May 2019

Molecular signatures of Calpain 10 isoforms sequences, envisage functional similarity and therapeutic potential

Bushra Chaudhry
Aga Khan University, bushra.chaudhry@aku.edu

Farina Hanif
Dow University of Health sciences, Karachi

Kausar Saboohi
University of Alberta

Molecular signatures of *Calpain 10* isoforms sequences, envisage functional similarity and therapeutic potential

Bushra Chaudhry*, Farina Hanif and Kausar Saboohi
Department of Biological & Biomedical Sciences, Aga Khan University Medical College, Stadium Road, Karachi, Pakistan

Abstract: *Calpain 10* plays a role in insulin secretion, action and susceptibility to type 2 diabetes. The mechanism through which it influences the insulin secretion and action is not completely defined. A structural bioinformatics approach is applied to envision its mechanism of action using available tools on NCBI (blastp and blastn), EMBL-EBI, Ensembl, Swiss Model Repository websites, I-TASSER, PROCHECK program and Discovery Studio software. Homology of domain I and II of *calpain10* (isoform *a*) was established with super family cysteine proteinase domains (*II a* and *II b*, e=1.30e-77, 1.00e-20). Remaining sequences of domain III and T from (isoform *a* and *c*) indicated some similarity (Avg. e=1.94e-37) to calpain large subunit domain III (PF01067), the isoform *g* (139 AA) showed similarity with a part of catalytic domain of cysteine protease super family (e-value 1.00e-20). Swiss-model repository for 3D structures of protein, showed structural resemblance of 29% with 1QXP template of mu-calpain, 27% with 1KFX of m-calpain and 32% with 2P0R of *calpain 9* in complex with leupeptin. Models prepared through I-TASSER confirmed through Ramachandran (RC) plots. The *calpain 10* isoforms *a, c* and *g* show partial structural and functional resemblance to m, mu and *calpain 9*. This information is useful to find new drugs for disease management.

Keywords: Cysteine proteases, *calpain 10* isoforms, conserved domain homology, 3D modeling structure function prediction, action mechanism, Type 2 diabetes, insulin secretion.

INTRODUCTION

Several drugs for the management of diabetes are available worldwide, but their effectiveness and inconsistency of side effects in >60% of patient population is a question, and no perfect remedy to give a comprehensive control is present to date. The commercially accessible drugs are classified as; Insulin, secretagogues, sensitizers, alpha-glucosidase inhibitors and peptide analogs (Adams and Norwood 2003, Lebovitz 2004). They may be used either individually or in combination but results are often not promising and efficacy varies inpatients. Prolonged use of these drugs results in the succession of co-morbidities, which add to complications in the disease (Dupree and Meyer 1980).

During past few decades, type II diabetes mellitus has attained epidemic levels rapidly due to the life style, environmental changes and dietary habits both in developed and developing countries. Discovery of a new drug is crucial to save the world population against the recent epidemic of diabetes and related complications. Structural bioinformatics (SBI) is the new field for drug designing and assist in the development of new drugs, through subsequent steps which start with the identification of a target gene, either with the help of genome scan or by genetic polymorphisms. Further three dimensional structures; conserved domain homology analysis; target–ligand binding analysis, protein- protein interactions and efficacy of specific protein against the target has been incorporated in current technological efforts.

Investigation of protein and gene function is a central question in molecular biology, biochemistry and genetics. Genes evolve from the same ancestral gene therefore retain similarity in their function in most of the cases. Finding known genes which have sufficient or less sequence identity with a query sequence is therefore a powerful way for predicting function of unknown proteins. In this paper we use computational techniques to generate 3D models for *calpain 10* isoforms (*a, c* and *g*). The selected isoforms are reported as major transcripts (Horikawa *et al*, 2000) and major protein bands through immuno-blots in our lab (unpublished data). We have studied the resources available in database sequences for functional prediction on our generated 3D models of theses isoforms and hence reporting comparison with existing 3D templates. The minor isoforms (*b, d, e, f and h*) of *Calpain 10* has already being reported from our lab (Saboohi *et al*, 2013). The modeling of minor isoforms was not included as it was beyond the length of this paper. The primary information about *Calpain 10* protein coding gene is located on chromosome 2 at the position of 2q37. 3. The Gene has official symbol CAPN10, identification ID-11132, OMIM-605286, and synonym is KIAA1845. Primary source for this gene is Hugo Nomenclature Committee (HGNC: 1477). It can be browsed on related sites such as Ensembl by typing ENSG00000142330 and at Human Protein Reference data base (HPRD: 5595)

**Corresponding author:* e-mail: bushra.chaudhry@aku.edu

Changes in expression level of the general calpain family members are known to be involved in the development of different diseases such as ischemic stroke, traumatic brain injury, rheumatoid arthritis, cataract formation, Alzheimer's disease, limb-girdle muscular dystrophy and type 2 diabetes mellitus (Hanis *et al.*, 1996, Horikawa 2006, Horikawa 2005, Horikawa *et al*., 2003, Horikawa 2002, Stumvoll *et al*., 2001, Hung and Wang 2001, Horikawa *et al*., 2000). *Calpain 10* isoforms (*a, c* and *g*) are universally found in all tissues of human body, are thought to be involved in calcium activated, non lysosomal, neutral, cysteine protease which partially proteolyses its substrates resulting in modulation of its own activity.

Genetic variation in *calpain10* gene has been found associated with 3 fold increased risk of type 2 diabetes. The severity of disease may vary in different population, (Horikawa 2006, Horikawa 2005,Turner *et al*., 2005, Horikawa *et al*., 2003, Horikawa 2002, Baier *et al*., 2000, Horikawa *et al*., 2000) according to the expression levels of its isoforms which exert their effects on integrated metabolism. The direct mechanism of action of *calpain 10* on the glucose metabolism is yet to be explored but suggestions of an indirect mechanism of action are reported (Saboohi *et al*., 2013). Its low expression is suggested to play integral role in causation of hyperglycemic condition. Its high expression has been reported in Huntington's disease, suggesting its possible role in neurological pathology (Gafni *et al*., 2004). Therefore predicting its functions by using bioinformatics tools is considered a feasible strategy to evaluate it as potential therapeutic candidate for better treatment modalities in future. This can be verified in wet lab experiments especially in diseased animal models or human cell lines by balancing the levels of deranged isoforms.

