Insilico Evolutionary studies of the Calreticulin protein in selected organisms

Sangeeta Sharma

Department of Biotechnology, D.A.V. College, Sector-10, Chandigarh

ARTICLE DETAILS

Article History
Published Online: 19 June 2018

Keywords
Evolution, *insilico*, Calnexin, Chaperone, Model systems, Computational tools.

ABSTRACT

Calreticulin has been implicated in the regulation of a variety of cellular functions. Much of the information of Calreticulin is pertaining to its structure and functions but the reason for its conservation during evolution remain obscure.

It has been postulated that the calreticulin protein and the gene are not found in yeast, suggesting that the protein must have emerged later during evolution. There is a considerable amino acid sequence similarity between calreticulin and calnexin, an integral ER membrane chaperone that is present in yeast. In the present study an *insilico* attempt has been made to study the evolutionary aspects of Calreticulin protein in selected organisms and a few model systems like *E coli* and *Arabidopsis thaliana* by using different online computational tools to analyze the conserved regions, domains, and the evolutionary relatedness in the selected organisms.

1. Introduction

It is essential to have insights in the evolutionary aspects of a particular protein of interest for studying the structure and function of a protein. Generally, it is the primary structure of a protein that dictates its fold and function, and cumulative changes in the sequence leads to the evolution of protein structures and functions [1].

Evolution of Calreticulin's function is quite varied and complicated because of the mutations in the gene. In comparison to present information about the structure and the functions of Calreticulin, little is known about its evolution [2].

Calreticulin is a highly conserved Ca^{2+}binding protein in eukaryotes comprising 3 distinct domains: a globular N-domain, a proline-rich middle (P) domain, and C-domain. The protein has an N-terminal signal sequence and a C-terminal KDEL endoplasmic reticulum retrieval signal (Figure1). The P-domain is a site of chaperone activity and oligosaccharide binding to both proteins. The N-terminal region of Calreticulin is a globular domain containing eight antiparallel β-strands [3]. This domain interacts with α-integrins [4] and DNA-binding site of steroid receptor [5]. The disulfide bond formed by cysteine residues in the N-domain may interact with P-domain to generate important chaperone function of Calreticulin [6].

There is a high degree of amino acid sequence similarity between the P-domain of calreticulin and calnexin, a chaperone that assists in protein folding and is situated on the integral ER membrane in yeast. The Calreticulin gene may have been derived from the calnexin gene by substitution of the N-domain and addition of the highly charged Ca^{2+}-binding C-domain.

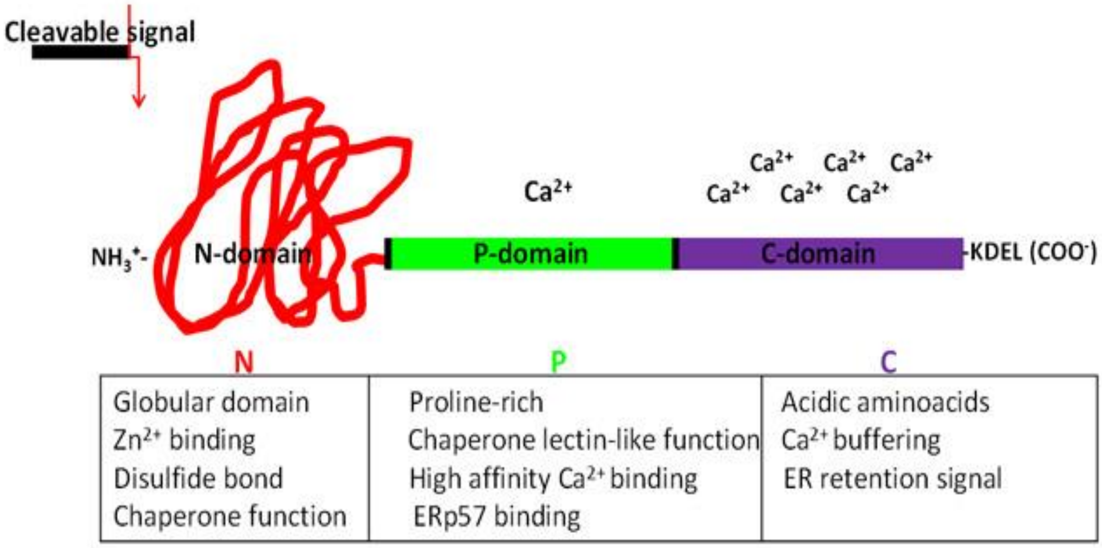


Fig 1: Calreticulin protein structure and putative functions of its domains
http://www.nature.com/scitable/topicpage/calreticulin-a-multifaceted-protein-14237270

Calnexin shares several regions of similarity with calreticulin, ranging from 42% to 78% identity (7, 8). The regions that are highly conserved among different Calreticulin's are the repeats A and B found in the P-domain of Calreticulin, both of which are present in Calnexin. The Calreticulin protein has not undergone much variations thus retaining its functional integrity during the course of evolution. There is 90% amino acid identity among human, mouse, rat, rabbit, and Aplysia Calreticulin (9).

