

Prediction of 3-Dimensional Structure of Cathepsin L Protein of *Rattus Norvegicus*

Sunil Kumar*, Priya Ranjan Debata and Prakash C. Supakar*

Institute of Life Sciences, Nalco Square, Bhubaneswar-751023, India

*Corresponding authors: Sunil Kumar, M.Sc., Institute of Life Sciences, Nalco Square,

Bhubaneswar-751023, India; Tel: 91-674-2301500;

Fax: 91-674-2300728; E-mail: sunil20051@rediffmail.com

Prakash C. Supakar, Ph.D, Institute of Life Sciences, Nalco Square,

Bhubaneswar-751023, India, Tel: 91-674-2302783;

Fax: 91-674-2300728; E-mail: pcsupakar@hotmail.com

Received July 23, 2008; Accepted August 25, 2008; Published September 13, 2008

Citation: Sunil K, Priya RD, Prakash CS (2008) Prediction of 3-Dimensional Structure of Cathepsin L Protein of *Rattus Norvegicus*. J Proteomics Bioinform 1: 307-314. doi:10.4172/jpb.1000039

Copyright: © 2008 Sunil K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Cathepsin L is a cysteine protease which degrades connective tissue proteins like collagen, elastin and fibronectin. Increase in the expression of cathepsin L in aged kidney leading to considerable loss of organ function in old age. Recently it has been reported that SARS-CoV or SARS-CoV spike protein-pseudotyped retroviruses utilize the enzymatic activity of endosomal cathepsin L protease for viral entry. A 3D structure of rat cathepsin L was constructed in this report through homology modeling using the X-ray structure of procathepsin L from *Homo sapiens* (PDB code: 1CS8). The homology modeling was done by using the MODELLER 9v2 software. The final model obtained by molecular mechanics and dynamics method and was assessed by PROCHECK and VERIFY 3D graph, which showed that the final refined model is reliable. The model could be further explored for characterizing the protein.

Keywords: Homology modeling; Cathepsin L; Protease; Aging; Structural bioinformatics

Abbreviations: EM: Energy Minimization; BLAST: Basic Local Alignment Search Tool

Introduction

Cathepsin L is a member of the papain superfamily of lysosomal cysteine proteases and is one of the most powerful endopeptidases. Its usual function is regulating cellular protein turnover in lysosome (Kirschke et al., 1995; Kazunobu et al., 2004; Kramer et al., 2007). It plays an important role with cathepsin B and H in the degradation of both endogenous and exogenous proteins. Cathepsin L, initially translated as preprocathepsin L, is then transferred through the Golgi as procathepsin L and stored in lysosomes as mature cathepsin L. (Chauhan et al., 1993). Over expression of procathepsin L in human melanoma cells increases their tumorigenicity and switches their phenotype from non-metastatic to highly metastatic (Nathalie et al., 2004). Therefore the enforced expression and secretion of procathepsin L by human melanoma cells arms them with the ability to inactivate complement-mediated cell lyses and contributes to tumor growth and metastasis (Frade et al.,

1998). Cathepsin L is found to be upregulated in rat kidney during aging (Debata et al., 2007). Cathepsin-L influences the expression of extracellular matrix in lymphoid organs and plays a role in the regulation of thymic output and of peripheral T cell number (Lombardi et al., 2005). It was reported that in human SARS-CoV or SARS-CoV spike protein-pseudotyped retroviruses utilize the enzymatic activity of endosomal cathepsin L protease for viral entry (Huang et al., 2006; Li et al., 2006). Cathepsin L is a lysosomal cysteine protease that digests proteins of both intracellular and extracellular origin. It is translated as a precursor protein pre-procathepsin L, transferred through the Golgi apparatus as procathepsin L and then stored in lysosomes as mature cathepsin L (Ishidoh and Kominami, 1998). It plays a diverse role in different organs and tissues such as maintenance of heart structure and function (Stypmann et al., 2002), epidermal differentiation, hair follicle morphogenesis and cycling (Benavides et al., 2002), development of type 1 diabetes in NOD mouse (Maehr et al., 2005), tumor

metastasis (Lah and Kos, 1998), thyroid function (Friedrichs et al., 2003), a modifier of extra cellular matrices in pre-ovulatory follicles (Baricos et al., 1988), degradation of basement membrane in kidney (Reiser et al., 2004), podocyte migration in nephrotic syndrome (Ohshita and Hiroi, 2006), regulation of thymic output and peripheral T cell number (Honey et al., 2002) and generation of MHC class II-bound peptide ligands presented by cortical thymic epithelial cells (Debata et al., 2007). Our previous study has documented that the expression of cathepsin L gene is significantly up-regulated in rat kidney during aging (Kim et al., 2004). A similar result was also demonstrated at protein level (Deval et al., 2001). The upregulation of cathepsin L is also found in skeletal muscle wasting in septic muscle (Deval et al., 2001) and in scrapie-infected Neuro2a cells (Zhang et al., 2003).

