Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Advances in Applied Science Research, 2014, 5(2):353-358

ISSN: 0976-8610
CODEN (USA): AASRFC

Evolutionary inferences on lysosomal membrane proteins (LMPs) defective in metabolic disorders: Cystinosin and Sialin

Manish Dwivedi[#] and Dwijendra K. Gupta

Center of Bioinformatics, University of Allahabad, Allahabad

ABSTRACT

Cystinosin and sialin proteins are the major constituent of lysosomal membrane and play an important role in the transport of the metabolite across the lysosomal membrane. Any mutation in these genes causes autosomal recessive lysosomal storage disorders named as cystinosis and salla disease. To understand cause of these metabolic disorders and their evolutionary relation, genomics based study of these proteins were facilitated by bioinformatics tools like BLAST, ClustalW, MEGA4, BioEdit (version 7.0.5.3) etc. This work established the molecular phylogenetic relationship among the different taxa of cystinosin and sialin. Structural properties were concluded by their amino acid frequency, hydrophobicity profile, entropy and conserved domains present. Both the protein showed that mean hydrophobicity for most of the positions in all the species was around zero and was basically hydrophobic in nature. These informations facilitate to explore the molecular basis of the lysosomal storage disorders and cell biologic features along with the subcellular localization of the proteins leading to predict the exact function in the lysosomal membrane and enable us to understand the critical regions of the proteins studied.

Keywords: Sialin, cystinosin, sequence, protein, phylogenetic.

INTRODUCTION

Metabolic disorders consist of any defect in metabolic reaction carried out inside the cells. Mostly these defects arise due to malfunction of the transporter proteins or mutation in their genes. Here we have reported investigations on lysosomal membrane transporter proteins, Sialin and cystinosin. Different bioinformatics softwares were employed to reveal core properties of the proteins defective in lysosomal storage disorders. The main functions of these proteins include transportation of the degraded and non-degraded lysosomal substrate across the lysosomal membrane. Lysosomes are principal site of intracellular digestion in mammalian cells. A defect in either the degradation or transport process can result in an accumulation of the undegraded substrate or the degradation product within the lysosomes, impairing the physiology of the cell and leading to a lysosomal storage disorder. Several such disorders have been described and are usually classified according to the molecule that has accumulated intra-lysosomally [1]. Most of these disorders are due to a defect in one of the lysosomal hydrolases, however, a subset exists which arises from a defect in one of the membrane transporters [2].

Membrane transporter protein sialin is recently characterized protein encoded by a gene SLC17A5 and causes sialic acid storage diseases (SASD) on mutation [3] where as CTNS encodes cystinosin protein causing cystinosis on mutation (Shotelersuk et al., 1998; Town et al., 1998; Attard et al., 1999; Thoene et al., 1999). This SLC17 family also comprises the synaptic vesicular glutamate transporters and putative sodium phosphate cotransporters representing the highest degree of homology [4]. Sialic acid includes a family of acidic nine-carbon monosaccharides, which are biosynthetic derivative of Nacetylneuraminic acid (Neu5Ac). These play an important role in many biological processes including cell – cell communication and recognition, pathogen and toxin binding [5], conformational stabilization and protease resistance, enhancement of water binding capacity, protein targeting, developmental regulation [6] regulation of signal transduction [7], and immune response [8].

Cystinosin is a lysosomal membrane protein but the way it is targeted to this organelle is still unknown. Cystinosis is an inherited lysosomal transport disorder characterized by imperfect transport of cystine out of lysosomes causing accumulation of intra-lysosomal cystine and it is the most common inherited cause of the renal Fanconi syndrome. The gene underlying cystinosis, CTNS, was identified using a positional cloning strategy (Town et al., 1998) which is highly glycosylated at the amino-terminal end and carries a GY-XXF lysosomal targeting motif in its carboxy tail. CTNS encodes a 367 amino acid protein, cystinosin, which comprises seven predicted transmembrane domains, a 128 amino acid N-terminal region bearing seven N-glycosylation sites and a 10 amino acid cytosolic C-terminus containing a tyrosine-based lysosomal sorting motif (GYDQL).

