Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue3, 2014, pp183-196
http://www.bipublication.com

In Silico Analysis of Human Calreticulin and Its Relationship with MHC Class I

Sérgio Suzart dos Santos[1*], Aldifran Ferreira da Silva [1,2], Rosimary de Jesus Gomes Turri[1], and José de Ribamar Sousa da Silva [1,3]

[1]Medicine Course at Ceuma University (UniCeuma), São Luís, Maranhão, Brazil
[2]Department of Morphology, Federal University of Maranhão (UFMA), São Luís, Maranhão, Brazil
[3]University Hospital of the Federal University of Maranhão (HUUFMA), São Luís, Maranhão, Brazil
[*]Corresponding author: Email: sergiosma@hotmail.com; Tel: +5598-32144233

[Received 11/10/2014, Accepted-25/10/2014]

ABSTRACT:

Calreticulin is member of a family of endoplasmic reticulum chaperones that fold newly synthesized polypeptides. This protein contains an N-terminal signal sequence and a C-terminal KDEL ER retrieval sequence responsible for targeting and retention of protein in the ER lumen. Computational analysis has become an indispensable bioinformatic approaches for the characterization of proteins regarding the physicochemical properties, prediction of signal peptides, protein-protein interaction network, among other in silico predictions. The results of the computational analysis revealed in this study corroborate those found in experimental studies previously conducted with human calreticulin, since the algorithms used in this work can be safely used to predetermine physicochemical and structural properties of protein analyzed in our study. The present work is aimed towards the theoretical basis of the physicochemical, structural and functional proprieties of calreticulin, using online computational tools usually utilized in in silico analysis, mainly because to date we have not found any computational analysis evaluating such parameters in this protein. The physicochemical properties of the human calreticulin was analyzed by using ExPASy's ProtParam tool and it was found that the molecular weight is 48,141.5 Da, its isoelectric point (pI) was found to be acidic protein. The instability index infers that this protein is unstable. Secondary structure prediction showed that random coil dominated all the other conformations. To infer the interactions of calreticulin with other proteins, we used STRING database v9.1. Using the tool STRING determined that the network of protein-protein interactions and could be found in human calreticulin interactions with various proteins, particularly those belonging to the class I MHC. From a scientific point of view, the results of protein-protein network analysis have clearly shown the interaction of human calreticulin with proteins involved in the later stages of class I assembly and antigen-presenting function of MHC Class I. Our study, to the best of our knowledge, demonstrates for the first time a set of results measured using computational analysis to assess structural and physicochemical nature of one the most important molecular chaperones of the endoplasmic reticulum envolved in a number of cellular processes and currently investigated its role in autoimmunity.

Keywords: in silico analysis, protein-protein interaction, physicochemical properties, secondary structure analysis, human calreticulin protein, human leukocyte antigen (HLA),

[I] INTRODUCTION

The endoplasmic reticulum (ER) is an organelle securely holds secrets beyond the biosynthesis of lipids, sterols and secreted proteins, but of sophisticated mechanisms of posttranslational

modification and trafficking that occur in a variety of secreted proteins and those associated with the cell membrane. Classically, transmembrane proteins and those secreted are translocated to the inner ER to undergo modifications in their original structure, assuming structural configurations that make them functional to be exported to the cytoplasm or released on the cell surface via secretory pathways. It is not surprising that for appropriate posttranslational folding to occur it is necessary to have the effective participation of chaperones residing in the ER, capable of assuring the quality of a recently synthesized protein [1,2].

Some lines of evidence suggest that calreticulin (CRT) is a ubiquitous intracellular chaperone protein found in most nucleated cells and commonly found in the lumen of the ER, where it has the principal role of regulating calcium stores, actively participating in the mechanism of signaling and protein processing. In addition, some studies point to the involvement of CRT in antigen presentation via class I MHC, because it participates in the formation of the loading complex, together with transporters associated with antigen processing (TAP), Erp57 (thiol-disulfide oxidoreductase) and tapasin [3-5].

An aspect of CRT yet to be unraveled is its probable participation in a gamut of autoimmune processes, including inhibition of complement activation, molecular mimicry, epitope spreading and the release of inflammatory mediators. Some reports point to the presence of anti-CRT antoantibodies in patients with Sjorgren's syndrome, systemic lupus erythematosus and rheumatoid arthritis [6-8].

In the last years, we have seen the emergence of computational methods that have been developed for predicting the primary, secondary and tertiary structures of proteins, as well as functional analyses, reducing the time needed to conduct experiments and allowing the more rapid acquisition of results. As far as physicochemical and structural characterizations of a protein, there is no doubt that in silico approaches help resolve these problems.

The aim of the present work was to extend the theoretical basis of the physicochemical, structural and functional proprieties of calreticulin, using online computational tools usually utilized in in silico analysis, mainly because to date we have not found any computational analysis evaluating such parameters in this protein.

[II] MATERIALS AND METHODS

2.1. Sequence retrieval

The complete amino acid sequence of calreticulin of Homo sapiens was retrieved from the Entrez protein database available at NCBI (http://www.ncbi.nlm.nih.gov/) with the accession number AAB51176. The length of the CRT protein was 417 amino acids.

2.2. Physicochemical properties analysis

For physicochemical characterization, molecular weight, isoelectric point (pI), total number of positive and negative residues, extinction coefficient [9, 10], instability index [11], aliphatic index [12], half-life [13] and grand average of hydropathy (GRAVY) [14] were assessed by Expasy's ProtParam server (http://web.expansy.org/protparam).