Human and mouse Calreticulin genes, their promoters and genomic organization has been thoroughly studied (10). The nucleotide sequence of the mouse gene showed that it is highly homologous to the human gene (9). Except for two introns, which, in the mouse, are twice the size of their human counterparts, the exon–intron organization of these genes is virtually identical. The human and mouse Calreticulin genes are relatively short and exist in a single copy (9,10).

In this study, the Calreticulin protein sequence was analyzed for the presence of conserved domains in Humans and significant orthologues were selected from BLASTp analysis (11) and were further studied for the functional annotations using UniProt and InterPro database (12). The sequence conservation was studied by performing a Multiple sequence alignment using Clustal Omega tool (13) in the selected orthologues and model organisms like *E coli* and *Arabidopsis thaliana.* A phylogram was then plotted to see the evolutionary relationships between the organisms selected in the analysis.

2. Methodology

BLASTp of *Homo sapiens* Calreticulin protein sequence (Url: blast.ncbi.nlm.nih.gov/Blast.cgi) was performed to find similar sequences by locating short matches between the two sequences using heuristic algorithm. The word size and non-redundant database were kept as default.

Conserved domains analysis using CDD & Interpro

(Url: http://www.ncbi.nlm.nih.gov/cdd/)

The Conserved domain database is a resource for the annotation of functional units in proteins. The Calreticulin protein domains were retrieved using CDD to provide insights into sequence/structure/function relationships.

InterPro (https://www.ebi.ac.uk/interpro/) : InterPro was used for finding all possible conserved regions of Calreticulin.

Functional annotations using UniProt (Url: http://uniprot.org/uniprot/)

UniProt is a comprehensive resource for protein sequence and annotation data. Functional analysis of Calreticulin was studied using UniProt.

Multiple Sequence Alignment and Conserved Region Analysis

ClustalΩ (Url: http://www.ebi.ac.uk/Tools/msa/clustalo/)

ClustalΩ uses guide trees and HMM profile techniques to generate alignments between three or more sequences. The significant hits of Blastp were further subjected to MSA using Clustal Omega tool.

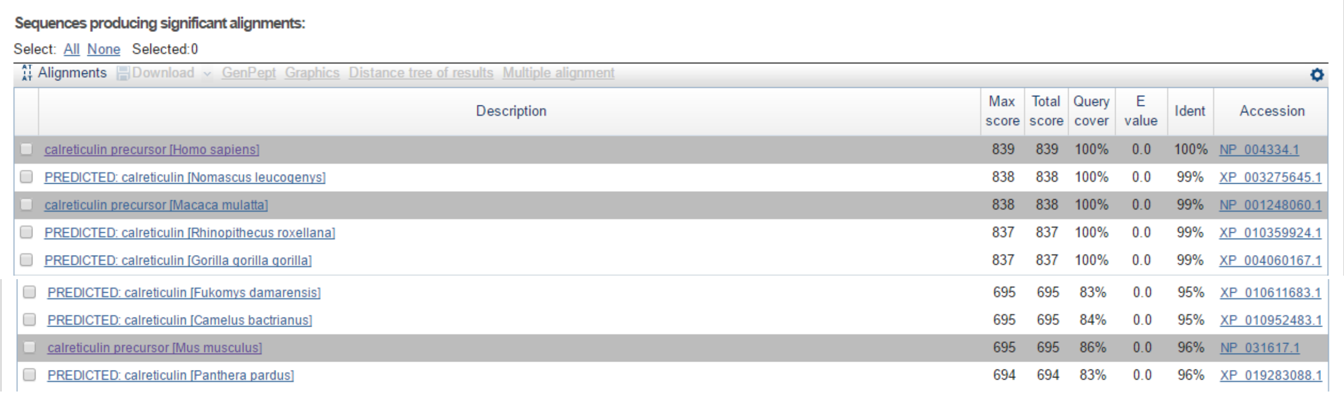
Phylogenetic analysis of Calreticulin using Treeview

(Url: http://taxonomy.zoology.gla.ac.uk/rod/treeview.html)

TreeView is a standalone bioinformatics tool that displays phylogenetic view of contents of NEXUS, PHYLIP, Clustal, or other format tree files. Clustal output generated from MSA was used as the input file for TreeView analysis.