This bioinformatics approach enlighten unrevealed aspects regarding the LMPs leading to metabolic disorder on mutation and made a firm platform to elucidate localization of proteins at cellular and tissue micro-anatomic level which is still unknown. We exploited our recent knowledge to determine the specific characteristics and inter-relation between sialin and cystinosin protein family and to establish the molecular phylogenetic relationship among the different taxa. It will assist to improve our understanding of their biogenesis and localization in the cells leading to recognize mechanism underlying the development of lysosomal storage disorders.

Abbreviations: *NCBI- National center for Biotechnology Information, BLAST-Basic local alignment search tool, BLOSUM-Blocks of amino acid substitution matrix.*

MATERIALS AND METHODS

Protein family members were searched by using blastp program *of BLAST* [9] in the protein database at *NCBI* [10]. *Homo sapiens* proteins amino acid sequence was selected as query. The sequences were examined individually and aligned by using *CLUSTALW* [11]. Entropy is then calculated as:

$$H(l) = -\mathrm{S} f(b,l) \ln(f(b,l))$$

where $H(l)$ = the uncertainty, also called ***entropy*** at position l, b represents a residue (out of the allowed choices for the sequence in question), and $f(b,l)$ is the frequency at which residue b is found at position l.

Table 1: Cystinosin and Sialin sequences with their length and NCBI accession code

Organism (Cystinosin) with NCBI Accession code	Length (aa)	Organism (Sialin) with NCBI Accession code	Length (aa)
Mus musculus gi|11967808|emb|CAC19455.1|	367	*Homo sapiens* gi|6912666|ref|NP_036566.1|	495
Bos Taurus gi|182639279|sp|A7MB63.1|	367	*Macaca mulatta* gi|109071731|ref|XP_001112868.1|	495
Homo sapiens gi|3036840|emb|CAA11021.1|	367	*Callithrix jacchus* gi|166092137|gb|ABY82116.1|	463
Rattus norvegicus gi|109491251|ref|XP_001080248.1|	392	*Ailuropoda melanoleuca* gi|301762398|ref|XP_002916619.1|	495
Caenorhabditis elegans gi|32565006|ref|NP_872022.1|	374	*Mus musculus* gi|27370146|ref|NP_766361.1|	495
Gallus gallus gi|50758292|ref|XP_415851.1|	377	*Rattus norvegicus* gi|57527612|ref|NP_001009713.1|	495
Canis familiaris gi|73967295|ref|XP_548340.2|	367	*Taeniopygia guttata* gi|224048541|ref|XP_002190807.1|	482
Aedes aegypti gi|157167697|ref|XP_001655585.1|	367	*Canis familiaris* gi|73973298|ref|XP_867399.1|	442
Caenorhabditis briggsae gi|215275218|sp|A8WN56.1|	403	*Xenopus (Silurana) tropicalis* gi|118403526|ref|NP_001072823.1|	496
Drosophila melanogaster gi|34223744|sp|Q9VCR7.2|	397	*Danio rerio* gi|115529331|ref|NP_001070195.1|	488
Ornithorhynchus anatinus gi|149641786|ref|XP_001508868.1|	605	*Salmo salar* gi|291190748|ref|NP_001167306.1|	492
Caligus clemensi gi|225717886|gb|ACO14789.1|	409	*Ovis aries* gi|57619234|ref|NP_001009742.1|	495
Trypanosoma cruzi gi|71422082|ref|XP_812021.1|	383	*Oryctolagus cuniculus* gi|284005471|ref|NP_001164764.1|	495
Perkinsus marinus gi|239885524|gb|EER09461.1|	302	*Papio anubis* gi|281182826|ref|NP_001162413.1|	495
Schistosoma mansoni gi|238665302|emb|CAZ36053.1|	359	*Gallus gallus* gi|71895023|ref|NP_001026257.1|	484
Ixodes scapularis gi|240999715|ref|XP_002404776.1|	266		
Pediculus humanus corporis gi|242006380|ref|XP_002424029.1|	356		
Culex quinquefasciatus gi|170044802|ref|XP_001850022.1|	399		

RESULTS AND DISCUSSION

Multiple sequence alignment and Conserved region search within the aligned sequence:
The Multiple alignments of sialin and cystinosin proteins were done by the *Culstal W* and resulted into an alignment of the sequences having 497 and 680 positions respectively. By statistical analysis of multiple aligned sequences, it was observed that sialin protein had higher frequency of alanine, phenylalanine, glycine, isoleucine, leucine, serine, threonine and valine while cystinosin was rich in phenylalanine, isoleucine, leucine, serine, valine and glycine. The multiple aligned sequence of sialin protein was found with higher number of conserved sites (234) than cystinosin which had 27 conserved sites. Parsimony informative sites (293), singleton sites (101) and variables (400) were higher in cystinosin in comparison to silain.

