260 nm)*</td>
<td>129370- 128120</td>
</tr>
<tr>
<td>Asp + Glu</td>
<td>143</td>
</tr>
<tr>
<td>Arg + Lys</td>
<td>135</td>
</tr>
<tr>
<td>Instability index</td>
<td>33.29</td>
</tr>
<tr>
<td>Grand average of hydropathicity</td>
<td>0.011</td>
</tr>
<tr>
<td>Aliphatic index</td>
<td>92.60</td>
</tr>
</tbody>
</table>

*First number is based on the assumption that both cysteine residues form cystine and the second number that both cysteines are reduced.

Preparation of Molecular Modelling Structure of ABCB11
For understanding the protein structure and function, preparation of 3D structure of protein is a vital process in biology [21]. Our studies show that heretofore has not been determined as the crystal structure of BSEP, and it is obligatory to provide the 3D model from this protein. For this, by submitting the sequence of ATP8B1 in I-TSSAR Web Server, 3D models of these proteins were obtained. The I-TSSAR Web Server generated five models and best model showed -1.82 values of overall C-score, 0.49±0.15 value of TM-Score and Exp. RMSD was 14.1±3.9 (Figure 1).

In the following, the physiochemical properties of ATP8B1 protein model were calculated in ProtParam tool (Table 1) [22].

Figure 1: The ribbon diagram of BSEP protein modelled in I-TSSAR Web Servers.

this model were computed. Hydrogen bonds were also detected by PIC web server [19] and WHAT IF web server [20]. Finally, the situation and relationships of these mutations and rare codons were evaluated in the structure of BSEP.

Results

Preparation of Molecular Modelling Structure of ABCB11
For understanding the protein structure and function, preparation of 3D structure of protein is a vital process in biology [21]. Our studies show that heretofore has not been determined as the crystal structure of BSEP, and it is obligatory to provide the 3D model from this protein. For this, by submitting the sequence of ATP8B1 in I-TSSAR Web Server, 3D models of these proteins were obtained. The I-TSSAR Web Server generated five models and best model showed -1.82 values of overall C-score, 0.49±0.15 value of TM-Score and Exp. RMSD was 14.1±3.9 (Figure 1).

In the following, the physiochemical properties of ATP8B1 protein model were calculated in ProtParam tool (Table 1) [22].

Table 1: In silico physico-chemical features of BSEP protein obtained from ProtParam tool.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>ATP8b1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical pl</td>
<td>6.17</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>146407.18</td>
</tr>
<tr>
<td>Sequence length</td>
<td>1321</td>
</tr>
<tr>
<td>Extinction coefficients (M-1 cm-1at 260 nm)*</td>
<td>129370- 128120</td>
</tr>
<tr>
<td>Asp + Glu</td>
<td>143</td>
</tr>
<tr>
<td>Arg + Lys</td>
<td>135</td>
</tr>
<tr>
<td>Instability index</td>
<td>33.29</td>
</tr>
<tr>
<td>Grand average of hydropathicity</td>
<td>0.011</td>
</tr>
<tr>
<td>Aliphatic index</td>
<td>92.60</td>
</tr>
</tbody>
</table>

*First number is based on the assumption that both cysteine residues form cystine and the second number that both cysteines are reduced.

membrane. 2 hits) and PF00005 (ABC_tran. 2 hits). Pfam is a comprehensive collection of protein domains and families [23]. These Pfam was analyzed in the Sherlocc program [13] and results show that any rare codon cluster was identified in these Pfam accession number of ABCB11 genes. Next, the nucleotide sequence of ABCB11 gene was analyzed in ATGme server [15]. Using the 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 2). Moreover, GC and AT contents of this gene were GC%: 45.44, AT%:54.56, calculated by this server.

In the following, RaCC server was used. By introduction of problematic residue codons as Pro, Ile, Arg and Leu, this result was refined. This analysis shows that ABCB11 gene has 48
Zarenehzad M. et al.

Figure 2: Schematic representation of the rare (orange) and highly rare (red) codons in ABCB11 gene.

These results demonstrate that MSS detected 6 clusters, Minmax detected 15 and sliding-window detected 10 clusters. It is important to note that the cluster length selected for MSS algorithms was 21 codons and for Minmax and sliding-window algorithms were 25 codons. The characteristics and position of these RCCs in the ABCB11 gene were calculated (data not shown).