MATERIALS AND METHODS

Functional assessment of *calpain 10* gene using Bioinformatics tools is determined in this study. We have used National Center for Biotechnology Information (NCBI), Ensembl genome browser and Swiss Model Repository Datasets for this study. Ensembl was used to search for the nucleotide and peptide sequences of different isoforms of the *calpain 10* gene. NCBI Basic Local Alignment Search Tool (LAST) was used for the analysis of nucleotide and peptide sequences according to NCBI tutorial, instruction and previously described method (Zhang *et al*., 2000). Advanced BLAST was used for sequence alignment analysis by Align Software (Marchler-Bauer *et al*., 2009). Homologous domain analysis was done by NCBI Conserved Domains Software (Marchler-Bauer *et al*., 2011, Marchler-Bauer and Bryant 2004). Three dimensional structures were predicted for *calpain 10* isoform *a, c* and *g* by depositing protein sequences in Swiss-Model Repository, a division of Swiss ExPASY- PROSITE on web (Kopp and Schwede 2004a, Kiefer *et al*., 2009) and on I-TASSER server. Quality of models was assessed by making RC plots using PROCHECK. We selected *calpain10* isoform *a (version NM_023083.2/NP_ 075571.1)*, as canonical sequence and analyzed the variation of sequence and structure among isoforms *c (NM_023085.3/NP_ 075573.2)* and *g (NM_023089.1)*. We would like to mention here that isoform g records have not found sufficient support of being a transcript recently when reviewed, yet we decided to involve this in the current study which is likely to become a part of other suitable peptide as there is no evidence yet if it degrades completely. The following tools were applied in the series of analyses presented in this paper.

Retrieving information of calpain 10 gene and its isoforms (a, c and g)

We have used Ensembl (http://www.ensembl.org/) to retrieve information of calpain10 gene and its isoforms (*a, c* and *g*). We extracted gene, mRNA and protein sequences from this resource. Briefly, Ensembl home page was accessed. Then "human" from popular genomes option was selected. A new page specifically for human genome appeared; CAPN10 for *calpain10* gene was written in the window and clicked for search. Summary results appeared which contained information by feature type and by species. The Feature type option provided gene, somatic mutation, transcript and variation which were available to select in left hand side navigation pane. By species option, similar information was also available. A new page appeared with a left hand side navigation pane which showed gene summary containing links for Splice variants, Supporting evidence, Sequence, External reference, Regulation, Comparative Genomes, Gene tree, Orthologoues, Paralogues, Protein families, Phenotypes, genetic variation and gene history. Nucleotide and amino acid sequences of *calpain 10* gene, mRNA and amino acid sequences of its isoforms were obtained from this site.

Nucleotide and protein sequence alignment

United States National Centre of Biotechnology Information (NCBI) data bases were accessed through (http://www.ncbi.nlm.nih.gov/). Medical Subject Headings (MeSH) was used to search all the medical and biological background information about *calpain 10* gene and its isoforms. We used NCBI blastn for nucleotide and blastp for protein sequence alignment. The protein sequences of calpain 10 isoforms searched were either obtained from Ensemble or from NCBI and analyzed by using different tool available on NCBI blast page (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For analysis, key word, *calpain 10* was entered in the search window of the home page. The available data present in the resource appeared in a new window. We have used Align tool for multiple alignment of *CAPN 10* isoforms. Alignment of

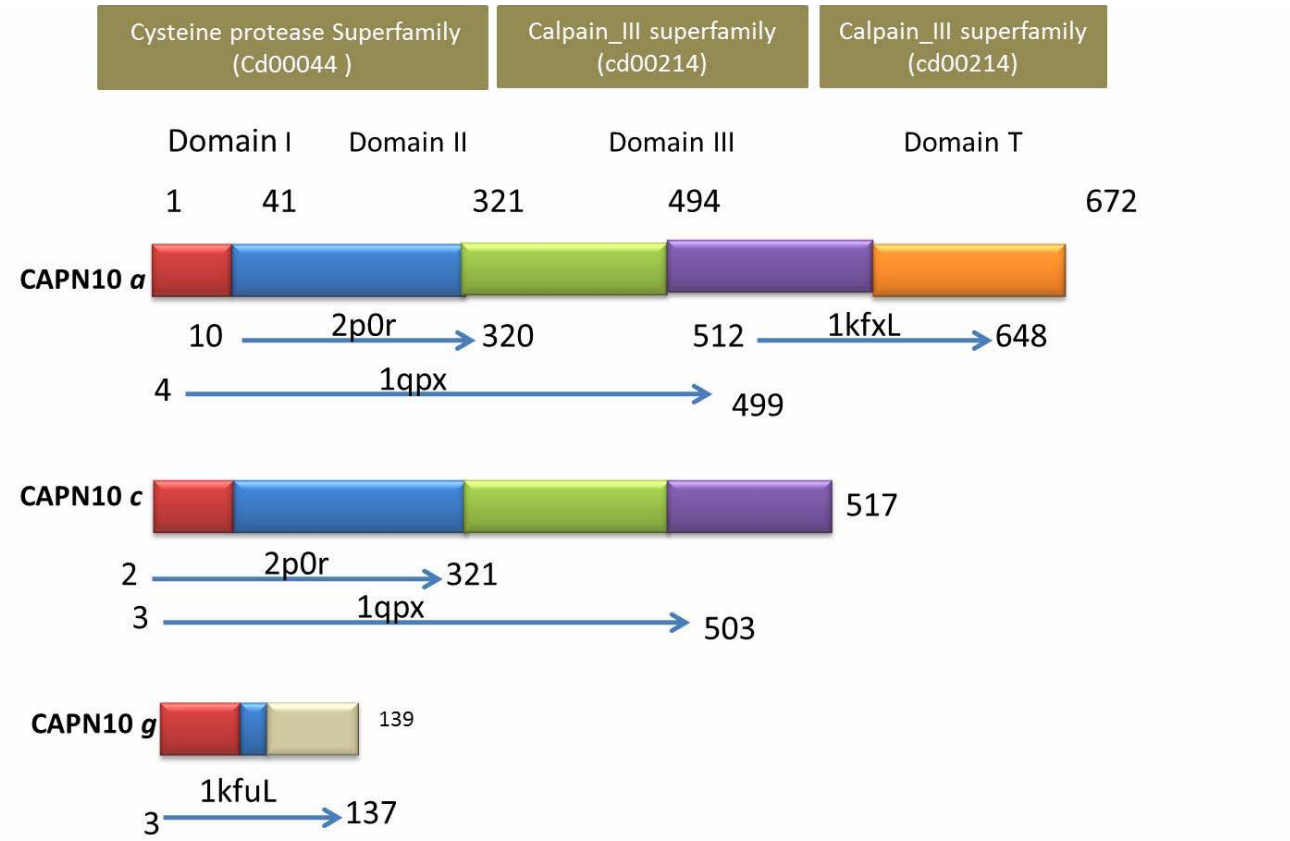

Fig. 1: Calpain10 isoforms conserved regions and templates: cd00044: Calpain super family domains IIa and IIb, cd00214 calpain_ III: C2 like sub domain III of super family, *calpain10 a, c & g* coloured boxes: red (Domain I a a1-41), Blue (Domain II, a.a 40-321), Green (Domain III, a a 322-494), purple and orange (Domain T, a.a 513-654), greenish block isoform *g* (47, specific substituted amino acids). Blue arrows indicate start and end limits of template