2.3. Protein signal peptide prediction

Determination of the presence of a signal peptide in the amino acid sequence was performed by Signal-3L server [15] (http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/).

2.4. Prediction of transmembrane regions

The TMPRED [16] (http://www.ch.embnet.org/software/TMPRED_form.html) and TMAP [17] (http://emboss.bioinformatics.nl/cgi-bin/emboss/tmap) servers were used to predict and analyze the transmembrane domains of CRT.

2.5. Prediction of subcellular localization of protein

The subcellular localization of human CRT was predicted employing CELLO v.2.5 (subCELlular LOcalization predictor) [18]

(http://cello.life.nctu.edu.tw/) and Hum-mPLoc 2.0 [19] (http://www.csbio.sjtu.edu.cn/bioinf/hum-multi-2/).

2.6. Prediction of disulfide bridges

The presence of disulfide bridges was analyzed using both the CYS-REC tool (linux1.softberry.com/berry.phtml?topic=cys_rec &group=programs&subgroup=propt) and DiANNA (DiAminoacid Neural Network Application) webserver prediction program [20] (http://clavius.bc.edu/~clotelab/DiANNA/), which predict the positions of cysteines, total number of cysteines present and the most probable bonding patterns between available cysteine residues.

2.7. Post-translational modification analysis

The NetPhos 2.0 server [21] was used for predictions of kinase-specific serine, threonine, and tyrosine phosphorylation sites (http://www.cbs.dtu.dk/services/ NetPhosK/). Potential N-, C- and O-glycosylations were predicted by the GlycoEP prediction server [22] (http://www.imtech.res.in/raghava/glycoep/submit_advance.html).

2.8 Secondary structure analysis

The SOPMA [23] (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), GOR IV [24] (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) and PREDATOR [25] (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_predator.html) algorithms were used to predict the secondary structure contents in full length human CRT.

2.9 Prediction of protein-protein interaction network

To explore possible relationship between CRT and any known proteins, we turned to the STRING database v9.1 (Search Tool for the Retrieval of Interacting Genes/Proteins) [26] (available at http://string-db.org/). STRING is a database of known and predicted protein interactions, which incorporates protein-protein interaction (PPI) information from a number of widely used interaction databases [27].

[III] RESULTS AND DISCUSSION

3.1. Retrieved sequence

CRT is a 46-kDa protein expressed in all cells of higher organisms, which exhibits a gene structure composed of nine exons and eight introns localized on chromosome 19 [6]. The sequence of human CRT used in this report was retrieved from the NCBI (National Center for Biotechnology Information) protein database in FASTA format under accession number AAB51176.

3.2. Primary protein sequence analysis

The physicochemical parameters of human CRT were predicted by using the ProtParam tool on ExPASy (the Expert Protein Analysis System). The predicted molecular weight of this protein is calculated as 48,141.5 Da, although it migrates with an apparent molecular mass of 60 kDa on SDS-PAGE [28].

The amino acid residue composition determined is given in Table 1, and the parameters molecular weight, isoelectric point (pI), total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, half-life and GRAVY were determined and tabulated as shown in Table 2.

The analysis of the composition of the amino acid sequence revealed that human CRT has a higher content of acidic amino acids (negatively charged: Asp, Glu) in relation to the number of basic amino acids (positively charged: Lys, Arg, His). With regard to the distribution of hydrophilic and hydrophobic amino acids in this protein, we determined a high hydrophilicity due to high content of polar residues (66.2%). In this work, the GRAVY index was negative, indicating that CRT interacts more efficiently with water, like a protein of hydrophilic nature. It is worth highlighting that this index indicates the water-solubility of proteins: a positive GRAVY value denotes that the protein is hydrophobic, while a negative value indicates hydrophilic [14].

The isoelectric point (pI) is the pH at which the surface of protein is covered with charge, but the net charge of the protein is zero [29]. The computed pI value revealed that human CRT is an acidic protein (pI 4.29), which allows its separation by anion-exchange chromatography [30]. Fliegel et al. [31] found a pI of 4.14 for CRT from rabbit skeletal muscle sarcoplasmic reticulum. With respect to the aliphatic index, computational analysis showed a predictive value of 59.14, revealing a slight thermostability of the protein, which is involved in a range of biological processes.

It has become clear that the aliphatic index is a measure of the relative volume occupied by aliphatic side chains (Ala, Val, Leu and Ile) [12]. A protein whose instability index is less than 40 is predicted to be stable [32]. Accordingly, the theoretically predicted value found in this study was above 40, indicating that the evaluated protein is unstable. Other physicochemical properties determined for human CRT are summarized in Table 2.

3.3. Protein signal peptide prediction

Secreted and cell-surface proteins possess an N-terminal signal peptide ranging between 15 and 30 amino acids long, where this sequence has a stretch of hydrophobic amino acids that is cut off during translocation of the protein across the membrane [33]. Using the computer analysis on the signal 3L server, the presence of a signal peptide and the location of the cleavage site of the human CRT sequence was predicted to be the first 17 amino acids (MLLSVPLLLGLLGLAVA) followed by a signal-peptidase cleavage site (between position 17 and 18). A study using the Epstein-Barr virus-transformed human Wil-2 B-cell line revealed that CRT has a 17-amino acid hydrophobic leader segment that is not present in the purified mature protein [6]. Regardless of the source of CRT (human, rabbit, mouse and other), the predicted results found in this report by computational analysis corroborate previous experimental studies [6, 34].