A conserved region search concluded that sialin had seven conserved domains with short length ranging 5-9 amino acids with average entropy of zero while cystinosin showed 4 conserved regions with long length upto 83 amino acids having average entropy between 0.6944 to 1.0170. These conservations had already been upheld by minimal entropy.

Hydrophobicity profile and hydrophobic moment:
The hydrophobicity profile plot of both the proteins showed that mean hydrophobicity of the protein in all the species were around zero except some species, occasionally it turned to be positive or negative (Figure.1). N-terminal domain and C-terminal domain were non-hydrophobic and maximum hydrophobicity were observed from positions 40 to 70, 200 to 250 and 460 to 485 positions in *Gallus gallus, Canis familiaris, Xenopus tropicalis* and *Ovis aries* subject to Sialin (Fig.1.a) and in case of cystinosin (Fig.1.b), N-terminal domains were more hydrophobic and c-terminal domains were non-hydrophobic in the most of the organisms along with exceptionally a long non-hydrophobic N-terminal domain in the *Ornithorhynchus anatinus* from position 1-230. *Culex quinquefasciatus* and *Trypanosoma cruzi* exhibits high jump in hydrophobicity from position 240 to 260 and 500 to 540 respectively.

(a)

(b)

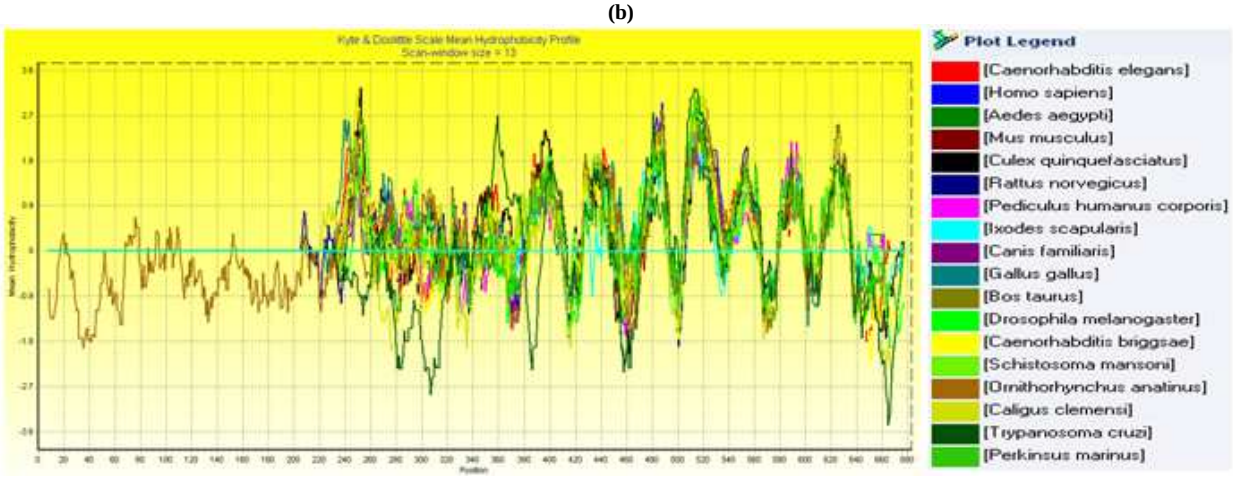


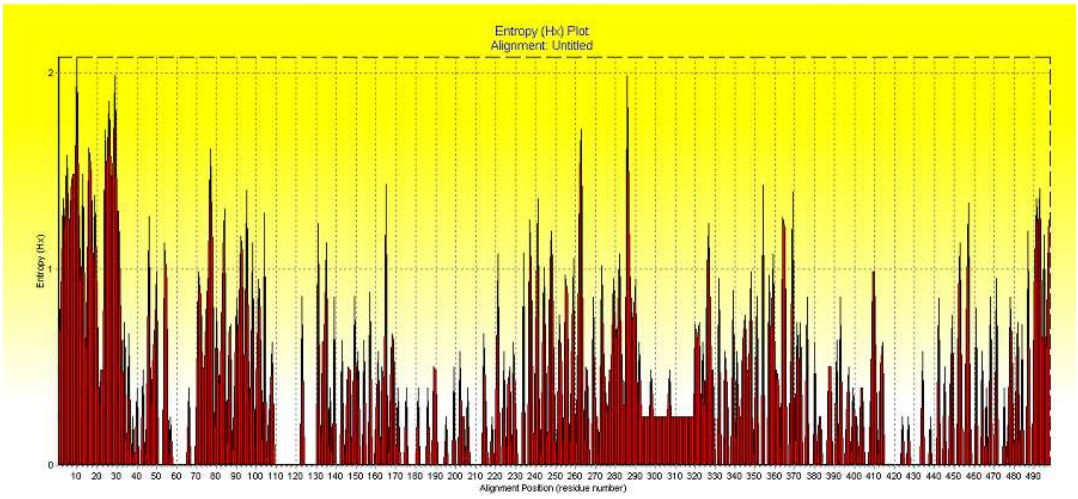
