

Characterization of Keratitis Disease Causing 56k Cysteine Protease Encoding Gene from *Serratia marcescens*

Anusha Bhaskar*, Akhilesh Ugrade and Kavitha P

Centre for Research and Development, PRIST University, Vallam, Thanjavur 614 403, India

Abstract

The present study aims at characterization of keratitis causing 56k protease encoding gene, and the homology modeling of this protease from *Serratia marcescens*. Keratitis is commonly caused by many bacteria such as *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Serratia marcescens*, viruses and fungi. The aim of the study is to isolate and culture *Serratia marcescens*, to identify it at the molecular level by amplifying its 16S rDNA and the 56k protease which is thought to be responsible for causing the disease. Both the 16S rDNA and the 56k protease gene were sequenced, and the same deposited in the GenBank with the accession numbers FJ588709 and FJ810078 respectively. The protein 56k protease was modelled and the protein was studied using bioinformatics approach.

Keywords: *Serratia marcescens*; Keratitis; Cysteine protease; Homology modeling

Introduction

Keratitis is an inflammation of the cornea, the outermost part of the eye. It usually is caused by a plethora of bacteria, fungi and virus. It is a serious condition in which blood vessels grow into the cornea. Bacterial keratitis is a sight-threatening process. A particular feature of bacterial keratitis is rapid progression and corneal destruction, which may be complete in 24-48 hrs with some of the more virulent bacteria. Corneal ulceration, stromal abscess formation, surrounding corneal edema, and anterior segment inflammation are characteristics of this disease.

Serratia marcescens is a Gram negative, bacillus shaped bacteria that belongs to the family Enterobacteriaceae [1]. It is differentiated from Gram negative bacteria, as it is able to perform casein hydrolysis. *Serratia marcescens* is a potential cause of infectious keratitis that appears to be associated with abnormal corneal surface, topical medications and contact lens wear. It can also cause refractory keratitis resulting in corneal perforation and blindness. *Serratia marcescens* produces a cysteine protease, in addition to metalloproteases. These bacterial proteases have a number of biological activities, such as degradation of tissue constituents and host defense-oriented proteins, as well as activation of zymogens through limited proteolysis.

The aim of the present work was to isolate the genomic DNA from *Serratia marcescens*, and to amplify the 56 kDa cysteine protease gene. The cysteine protease gene from *Serratia marcescens* was for the first time amplified and sequenced in our lab, and as it was a preliminary study we were able to perform only a partial amino acid sequence. Since this is the first data regarding the sequence of cyteine protease of *Serratia marcescens*, we thought it necessary to study its structure. The structure of cysteine protease for *S. marcescens* has not yet determined experimentally (X-ray of NMR), therefore models are built following homology modeling protocol through computational methods. The structure prediction by homology modeling can help in understanding the 3D structure of the given protein and this in turn, will help to unravel its function. Therefore, in this paper we report for the first time, the *in silico* analysis, and homology modeling studies of cysteine protease was performed, as the 3D structures for these proteins are not available.

Materials and Method

Corneal scrapings were collected after instillation of 4% lignocaine, without preservative under aseptic conditions, from each ulcer by an ophthalmologist using a sterile Bard Parker blade. Scrapings were performed under magnification of slit-lamp or operating microscope. Leading edge and base of each ulcer were scraped initially and the material obtained, were directly inoculated onto the surface of solid media such as sheep blood agar, chocolate agar and Sabouraud Dextrose Agar (SDA) and cultured. The battery of samples was processed with traditional isolation processes and was subjected to molecular identification by extracting genomic DNA. Isolation of DNA from *S. marcescens* was performed by CTAB/NaCl method [2-6]. Identification of these isolates was performed by amplification and sequencing of the 16S rDNA gene, followed by comparison of the sequence with the sequences in the GenBank DNA sequence database, using BLAST and submitted to NCBI GenBank (Accession number FJ588709).

Amplification of 56k Cysteine Protease

The amplification of 56k cysteine protease was performed using specific primers (Fwd seq: aagcgcataaacgattgg, and Rev seq: acgcatggcgatgataca). The gene was sequenced and the sequence was deposited to NCBI GenBank (Accession No: FJ810078). The same sequence was used for the further studies *in silico* homology modeling analysis studies.

Sequence Alignment and Structure Prediction

The FASTA sequence of the query protein was retrieved from NCBI Entrez sequence search (<http://www.ncbi.nlm.nih.gov>). Following BLASTp run (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

***Corresponding author:** Anusha Bhaskar, Centre for Research and Development, PRIST University, Vallam, Thanjavur 614 403, India, E-mail: dranushaparthiban@gmail.com

Received July 12, 2012; **Accepted** September 07, 2012; **Published** September 09, 2012

Citation: Bhaskar A, Ugrade A, Kavitha P (2012) Characterization of Keratitis Disease Causing 56k Cysteine Protease Encoding Gene from *Serratia marcescens*. J Proteomics Bioinform 5: 226-229. doi:10.4172/jpb.1000241

Copyright: © 2012 Bhaskar A, 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.