3. Results and Discussions

Calreticulin protein [*Homo sapiens*] was subjected to Blastp analysis with default settings but in the algorithm parameters, maximum target sequences were changed from 100 to 500 sequences to extract a greater number of hits since the regular Blastp did not retrieve enough significant matches as most of the hits were experimentally not validated. The BLAST output retrieved significant hits for the orthologues viz *Danio rerio, Mus musculus* and *Macaca mulatta*. The selected significant orthologous are shown in Table1:



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	calreticulin precursor [Homo sapiens]	839	839	100%	0.0	100%	NP_004334.1
☐	PREDICTED: calreticulin [Nomascus leucogenys]	838	838	100%	0.0	99%	XP_003275645.1
☐	calreticulin precursor [Macaca mulatta]	838	838	100%	0.0	99%	NP_001248060.1
☐	PREDICTED: calreticulin [Rhinopithecus roxellana]	837	837	100%	0.0	99%	XP_010359924.1
☐	PREDICTED: calreticulin [Gorilla gorilla gorilla]	837	837	100%	0.0	99%	XP_004060167.1
☐	PREDICTED: calreticulin [Fukomys damarensis]	695	695	83%	0.0	95%	XP_010611683.1
☐	PREDICTED: calreticulin [Camelus bactrianus]	695	695	84%	0.0	95%	XP_010952483.1
☐	calreticulin precursor [Mus musculus]	695	695	86%	0.0	96%	NP_031617.1
☐	PREDICTED: calreticulin [Panthera pardus]	694	694	83%	0.0	96%	XP_019283088.1

☐ PREDICTED: calreticulin-like isoform X3 [Sinocyclocheilus grahami]	567	567	90%	0.0	74%	XP_016101559.1
☐ PREDICTED: calreticulin-like [Clupea harengus]	566	566	86%	0.0	74%	XP_012683792.1
☐ PREDICTED: calreticulin-like [Larimichthys crocea]	566	566	85%	0.0	75%	XP_019109803.1
☐ unnamed protein product [Oncorhynchus mykiss]	566	566	89%	0.0	70%	CDQ56868.1
☐ PREDICTED: calreticulin-3 [Haplochromis burtoni]	566	566	86%	0.0	75%	XP_005938461.1
☐ calreticulin [Ixodes persulcatus]	566	566	86%	0.0	74%	AAR29957.1
☐ PREDICTED: calreticulin-like [Sinocyclocheilus rhinocerous]	566	566	85%	0.0	74%	XP_016421984.1
☐ hypothetical protein CAPTEDRAFT_21877 [Capitella teleta]	566	566	86%	0.0	74%	ELU06446.1
☐ PREDICTED: calreticulin-like [Salmo salar]	565	565	90%	0.0	70%	XP_013996614.1
☐ calreticulin [Danio rerio]	565	565	86%	0.0	74%	AAF13700.1
☐ PREDICTED: calreticulin-like [Sinocyclocheilus grahami]	565	565	85%	0.0	74%	XP_016090164.1
☐ hypothetical protein cypCar_00034818 [Cyprinus carpio]	565	565	80%	0.0	78%	KTG40598.1
☐ PREDICTED: calreticulin-like [Austrofundulus limnaeus]	565	565	86%	0.0	75%	XP_013885892.1

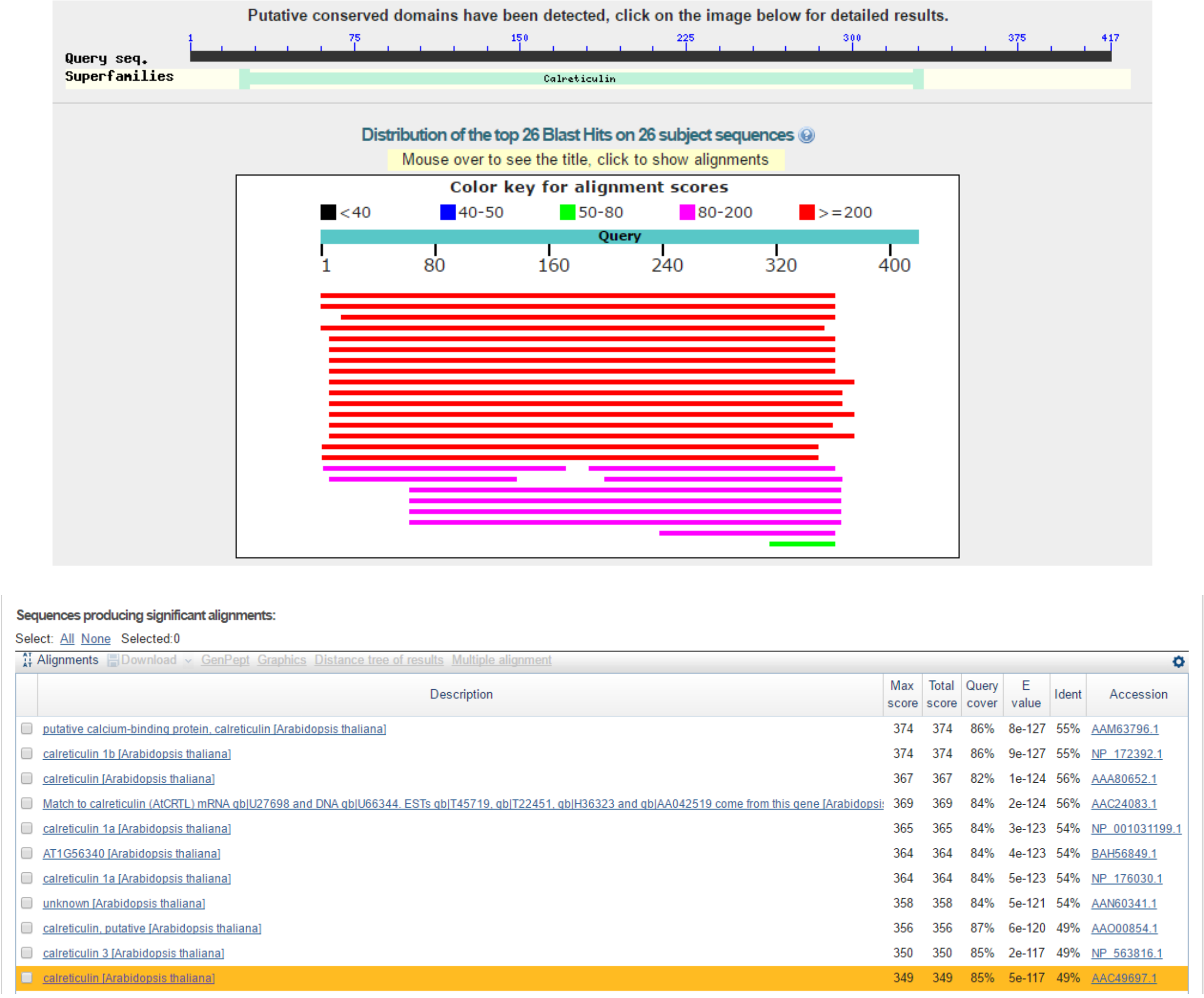
Fig 2: Retrieval of similar protein sequences of Calreticulin [*Homo sapiens*] using BLASTp