3.4. Prediction of transmembrane regions

In this study, we used two algorithms designed to predict transmembrane domains (TMpred and TMAP). TMpred algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins [16], while TMAP is a computational method for predicting transmembrane segments from multiple sequence alignments [17].

According to the TMpred analysis, there is one possible inside to outside helix (i→o) located from 5 to 21 residues and one possible outside to inside helix (i→o) identified from 1 to 17 residues (Figure 1). Additionally, computational analysis also revealed that most of this protein has a long hydrophilic chain located in the cytoplasm.

Another computational method used in this study to predict transmembrane domains was TMAP, which identified one putative transmembrane segment, which includes amino acid positions 4 to 25 (Figure 2).

It is not surprising that the positions and numbers of amino acids varied according to the algorithm used in this study. This is probably due to the fact that while the TMpred uses a combination, TMAP utilizes an algorithm to identify transmembrane segments of aligned sequences.

3.5. Prediction of subcellular localization of protein

The human cell is organized into several compartments of different cellular processes. Computational prediction of subcellular localization provides a quick and inexpensive means of gaining insight into protein function. In our study, we decided to use two computational tools (CELLO v.2.5 and Hum-mPLoc 2.0), capable of determining the highest probably of finding a protein in a particular cellular compartment.

As seen in Table 4, CELLO suggested that the subcellular localization of human CRT could be the endoplasmic reticulum, which was corroborated by the Hum-mPLoc tool. Data provided by computational analysis confirm the

results previously observed in experiments carried out with indirect immunofluorescence and cultured cells from a wide variety of tissues, demonstrating that CRT is located in the endoplasmic reticulum [35].

3.6. Prediction of disulfide bridges

Disulfide bonds in proteins are formed between the thiol groups of cysteine residues, which play an important role in the folding and stabilization of protein structure [36]. Keeping this in mind, prediction of disulfide bonds by CYS-REC tool showed that human CRT contains three cysteines at positions 105, 137 and 163. When disulfide bond prediction was performed using DiANNA, an optimal disulfide bond reliability was found between Cys-105 and Cys-137 (score = 0.99969), between Cys-105 and Cys-163 (score =0.99926) and between Cys-137 and Cys-163 (score = 0.9964), and Cys-105 showed a tendency to form an intramolecular disulfide bridge with Cys-137 (Cys-105-Cys-137, found in the N-terminal domain).

Three cysteine residues in the N-terminal domain are conserved in the majority of CRT protein sequences, but the number and location of disulfide bonds vary between the samples analyzed, as observed in the studies performed in human placenta (Cys-88 and Cys-120 residues) [37], bovine brain (Cys-120 and Cys-146) [38], and barley CRT (Cys-136 and Cys-162) [39].

3.7. Post-translational modification analysis

Protein glycosylation is an important posttranslational modification process that protects glycoproteins against proteolytic degradation, and it is involved in various cellular functions, such as intracellular trafficking, protein folding and cell-cell interactions, as well as other cellular events [40]. It has been reported in a few studies that CRT is glycosylated in hamster ovary [41] and bovine brain [38], but CRT glycosylation does not occur in lymphoid cells from mouse or human [6, 42].

In this report, potential serine, threonine and tyrosine phosphorylation sites were identified using the prediction NETPhos 2.0 servers, which recognized twenty-three phosphorylation sites successfully predicted by the neural network, which were distributed as follows: eleven predicted serine (residues 40, 53, 85, 116, 126, 189, 195, 214, 231, 300 and 304), four predicted threonine (residues 68, 96, 141, and 181), and eight predicted tyrosine (residues 109, 182, 271, 285, 299, 306, 308, and 338) phosphorylation sites. The results can also be interpreted from the graphical representation of the serine, threonine and tyrosine phosphorylation sites (Figure 3). As described previously in some studies, phosphorylation is an event that can happen at multiple sites in a protein [43], where it is an essential regulatory process in the cell cycle and activity of transcriptional factors [44, 45].

Using GlycoEP prediction server, we found that human CRT is predicted to be a glycoprotein with five putative N-glycosylation sites located at Asn-102, Asn-154, Asn-188, Asn-310 and Asn-344. A previous study by Matsuoka et al. [38] revealed that bovine CRT showed a single glycosylated amino acid residue, Asn-162, in contrast to the results shown here for CRT from human fibroblasts. On the other hand, the results obtained for the prediction of O-glycosylation sites showed that there were four potential sites at Ser-69, Thr-229, Ser-231 and Ser-323. No potential site was predicted for C-glycosylation (corresponding to the amino acid tryptophan). In previous studies [6, 46], no evidence of glycosylation or phosphorylation of human CRT was observed. We believe that the results of these studies were contradictory to those found in the present study due to the different tissues or species used to isolate the CRT protein.

3.8. Secondary structure analysis

The primary structure of a given protein tends to form regular formations, resulting in the secondary structure. The secondary structure of protein is intrinsically related to the bridges of hydrogen bonds that occur between the backbone amide and carboxyl groups.