mRNA and amino acid sequences of *calpain 10* isoforms were analyzed. The blast nucleotide algorithm (blastn) first breaks the query sequence into the short subsequences (words) and then identifies the exact matches. In contrast standard protein-protein blast (blastp) identified a query sequence as well as other similar sequences in protein database. *Calpain 10* isoform *a* was used as canonical sequence to determine the alignment analysis of *calpain 10 c* and *g*. We also analyzed conserved domains of calpain 10 with other proteins by Conserve domain tool in NCBI following already reported method.

Protein modeling and structure analysis

ExPASy-PROSITE (http:// prosite expasy.org/) is an exclusive web server for 3D protein modeling and structure analysis. 3D structures of *calpain.10* isoforms were obtained from Swiss-model ExPASy-PROSITE (http://swissmodel.expasy.org/workspace) by uploading of amino acid sequences of all three isoforms, isoform *a* (Q9HC96-1), isoform *c* (Q9HC96-3), Isoform *g* (Q9HC96-7) in workplace available on the site. Login ID was created on home page of this site. Once registration was completed amino acid sequence of a protein were uploaded in the specific window. Results of 3-dimentional protein structure were received on email.

Further, I-TASSER (Iterative Threading ASSEmbly Refinement) is an on line server which uses amino acids sequences to predict three D structure models of protein molecules. Briefly, login ID was created on I-TASEER (Zhang lab.ccmb.med.umich.edu\I-TASEER). Once ID was created sequences of the three isoforms were pasted individually in the window available on the server and I-TASEER was run on them. Results generated by the server were retrieved by clicking on the link provided by I-TASEER via email. To identify conserved domains three-D models that were obtained from I-TASEER were open as PDB file in Discovery Studio software. Domains were identified by high lighting the sequence of only those amino acids that comprises conserved domains. The software then simultaneously highlights that specific sequence on the model

Evaluation and validation of protein model

The three-D models of isoforms *a, c* and *g* that we retrieved from I-TASSER were uploaded on PDB sum Generate site (http://www.ebi.ac.uk\thornton-srv/database

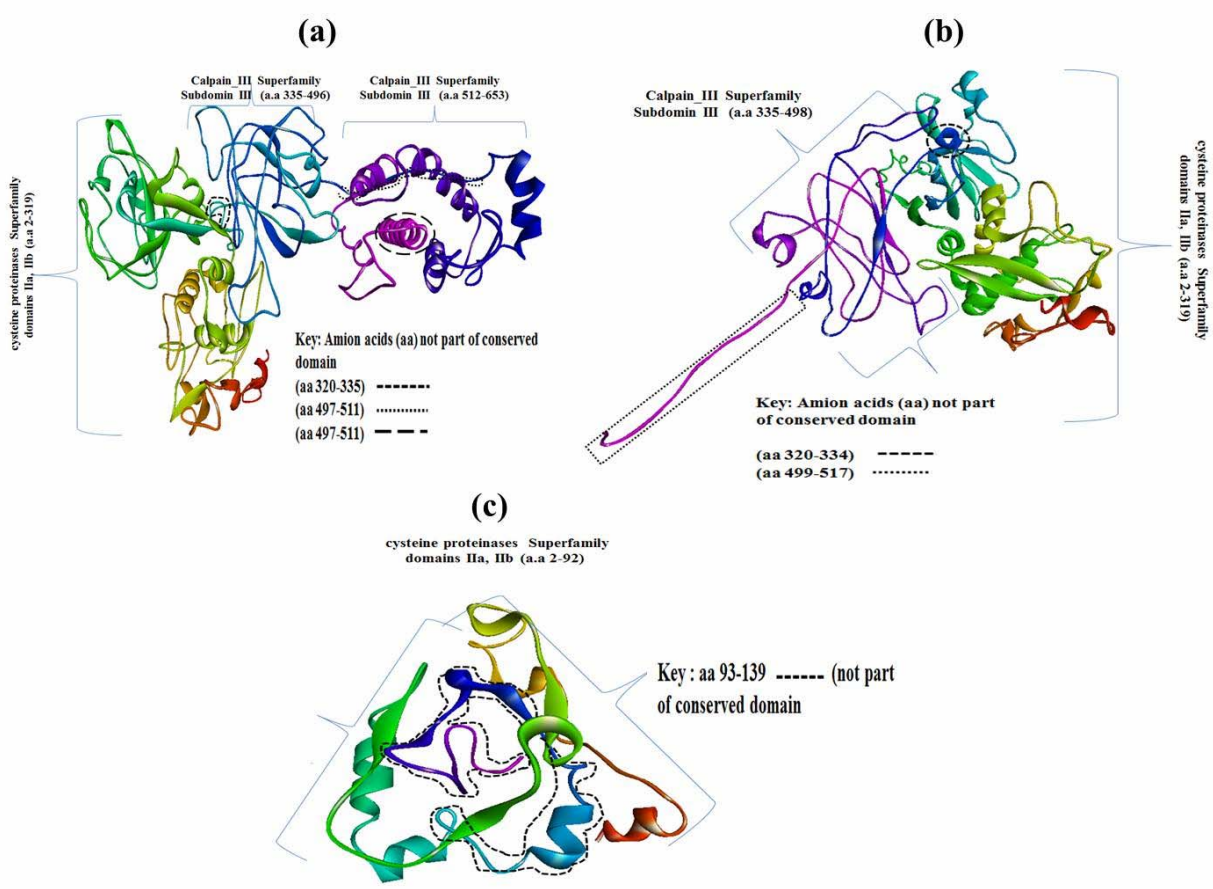

Fig. 2: 3D model of Calpain 10 a, c and g showing conserved domains. 3D Model was obtained using I-TASSER and Conserved regions determined using CDART and were identified on the model using Discovery Studio software.