The remaining significant hits were checked for the conserved domains and it was found that the domains were lacking in them and therefore were not considered in the analysis. Some orthologues had Ca+ binding sites and were also not considered in the present study.

S. No.	Accession Number	Name	Identity	E-value
1	AAF13700.1	*Danio rerio*	74%	0
2	NP_031617.1	*Mus musculus*	96%	0
3	NP_001248060	*Macaca mulatta*	99%	0

Table 1: Orthologues with significant E-Values

To check for the conservation of Calreticulin protein in plants, an organism specific BlastP was performed for *Arabidopsis thaliana* (Table 2) which is a model organism in plant biology.



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	putative calcium-binding protein, calreticulin [Arabidopsis thaliana]	374	374	86%	8e-127	55%	AAM63796.1
☐	calreticulin 1b [Arabidopsis thaliana]	374	374	86%	9e-127	55%	NP_172392.1
☐	calreticulin [Arabidopsis thaliana]	367	367	82%	1e-124	56%	AAA80652.1
☐	Match to calreticulin (AtCRTL) mRNA gb|U27698 and DNA gb|U66344. ESTs gb|T45719, gb|T22451, gb|H36323 and gb|AA042519 come from this gene [Arabidopsi	369	369	84%	2e-124	56%	AAC24083.1
☐	calreticulin 1a [Arabidopsis thaliana]	365	365	84%	3e-123	54%	NP_001031199.1
☐	AT1G56340 [Arabidopsis thaliana]	364	364	84%	4e-123	54%	BAH56849.1
☐	calreticulin 1a [Arabidopsis thaliana]	364	364	84%	5e-123	54%	NP_176030.1
☐	unknown [Arabidopsis thaliana]	358	358	84%	5e-121	54%	AAN60341.1
☐	calreticulin, putative [Arabidopsis thaliana]	356	356	87%	6e-120	49%	AAO00854.1
☐	calreticulin 3 [Arabidopsis thaliana]	350	350	85%	2e-117	49%	NP_563816.1
☐	calreticulin [Arabidopsis thaliana]	349	349	85%	5e-117	49%	AAC49697.1

Fig 3: Organism specific Blastp result of Calreticulin in *Arabidopsis thaliana*

S. No.	Accession Number	Name	Identity	E-value
1.	NP_497641.2	*Arabidopsis thaliana*	49%	5e-117

Table 2: Model organism in plant biology with significant E-Value

Similarly, Organism specific blast for the remaining model organisms like *E. coli, Saccharomyces cerevisiae, Drosophila melanogaster* and *Caenorhabditis elegans* also did not retrieve the conserved N, P and C-domains.