SOPMA (Self-Optimized Prediction Method with Alignment), GOR IV (Garnier-Osguthorpe-Robson method) and PREDATOR online servers were used to predict the secondary structure, and the percentages of alpha helices, beta sheets, extended strands and random coils of human CRT are presented in Table 3. Regardless of the algorithm used in this study, our result revealed that random coils dominated among secondary structure elements followed by alpha helix and extended strand, but beta turns were exclusively predicted by the SOPMA method. Interestingly, the predicted secondary structure of the CRT from the ant Polyrhachis vicina Roger (Hymenoptera: Formicidae) using the SOPM algorithm [47] exhibited a high number of random coils (71.63%), followed by extended strands (14.66%) and alpha helices (13.70%). This observation is consistent with our computational analysis data, given that whether from ant or human fibroblasts, CRT in both cases showed a predominance of random coils.

It is interesting to note that if a particular protein contains amino acids such as isoleucine or charged residues such as glutamic acid and aspartic acid in its backbone, repulsion forces between these groups enable the polypeptide chain to assume a random coil configuration due to regions of proteins that are not identifiably organized as helices or pleated sheets [48]. As previously seen in Table 1, the protein analyzed in our study had a slightly higher content of charged amino acids, which may contribute to the high number of random coils predicted by different computational algorithms used.

3.9. Prediction of protein-protein interaction network

As we know, a considerable number of molecules described such as proteins, carbohydrates, lipids, nucleic acids and drugs, among others, are able to interact with other molecules and produce specific biological functions essential for the maintenance of the cell. Some interactions can produce complex networks represented by nodes (e.g., metabolites and macromolecules) and edges (physical, biochemical and functional interactions between the molecules). In this type of network, all kinds of the molecular interactions can be referred to as an interactome. For the systematic analysis of the CRT interactome, the STRING database was used for the purpose of recognizing probable predicted protein–protein associations comprising known and predicted physical and functional associations derived from genomic context, high-throughput experiments, conserved co-expression, and previous knowledge. To avoid spurious interactions in our data set, we considered only the hits with the number of interactors set at 20 and high stringency level with a confidence score of 0.7.

In our predicted high-confidence interaction network, nine proteins identified in this study belong to the repertoire of MHC class I subtypes. It should be noted that in interactome analysis, we find a predominance of HLA-B belonging to the alleles HLA-B57, HLA-B*3501, HLA-B41, HLA-B*15 and HLA-Cw, whose associations between these types and particular diseases have been revealed in studies in patients with hepatitis [49, 50], inflammatory bowel disease [51, 52], AIDS [53, 54] and autoimmune diseases [55, 56]. If we look at, for example, the probable interrelation between HLA-B57 and certain diseases, there are some works suggesting that this type of HLA has a protective effect in individuals infected with HIV, since these patients can display low viremia and high CD4 counts, as well as mild or no symptoms during infection with HIV-1 virus [54, 57].

Taken together, the results of the interactome analysis strongly indicated interactions of human CRT with class I HLA, which is not surprising, since CRT plays important roles in the folding of MHC class I molecules. As we can see in Figure 4, proteins involved in the assembly process of MHC class I molecules, such as calnexin (CANX), tapasin (TAPBP), protein disulfide isomerase family A (PDIA3) and beta2-microglobulin (B2M), are integrative elements of the CRT

network. Other proteins involved in metabolic processes are part of this network, but are not discussed because they are not intrinsically related to class I MHC.

[IV] CONCLUSION

The present work emphasizes on the study of physicochemical properties and structural analysis of human calreticulin. Primary structure analysis has revealed that this chaperone is an acidic, unstable and hydrophilic protein. In addition, it was possible to find a predicted signal peptide in the first 17 amino acids and recognized that most of calreticulin has a hydrophilic chain located in the cytoplasm. In our study the prediction of disulfide bonds using different algorithms was able to identify three cysteines located in the N-terminal domain and, at the same time, an independent approach identified several potential glycosylation sites. Secondary structure analysis infers that random coil is the most predominating structure in our protein studied. Finally, in our predicted high-confidence interaction network, we observed a strong correlation between human calreticulin and some subtypes belonging to the repertoire of MHC class I.

ACKNOWLEDGEMENTS

We are highly thankful Dr Albert Leyva for their cordial advice and suggestions.

REFERENCES

1. Kadowaki H, Nishitoh H (2013). Signaling pathways from the endoplasmic reticulum and their Roles in Disease. Genes, 4:306-333.
2. Lewis MJ, Mazzarella RA, Green M (1985). Structure and assembly of the endoplasmic reticulum: The synthesis of three major endoplasmic reticulum proteins during lipopolysaccharide-induced differentiation of murine lymphocytes. J. Biol. Chem, 260:3050-3057.
3. Ortmann B, Androlewicz M, Cresswell P (1994). MHC class I/ß2-microglobulin complexes associate with TAP transporters before peptide binding. Nature, 368:864–867.
4. Ortmann B, Copeman J, Lehner PJ et al (1997). A critical role for tapasin in the ssembly and function of multimeric MHC class I-TAP complexes. Science, 277: 1306-1309.
5. Sadasivan B, Lehner PJ, Ortmann B, Spies T, Cresswell P (1996). Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. Immunity, 5:103–114.
6. McCauliffe DP, Lux FA, Lieu TS et al (1990) Molecular cloning, expression, and chromosome 19 localization of a human Ro/SS-A autoantigen. J, Clin. Invest., 85: 1379-1391.
7. Kishore U, Sontheimer RD, Sastry KN et al (1997). The systemic lupus erythematosus (SLE) disease autoantigen—calreticulin can inhibit C1q association with immune complexes. Clin. Exp. Immunol., 108:181-190.
8. Holoshitz J, De Almeida DE, Ling S (2010). A role for calreticulin in the pathogenesis of rheumatoid arthritis. Ann N Y Acad Sci.,1209:91-98.
9. Pace CN., Vajdos F, Fee L et al (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci.,11:2411-2423.
10. Gill SC. and von Hippel PH (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem., 182:319-326.
11. Guruprasad K, Reddy BVB, Pandit, MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng., 4:155-161.
12. Ikai, A (1980). Thermostability and aliphatic index of globular proteins. J. Biochem., 88:1895–1898.