In this communication, an effort was made to generate a three-dimensional (3D) model of cathepsin L protein based on the available template crystal structure of procathepsin L from protein data bank (PDB code: 1CS8) (Berman et al., 2000). The structural information of cathepsin L could prove useful to further characterizing the protein.

Materials & Methods

Comparative modeling of rat cathepsin L

The amino acid sequence of rat cathepsin L was retrieved from the sequence database of NCBI (www.ncbi.nlm.nih.gov) (ID: AAH63175). It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank, hence the present exercise of developing the 3D model of the rat cathepsin L was undertaken.

BLAST (Altschul et al., 1990) search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. Sequences were aligned and the one that showed the maximum identity with high score and lower e-value and 73% sequence identity was used as a reference structure to build a 3D model for rat cathepsin L. The rat cathepsin L structure was modeled by means of comparative modeling procedure using the 1CS8 as the template. The rat cathepsin L sequence was submitted to Genesilico protein fold-recognition metaserver. Fold-recognition server Fugue and 3D PSSM reported 1CS8 as the best template with highly significant score. The sequence alignment of rat cathepsin L and 1CS8 was carried out using the CLUSTAL W (Thompson et al., 1994) (<http://www.ebi.ac.uk/clustalw>) program. The alignment was manually refined at some loop regions of the template. The academic version of MODELLER 9v2 (Sali and Blundell, 1993) (<http://www.salilab.org/modeler>) was used for model building.

Backbone of the core regions of the protein were transferred directly from the corresponding coordinates of 1CS8. Side chains confirmation for backbone residues was generated automatically by homology. Out of 20 models generated by MODELLER, the one with the best G-score of PROCHECK (Laskoswki et al., 1993) and with the best VERIFY3D (Luthy et al., 1992) profile was subjected to energy minimization. The distance-dependent dielectric constant $\epsilon = 1.0$ and non binding cutoff of 14 Å. CHARM (Brooks et al., 1983) force field and CHARM-all-atom charges were used for the energy minimization. Initially an 800 step steepest descent algorithm was used to remove close Van der waals contacts, followed by the 1000 iteration conjugate gradient minimization until the maximum derivative is less than 20.0 kcal.mol⁻¹.nm⁻¹. All hydrogen atoms were included during the calculation. The above energy minimization was started with the core main chain, then all the core side chains. All calculations were performed by using Accelrys DS Modeling 2.0, (Accelrys Inc. San Diego, CA 92121, USA) software suite. During these steps the quality of the initial model was improved. VERIFY3D was used to check the residue profiles of the three-dimensional models. In order to assess the stereo-chemical qualities of the three dimensional models PROCHECK analysis was performed and Ramachandran plot was drawn.

Results and Discussion

BLAST (Altschul et al., 1990) search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. Based on the maximum identity with high score and lower e-value in the BLAST search, 1CS8 (PDB code) is used as the structural template for modeling the rat cathepsin L protein. The sequence – structure alignment used for model building shown in Figure 1. The alignment is characterized by some insertions and deletions in the loop regions. Since the first 17 residues from the N-terminal end did not have corresponding equivalent regions in 1CS8, the modeling was carried out from the 18th to the 317th residue, followed by a rigorous refinement of the model by means of EM and the final stable structure of the rat cathepsin L obtained is shown in Figure 2. The model has 89% of the residues in the most favored regions of the Ramachandran Map Figure 3 with a PROCHECK G-score value of 0.03 and a satisfactory VERIFY-3D profile. The predicted 3-D model of rat cathepsin L protein will be very useful while studying the real structure of the protein.