/pdbsum/generate) along with the email address. PDB sum then performed full structural analysis and provide RC plots of the uploaded protein models using PROCHECK program.

RESULTS

Nucleotide sequence alignment

Sequences of *calpain 10* isoforms (*a, c and g*) indicated the difference in the structure. Isoform *c*, 2197 nucleotide stretch showed overall 82% (e-value, 0.0) with isoform *a* sequence. Similarity was distributed as 100% identity from 1-1464 nucleotides (e-value 0.0) and a 462 nucleotide sequence stretch gap was found as missing nucleotide. While the remaining sequence from 1938-2662 again showed 100% identity (e-value, 0.0).

The isoform *g* showed 17% (e-value, 0.0) nucleotides similarity with isoform *a* sequence. It aligned from 1-453 with 20-472 sequences of form *a*. Rest of the sequences is different from isoform *a* as well as isoform *c*.

Protein sequence alignment

Protein sequence showed similar trend as of nucleotide sequence similarity. Taking isoform *a* (672 amino acids) protein as reference and aligning with isoform *c* (517 amino acid), showed 91% homology (e-value 0.0) in initial 487 amino acids with 0% gap. A stretch of amino acids from 428-582 was missing from isoform *c* and then again 65%identity (e-value 8e-52) was found in last 182 in initial 487 amino acids with 0% gap. A stretch of amino acids with 428-582 was missing from isoform *c* and then again 65% identity (evalue 8e-52) was found in last 182 amino acids with 15% gaps. Isoform *g* (139 amino acids) showed 97% similarity (e-value 3e-53) from 1-97 amino acids with isoform *a*. Amino acids sequence from 93-139 are found different from isoform *a* and rest of the peptide sequence is missing, Indicating difference of isoform *g* sequences. Isoform *g* showed the same similarity when aligned with isoform *c*.

Conserved domain analysis

Search for similar domain architecture for *calpain 10a* (672 amino acids) and those of *c* and *g* was performed through NCBI conserved domain analysis (Marchler-Bauer *et al*., 2009, Marchler-Bauer and Bryant 2004). The results are presented in (fig. 1 and fig. 2).

Results indicated that first two domains of isoform *a* (domain I consisted of 32 amino acids & domain II of 270 amino acids) also present in all the isoforms, hybridized

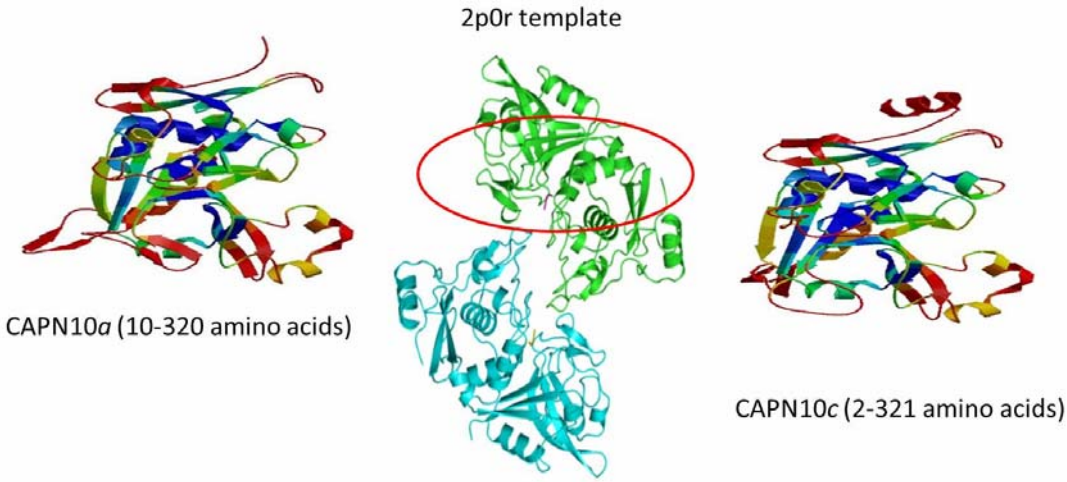

Fig. 3: 3D Structure of Calpain 10 (CAPN10) isoform *a* & *c*: created from Swiss model-ExPASy (http://swissmodel.expasy.org). PDB template 2p0r hybridized 32% and 31.8% with CAPN10 *a* and *c* respectively. Estimated per-residue inaccuracy visualized using a color gradient from blue (more reliable regions) to red (potentially unreliable).

with calpains of calcium dependent cysteine proteinase super family domain IIa and IIb. The conserved domain length was 315 amino acids (a. a) (2-319 a. a, bit score: 324.67, e-value: 1.25e-106). These conserved domains function broadly in cytoskeleton remodeling processes, cell differentiation, apoptosis and signal transduction.

The calpain isoform *a*, has two more consecutive domains, domain III and T domain. Domain III consisted of 186 amino acid starting from 304-490 amino acids and T domain is of 181 amino acids starting from 491-672 amino acid, were similar to the super family calpain large subunit, domain III (243 a. a, starting from 327-570). This domain III is in 80 kDa large subunit, consisted of two sub domains which are calmodulin domains involved in binding with calcium and activating the domain II to make interaction with phospholipids and translocation to cytoplasmic/nuclear membranes. Conserved domain included sub domain III of typical and atypical calpains. The conserved domain length was 150 amino acids (335-496 a. a, bit-score: 164.39,e-valueof 4.74e-48). The next conserved region including domain T was of 150 amino acids (512-653 a. a, bit score: 161.70 e-value: 5.45e-47), fig. 2a.

The Isoform *c* contained three domains. First two domains were similar to the domains of calcium dependent cysteine protease super family. The conserved domain length was 315 amino acids (2-319 a. a, bit score: 321.59, e-value: 2.30e-107) while third domain was similar to the C2 like sub domain III, as in isoform *a.* the conserved domain length was 150 amino acids (335-498 a. a, bit score: 160.93 e-value: 1.37e-47), fig. 2b.

The form *g* of *calpain 10* consisted of a shorter peptide (139 amino acids) and a portion of 91 Amino acids hybridized with proteolytic domain II of calcium dependent cysteine protease super family. The conserved domain length was 315, bit score: 105.49, e-values: 8.13e-29), fig. 2c.