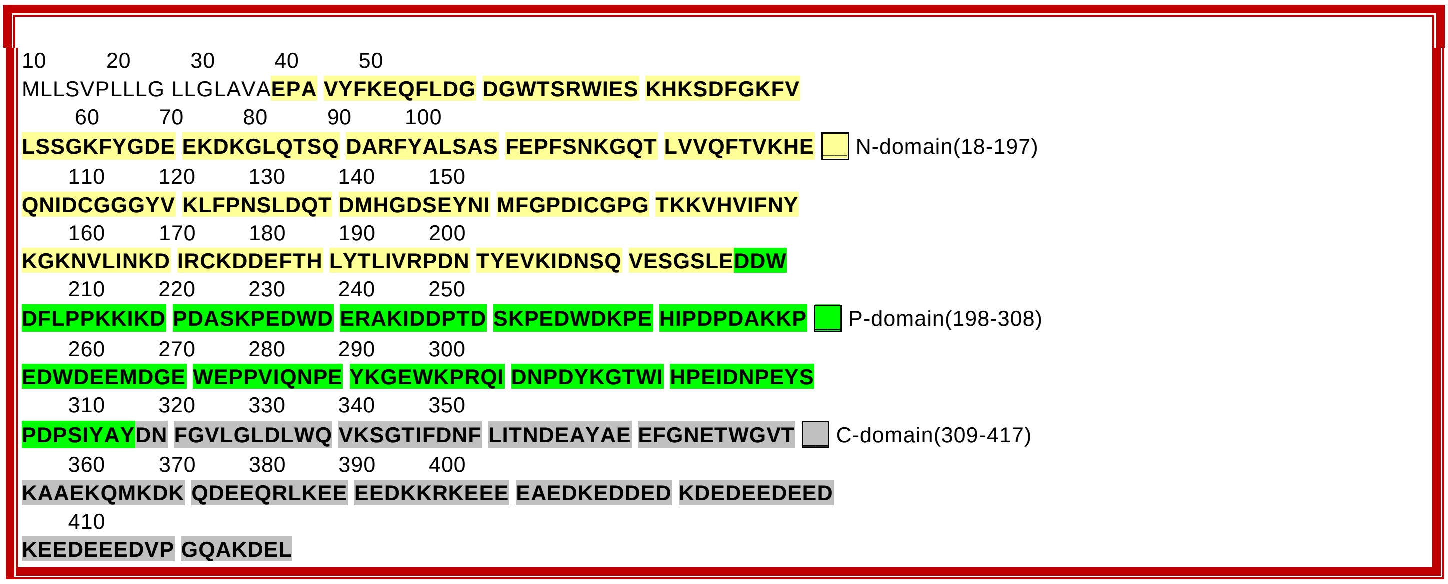
The *Caenorhabditis elegans* Calreticulin protein was seen to be involved in stress response, fertility, and Ca+ ion regulator in Endoplasmic Reticulum.

The organism specific Blastp for *Saccharomyces cerevisiae* retrieved Calnexin and Calreticulin like protein involved in folding and quality control of glycoproteins. It did not show Ca+ binding in yeast.

For *E. coli* no significant match was reported. Similarly, for *Drosophila* no conserved domain or Ca+ ion binding sites was reported.

Conserved Domains were predicted for Calreticulin [*Homo sapiens*] and N, P and C-domains were found along with Zinc binding sites. Furthermore, InterPro was used for finding all possible conserved regions of Calreticulin.

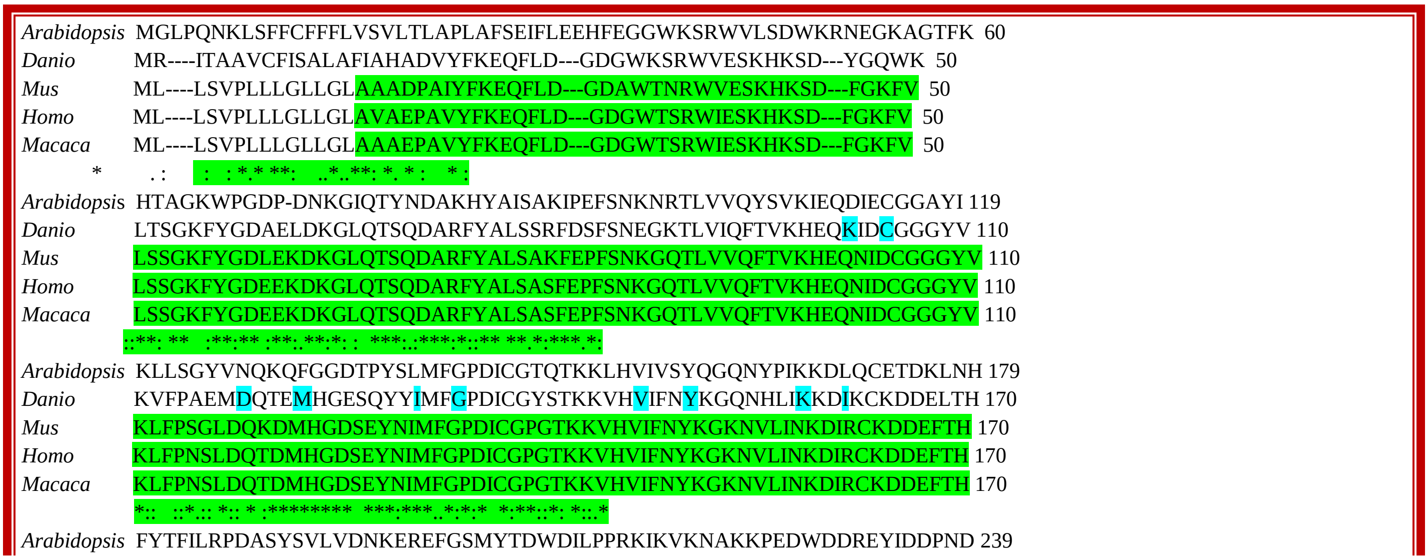
Interpro analysis for Calreticulin shows that P-domain is highly conserved (viz. 201-315) in *Homo sapiens*, *Macaca mulatta, Mus musculus, Dario rerio* whereas it exists at 204-319 for *Arabidopsis thaliana*



```
10         20         30         40         50
MLLSVPLLLG LLGLAVAEPA VYFKEQFLDG DGWTSRWIES KHKSDFGKFV
        60         70         80         90        100
LSSGKFYGDE EKDKGLQTSQ DARFYALSAS FEPFSNKGQT LVVQFTVKHE  🟨 N-domain(18-197)
       110        120        130        140        150
QNIDCGGGYV KLFPNSLDQT DMHGDSEYNI MFGPDICGPG TKKVHVIFNY
       160        170        180        190        200
KGKNVLINKD IRCKDDEFTH LYTLIVRPDN TYEVKIDNSQ VESGSLEDDW
       210        220        230        240        250
DFLPPKKIKD PDASKPEDWD ERAKIDDPTD SKPEDWDKPE HIPDPDAKKP  🟩 P-domain(198-308)
       260        270        280        290        300
EDWDEEMDGE WEPPVIQNPE YKGEWKPRQI DNPDYKGTWI HPEIDNPEYS
       310        320        330        340        350
PDPSIYAYDN FGVLGLDLWQ VKSGTIFDNF LITNDEAYAE EFGNETWGVT  ▨ C-domain(309-417)
       360        370        380        390        400
KAAEKQMKDK QDEEQRLKEE EEDKKRKEEE EAEDKEDDED KDEDEEDEED
       410
KEEDEEEDVP GQAKDEL
```