Primary Structure Prediction

For physio-chemical characterization, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average of hydropathy (GRAVY) were computed using the ExPasy ProtParam server [7-9].

Secondary Structure Prediction

Secondary structure of this protein was predicted using the FASTA sequences of 56 k protease, and predicted using GOR IV and SOPMA [10,11].

Homology Modeling

The protein sequence was subjected for comparative homology modeling via Swiss model [12] and ESyPred3D (via The MODELLER 9v7 to generate putative 3D model [13]. The Swiss model performs the sequence alignments and searches for the putative template protein for generating the 3D model.

Sequence Subjected for Modeling

>target

MYQLQFTNLVYDTRKLTHTLEQTIINLFIGNWSNHQLQK-SICIRHGDDTSHNRYHILLIDTAHQRIKFSSIDNEGIIYILDYDD-AQHILMQPSSKQIGTSTRPIVYERLVS

Template Selection

The template selection was done by performing BLAST against PDB database, to find the structural homologs. The template was selected based on the following criteria PROCHECK and PROSA analysis.

Validation

The validation of the modeled structure was carried out using PROCHECK [14], which calculates the main-chain torsion angles, i.e. the Ramachandran Plot [15]. The models were further checked with WHAT check [16]. Structural analysis was performed and figures representations were generated with Swiss PDB Viewer [17]. The root-mean-square deviation (RMSD) values were calculated using modeler, by fitting the carbon backbone of the predicted. Finally, the all-atom models were subjected to a validation using Ramachandran plot by NIH MBI server.

Results and Discussion

Cysteine protease has been found in viruses and prokaryotes, as well as in higher organisms such as plants and mammals, including humans [18]. The cysteine proteases produced by pathogenic bacteria are considered important virulence factors and have their role in the development of many diseases [19]. The precise control of proteolytic processes is essential for appropriate functioning of cells and whole organisms. This is achieved at many levels, from regulation of protease expression, secretion and maturation through specific degradation of mature enzymes, to blockage of their activity by inhibition. A search for similar sequences using BLAST showed that the Staphostatin B and Cysteine protease from *S aureus* showed 95.9% similarity while that from *S warneri* showed only 56.6% similarity (Table 1).

Primary Structure Prediction

In this study, primary structure of 56k protease of *Serratia marcescens* were predicted using ExPasy's ProtParam server ([\[expasy.org/cgi-bin/protparam\]\(http://expasy.org/cgi-bin/protparam\)\), using the gene sequence retrieved from GenBank \(Accession no: FJ810078\) and the results are shown in Table 2. Results showed that 56k protease had 110 amino acid residues and the estimated molecular weight to be 12846.5. The calculated isoelectric point \(pI\) is useful, at pI, the solubility is least and the mobility in an electric field is zero. Isoelectric point \(pI\) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. The calculated isoelectric point \(pI\) was computed to be 6.35. The computed value, 7 indicates that the protein is acidic. The very high aliphatic index \(101.91\) indicates that this protein is stable for a wide range of temperature range. While the instability index \(37.3\) provides the estimate of the stability of protein in a test tube, The Grand Average of Hydropathy \(GRAVY\) value which is as low as -0.345 indicates better interaction of the protein with water.](http://</p>
</div>
<div data-bbox=)

Secondary Structure Prediction

The secondary structure is composed of alpha helix and beta sheets and the secondary structure is predicted using GOR IV and SOPMA. Table 3 presents the comparative analysis of GOR IV and SOPMA, from which it is clear that random coil is predominantly present, when the structure was predicted both by SOPMA and GOR, followed by extended strand and alpha helix.

Tertiary Structure Prediction

Swiss Model was used to study the 3D modeling of the 56 k protease. The assessment of the predicted models generated by Swiss-MODEL is shown in Figure 1. However, earlier studies have shown that the majority of cysteine protease profragment share similar fold and consist of two parts. In this study, we have modeled cysteine protease using SWISS-Modeller 9v7 (Figure 2) and the secondary structure as predicted by GOR IV and SOPMA are compared in Table 2. This

Accession number	Description	Score
ADL22850.1	Staphostatin B	95.9
NP645747.1	Cysteine protease (<i>Staphylococcus aureus</i>)	95.9
CAD43737.1	Cysteine protease (<i>Staphylococcus warneri</i>)	56.6

Table 1: BLAST results of cysteine protease.