Protein model analysis

In order to predict the structure and function of these isoforms of *capain 10*, three dimensional structure were analyzed by uploading polypeptide sequences of *calpain 10* isoforms *a*, *c* and *g* to Swiss-model repository of ExPASy prosite (Kiefer *et al*., 2009, Kopp and Schwede 2004a, Kopp and Schwede 2004b).The results indicated the hybridization of all three isoforms to its template through ExPASy as given below:

Isoform a

Polypeptide sequence of Isoform *a* hybridized to three different building models (figs. 3-5). A stretch of 5-499amino acids has been hybridized to 1qxpB template.1qxpB is a mu-like calpain and showed 29% similarity with it (e-value0.00e-1, QMEAN Z-score-5). Another 10-320 amino acid stretch showed 32% similarity with2p0r template which is a sequence stretch of calpain 9 (e-value 0.00e-1, QMEAN Z-score-4.63). Another stretch of 512-649 amino acids was found 27% similar with 1kfxL template, a sequence stretch of m-calpain large subunit (e-value 1.6e-29, QMEAN Z-score-3.72). The low QMEANZ scores represent either only for membrane proteins or indicate models of poor quality which generally are expected to reach such scores.

Isoform c

Polypeptide sequence of isoform *c* was hybridized with two different building models (figs.3-4). A stretch of 2-503 amino acids was found 28.2% similar to 1qxpB template which is a stretch of nucleotides from mu-calpain (e-value 0.00e-1, QMEAN Z-score-5.47) while the other 2-321 amino acid stretch was found 32.4%

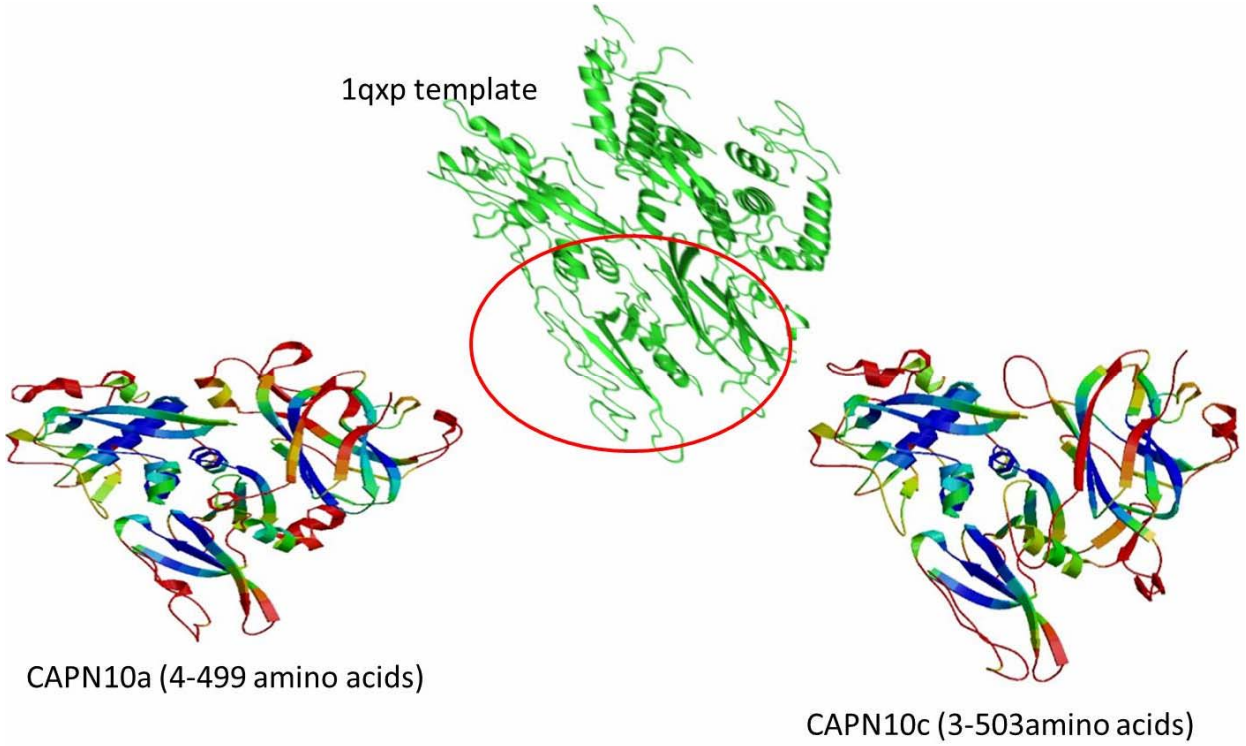

Fig. 4: 3D Structure of Calpain 10 (CAPN10) isoform *a & c:* created from Swiss model-ExPASy (http://swissmodel.expasy.org). PDB template 1qxp hybridized 29% and 28.2% with CAPN10*a* and *c* respectively. Estimated per-residue inaccuracy visualized using a color gradient from blue (more reliable regions) to red (potentially unreliable).

similar as 2p0rA template , a part of calpain 9 (e-value 0.00e-1, QMEAN Z-score-4.48).

Isoform g

Polypeptide sequence of isoform *g* was hybridized with only one building model (fig. 5). A stretch of 3-137 amino acids was found to 26.3% similar as 1kfuL template (e-value 1.3e-39), which is a part of m-calpain form II while similarity with m-calpain form I (1kfxL, e-value 5.7e-38QMEAN Z-score-5.737)

Validation of protein models

3D models of *calpain 10a, c* and *g* that we salvaged from I-TASSER server was evaluated for authentication using RC plot. Briefly, about 93.3%, 94.6% and 98.3% residues (non-glycine and non-proline) of *calpain 10 a, c* and *g* are in acceptable regions. The distribution of amino acids with respect to RC plot is given in table 1.

DISCUSSION

To envisage *calpain 10* possible involvement in disease pathogenesis and their possible potential as future drug targets, a structural bioinformatics approach was attempted in this study. This approach provided useful information for their potential to approach new developments. The importance of *calpain10* involvement in the development and etiology of diabetes has increased in the last decade. We report computational analysis of three isoforms of *calpain 10* (*a, c and g*) in this study. Our results predicted the possible functions of (*calpain 10a, c* and *g*) isoforms are similar to protease domains of three family members (*mu, m, calpain 9*) of calpain superfamily. These are involved in pathogenesis of diseases. However the expression levels of calpain 10 isoforms in health and diseases is almost unknown. RNA interference assays may help to identify the specific target for drug development, avoiding deleterious side effects of unspecific protease inhibition.

