In Silico Structure Analysis of Type 2 Diabetes Associated Cysteine Protease Calpain-10 (CAPN10)

Rajneesh Prajapat*, Ijen Bhattacharya

Department of Medical Biochemistry, Faculty of Medical Sciences, Rama Medical College, Rama University, India

Copyright©2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Calpain-10 (CAPN10) is a cysteine protease that is known to hydrolyze specific substrates significant for calcium-regulated signaling pathways and it's activated by intracellular calcium (Ca^{2+}). The calpain10 is known to be involved in the cellular degenerative processes that characterize several diseases such as cancer, stroke and heart attack. The role of calpain10 was recently identified and associated with diabetes mellitus type 2. In this paper, the homology modelling procedure was used to determine the 3D structure of human calpain10 (AAH07553). The \( \mu \)-calpain (1QXP) of Rattus norvegicus was selected as a template for the construction of calpain10 model. Ramachandran plot of calpain10 (AAH07553) has only 55.9% residues in the most favored regions, while template \( \mu \)-calpain (1QXP) has 69.3% residues in the most favored regions. The model was validated by using protein structure tools RAMPAGE and Prochek for reliability. 3D structure of calpain10 suggested its active site remains conserved among family members and the major interactions are similar to those observed for the template (1QXP).

Keywords Calpain 10, Type II Diabetes, Homology Modelling

1. Introduction

The type 2 diabetes (T2D) is triggered by environmental, genetic risk factors. T2D is characterized by impaired insulin stimulated glucose uptake in muscle; improved hepatic glucose production and altered glucose induced insulin secreton [10,33]. The CAPN10 gene encodes calpain10. The CAPN10 abundant in the mitochondria and it is involved in apoptosis and age related diseases, plus release and storage of Ca^{2+} [22]. Approximate 14 isoforms of calpains (calcium-dependent cysteine proteases) have been characterized that present in multiple tissues [29, 28, 23] such as pancreas, liver and skeletal muscle and involved in diabetes [29]. The calpains are non-lysosomal cysteine proteases [26] that are activated by intracellular Ca^{2+}. An increase in the concentration of mitochondrial Ca^{2+} can initiate a series of proteolytic signals and participate in various signal transduction pathways [11] that can cause irreversible damage to cells. Therefore, the over expression of this enzyme is responsible for mitochondrial dysfunction [2].

The fine mapping of diabetes related genes suggested that the CAPN10 may serve as an important T2D susceptibility gene [31; 7]. A variation in the gene encoding the cysteine protease calpain10 (CAPN10) is associated with type 2 diabetes [27, 25].

New findings support the association of polymorphisms with T2D and / or related quantitative phenotypes [13]. The genetic variation at CAPN10 in different human populations over a range of phenotypes related to T2D. The CAPN10 influences insulin sensitivity and glucose homeostasis in nondiabetic members of kindreds at high risk for Type 2 diabetes mellitus (T2DM) [9].

Homology modelling refers to the modelling of protein 3D structure using a known experimentally determined structure of homologous protein as a template. Homology modelling provides structural information, that is important to understand of protein function, dynamics, interactions with ligands [3]. In the present study, we used protein homology modelling [15] to determine the structure of human calpain 10.

2. Materials and Methods

Operating System: In the present study, the Intel (R) Core (TM) i3-370 M CPU @ 2.40 GHz and 32 bit operating system (HP ProBook) was used.

Retrieval of Sequences and Sequence Alignment

The amino acid sequence of human calpain10 (AAH07553) was retrieved from GenBank-NCBI (www.ncbi.nlm.nih.gov) in the FASTA format. The calpain10 sequence was identified as one of the 46 members of the C2 family of proteases according to the MEROPS.
In order to build a model of protein domain, Multiple Sequence Alignment was performed between full length calpain10 (AAH07553) sequence via the BLASTp alignment tool [1, 34] to find out the related homologous. The PDB file of calpain10 (AAH07553) was generated by using 3D-JIGSAW protein comparative modelling servers. The PDB file of the query and homologous template sequence were further utilized for 3D model energy validation [12].