Fig 4: N, P, C Domains of Calreticulin using UniProt

Since only P domain was observed in InterPro results so UniProt was referred for functional annotations of Calreticulin to find N and C-domains. Uniprot results shows that N-domain exists at positions 18-197, P-domain at positions 198-308 and C-domain positioned at 309-416 (in *Mus musculus*) and 309-417 (*Homo sapiens*).



```
Arabidopsis  MGLPQNKLSFFCFFFLVSVLTLAPLAFSEIFLEEHFEGGWKSRWVLSDWKRNEGKAGTFK  60
Danio        MR----ITAAVCFISALAFIAHADVYFKEQFLD---GDGWKSRWVESKHKSD---YGQWK  50
Mus          ML----LSVPLLLGLLGLAAADPAIYFKEQFLD---GDAWTNRWVESKHKSD---FGKFV  50
Homo         ML----LSVPLLLGLLGLAVAEPAVYFKEQFLD---GDGWTSRWIESKHKSD---FGKFV  50
Macaca       ML----LSVPLLLGLLGLAAAEPAVYFKEQFLD---GDGWTSRWIESKHKSD---FGKFV  50
                 *     . :     :  : *.* **:   ..*..**: *. * :   * :
Arabidopsis  HTAGKWPGDP-DNKGIQTYNDAKHYAISAKIPEFSNKNRTLVVQYSVKIEQDIECGGAYI 119
Danio        LTSGKFYGDAELDKGLQTSQDARFYALSSRFDSFSNEGKTLVIQFTVKHEQKIDCGGGYV 110
Mus          LSSGKFYGDLEKDKGLQTSQDARFYALSAKFEPFSNKGQTLVVQFTVKHEQNIDCGGGYV 110
Homo         LSSGKFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDCGGGYV 110
Macaca       LSSGKFYGDEEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDCGGGYV 110
            ::**: **   :**:** :**:.**:*: :  ***:..:***:*:.:** **.*:***.*:
Arabidopsis  KLLSGYVNQKQFGGDTPYSLMFGPDICGTQTKKLHVIVSYQGQNYPIKKDLQCETDKLNH 179
Danio        KVFPAEMDQTEMHGESQYYIMFGPDICGYSTKKVHVIFNYKGQNHLIKKDIKCKDDELTH 170
Mus          KLFPSGLDQKDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIRCKDDEFTH 170
Homo         KLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIRCKDDEFTH 170
Macaca       KLFPNSLDQTDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIRCKDDEFTH 170
             *::   ::*.:: *:: * :********  ***:***..*:*:*  *:**::*: *:..*
Arabidopsis  FYTFILRPDASYSVLVDNKEREFGSMYTDWDILPPRKIKVKNAKKPEDWDDREYIDDPND 239
```

```
Danio       LYTLILRPDQTYEVKIDNEKVESGSLEEDWDFLPPKKIKDPEAKKPDDWDDRAKIDDPED 230
Mus         LYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDAAKPEDWDERAKIDDPTD 230
Homo        LYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPTD 230
Macaca      LYTLIVRPDNTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPTD 230
            : **:*:*** :*.* :**.: * **:  ***:***:***  :* **:***:*  **** *
Arabidopsis VKPEGFDSIPREIPDRKAKEPEDWDEEENGLWEPPKIPNFAYKGPWKAKRIKNPNYKGKW 299
Danio       TKPEDWDK-PENIPDPDAKKPDDWDEDMDGEWEPAMIPNPEYKGEWKPKQIDNPSYKGTW 289
Mus         SKPEDWDK-PEHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNPDYKGTW 289
Homo        SKPEDWDK-PEHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNPDYKGTW 289
Macaca      SKPEDWDK-PEHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNPDYKGTW 289
            ***.:*. *..*** .**:*:****: :* ***  * *  *** ** ::*.**.***.*
Arabidopsis KNPWIDNPEFEDDPDLYVLKSIKYAGIEVWQVKAGSIFDNILICDDPAYARSIVDDYFAQ 359
Danio       VHPEIDNPEYSADDAIYKFDSIGVLGLDLWQVKSGTIFDNFLITDDVEEAEKFGTDTWGA 349
Mus         IHPEIDNPEYSPDANIYAYDSFAVLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGV 349
Homo        IHPEIDNPEYSPDPSIYAYDNFGVLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGV 349
Macaca      IHPEIDNPEYSPDPSIYAYDNFGVLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGV 349
            :* *****:. *  :*  ..:  *:::****:*:****:** :*   *..: :.:
Arabidopsis HRESEKELFAEAEKERKAREDEEARIAREEGERRRKERDHRYGDRRRRYKRPNPRD---- 415
Danio       TKGPEKKMKDQQEEEERKKREEEEKSKKDDNEEDEEDEDEDEPEEDDHTEEPPEEEEEGE 409
Mus         TKAAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDKEDDDDR-DEDEDEEDEK-EEDEEE 407
Homo        TKAAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDKEDDEDK-DEDEEDEEDKEEDEEED 408
Macaca      TKAAEKQMKDKQDEEQRLKEEEEDKKRKEEEEAEDKEDDEDK-DEDEEDEEDKEEDEEED 408
            :  **::  : ::*.: :.:** :  ::: *   :: *.  :.   :   .:
Arabidopsis YMDDYHDEL         424
Danio       DDALPKDEL         418
Mus         SPGQAKDEL         416
Homo        VPGQAKDEL         417
Macaca      VPGQAKDEL         417
            :***
```