Evaluation of Some Mutation Associated with Rare Codons

Some mutations in ABCB11 gene which are responsible for PFIC2 disease were pre-

recent studies show that the cluster length selected for MSS algorithms was 21 codons and for Minmax and sliding-window algorithms were 25 codons. The characteristics and position of these RCCs in the ABCB11 gene were calculated (data not shown).

Evaluation of Some Mutation Associated with Rare Codons

Some mutations in ABCB11 gene which are responsible for PFIC2 disease were pre-

recent studies show that the cluster length selected for MSS algorithms was 21 codons and for Minmax and sliding-window algorithms were 25 codons. The characteristics and position of these RCCs in the ABCB11 gene were calculated (data not shown).

Evaluation of Some Mutation Associated with Rare Codons

Some mutations in ABCB11 gene which are responsible for PFIC2 disease were pre-

recent studies show that the cluster length selected for MSS algorithms was 21 codons and for Minmax and sliding-window algorithms were 25 codons. The characteristics and position of these RCCs in the ABCB11 gene were calculated (data not shown).

Evaluation of Some Mutation Associated with Rare Codons

Some mutations in ABCB11 gene which are responsible for PFIC2 disease were pre-
Rare Codon Evaluation of ABCB11 Gene

Figure 3: Representation the rare codon of Arg, Leu, Ile, and Pro in the ABCB11 gene. These residues display in red, blue, green, orange, and red, respectively.
Figure 4: The representation of RCCs location in ABCB11 gene using MSS algorithm (A), minmax algorithm (B), and sliding window method (C).
residues with other residues.

Another missense mutation (1295 G>C) was also detected resulting in PFIC (Figure 7) [33]. In this mutation, the codon sequence of Arg470 residue (a residue with rare co-

residues with other residues.

Another missense mutation (1295 G>C) was also detected resulting in PFIC (Figure 7) [33]. In this mutation, the codon sequence of Arg470 residue (a residue with rare co-
Table 2: The 1-4 numbers are the position of rare codon in ABCB11 gene that mutated to the non-rare codon and 5-12 numbers are the position of non-rare codons that mutated to the rare codon in the PFIC2 patients and analyzed.

<table>
<thead>
<tr>
<th>Number</th>
<th>Exon</th>
<th>Nucleotide Change</th>
<th>Predicted Protein Effect</th>
<th>Location in Protein</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>c.1244G&gt;A</td>
<td>R415Q</td>
<td>NBF</td>
<td>[25]</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>c.1295G&gt;C</td>
<td>R432T</td>
<td>NBF</td>
<td>[26]</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>c.1409G&gt;A</td>
<td>R470Q</td>
<td>Adj WA</td>
<td>[27]</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>c.3724C&gt;A</td>
<td>L1242I</td>
<td>WB</td>
<td>[27]</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>c.1168G&gt;C</td>
<td>A390P</td>
<td>NBF</td>
<td>[27]</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>c.1779T&gt;A</td>
<td>S593R</td>
<td>NBF1</td>
<td>[28]</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>c.1388C&gt;T</td>
<td>T463I</td>
<td>WA</td>
<td>[27]</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>c.2130T&gt;C</td>
<td>P710P</td>
<td>IC3</td>
<td>[29]</td>
</tr>
<tr>
<td>9</td>
<td>20-21</td>
<td>c.2576C&gt;G</td>
<td>T859R</td>
<td>IC4</td>
<td>[27]</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>c.2776G&gt;C</td>
<td>A926P</td>
<td>IC5</td>
<td>[30]</td>
</tr>
<tr>
<td>11</td>
<td>23</td>
<td>c.2944G&gt;A</td>
<td>G982R</td>
<td>TM11</td>
<td>[31]</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>c.3346G&gt;C</td>
<td>G1116R</td>
<td>WA</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: A) the ribbon diagram of BSEP protein, with location of Arg\textsuperscript{415} residue (rare codon). The Arg\textsuperscript{415} residue forms the hydrogen bond with Gln\textsuperscript{414} B) mutation Arg\textsuperscript{415} to Gln\textsuperscript{415}.