Figure 1- Mean hydrophobicity profile of Sialin (a) and Cystinosin (b) developed by Kyte and Dolittle method where Y-axis represents the hydrophobicity along the amino acid positions on X-axis

This profile concluded that these proteins basically comprised more hydrophobic region in comparison to non-hydrophobic region as number of positions were across show above mean hydrophobicity in the organisms studied here. The protein sialin and cystinosin can be treated as hydrophobic proteins and drug can be designed accordingly for respective lysosomal storage disorders.

Entropy plot:

An entropy plot, measure of the lack of the information content and the amount of variability, was generated for all the aligned positions. The entropy plot of silain and cystinosin (Figure.2) rarely touches a scale of two and showed minimal entropy at several positions while at some of the positions, entropy rarely crosses a scale of one, which is a sign of better alignment in the region. The lower values of entropy lead to the lack of randomness in the proteins sequences resulting in to conservedness at many positions in the proteins. For the lysosomal disorders this conservedness of the proteins can be taken up as marker positions for further analysis.

(a)



(b)

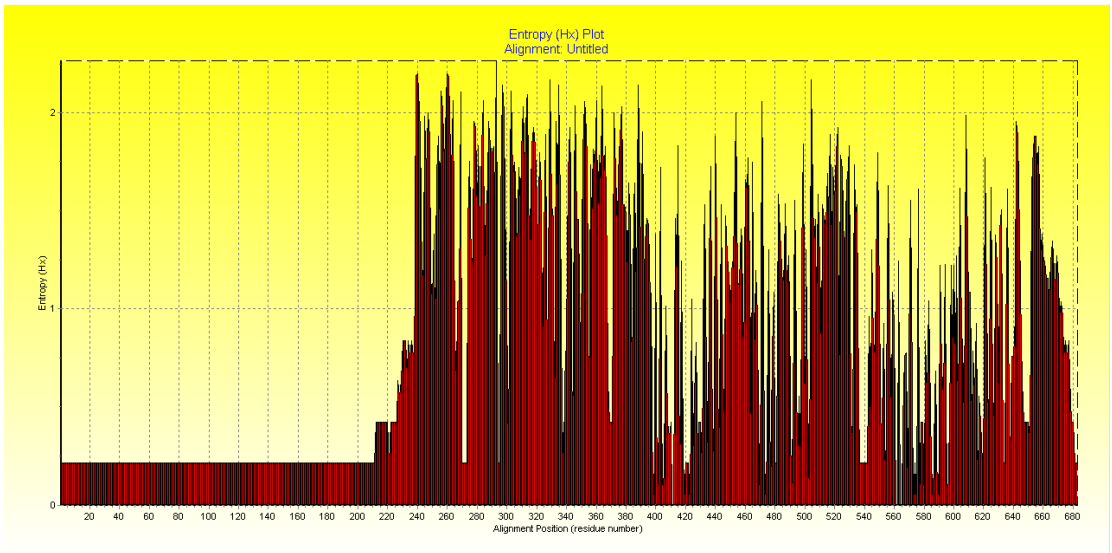


Figure: 2. Entropy (Hx) of the Sialin (a) and Cystinosin (b) protein on Y-axis and their corresponding residue number plotted on X-axis inferring the better alignment and conservedness of the sequences