Validation of the model was carried out after the refinement process using Ramachandran Map calculations computed with the PROCHECK program. The Φ and Ψ distributions of the Ramachandran Map of non-glycine, non-proline residues are summarized in Figure 3 and table 1. The model has 89% of the residues in the most favored regions

```

1CS8      SLTFDHSLEAQWTKWKAMHNRLYGMNEEGWRRRAVWEKNNMKMIELHNQVEYREGKHSFTMAM 60
rat5      TPKEFDQTFNAQWHQWKSTHRRLYGTNEEEWRRRAVWEKNMRMIQLHNGEYSNGKHGFTMEM 60
          : . * * : : : * * * : * * : * . * * * * * * * * * * * * * * * * * * * * * *
1CS8      NAFGDMTSEEFQVMNGFQNRKPRKGKVFQEPLFYEAPRSVDWREKGYVTPVKNQGQCGS 120
rat5      NAFGDMTNEEFQIVNGYRHQKHKKGRLFQEPLMLQIPKTVDWREKGCVTPVKNQGQCGS 120
          * * * * * * . * * * * * : : * * : : : * * : * * : * * * * * * : : * * : * * * * * * * * * * * * *
1CS8      CWAFSATGALEGQMFRKTGRLLISLSEQNLVDCSGPQGNEGCNGGLMDYAFQYVQDNGGLD 180
rat5      CWAFSASGLEGQMFLKTKGLLISLSEQNLVDCSHDQGNQGCNGGLMDFAFQYIKENGGGLD 180
          * * * * * * : * . * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
1CS8      SEESYPYEATEESCKYNPKYSVANDAGFVDIPQEKALMKAVATVGPISVAIDAGHESFL 240
rat5      SEESYPYEAKDGSCKYRAEYAVANDTGFVDIPQEKALMKAVATVGPISVAMDASHPSLQ 240
          * * * * * * * * . : * * * * . : : * * * * : * * * * * * * * * * * * * * * * * * * * * * * * * *
1CS8      FYKEGIYFEPDCSSEDMDHGVLVVGYGFEESTESDNNKYWLVKNSWGEWGMGGYVKMAKD 300
rat5      FYSSGIYYEPNCSSKDLDHGVLVVGYGEGTDSNKKYWLVKNSWGKEWGMGDGYIKIAKD 300
          * * . . * * * : * * : * * * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
1CS8      RRNHCGIASAASYPTV- 316
rat5      RNNHCGLATAASYPIVN 317
          * . * * * * * : * : * * * * * *
    
```

Figure 1: Sequence alignment of cathepsin L from Rat with PROCATHEPSIN L of Homo sapiens (PDB code 1CS8) done using CLUSTAL W server that was subsequently submitted to MODELLER. The conserved regions are indicated by ‘*’.