N-Domain: 18 … 197, P-Domain: 198 … 308, Zinc binding sites: 112,115,128,132, 140,143,156,160, 168,171,184,188, 196,199,212,216, 224,227,240,244, 252,255,268,272, 308,311,324,328, 335, 338, 351, 355.

Fig 5: ClustalΩ result of Calreticulin protein showing the conserved regions in selected Orthologues

The significant hits retrieved from Blastp analysis were considered for the Multiple sequence alignment. The results show the presence of conserved regions which have been identified as the N, P and C-domains of Calreticulin protein.

a) In orthologous viz. *Homo sapiens, Macaca mulatta* and *Danio rerio* N, P and C-domains were found at following positions:
   i. N-domain positions from 18 to 197 with a length of 180.
   ii. P-domain positions from 198 to 308 with a length of 111.
   iii. C-domain positions from 309 to 417 with a length of 109

In *Mus musculus*,

i. N-domain positions from 18 to 197 with a length of 180.
ii. P-domain positions from 198 to 308 with a length of 111.
iii. C-domain positions from 309 to 416 with a length of 108.

b) In organism specific (*Arabidopsis thaliana*) Blast, these domains were found at:
   i. N-domain positions from 18 to 197 with a length of 180.
   ii. P-domain positions from 198 to 308 with a length of 111.
   iii. N-domain positions from 309 to 416 with a length of 108.

The Clustal output file was taken as the input for TreeView which generated Phylogram of the selected orthologues.

TreeView output shows the evolutionary relationships of Calreticulin protein amongst the selected orthologues, *Homo sapiens* are more related to *Macaca mulatta* and subsequently to *Mus musculus*. Similarly, *Danio rerio* and *Arabidopsis thaliana* are related to each other and form a clade.

