http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 3 (4): 639-641, 2000

Amino Acid Sequence Homology and Modeling of Cysteine Protease Cathepsin S

Rukhshan Khurshid* and Asmat Saleem**
*Department of Biochemistry, Fatima Jinnah Medical College, Lahore
**Department of Protein Chemistry, HEJ Research Institute of Chemistry, University of Karachi, Pakistan

Abstract: There is an increasing evidence for the role of cysteine proteases cathepsin B, L and S in cancer progression. Primary structure of human cathepsin S was aligned with other human cathepsins. According to the alignment, the amino acid sequences of cathepsin S shows 43.3, 22, 29, 42, 56.6, 37.7, 43 and 33% identity with papain and human cathepsin W, B, H, L, O, K and C. Amino acid sequence of human cathepsin L shares the highest degree of identity with cathepsin S, whereas low sequence identity was observed in case of cathepsin W. The 3D model of cathepsin S has been constructed using the crystal co-ordinates of human omega protease. The structure of cathepsin S shows that the molecule is folded into two distinct domain designated as left hand, that mainly consists of beta pleated sheet, and right hand, that consist of alpha helix interacting with each other through the extended polar interface. The active site of protease was predicted and mode of binding with substrate was discussed in the light of constructed model.

Key words: Amino acid sequence, three-dimensional structure, cathepsin S

Introduction

Like cysteine proteases, cathepsin S synthesizes as preproenzyme which is processed to the corresponding proenzyme and is targeted to lysosome by mannose 6-phosphate signal. The enzyme is further processed to mature forms, which consist of single polypeptide chain. In some cases, precursors of these lysosomal proteases escape from the processing pathway, continue along the secretary route, enter the storage granules and are finally released into extracellular space (Petanceska and Devi, 1992).

These cysteine proteases including cathepsin S are implicated in many diverse cellular events, including receptor internalization, protein turnover, neuronal degeneration as well as regeneration during development (Kornfeld and Mellmann, 1989) and antigen processing (Golde *et al.,* 1992), bone remodeling (Tezuka *et al.*, 1994) and prohormone processing (Krieger and Hook, 1991). In addition, these appear to be involved in a variety of pathological processes such as glomerulonephritis, Alzheimer's disease, cancer invasion and metastasis (Sloane, 1990).

It is hypothesized that protease secreted from transformed cells, degrade extracellular matrix and thus contribute to tumor invasion. Tumor invasion and degradation is thought to occur in extracellular environment where pH is neutral. It is observed that cathepsin S is active at this pH, and plays a role in matrix degradation (Portnoy *et al.,* 1986). Cathepsin S has been identified and isolated from breast cancerous tissue employing gel filtration and reverse phase HPLC (Khurshid, 2000).

Purpose of study is to show homology of cathepsin S with other cathepsins of homo-sapines. 3 dimensional structure of cathepsin S constructed to identify the active site, specify the mode of binding of substrate and elucidate the enzymatic mechanism of cathepsin S.

Materials and Methods

Amino acid sequences of cathepsins (B,H,L,O,K,C) except cathepsin W (Linnevers *et al.,* 1997) were taken from Swiss-Port protein sequence data bank (Bairoch and Apweiler, 1997). A model of of cathepsin S constructed using Omega protease as templates by program Modeler (Sali and Blundell, 1993).

Results and Discussion

According to the alignment, the amino acid sequences of cathepsin S shows 43.3, 22, 29, 42, 56.6, 37.7, 43 and 33% identity with papain and human cathepsin W, B, H, L, O, K and C. Amino acid sequence of human cathepsin L shares the highest degree of identity with cathepsin S, whereas low sequence identity was observed in case of cathepsin W (Fig. 1a, b).

Fig. 1a: Ribbon diagram representing 3D structure of cystatin taken from protein data bank

Fig. 1b: Ribbon plot of cathepsin S of polypeptide chain folded into 7 hilices, 18 beta sheets and 25 turns containing proline

Fig. 2: Ribbon diagram of cathepsin S representing non-polar amino acid mostly in hydrophobic core