Phylogeny:

The phylogenetic trees were constructed by using Neighbour –joining method (Figure.3) of MEGA4 tool .The tree showed different organisms on tree nodes branched on the basis of their proteins. *Danio rario*, *Salmo salar, Xeopus (Silurana) tropicalis* and *Perkinsus marinus, Trypanosoma cruzi* were made a totally diverged branch from the main tree among the sialin and cystinosin proteins respectively. In the phylogeny tree of both the proteins, node for vertebrates (for silain: *Canis familiaris, Mus musculus, Callithrix jacchus, Homo sapiens, Gallus gallus* and for Cystinosin: *Homo sapiens,Mus musculus, Gallus gallus*) were supported by the very high bootstrap value i.e. 100%

where as node for the invertebrates were supported by the lower bootstrap values. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed.

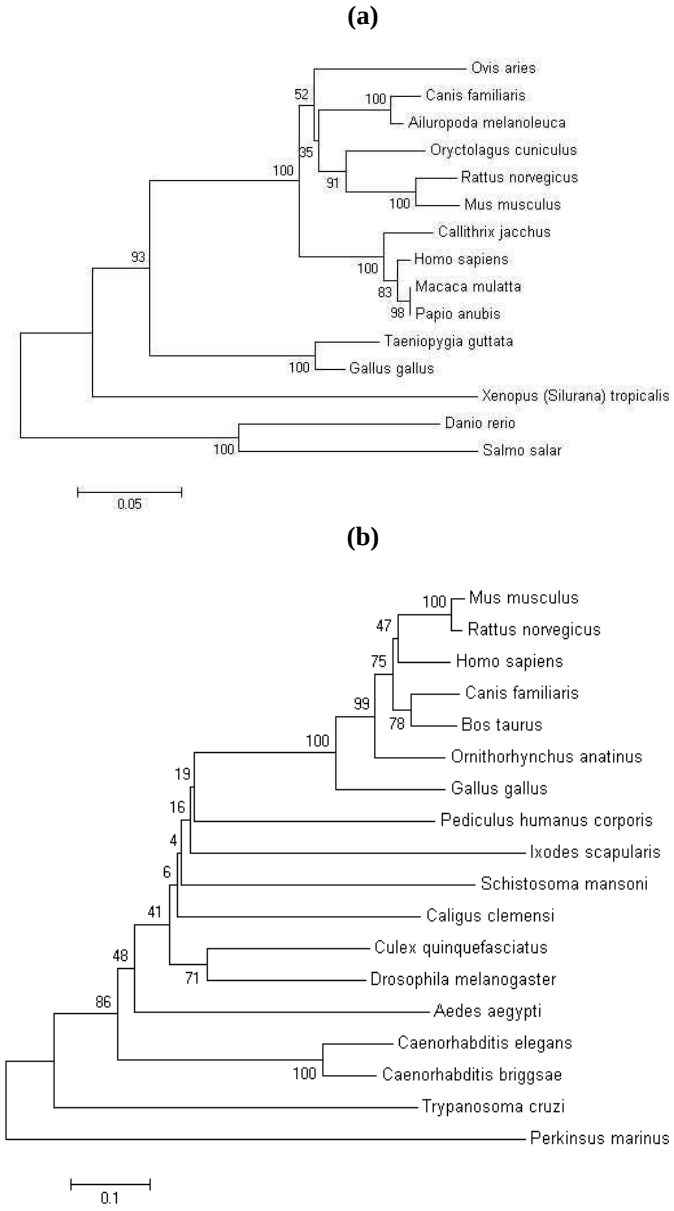


Figure 3- Bootstrap original phylogenetic tree of Sialin (a) and Cystinosin (b) proteins created by Neighbour-joining method showing bootstrap support values on the nodes . The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the *phylogenetic tree*. All the characters were given equal weights

There were a total of 405 and 223 positions in the final dataset of sialin and cystinosin respectively. This tree gave an idea about the evolutionary order of the proteins. This phylogeny was not seemed to be completely consistent with the current view of taxonomy perhaps due to use of a specific protein rather than complete genomes.