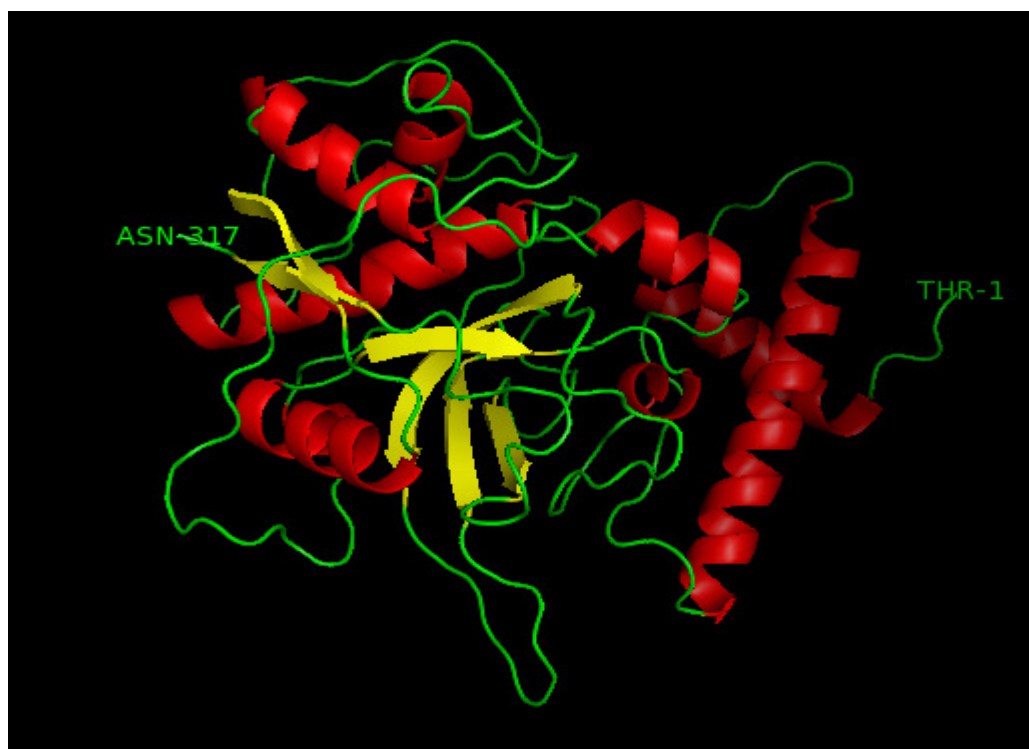


Figure 2: The final 3D structure of rat cathepsin L. The α -helices and β -sheets are represented by red helices and yellow ribbons respectively.

of the Ramachandran Map with a PROCHECK G-score value of 0.03 and a satisfactory VERIFY3D profile.

The structural superimposition of C trace of template and rat cathepsin L is shown in Figure 4. The weighted root mean square deviation of C trace between the template and final refined model was 0.43 Å which suggest that the model is reliable.

The amino acid sequences of template and final structure are generated using JOY server (protein sequence-structure representation and analysis (Mizuguchi et al., 1998), were aligned using CLUSTALW. Given their PDB files, secondary structures were also analyzed and compared by the JOY program. The secondary structures of template and final model of rat cathepsin L are highly conserved which showed that final model is highly reliable as shown in Figure 5.

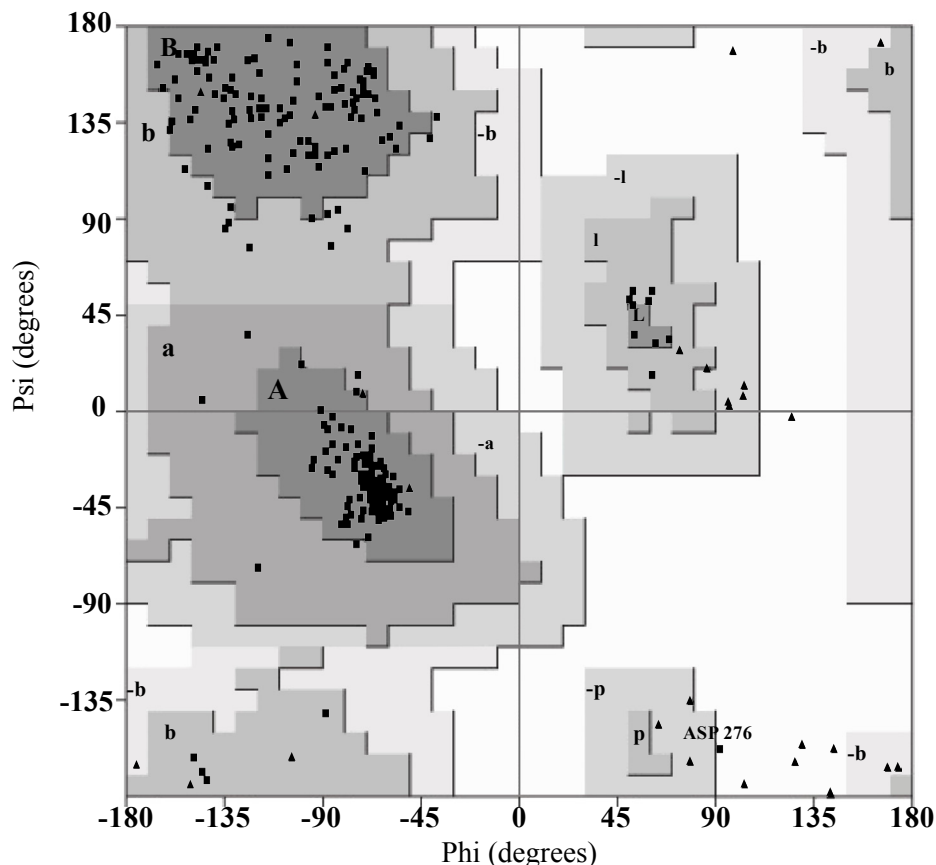


Figure 3: Ramachandran’s Map of rat cathepsin L protein. The plot calculation on 3D model of rat cathepsin L protein was calculated with the PROCHECK program.