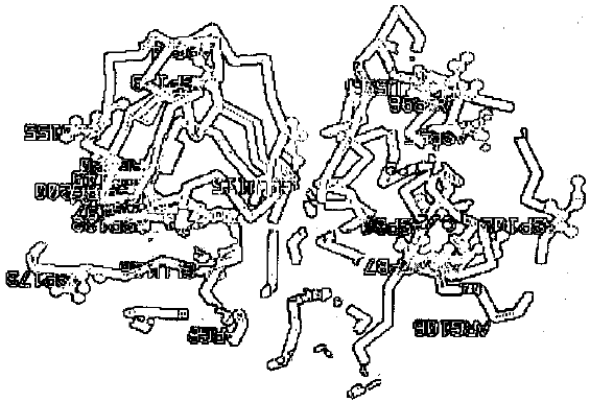

Fig. 3: Ribbon diagram of cathepsin S model shows polar charged groups on surface of enzyme

Fig. 4: Schematic represe ntation showing the catalytic domain (Cys^{25} , His^{164}) of cathepsin S

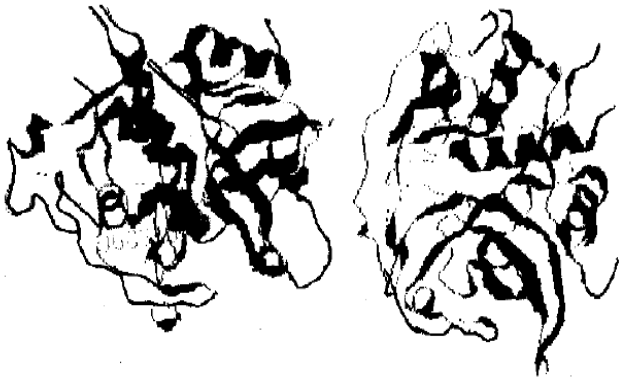

Fig. 5: Ribbon diagram of cathepsin B with cys^{25} and his^{200} present in active center

Fig. 6: Ribbon diagram of papin with catalytically active site of Cys^{25} and His^{159}

The comparison of cysteine proteases facilitated by arbitrarily dividing the sequences into three region- an amino terminal region, a central and a carboxyl terminal region. Takio *et al.* (1983) reported that both N and C terminal parts of these cysteine proteases are conserved more than the central region. An almost identical 11 residue sequence is found in all cysteine proteases in the vicinity of active site cystenyl residue. Sequences in the central region are more difficult to align because identity is low. Degree of identity among sequences is high in carboxyl terminal region.

The amino acid residue cys^{25} and his^{164} are proposed to be important for catalytic activity. It was observed that in all cysteine proteases, amino acid directly flanking cys^{25} is highly conserved whereas his^{164} is not identical. The alignment of cathepsin S with other cathepsin clearly shows the difference in cathepsins. Important difference at position 12 where cysteine was found in cathepsin S but not in other cysteine proteases. Residue probably formed disulfide bridge at position 110 and was not found in other cathepsins. A peculiarity of the sequence of cathepsin S is found around the active site cysteine residue. Usually in human and other mammalian cathepsin, the first residue upstream from the catalytic cysteine residue is a serine residue. In cathepsin S however a hydrophobic residue alanine replaces this residue. Neutral nature of cathepsin S may be due to this amino acid. The

enzyme range in size and alignment with other enzyme imply that cathepsin S belongs to branch one enzyme of papain rather than branch 2. A rule based structural model of human cathepsin S was constructed on the basis of known Omega protease, using Modeler program. Omega protease used as template. The over all structure of human cathepsin S is similar to omega protease. This shows 44.3% identity. Among 216 residue of mature cathepsin S, 212 carbon alpha atoms can be superimposed onto the corresponding carbon alpha atom of omega protease. Resolution of model of cathepsin S is 2.5A.

Cathepsin S is a compact molecule with little empty space inside. Main chain is folded into 7 helices, 18 beta strands and 25 turns. Proline is in turns and this disrupts alpha helices because of its rigid 5 membered ring. The structure of cathepsin S shows that the molecule is folded into two distinct domain designated as left hand, that mainly consists of beta pleated sheet, and right hand, that consist of alpha helix interacting with each other through the extended polar interface (Fig. 2).

It was predicted that conserved regions tend to form beta sheet. Model shows that left side domain is formed by N terminal half of polypeptide chain and the right by the C terminal domain. Terminal carboxylate group would near to the exposed imidazole ring of his^{164} and thus more interactive electrostatically. This would assist in fixing the substrate with respect to active site. Cathepsin S contains highly packed non-polar interior regions consisting of leucine, valine, methionine and phenlalanine (Fig. 3) whereas the polar region is mainly on the surface and consists of aspartate, glutamate, arginine and lysine (Fig. 4). Two disulfide and 123 H bonds further stabilizes the enzyme. Although 6 cystiene residues are selected for model, but in structure only 4 cysteine residue are presented to be involved in disulfide bridges whereas the remaining cysteine residue have free SH group. One at side i.e., cys^{25} bears the active site SH group of enzyme and other at the surface of lobe.