Discussion

Three types of PFIC which are referred to autosomal-recessive liver disorders are related to mutations in hepatocellular transport-system genes [32, 33]. Mutations in ABCB11 gene have a variety in phenotype of autosomal recessive cholestasis liver diseases [1]. Liver disease in BSEP deficiency attributes to intrahepatocytic accumulation of toxic bile salts and failed the secretion of toxic bile salts [27]. Our comprehensive evaluation of these ABCB11 mutations show that all of these mutations that are resulted in cholestasis liver diseases are scattered throughout the gene and...
are not concentrated in special regions of this gene. This tells that the protein structure and function of BSEP has a very high sensitivity to the mutations and structural changes. This shows that any mutation in this structure can have destructive effects on the structure and function of BSEP. Re-evaluation of this mutation can provide a new approach in the study of these patients and design of new drugs. In this regard, considering the hidden information as “rare” codons that are infrequently used by cells and the specific roles of these rare codons in the proper folding of proteins is critical.

We have previously conducted the identification of detection of rare codons and molecular modelling of some proteins in our lab and have a good experience in these techniques [34-39]. For better evaluation of ABCB11 gene in this study, the detecting and studying of rare codons were conducted. In the following, the relation of some ABCB11 mutation with rare codons was studied. For the detection of rare codons, the following web server was used. For Pfam detection, the UniProt database identified two Pfams for ABCB11 as PF00664 (ABC_mem-

---

**Figure 7:** A) the ribbon representation of BSEP protein, with location of Arg$^{432}$ residue (rare codon). B) Mutation of Arg$^{432}$ to Thr$^{432}$.

**Figure 8:** A) the ribbon diagram of BSEP protein, with location of Arg$^{470}$ residue (rare codon). The Arg$^{470}$ residue forms the hydrogen bond with Asp$^{473}$ B) mutation Arg$^{470}$ to Gln$^{470}$. 

---
brane. 2 hits) and PF00005 (ABC_tran. 2 hits). These Pfams were analyzed in the Sherloc program that identified no rare codon clusters in the ABCB11 gene. In the following, this sequence nucleotide of this gene was analyzed in the ATGme web server that detected the 69 rare codon and 10 highly rare codons that may have a critical role in proper folding of protein chain. In addition, this gene was analyzed in the RaCC server detecting 48 rare codons for Arg, 18 rare codons for Ile, 11 rare codon for Leu and 11 rare codons for Pro. Finally, using LaTcOm web tool, the RCCs of this gene were detected. Results showed a large number of RCCs in the ABCB11 gene in these three algorithms.

The overall evaluation of rare codons of ABCB11 gene showed a large number of rare codons and rare codon clusters. This connotes that the protein structure of ABCB11 has important hidden features that need to guarantee the proper folding of this protein. For this reason, a large number of rare codons slowly had done the overall folding rate of this pump so that the final protein has a correct structure and function. However, these results show a large number of the rare codons of Arg and with large number of non-covalent hydrogen bonds play a special role in the correct folding of ABCB11. Finally, we focus on some rare codons related to PFIC2. In this regard, 3D molecular modelling of ABCB11 was conducted in I-TSSAR Web Server.

The precise analysis reveals that four rare codons were mutated in PFIC2 disease. These four rare codons are distributed in different regions in the structure of BSEP protein (Figure 10).


Structural analysis demonstrates that these rare codons form some hydrogen bonds with other residues disrupted with mutation in PFIC2 patients. This mutation with the disruption of these hydrogen bonds or change in the protein folding rate affects the protein folding that may disrupt the proper structure and function of ABCB11. It shows the critical role of these residues in the process of protein folding. However, other hypotheses should be considered too.

Table 2 shows new rare codons which are...
caused by some mutations of PFIC2 patients. In these patients, non-rare codons were converted to rare codons that interfere with suitable protein folding rate. These mutations change the hydrogen bond network affecting the structure and function of BSEP protein. On the other hand, these mutations changed the protein folding rate interfering with correct protein folding. Besides this, these new residues have a different side chain in comparison with original residues that may create the structural repulsion interfering with proper folding and functional activity of BSEP. Finally, these mutations either by change of folding rate or by change of hydrogen interaction have a negative effect on the BSEP and result in PFIC2 disease. Meanwhile, some in-vivo and in-silico evidence as molecular docking and evaluating these mutations is needed for our theoretical study confirmation. Our data showed that rare codon positions might have an essential role in folding and activity of BSEP. This study may also provide new insights into drug design for the treatment of PFIC2, in the future.

**Acknowledgment**

The study was performed as a PhD thesis in
Data Sharing Statement: All data were extracted from database of clinics and hospitals affiliated to Shiraz University of Medical Sciences.

Conflict of Interest
There is no conflict of interest for this study.

References
20. Vriend G. WHAT IF: a molecular modeling and