Residue in most favored regions	89.0%
Residue in the additionally allowed zones	10.7%
Residue in the generously regions	0.00%
Residue in disallowed regions	0.3%
Non-glycine and non-proline residues	100. %

Table 1: Ramachandran plot calculation for 3D model of rat cathepsin L computed with the PROCHECK program.

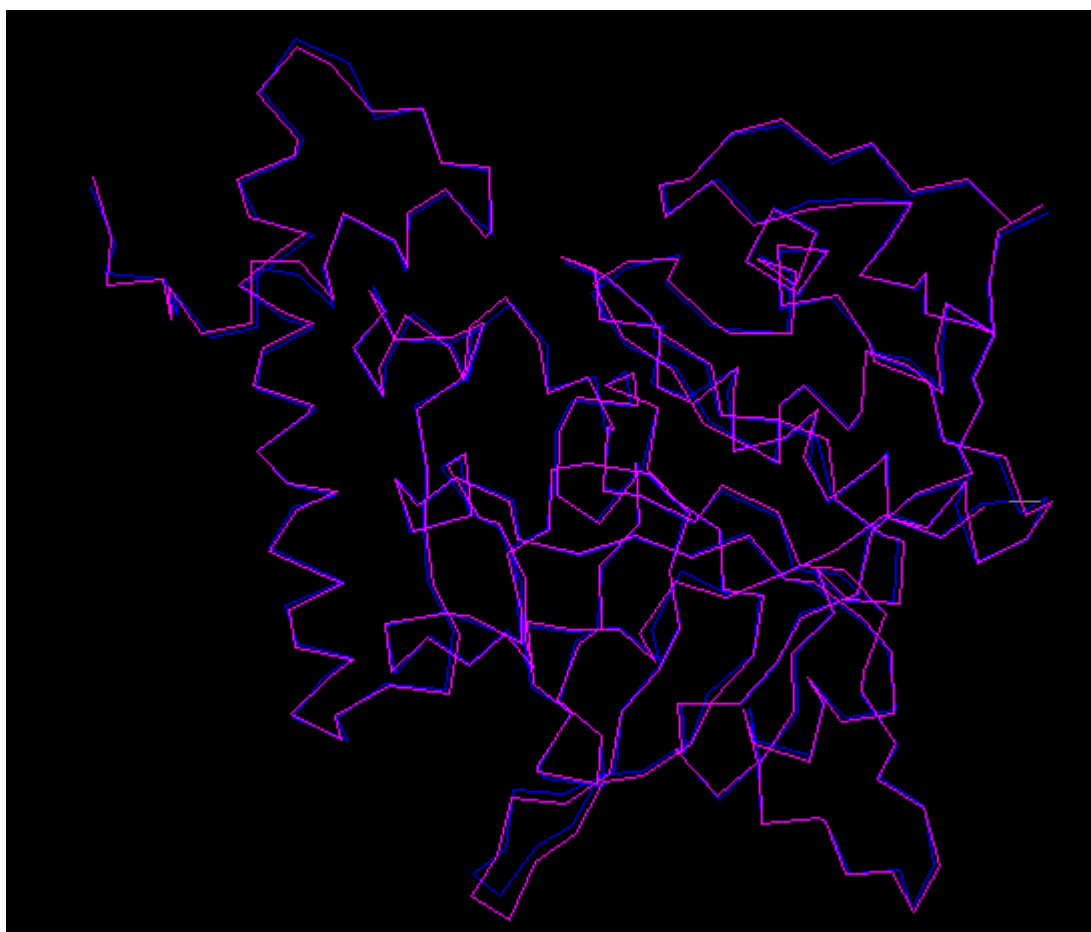


Figure 4: Superimposition of C α trace of cathepsin L (represented in blue color) and 1CS8 (represented in pink color).

Conclusion

In this report, a molecular model of rat cathepsin L protein has been constructed through homology modeling which could be used for further characterization.

Acknowledgement

This work was supported by funds from Distributed Information Sub-Center (BT/B1/04/058/2002), Department of Biotechnology, Ministry of Science and Technology, Government of India and Institutional funds of the Institute of Life Sciences, Bhubaneswar, India.