CONCLUSION

This work represented the first comparative genomics based evolutionary analysis of the sialin and cystinosin proteins across family of organisms with special reference to mammals. This comparative analysis on sequences inferred that both the proteins were basically hydrophobic in nature with their specific conserved regions that can be used as marker to identify and analyze the lysosomal disorders. These proteins showed very slight randomness in their sequences as they had very low entropy. Sialin, comparatively has short conserved domains but in greater number which help in the analysis of metabolic disorders. Amino acid composition and its frequency within the sequences and their hydrophobicity profile also assist to predict the exact function of the proteins in the lysosomal

membrane and can allow a better understanding of their critical regions. These approaches help to reveal the molecular basis of the SASD and cystinosis diseases and cell biologic features of responsible proteins for these metabolic disorders and their sub-cellular localization.

Acknowledgement
MD is thankful to Department of Science and Technology, New Delhi for a research fellowship. The work has been supported by a DBT-BIF Grant to DKG under its BTISNet scheme.

REFERENCES

[1] Neufeld, E. F, *Annu Rev Biochem,* **1991**, 60, 257-80.
[2] Mancini, G. M., Havelaar, A. C., and Verheijen, F. W., *J Inherit Metab Dis*., **2000**, 23(3), 278-92.
[3] Verheijen F W et al., *Nat Genet*. **1999**, 23(4):462-465.
[4] Natalia Y et al., *Neurobiology of Disease*, **2005**, 19:351 – 365.
[5] Kelm S & Schauer R., *Int Rev Cytol.* **1997**, 175:137-240.
[6] Paulson J.C. *Trends Biochem Sci.* **1989**, 14(7):272-276.
[7] Grow WA & Gordon H., *Cell Tissue Res.* **2000**, 299(2):273-279.
[8] Crocker et al., *Immunology,* **2001**, 103, 137 – 145.
[9] Altschul et al., *Nucleic Acids Research,* **1997**, 25:3389-3402.
[10] *http://www.ncbi.nlm.nih.gov/entrez* (National Centre for Biotechnology Information).
[11] Higgins D. et al., *Nucleic Acids Research,* **1994**, 22:4673-4680.
[12] Tamura K, Dudley J, Nei M & Kumar S., *Molecular Biology and Evolution*, **2007**, 24:1596-1599.
[13] Hall, T.A., *Nucl. Acids. Symp. Ser.* **1999**, 41:95-98.
[14] Altschul S.F. and Gish G., *Enzymol.* **1996**, 266:460-480.
[15] Saitou N & Nei M., *Molecular Biology and Evolution,* **1987**, 4:406-425.
[16] Natalia Y et al., *Neurobiology of Disease*, **2005**, 19:351 – 365.
[17] Zuckerkandl E & Pauling L., *Academic Press, New York,* **1965**, 97-166.
[18] Aula P et al, *Arch Neurol.,* **1979**, 36(2):88-94.
[19] Blom H.J. et al., *Biochem. J.,* **1990**, 268, 621 – 625.
[20] Eisenberg D. E. Schwarz, M. Komaromy and R.Wall, *J. Mol Biol.,* **1984**, 179(1):125-42.
[21] M Dwivedi and D K Gupta, *Journal of Computing*, 2011, 3: 6, 13, **1984**, 179(1):125-42.0-134.
[22] Felsenstein J., *Evolution,* **1985**, 39:783-791.
[23] Fukuda, M. N., Dell, A., and Scartezzini, P., *J. Biol. Chem.*, **1987**, 262:7195-7206.
[24] Kyte J and Doolittle RF., *J. Mol. Biol.,* **1982**, 157:105.
[25] A Mani, M Dwivedi, V Tripathi and D.K. Gupta, *European Journal of Experimental Biology*, **2011**, 1 (1):148-155.
[26] Laura M, Hannes V and Richard J., *The Journal of Neuroscience*, **2009**, 29(49):15355– 15365.
[27] M Dwivedi, V Tripathi, A Mani and D K Gupta, *BVICAM'S International Journal of Information Technology*, **2010**, 3: 2-1.
[28] Mancini et al., *J. Clin. Invest.,* **1991**, 87, 1329 – 1335.