**Molecular Modelling of Calpain10**

The UCLA-DOE server provides a visual analysis of the quality of a putative crystal structure of proteins. Verify 3D expects this crystal structure to be submitted in PDB format [17]. RAMPAGE program was used for visualizing and assessing the Ramachandran plot of calpain10 (AAH07553) and its homologous template [16]. The validation of structure models was performed by using PROCHECK [14]. The model was selected on the basis of various factors such as overall G-factor, number of residues in core that fall in generously allowed and disallowed regions in Ramachandran plot (Fig 1; Fig 2). The model was further analyzed by QMEAN [4, 21] and ProSA [32]. ProSA was used for the display of Z-score and energy plots.

The 3D structure selected as a template for constructing the model of calpain10 was μ-calpain from Rattus norvegicus (PDB access code: 1QXP), which has a length of 900 amino acid residues and a resolution of 2.8 Å by X-ray crystallography [8]. It has 31% identity and 46% similarity with the target sequence (1QXP). The quality of folding was checked using Verify3D [17], and the system energy and quality of 3D alignment were monitored using Atomic Non-Local Environment Assessment (ANOLEA) [18, 19]. Produced models of query and homologous template were ranked on QMEAN server and utilized for the ribbon structural model construction (Fig. 9, Fig. 10). QMEAN locate the position of different amino acids present in the active site of proteins and estimates per-residue error [4].

### 3. Results and Discussion

**Building of Protein Model:** Sequence alignment of calpain10 (AAH07553) protein by using the BLAST, revealed sequence homology with μ-calpain (1QXP) (ID= 31%), which was selected as template for the model building of calpain10 (AAH07553) protein. Total 227 residues (45% of query sequence) have been modelled with 99% confidence by the single highest scoring template. To build the model, BLAST was done with the maximum E-value allowed for template being 5e⁻⁶².

**Model Reputation:** The calpain10 (AAH07553) model showed good stereo chemical property in terms of overall G-factor value of -0.65. G-factor value indicating that geometry of the model corresponds to the probability conformation with 55.9% residues in the core region of the Ramachandran plot showing high accuracy of model predicted [24]. The number of residues in allowed and generously allowed region was 26.5% and 17.6%, respectively, and none of the residues were present in the disallowed region of the plot (Figure 1). A similar approach was also used for template μ-calpain (1QXP) and its Ramachandran plot has 69.3% residues in favoured region, 20.1% in the allowed region and 10.6% in the outlier regions (Figure 2).

The Ramachandran plot of calpain10 (AAH07553) has only 55.9% residues in the most favoured region and homologous template μ-calpain (1QXP) has 69.3% residues in most favoured regions therefore μ-calpain (1QXP) is more stable than calpain10 (AAH07553) (Table 1).

A good quality Ramachandran plot has over 90% in the most favoured regions [20] therefore both target and template model cannot be included in the good quality class. Therefore the energy minimization of both models should be required to enhance the stability by using the standard protocols of combined application of simulated annealing, conjugate gradient and steepest descent.
Figure 1a. Ramachandran plot of 3D model of calpain10 (AAH07553)
Figure 1b. Non-proline residues and non-glycine residue regions
Figure 2a. Ramachandran plot of 3D model of homologous template μ-calpain (1QXP)
Figure 2b. Non-proline residues and non-glycine residue regions

Table 1. Results summary of the Ramachandran plot

<table>
<thead>
<tr>
<th>Accession No</th>
<th>Protein</th>
<th>Description</th>
<th>Residues (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Favourite regions</td>
</tr>
<tr>
<td>AAH07553</td>
<td>calpain10</td>
<td>Homo sapiens</td>
<td>55.9</td>
</tr>
<tr>
<td>1QXP</td>
<td>μ-calpain</td>
<td>Rattus norvegicus</td>
<td>69.3</td>
</tr>
</tbody>
</table>
The verified 3D high score 0.43 for calpain10 (AAH07553) indicates that the environmental profile of the model is good (Figure 3). A profile score above zero in the Verify 3D graph [5, 17] corresponds to an acceptable environment of the model. In the Verified 3D plot, 25.81% of the residues had an averaged 3D-1D score $\geq$ 0.2.