13. Bachmair A, Finley D, Varshavsky A (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science, 234:179-186.
14. Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157(1):105–132.
15. Shen H-B, Chou K-C (2007). Signal-3L: a 3-layer approach for predicting signal peptides. Biochem. Biophys. Res. Commun., 63:297-303.
16. Hofmann K and Stoffel W (1993). TMbase - A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler, 374:166.
17. Persson, B, Argos, P (1994). Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol., 237182–192.
18. Yu CS, Chen YC, Lu CH, Hwang JK (2006). Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics.64:643-651.
19. Shen H-B, Chou K-C (2007). Hum-mPLoc: an ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites. Biochem. Biophys.Res. Commun., 355:1006-1011.
20. Ferre F, Clote P (2005). DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res, 33:W230–W232.
21. Blom N, Gammeltoft S, Brunak S (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol, 294(5): 1351-1362.
22. Chauhan JS, Rao A, Raghava GPS (2013). In silico Platform for Prediction of N-, O- and C-Glycosites32. in Eukaryotic Protein Sequences. Plos One, 8(6):e67008.
23. Georjon C, Deleage G (1995). SOPMA: significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments. CABIOS, 11: 681–4.
24. Garnier J, Gibrat J-F, Robson B (1996). GOR secondary structure prediction method version IV. Meth Enzymol, 266:540-553.
25. Frishman D, Argos P (1996). Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. Protein Eng, 9:133-142.
26. Franceschini A, Szklarczyk D, Frankild S et al (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res., 41:808-815.
27. von Mering C, Huynen M, Jaeggi D et al (2003). STRING: a database of predicted functional associations between proteins. Nucleic Acids Res, 31:258-261.
28. Kishore U, Sontheimer RD, Sastry KN et al (1997). Release of calreticulin from neutrophils may alter C1q-mediated immune functions. Biochem. J., 322:543–550.
29. Bjellqvist B, Hughes GJ, Pasquali CH et al (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis, 14:1023-1031.
30. Golovchenko NP, Kataeva IA, Akimenko VK (1992). Analysis of pH-dependent protein interactions with gel filtration medium. J. Chromatogr. A , 591(1-2): 121–128.
31. Fliegel L, Burns K, MacLennan DH, Reithmeier RAF, Michalak M (1989). Molecular cloning of the high affinity calcium binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum. J. Biol. Chem., 264, 21522-21528.
32. Guruprasad K, Reddy BV, Pandit MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng., 4: 155–161.
33. Nakai, K (2000). Protein sorting signals and prediction of subcellular localization. Adv. Protein Chem., 54:277–344.

34. Waser M, Mesaeli N, Spencer C, Michalak M (1997). Regulation of calreticulin gene expression by calcium. J. Cell Biol., 138:547–557.
35. Ostwald TJ, MacLennan DH (1974). Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum. J. Biol. Chem. 249:974–979.
36. Sevier CS, Kaiser CA (2002). Formation and transfer of disulphide bonds in living cells. Nat. Rev. Mol Cell. Biol., 3: 836–847.
37. Højrup P, Roepstorff P, Houen G (2001). Human placental calreticulin. Eur. J. Biochem, 268:2558-2565.
38. Matsuoka K, Seta K, Yamakawa Y et al (1994). Covalent structure of bovine brain calreticulin. Biochem. J, 298:435–442.
39. Chen F, Hayes PM, Mulrooney DM, Pan A (1994) Identification and characterization of cDNA clones encoding plant calreticulin in barley. Plant Cell, 6: 835-843.
40. Haltiwanger RS, Lowe JB (2004). Role of glycosylation in development. Annu Rev Biochem., 73:491-537.
41. Jethmalani SM, Henle KJ, Kaushal GP (1994). Heat shock-induced prompt glycosylation. Identification of P-SG67 as calreticulin. J Biol Chem, 269:23603-23609.
42. Lewis MJ, Mazzarella RA, Green M (1985). Structure and assembly of the endoplasmic reticulum: The synthesis of three major endoplasmic reticulum proteins during lipopolysaccharide-induced differentiation of murine lymphocytes. J. Biol. Chem., 260:3050-3057.
43. Gunawardena J (2005). Multisite protein phosphorylation makes a good threshold but can be a poor switch. Proc. Natl. Acad. Sci. U.S.A, 102:14617–14622.
44. Komeili A, O'Shea EK (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science, 284:977–980.
45. Kim SY, Song EJ, Lee KJ, Ferrell JE Jr (2005). Multisite M-phase phosphorylation of Xenopus Wee1A. Mol Cell Biol, 25:10580–10590.
46. Denning, G.M., Leidal, K.G., Holst, V.A et al (1997). Calreticulin biosynthesis and processing in human myeloid cells: demonstration of signal peptide cleavage and N-glycosylation. Blood , 90:372-381.
47. Liu LP, Dang L, Xi GS, Wang F (2013). Molecular Cloning, Characterization and expression analysis of calreticulin gene in the ant Polyrhachis vicina Roger (Hymenoptera: Formicidae). Sociobiology, 60:355-361.
48. Smith LJ, Fiebig KM, Schwalbe H, Dobson CM (1996). The concept of a random coil Residual structure in peptides and denatured proteins. Folding & Design, 1:R95–R106.
49. Fanning LJ, Kenny-Walsh E, Shanahan F (2004). Persistence of hepatitis C virus in a white population: associations with human leukocyte antigen class 1. Hum Immunol, 65:745-51.
50. Tripathy AS, Shankarkumar U, Chadha MS, Ghosh K, Arankalle VA (2009). Association of HLA alleles with hepatitis C infection in Maharashtra, western India. Indian J. Med. Res, 130:550-555.
51. Adhiah AH, Jasim HA, Fadhil AM (2008). HLA antigens and inflammatory bowel disease in a sample of Iraqi patients. East Mediterr Health J., 14:1155-63.
52. Zhang H, Liu S, Liu Z, Li J (2008). Expression of iKIR-HLA-Cw in patients with inflammatory bowel disease. Life Sci J, 5:17-22.
53. Pereyra, F., Jia, X., McLaren, P.J. et al (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330:1551–1557.
54. Vaidya, SA, Streeck H, Beckwith N et al (2013). Temporal effect of HLA-B*57 on viral control during primary HIV-1 infection. Retrovirology, 10:139.