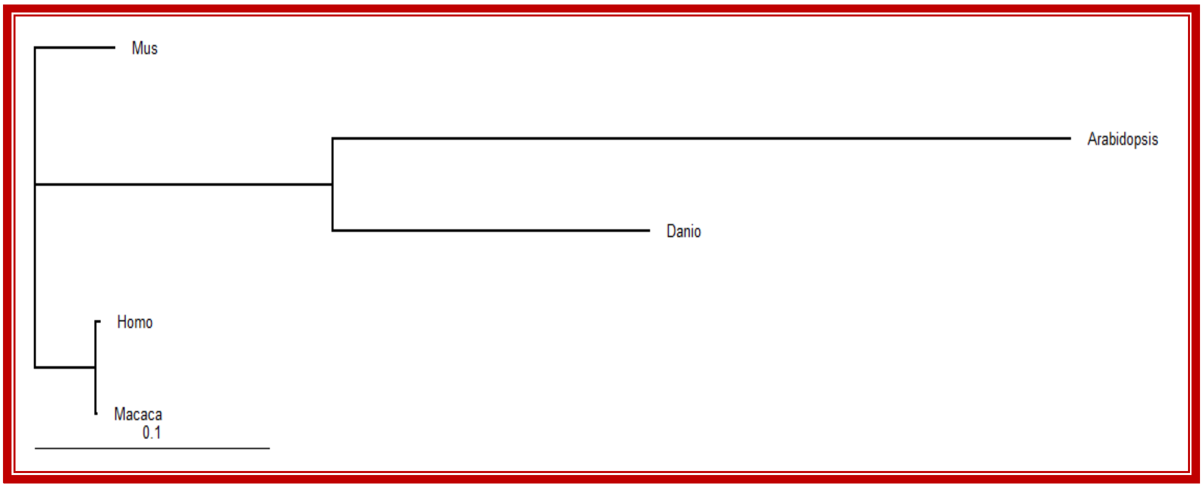


Fig 6: Phylogram depicting the evolutionary relatedness of Calreticulin in selected orthologues using TreeView

4. Conclusion

The present work emphasizes on the study of conserved regions and evolutionary analysis of Human Calreticulin protein using online computational tools.

Conserved domain analysis and functional annotations in the selected Orthologues viz. *Homo sapiens, Mus musculus, Macaca mulatta, Danio rerio* and *Arabidopsis thaliana* for Calreticulin protein confirmed the findings that N, P, C-domains, and Calcium binding sites are highly conserved in selected orthologues. On the other hand the other model organisms like *E. coli, Saccharomyces cerevisiae, Drosophila melanogaster* and *Caenorhabditis elegans* did not retrieve the conserved N, P and C-domains. The *Caenorhabditis elegans* Calreticulin protein was seen to be involved in stress response, fertility, and Ca+ ion regulator in Endoplasmic Reticulum.

The *Saccharomyces cerevisiae* Organism specific Blast retrieved Calnexin and Calreticulin like protein involved in folding and quality control of glycoproteins.

TreeView results indicate that Calreticulin protein has not undergone much variations in the selected Orthologues during the course of evolution and thus retaining its functional integrity.

References

1. Kinch L.N. & Grishin N.V. 2002. Evolution of protein structures and functions. Curr. Opin. Struct. Biol. 12: 400–408.
2. Pinto R.D., Moreira A.R., Pereira P.J. & Dos Santos N.M. 2013.Molecular cloning and characterization of sea bass (Dicen-trarchus labrax, L.) calreticulin. Fish Shellfish Immunol. 34:1611–1618.
3. M. Michalak, J. M. R. Parker, and M. Opas, "Ca2+ signaling and calcium binding chaperones of the endoplasmic reticulum," Cell Calcium, vol. 32, no. 5-6, pp. 269–278, 2002.
4. M. V. Rojiani, "In vitro interaction of a polypeptide homologous to human Ro/SS-A antigen (Calreticulin) with a highly conserved amino acid sequence in the cytoplasmic domain of integrin α subunits," Biochemistry, vol. 30, no. 41, pp. 9859–9866, 1991.
5. K. Burns, B. Duggan, E. A. Atkinson et al., "Modulation of gene expression by calreticulin binding to the glucocorticoid receptor," Nature, vol. 367, no. 6462, pp. 476–480, 1994.
6. M. Michalak, E. F. Corbett, N. Mesaeli, K. Nakamura, and M. Opas, "Calreticulin: one protein, one gene, many functions," Biochemical Journal, vol. 344, no. 2, pp. 281–292, 1999.
7. V. Martin, J. Groenendyk, S. S. Steiner et al., "Identification by mutational analysis of amino acid residues essential in the chaperone function of calreticulin," The Journal of Biological Chemistry, vol. 281, no. 4, pp. 2338–2346, 2006.
8. K.-H. Krause and M. Michalak, "Calreticulin Meeting Review," Cell, vol. 88, no. 4, pp. 439–443, 1997.
9. L. W. Tjoelker, C. E. Seyfried, R. L. Eddy Jr. et al., "Human, mouse, and rat calnexin cDNA cloning: identification of potential calcium binding motifs and gene localization to human chromosome 5," Biochemistry, vol. 33, no. 11, pp. 3229–3236, 1994.
10. K. Nakamura, A. Zuppini, S. Arnaudeau et al., "Functional specialization of calreticulin domains," The Journal of Cell Biology, vol. 154, no. 5, pp. 961–972, 2001.
11. Stephen F.Altschul[1], WarrenGish[1]Basic local alignment search tool ; JMB, Volume 215, Issue 3, 5 October 1990, Pages 403-410
12. Blum M, Chang H, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, The InterPro protein families and domains:20 years on; *Nucleic Acids Research, Nov 2020, (doi: 10.1093/nar/gkaa977)*
13. Sievers F., Wilm A., Dineen D., Gibson T.J.*et al* (2011) Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539