The isoforms *a* and *c* are 82%, while *a* and *g* are 17% similar to each other in their nucleotide sequences and possessed similar pattern in conserve domain homology. Conserved domain analysis of isoform *a*, showed two types of proteolytic domains: 1) the first two domains belong to the cysteine protease super family, 2) two consecutive domains are similar as the C2 like sub domain III of superfamily. It is interpreted from the above observation that *calpain10a, c* and *g* isoforms may have a role of cysteine protease because of their domain I & II similarity with the protease superfamily domains. Further, Domains III and T of *calpain 10a* have similarity with C2 like subdomain III of mu-like calpain (*calpain1*). Calpains 1 and 2 both have multiple known targets for pathology of diseases.

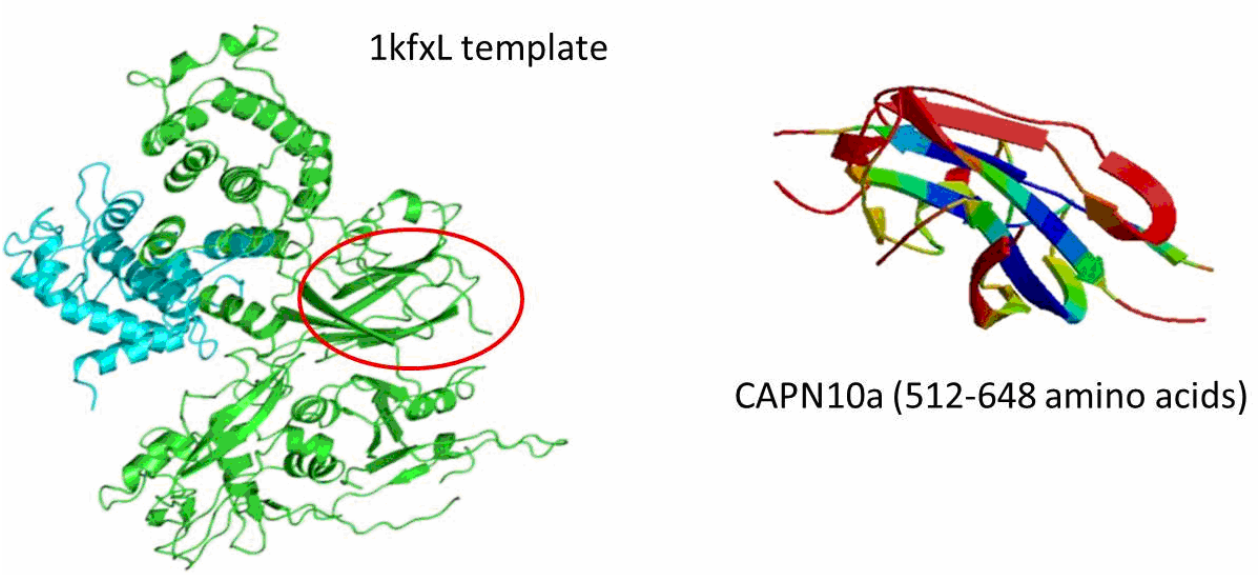

Fig. 5: 3D Structure of Calpain 10 (CAPN10) isoform *a:* created from Swiss model-ExPASy (http://swissmodel.expasy.org). PDB template 1kfxL hybridized 27% with CAPN10*a*. Estimated per-residue inaccuracy visualized using a color gradient from blue (more reliable regions) to red (potentially unreliable).

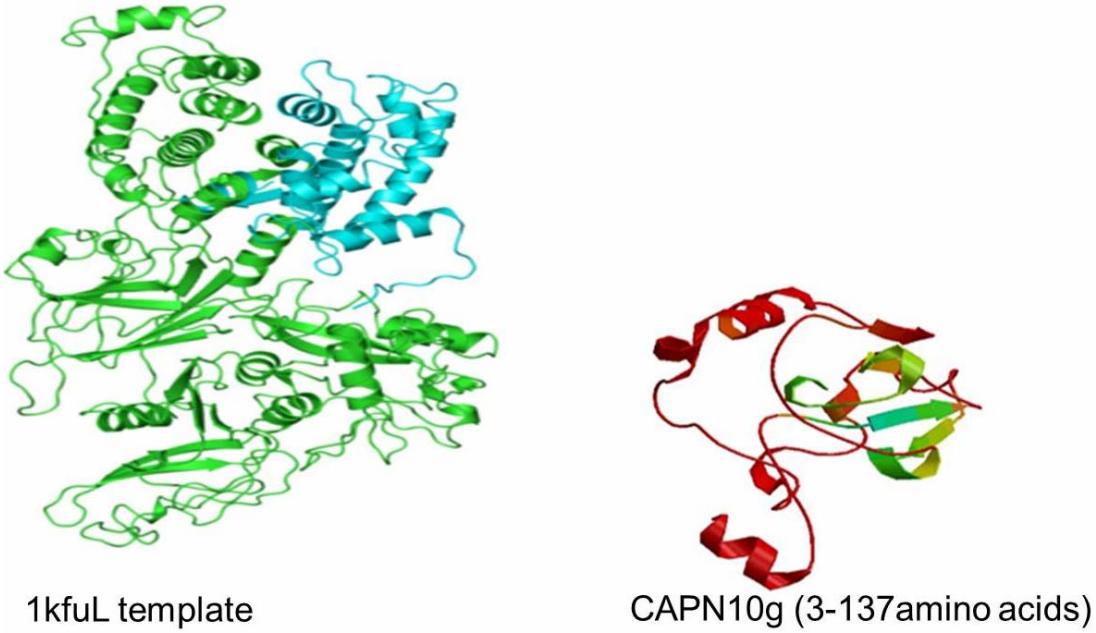

Fig. 6: 3D Structure of Calpain 10 (CAPN10) isoform *g* created from Swiss model-ExPASy (http://swissmodel.expasy.org). PDB template 1kfxL hybridized 26.3% with CAPN10*g*. Estimated per-residue inaccuracy visualized using a color gradient from blue (more reliable regions) to red (potentially unreliable).