55. Cho WK, Jung MH, Choi E-J, Choi H-B, Kim T-G, Suh B-K (2011). Association of HLA alleles with autoimmune thyroid disease in Korean children. Horm Res Paediatr, 76:328-334.
56. Bahcetepe N, Kutlubay Z, Yilmaz E, TuzunY, ErenB (2013). The role of HLA antigens in the aetiology of psoriasis. Med Glas (Zenica), 10:339-342.
57. Altfeld M, Addo MM, Rosenberg ES et al (2003). Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS, 17:2581-2591.

Figure 1. The prediction of transmembrane domains by TMPRED server. X-axis represents protein length from N- to C-terminal end. Y-axis represents the score for transmembrane area (> 500 was considered significant) computed by TMPRED. Solid and dashed lines depict inside to outside and outside to inside orientations of the helices, respectively.

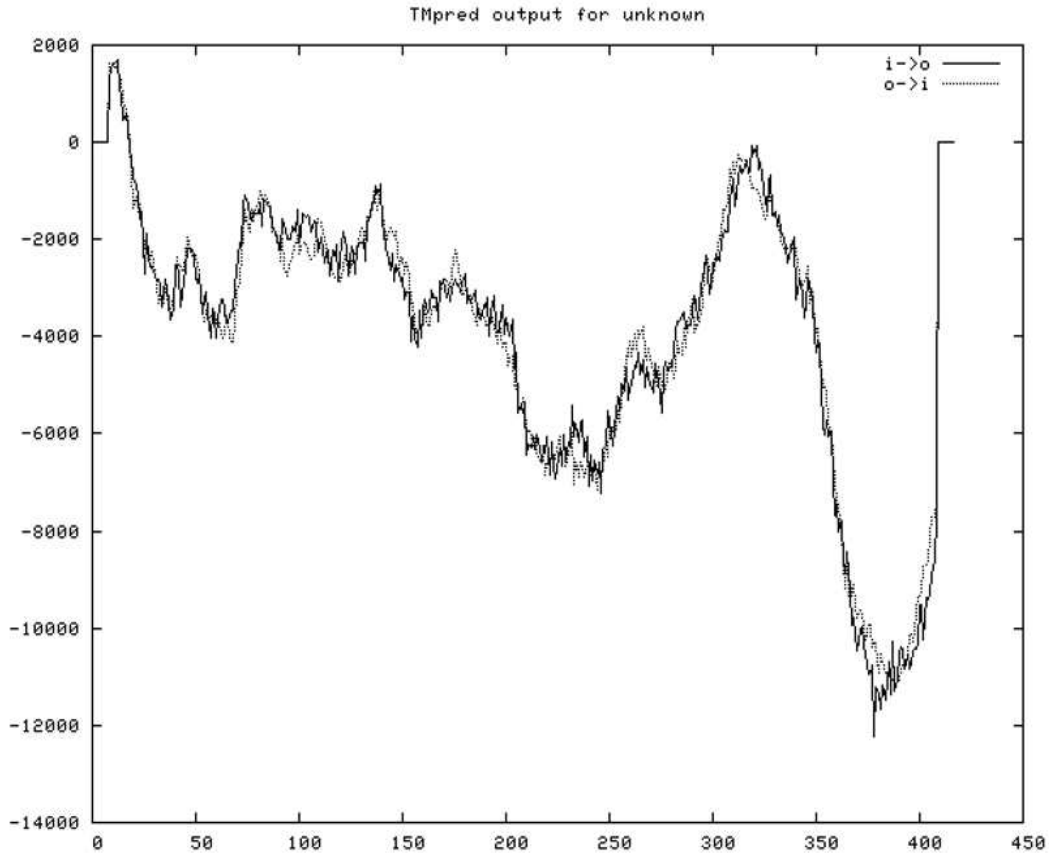

Figure 2. The prediction of transmembrane domains by TMAP server. X-axis represents protein length from N- to C-terminal end. Y-axis represents the score computed by TMAP algorithm. Black bar at the top indicates the one putative transmembrane segment identified, which spans amino acid positions 4 to 25.