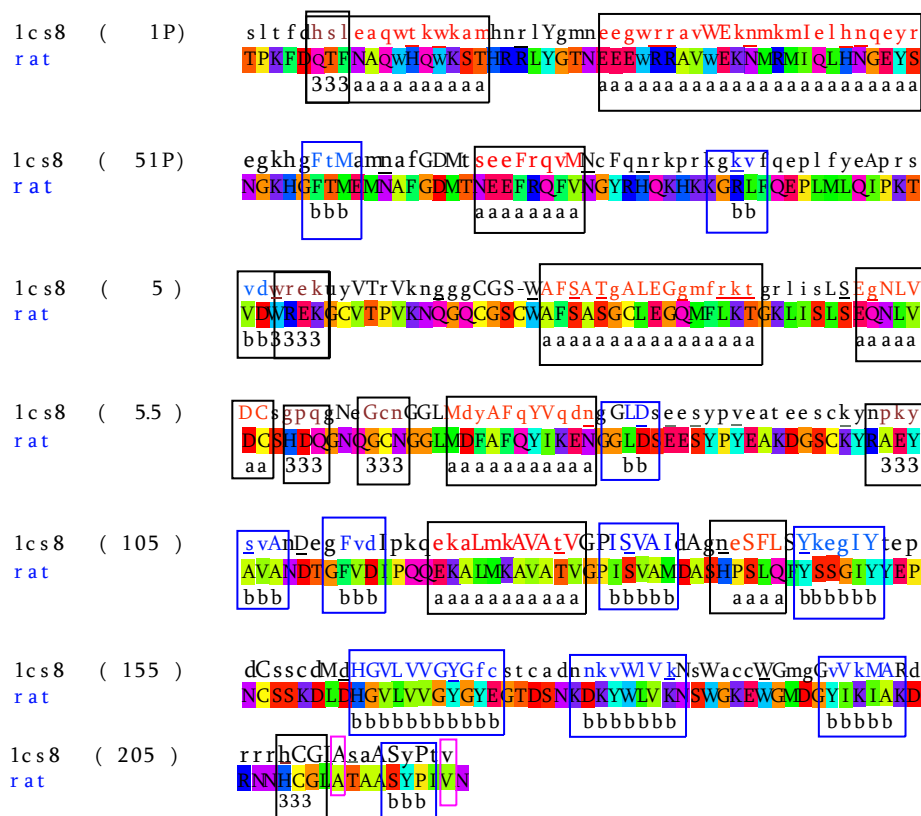


Figure 5: Structure based sequence alignment of template and final structures of the rat cathepsin L using JOY program. The key to the JOY annotation is as follows: lowercase red letter, α -helix; lowercase blue letter, β -strand; lowercase maroon letter, 3_{10} -helix; uppercase letter, solvent-inaccessible residue; lowercase letter, solvent-accessible residue; italic lowercase letter, positive ϕ .

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
- Baricos WH, Zhou Y, Mason RW, Barrett AJ (1988) Human kidney cathepsin B and L Characterization and potential role in degradation of glomerular basement membrane. *Biochem J* 252: 301-4. » [Pubmed](#) » [Google Scholar](#)
- Benavides F, Starost MF, Flores M, Gimenez-Conti IB, Guenet JL, et al. (2002) Impaired hair follicle morphogenesis and cycling with abnormal epidermal differentiation in nackt mice, a cathepsin L-deficient mutation. *Am J Pathol* 161: 693-703. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235-242. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* 4: 187-217. » [CrossRef](#) » [Google Scholar](#)
- Chauhan SS, Popescu NC, Ray D, Fleischmann R, Gottesman MM, et al. (1993) Cloning, genomic organization, and chromosomal localization of human cathepsin L. *J Biol Chem* 268: 1039-45. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
- Debata PR, Panda H, Supakar PC (2007) Altered ex-