The active sites of the proteins depend on the three dimensional structure and ionic bonds. Our analysis for these three isoforms sequence, through Swiss-model repository hybridized with already existing sequences in databases which served as templates. The templates were experimentally proven peptides and were of four different kinds. These were 1qxp (a mu-like calpain), 2p0r (structure of human calpain 9 in complex with leupeptin) 1kfxL (structure of human m-calpain form I) and 1kfuL template (a crystal structure of human m-calpain form II). Overall sequence identity was not significant as portion of the isoform had similarity with templates.

The three isoforms of *calpain 10* showed structural similarities with the n-terminal end of the members of typical calpain family (*m, mu and calpain 9*). Among these, m and mu are the classical members while calpain 9 is an atypical calpain. This is because of having T-

domain (150 amino acids) which covers 22% of the total sequences of the protein. The first 320 amino acids of *Calpain 10* isoform *a* showed 32% identity to 2p0r template which is a part of calpain 9 and 500 amino acids showed 29% sequence identity to the 1qxp template which is a portion of mu-calpain. A stretch of (137 amino acids) starting from 512-649 amino acids showed structural identity of 27% with 1kfxL sequences. The protein folding looks quite similar with calpain 9, which is involved in causing peptic ulcer. Further, greater than 50% down regulation have been reported in rats' heart during cardiac hypertrophy.

Conserved domain analysis of this stretch (137 amino acids) starting from 512-649 amino acids showed its similarity with C2 like subdomain III which plays a key role in calcium induced activation of calpain and facilitates its interaction with phospholipids resulting in translocation to cytoplasmic/ nuclear membrane. Swiss repository structural analysis of *calpain10c* isoform showed its similarity with two template peptides, 1qxp (mu- calpain) and 2p0r (Calpain 9), which are similar as shown by *calpain 10a*. It lacks the second domain similar to C2 like domain III which is present in *calpain10a*.

Based upon observed homology found with low z scores in modeling indicate week representation of the overlay functions from domains I, II and III of these three isoforms. It was inferred that *calpain 10* may have little involvement in similar functions under normal physiological state. Normal expression and function of *calpain10* isoforms especially *a* and *c* seems essential for the normal glucose metabolism and insulin receptor activation at times of low calcium availability when *m, mu* and *calpain 9* are in repressed condition. *Calpain10* is also reported to regulate purified protein kinase C activity in vitro (Bollag *et al*., 1986). A lower expression level of *calpain 10* gene because of variants is consistence with insulin levels. This represents its regulatory role in pancreatic beta cells for insulin secretion and apoptotic activity in these pancreatic cells. This can lead to up regulation of protein kinase C activity which directly phosphorylate serine residues of insulin receptor resulting in the reduction of their tyrosine kinase activity causing reduced insulin signaling, and develop insulin resistance. Down regulation of protein kinase C activity is an important factor in maintaining proper phosphorylation of insulin receptors. The insulin resistance was reversed with the addition of PKC inhibitors in insulin resistance tissues such as fat and skeletal muscles (Itani *et al*., 2000, Griffin *et al*., 1999, Muller *et al*., 1991, Maegawa *et al*., 1991, Caro *et al*., 1987). Low levels of *calpain10* have been found associated to cholesterol levels, metabolic syndrome and polycystic ovary syndrome (Saez *et al*., 2007). Therefore increasing the *calpain 10a* and *c* level carefully, may also reverse the level of protein kinase C and hence lowering of insulin resistance or increasing insulin sensitivity. Further binding of isoform *a* with phospholipids may lead to better insulin secretion as well. Similarity of both isoforms *a* and *c* with *calpain 9* indicate its involvement in functions of stomach and small intestine as of *calpain 9*, being exist at both locations in humans. Lower expression of *calpain 9* is involved in developing gastric cancer (Davis *et al*., 2007). Therefore absence of these two isoforms may lead to developing similar pathologies in stomach and small intestine. However bringing their levels to normal as of healthy person may have a therapeutic potential. Therefore by increasing the level of these isoforms in stomach may be helpful in controlling the diabetes complication of stomach.

Structural analysis of *Calpain 10g* the shortest protein showed 26% identity (Sequences from 3-137 amino acids) with 1kfxL template which is *m-calpain* form *I* and *II*. The *m-Calpain* form *II* is typical calpain and is involved in cardiovascular disease, acute neurological injury, and Alzheimer and cataract formation. Mu-calpain is associated with the positive regulation of cell proliferation and proteolysis (Randriamboavonjy and Fleming 2009). Increased activity of mu calpain due to elevated platelet concentration in diabetes result in hyper agree ability leading to prothrombotic state (Dean 2010). Whereas, *m-calpain* is known to have association with protein auto-processing, responding to hypoxia, proteolysis and myoblast fusion. Therefore isoform *g* falling in the twilight zone of homology may play a similar but week role in pathogenesis as predicted to exhibit protease like activity only because of having merely papain-like domain.

Normal expression and function of *calpain 10* isoforms especially *a*, and *c* looks essential for the normal glucose metabolism and insulin receptor activation. These isoforms belong to the atypical calpain family which does not need excessive Ca^{+2} ions and they require only trace amount of calcium for their activity. Therefore at the time of low calcium availability, they regulate PKC driven insulin receptor activation and their down regulation may cause diabetes and may favors the environment of co-morbid complexity. *Calpain-1* and *calpain-2* show optimal activities when present in a hetrodymeric forms with a small subunit of 30 kDa. Structurally *calpain 10* is less likely to interact with the small subunit because of its non hetrodymeric existence. The domain IV of general calpains is a calmodulin-like domain that is critical for heterodimerization, is absent in calpain10 (Lebovitz 2004, Adams and Norwood 2003, Horikawa 2000, Dupree and Meyer 1980). Instead there is domain T which makes it atypical form of calpain and its function is not known yet. Therefore our results raise the possibility that *calpain10* might occur at specific intracellular locations such as membrane bound or become functional under specific physiological conditions such as oxidative or low

inflammatory situation when its regulatory proteins are available.

CONCLUSION

Our data concluded that form *a and c* have low level similarity with m, mu and *calpain 9*. Therefore these isoforms may compensate similar functional activities at the time when template levels are down in the cells. The presence of C2 like subdomain III of mu-like calpain (calpain 2) in calpain 10 isoforms may plays a vibrant role in calcium induced activation of their own or other proteins. Isoform *g*, the shortest protein, is predicted of having papain-like domain and exhibiting protease like activity only. Further carefully addressed in vivo studies on animal models or cell lines are required to determine predicted mechanism of action of these isoforms. It will help to establish the treatment modalities in diabetes and its complications by adjusting the levels of *calpain 10* isoforms, because generally caplains over activation has been linked to sever diseases which may influence the therapeutic correction of diabetes complications.