The Errat is a program for verifying protein structures determined by crystallography. On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value [6]. Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3A) the average overall quality factor is around 9. The overall quality factor for calpain10 (AAH07553) was 33.333 (Figure 4).

Model Validation: Potential problems of protein structures based on energy plots are easily seen by ProSA and are displayed in a three-dimensional manner. ProSA was used to check the three dimensional model of calpain10 (AAH07553) and $\mu$-calpain (1QXP) proteins for potential errors (Figure 5 and Figure 6). The ProSA web $z$-scores of protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The plot shows only chains with less than 1,000 residues and a $z$-score of 10. The $z$-scores of calpain10 (AAH07553) and $\mu$-calpain (1QXP) are highlighted as large dots.

The calpain10 (AAH07553) ProSA $Z$-score was -5.96 for [Figure 5] and for template $\mu$-calpain (1QXP) ProSA $Z$-score was -1.80, that indicates the overall model quality of target and template (Fig 6) measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations [30].
Figure 5. ProSA web service analysis of calpain10 (AAH07553) overall model quality (a) and local model quality (b).

Figure 6. ProSA web service analysis of μ-calpain (1QXP) overall model quality (a) and local model quality (b).
The QMEAN score of the calpain10 (AAH07553) model was 0.19 and the Z-score was -5.96 (Fig. 7) and for the template \( \mu \)-calpain (1QXP) QMEAN score was 0.60 with Z-score was -1.80 (Figure 8), which very close to the value of 0 and this shows the fine quality of the both models [32; 24]. The estimated reliability of the model was expected to be in between 0 and 1 and this could be inferred from the density plot for QMEAN scores of the reference set. A comparison between the normalized QMEAN score (0.40) and protein size in non-redundant set of PDB structures in the plot revealed different set of Z-values for different parameters such as C-beta interactions (-3.36), interactions between all atoms (-3.44), solvation (-7.62), torsion (-5.33), SSE agreement (-3.65) and ACC agreement (-4.05) (Figure 7; Table 2).
Figure 8a. The density plot for template μ-calpain (1QXP) showing the value of Z-score and QMEAN score.

Figure 8b. Plots showing the QMEAN value as well as Z-score.

Table 2. Table showing the QMEAN value as well as Z-score of calpain10 and μ-calpain.

<table>
<thead>
<tr>
<th>Accession No / PDB access code</th>
<th>Protein</th>
<th>Z - Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>QMEAN</td>
</tr>
<tr>
<td>AAH07553 calpain10</td>
<td>-5.96</td>
<td>-3.36</td>
</tr>
<tr>
<td>1QXP μ-calpain</td>
<td>-1.80</td>
<td>-0.12</td>
</tr>
</tbody>
</table>
4. Conclusions

The human calpain10 (AAH07553) protein model was obtained through homology modelling and the main interactions are similar to those observed for template μ-calpain (1QXP). The calpain10 (AAH07553) model showed overall G-factor value of -0.65 with 55.9% residues in the favoured region of Ramachandran plot and its template μ-calpain (1QXP) had 69.3% residues in favoured region that indicates high accuracy of model predicted. The calpain10 (AAH07553) ProSA Z-score was -5.96 and -1.80 for its template μ-calpain (1QXP), that indicates overall model quality and measures the deviation of total structural energy with respect to an energy distribution derived from random conformations. We hope that, these results will be useful for the designing of inhibitors for calpain-10 and understanding of the mechanism of inhibition at the molecular level.

REFERENCES