- pression of trefoil factor 3 and cathepsin L gene in rat kidney during aging. *Biogerontology* 8: 25-30. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Debata PR, Panda H, Supakar PC (2007) Altered expression of trefoil factor 3 and cathepsin L gene in rat kidney during aging. *Biogerontology* 8: 25-30. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Deval C, Mordier S, Obled C, Bechet D, Combaret L, et al. (2001) Identification of cathepsin L as a differentially expressed message associated with skeletal muscle wasting. *Biochem J* 360: 143-50. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. Frade R, Rodrigues-Lima F, Huang S, Xie K, Guillaume N, et al. (1998) Procathepsin-L, a proteinase that cleaves human C3 (the third component of complement), confers high tumorigenic and metastatic properties to human melanoma cells. *Cancer Res* 58: 2733-2736. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Friedrichs B, Tepel C, Reinheckel T, Deussing J, Figura KV, et al. (2003) Thyroid functions of mouse cathepsins B K and L. *J Clin Invest* 111: 1733-45. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Honey K, Nakagawa T, Peters C, Rudensky A (2002) Cathepsin L regulates CD4⁺T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. *J Exp Med* 195: 1349-1358. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Huang IC, Bosch BJ, Li F, Li W, Lee KH, et al. (2006) SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *J Biol Chem* 281: 3198-203. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
14. Ishidoh K, Kominami E (1998) Gene regulation and extracellular functions of procathepsin L. *Bol Chem* 131-5. » [Pubmed](#) » [Google Scholar](#)
15. Tsunemoto K, Osatomi K, Nozaki Y, Hara K, Ishihara T (2004) Molecular characterization of cathepsin L from hepatopancreas of the carp *Cyprinus carpio*. *Comp Biochem Physiol B Biochem Mol Biol* 137: 107-114. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Kim CH, Park DU, Chung AS, Zou Y, Jung KJ, et al. (2004) Proteomic analysis of post-mitochondrial fractions of young and old rat kidney. *Exp Gerontol* 39: 1155-68. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Kirschke H, Barrett AJ, Rawlings ND (1995) Proteinases 1: lysosomal cysteine proteinases. *Protein Profile* 2: 1581-1643. » [Pubmed](#) » [Google Scholar](#)
18. Kramer G, Paul A, Kreuzsch A, Schu"lwer S, Wiederanders B, et al. (2007) Optimized folding and activation of recombinant procathepsin L and S produced in *Escherichia coli*. *Protein Expr Purif* 54: 147-156. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
19. Lah TT, Kos J (1998) Cysteine proteinases in canochemical quality of protein structures. *J Appl Cryst* 26: 283-291.
20. Li F, Berardi M, Li W, Farzan M, Dormitzer PR, et al. (2006) Conformational States of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J Virol* 80: 6794-800. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
21. Lombardi G, Burzyn D, Mundinano J, Berguer P, Bekinschtein P, et al. (2005) Cathepsin-L influences the expression of extracellular matrix in lymphoid organs and plays a role in the regulation of thymic output and of peripheral T cell number. *J Immunol* 174: 7022-32. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
22. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83-85. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
23. Maehr R, Mintern JD, Herman AE, Lennon-Dumenil AM, Mathis D, et al. (2005) Cathepsin L is essential for onset of autoimmune diabetes in NOD mice. *J Clin Invest* 115: 2934-43. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
24. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics* 14: 617-623. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
25. Rousselet N, Mills L, Jean D, Tellez C, Bar-Eli M, et al. (2004) Inhibition of tumorigenicity and metastasis of human melanoma cells by anti-cathepsin L single chain variable fragment. *Cancer Res* 64: 146-51. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
26. Ohshita T, Hiroi Y (2006) Cathepsin L also plays a role in the lysosomal degradation of L-lactate dehydrogenase. *Biosci Biotechnol Biochem* 70: 2254-61. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
27. Reiser J, Oh J, Shirato I, Asanuma K, Hug A, et al. (2004) Podocyte Migration during nephrotic syndrome requires a coordinated interplay between cathepsin L and integrin. *J Biol Chem* 279: 34827-32. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
28. Sali A, Blundell TL (1993) Comparative protein model-

- ing by satisfaction of spatial restrains. *J Mol Biol* 234: 779-815. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
29. Stypmann J, Glaser K, Roth W, Tobin DJ, Petermann I, et al. (2002) Dilated cardiomyopathy in mice deficient for the lysosomal cysteine peptidase cathepsin L. *Proc Natl Acad Sci USA* 99: 6234-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
30. Salustri A, Camaioni A, Di Giacomo M, Fulop C, Hascall VC (1999) Hyaluronan and proteoglycans in ovarian follicles. *Hum Reprod Update* 5: 293-301. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
31. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
32. Zhang Y, Spiess E, Groschup MH, Burkle A (2003) Up-regulation of cathepsin B and cathepsin L activities in scrapie-infected mouse Neuro2a cells. *J Gen Virol* 84: 2279-83. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)