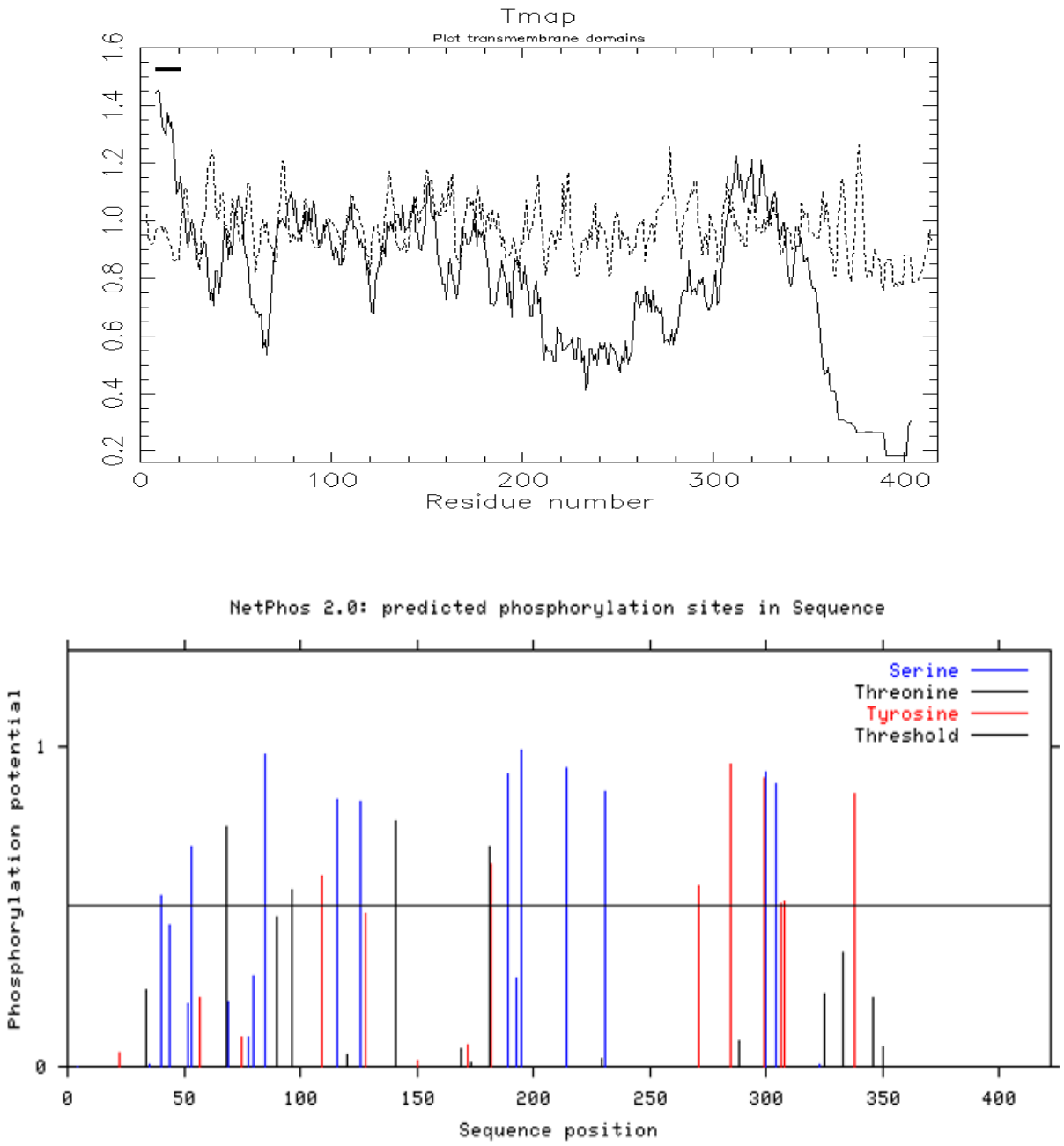

Figure 3. Graphical representation of the predicted phosphorylation sites of human calreticulin protein. Residues above the threshold (score >0.5) show strong possibility of phosphorylation sites while residues that are below or around the threshold, are not supposed to be phosphorylated. The horizontal line in the middle shows the threshold while vertical lines indicate the score.