Active side cleft is located at top of the polar interface face region (Fig. 5). As we know that cys^{25} and his^{164} are catalytically active and participate in peptide bond hydrolysis. Catalysis proceeds through thiol ester. Nucleophilicity of thiol group is enhanced by the adjacent histidine that serves as proton acceptor (Cys-CH-S-C-R). Stryer (1992) proposed that for hydrolysis of peptide substrate, substrate enzyme complex is formed in the intermediate stage. At this stage carboxyl component of substrate is esterified to the thiol group of cys^{25} and his^{164} that accepts a proton from cysteine markedly enhances the nucleophilicity of Cys-SH as cysteine attacks the carbonyl carbon atom of substrate. The resulting positively charged histidine is stabilized by the electrostatic interaction with negatively charged asparagine. Kirschke and Wiederanders (1994) reported that Cysteine, Histidine and Asparagine form the catalytic triad, that is the heart of catalytic action of all cysteine proteases. They also observed that cathepsin S is a real endopeptidase since its binding cleft is not satirically hindered.

An X-ray study shows that 3D structure of papain (Fig. 6) and cathepsin S is vary similar but differ from cathepsin B. As both have cys^{25} and his^{164} as catalytic triad with single polypeptide chain whereas cathepsin B, have two chains with different active sites (Takio *et al.,* 1983).

Result not only reveals the structural features of cathepsin S but also provide insight into the evolutionary relationship among mammalian proteases.

References

Bairoch, A. and R. Apweiler, 1997. The Swiss-Prot protein sequence data bank and its supplement TrEMBL. Nucl. Acids Res., 25: 31-36.

Golde, T.E., S. Estus, L.H. Younkin, D.J. Selkoe and S.G. Younkin, 1992. Processing of the amyloid protein precursor to potentially amyloidogenic derivatives. Science, 255: 728-730.

Khurshid, R., 2000. Isolation identification of cysteine protease cathepsin S from breast cancerous tissue. Pak. J. Biol. Sci., 3: 508-510.

Kirschke, H. and B. Wiederanders, 1994. Cathepsin S and related lysosomal endopeptidases. Methods Enzymol., 244: 500-511.

Kornfeld, S. and I. Mellmann, 1989. Magnesium 6 phsophate receptor and lysosomal enzyme targeting. Annu. Rev. Cell Biol., 5: 483-490.

Krieger, T.J. and V.Y. Hook, 1991. Purification and characterization of a novel thiol protease involved in processing the enkephalin precursor. J. Biol. Chem., 266 : 8376-8383.

Linnevers, C., S.P. Smeekens and D. Bromme, 1997. Human cathepsin W, a putative cysteine protease predominantly expressed in $CD8^+$ T lymphocytes. FEBS Lett., 405: 253-259.

Petanceska, S. and L. Devi, 1992. Sequence analysis, tissue distribution and expression of rat cathepsin S. J. Biol. Chem., 267: 26038-26043.

Portnoy, D.A., A.H. Erickson, J. Kochan, J.V. Ravetch and J.C. Unkeless, 1986. Cloning and characterization of a mouse cysteine proteinase. J. Biol. Chem., 261: 14697-14703.

Sali, A. and T.L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol., 234: 779-815.

Sloane, B.F., 1990. Cathepsin B and cystatins: Evidence for a role in cancer progression. Semin. Cancer Biol., 1: 137-152.

Stryer, L., 1992. Biochemistry. 3rd Edn., W H Freeman Co., New York, pp: 141-142.

Takio, K., T. Towatari, N. Katunuma, D.C. Teller and K. Titani, 1983. Homology of amino acid sequences of rat liver cathepsins B and H with that of papain. Proc. Natl. Acad. Sci. USA., 80: 3666-3670.

Tezuka, K., Y. Tezuka, A. Maejima, T. Sato and K. Nemoto *et al*., 1994. Molecular cloning of a possible cysteine proteinase predominantly expressed in osteoclasts. J. Biol. Chem., 269: 1106-1109.