REFERENCES

Adams MIH and Norwood N (2003). Core Concepts in Pharmacology. Prentice Hall, Englewood Cliffs, NJ, USA. ISBN 0-13-089329-3.

Baier LJ, Permana PA, Yang X, Pratley RE, Hanson RL, Shen GQ, Mott D, Knowler WC, Cox NJ, Horikawa Y and Oda N (2000). A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J. Clin. Invest.,* **106**(7): R69-73.

Bollag GE, Roth RA, Beaudoin J, Mochly-Rosen D and Koshland DE (1986). Protein kinase C directly phosphorylates the insulin receptor in vitro and reduces its protein-tyrosine kinase activity. *Proc. Natl. Acad. Sci.* USA., **83**(16): 5822.

Caro JF, Sinha MK, Raju SM, Ittoop O, Pories WJ, Flickinger EG, Meelheim D and Dohm GL (1987). Insulin receptor kinase in human skeletal muscle from obese subjects with and without noninsulin dependent diabetes. *J. Clin. Invest,* **79**(5): 1330-1337.

Davis TL, Walker JR, Finerty PJ Jr, Mackenzie F, Newman EM and Dhe-Paganon S (2007). The crystal structures of human calpains 1 and 9 imply diverse mechanisms of action and auto-inhibition. *J. Mol. Biol.,* **366**(1): 216-29.

Dean WL (2010). Role of platelet plasma membrane Ca-ATPase in health and disease. *World J. Biol. Chem*., **1**(9): 265-70.

Dupree EA and Meyer MB (1980). Role of risk factors in complications of diabetes millitus. Am. J. Epidemiol., **112**(1): 100-112.

Gafni J, Hermel E, Young JE, Wellington CL, Hayden MR and Ellerby LM (2004). Inhibition of calpain cleavage of huntingtin reduces toxicity: Accumulation of calpain/caspase fragments in the nucleus. *J. Biol. Chem*., **279**(19): 20211-20220.

Griffin ME, Marcucci MJ, Cline GW, Bell K, Barucci N, Lee D, Goodyear LJ, Kraegen EW, White MF and Shulman GI (1999). Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. *Diabete*s., **48**(6): 1270-1274.

Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B,Spielman RS, Gogolin-Ewens KJ and Shephard JM (1996). A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat. Genet*., **13**(2): 161-166.

Horikawa Y (2002). Calpain 10 as a susceptibility gene of type 2 diabetes. Nihon Rinsho. *JPN. J. CLIN. MED.*, **60**(Suppl7): 489-497.

Horikawa Y (2005). Type 2 diabetes and genetic variations of calpain 10 gene. Nihon Rinsho. *JPN. J. CLIN. MED*., **63**(Suppl 2): 155-9.

Horikawa Y (2006). Calpain-10 (NIDDM1) as a Susceptibility Gene for Common Type 2 Diabetes. *Endocr J*., **53**(5): 567-576.

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE and del Bosque-Plata L (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.,* **26**(2): 163-75.

Horikawa Y, Oda N, Yu L, Imamura S, Fujiwara K, Makino M, Seino Y, Itoh M and Takeda J (2003). Genetic variations in calpain-10 gene are not a major factor in the occurrence of type 2 diabetes in Japanese. *J. Clin. Endocrinol. Meta*b., **88**(1): 244-247.

Huang Y and Wang KK (2001). The calpain family and human disease. *Trends. Mol. Med.,***7**(8): 355-62.

Itani SI, Zhou Q, Pories WJ, MacDonald KG and Dohm GL (2000). Involvement of protein kinase C in human skeletal muscle insulin resistance and obesity. *Diabetes,* **49**(8): 1353-1358.

Kiefer F, Arnold K, Künzli M, Bordoli L and Schwede T(2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res*., **37**(Database issue): D387-92.

Kopp J and Schwede T (2004). Automated protein structure homology modeling: A progress report. *Pharmacogenomics.*, **5**(4): 405-16.

Kopp J and Schwede T (2004). The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.*, **32**(Database issue): D230-4.

Lebovitz HE(2004). Therapy for Diabetes Mellitus and Related Disorders, 4th ed., American Diabetes Association, Alexandria, Va, p.543.

Maegawa H, Shigeta Y, Egawa K and Kobayashi M (1991). Impaired auto-phosphorylation of insulin

receptors from abdominal skeletal muscles in nonobese subjects with NIDDM. *Diabetes*, **40**(7): 815-819.

Marchler-Bauer A and Bryant SH (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.*, **32**(Web Server issue): W327-31.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M and He S (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.*, **37**(Database issue): D205-10.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR and Gwadz M (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*., **39**(Database issue): D225-9.

Müller HK, Kellerer M, Ermel B, Mühlhöfer A, Obermaier-Kusser B, Vogt B and Häring HU (1991). Prevention by protein kinase C inhibitors of glucose-induced insulin-receptor tyrosine kinase resistance in rat fat cells. *Diabetes*, **40**(11): 1440-8.

Randriamboavonjy V and Fleming I (2009). Insulin, insulin resistance, and platelet signaling in diabetes. *Diabetes Car*e, **32**(4): 528-530.

Sáez ME, Martínez-Larrad MT, Ramírez-Lorca R, González-Sánchez JL, Zabena C, Martinez-Calatrava MJ, González A, Morón FJ, Ruiz A and Serrano-Ríos M (2007). Calpain-5 gene variants are associated with diastolic blood pressure and cholesterol levels. *BMC Medical Genetics*, **8**(1): 1.

Sboohi K, Bano S and Chaudhry B (2013). Identification of Calpain 10 Isoforms (b, d, e, f & h) Conserved Regions and Possible Functional Prophecy through Bioinformatics. *Journal of Basic & Applied Sciences*, **9**: 489-495.

Stumvoll M, Fritsche A, Madaus A, Stefan N, Weisser M, Machicao F and Häring H (2001). Functional significance of the UCSNP-43 polymorphism in the CAPN10 Gene for proinsulin processing and insulin secretion in nondiabetic Germans. *Diabetes*, **50**(9): 2161-2163.

Turner MD, Cassell PG and Hitman GA (2005). Calpain-10: from genome search to function. *Diabetes Metab. Res. Rev.*, **21**(6): 505-14.

Zhang Z, Schwartz S, Wagner L and Miller Webb (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, **7**(1-2): 203-214.