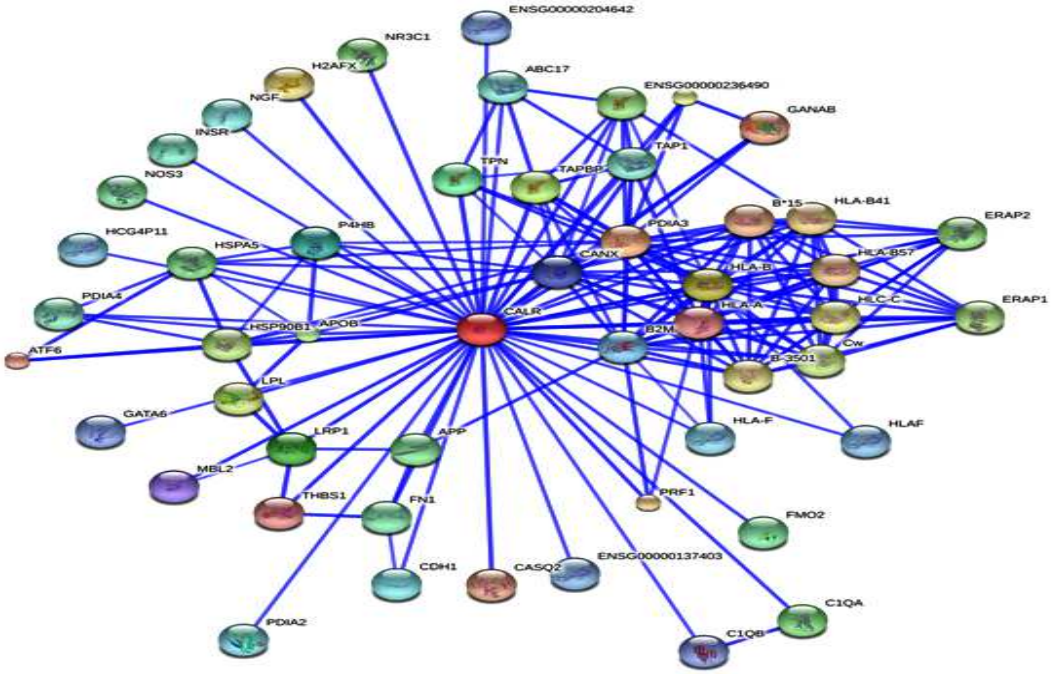

Figure 4. Model-predicted functional interaction network of the human calreticulin protein (CALR) and its probable functional partners as displayed by the STRING 9.1 database. The interactions are shown in a "confidence view," in which the stronger associations are represented by thicker lines. The proteins are identified by their gene names located near each sphere. PPIs with score greater than 0.7 were selected to construct PPI networks.

Table 1. Amino acid composition of the human calreticulin protein computed using Expasy's ProtParam tool

S. No.	AMINO ACID	NUMBER	PERCENT
1.	Ala (A)	16	3.8%
2.	Arg (R)	8	1.9%
3.	Asn (N)	16	3.8%
4.	Asp (D)	55	13.2%
5.	Cys (C)	3	0.7%
6.	Gln (Q)	15	3.6%
7.	Glu (E)	54	12.9%
8.	Gly (G)	27	6.5%
9.	His (H)	7	1.7%
10.	Ile (I)	19	4.6%
11.	Leu (L)	26	6.2%
12.	Lys (K)	42	10.1%
13.	Met (M)	5	1.2%
14.	Phe (F)	18	4.3%
15.	Pro (P)	27	6.5%
16.	Ser (S)	20	4.8%
17.	Thr (T)	15	3.6%
18.	Trp (W)	11	2.6%
19.	Tyr (Y)	14	3.4%

20.	Val (V)	19	4.6%
21.	Pyl (O)	0	0.0%
22.	Sec (U)	0	0.0%

Table 2. Physicochemical Parameters computed using Expasy's ProtParam tool

S. No.	PARAMETERS	PREDICTED VALUE
1.	Number of amino acids	417
2.	Molecular weight	48141.5 Da
3.	Theoretical pI	4.29
4.	Total number of negatively charged residues(Asp+Glu)	109
5.	Total number of positively residues (Arg+Lys)	50
6.	Extinction coefficient (Abs_{280} = 1.693)	81,485
7.	Extinction coefficient (Abs_{280} = 1.690)	81,360
8.	Estimated half-life (mammalian reticulocytes, in vitro)	30 hours
9.	Instability index	47.56
10.	Aliphatic index	59.14
11.	Grand average of hydropathicity (GRAVY)	-1.104

Table 3. Comparative analysis of secondary structures of calreticulin computed using SOPMA, GOR IV and PREDATOR

SECONDARY STRUCTURE	SOPMA	GOR IV	PREDATOR
Alpha helix	29.50%	26.14%	22.54%
3_{10} helix	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%
Extended strand	14.63%	18.47%	13.91%
Beta turn	3.84%	0.00%	0.00%
Bend region	0.00%	0.00%	0.00%
Random coil	52.04%	55.40%	63.55%
Ambigous states	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%

Table 4. Subcellular localization prediction of human calreticulin by CELLO version 2.5

SUPPORT VECTOR MACHINE (SVM)	LOCALIZATION	RELIABILITY
Amino acid composition	Endoplasmic reticulum	0.807
N-peptide composition	Endoplasmic reticulum	0.988
N-peptide composition	Endoplasmic reticulum	0.921
Physicochemical composition	Endoplasmic reticulum	0.905
Neighboring sequence composition	Endoplasmic reticulum	0.960

CELLO prediction	Endoplasmic reticulum	4.582*
	Nuclear	0.081
	Cytoplasmic	0.062
	Vacuole	0.058
	Mitochondrial	0.049
	Extracellular	0.036
	Plasma membrane	0.032
	Chloroplast	0.025
	Cytoskeletal	0.021
	Golgi	0.020
	Peroxisomal	0.020
	Lysosomal	0.014