Name	Accession number	Sequence Length	M. Wt	pI	-R	+R	EC	II	AI	GRAVY
Cysteine protease	FJ810078	110	12843.5	6.35	9	11	14440	37.33	101.91	-0.345

Table 2: Parameters computed using ExPasy's ProtParam tool.

Secondary structure	SOPMA	GOR
Alpha helix	10.91%	25.45%
310 helix	0.00%	0.00%
Pi helix	0.00%	0.00%
Beta bridge	0.00%	0.00%
Extended strand	41.82%	26.36%
Beta turn	7.27%	0.00%
Bend region	0.00%	0.00%
Random coil	40.00%	48.18%
Ambiguous states	0.00%	0.00%
Other states	0.00%	0.00%
Sequence length	110	110

Table 3: Secondary structure of cysteine protease by SOPMA and GOR IV.

structure has been further validated using Ramachandran plot by NIH MBI server (<http://nihserver.mbi.ucla.edu/SAVES/>) (Figure 3). Structural analysis was performed and representations were generated using Swiss PDB Viewer.

Protein Structure Validation

Ramachandran plot

The Ramachandran plot for the predicted structure performed using the NIH MBI server, shows that the protein structure holds totally 109 amino acids, in which 95 amino acids are in favored region, 4 amino acids in additionally allowed region, 0 amino acids in generously allowed region and 0 amino acids in disallowed region.

PROCHECK program

All proteases share in common, the general mechanism of a nucleophilic attack on the carbonyl-carbon of an amide bond. This results in a general acid-base hydrolytic process that disrupts the covalent bond. Different proteases utilize different strategies to generate the nucleophile and to juxtapose the nucleophile with the targeted bond. These distinctions serve as a useful classification scheme, and on this basis proteases can be grouped into four major classes: serine, cysteine, aspartate, and metalloproteins. Serine and cysteine proteases utilize their HO- and HS- side chains, respectively, directly as nucleophiles. Although not identical, the catalytic mechanisms of serine and cysteine proteases are remarkably similar. The latter two groups of enzymes utilize aspartate residues and heavy metals, respectively, to immobilize and polarize a water molecule so that the oxygen atom in water becomes the nucleophile. By the validation and 3D Errat done by using <http://>

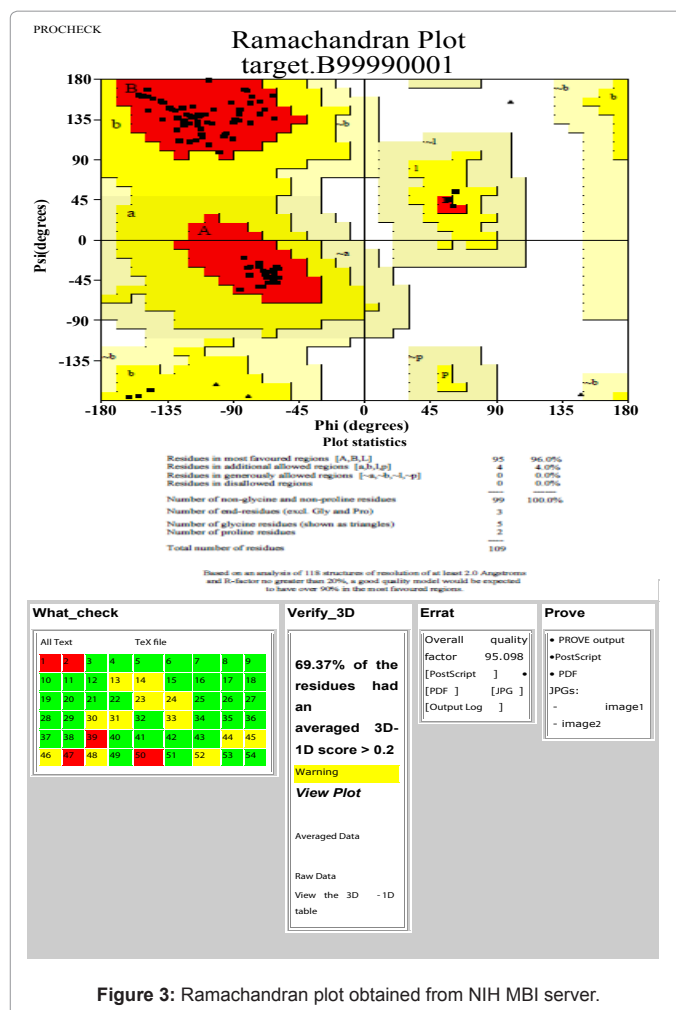


Figure 3: Ramachandran plot obtained from NIH MBI server.

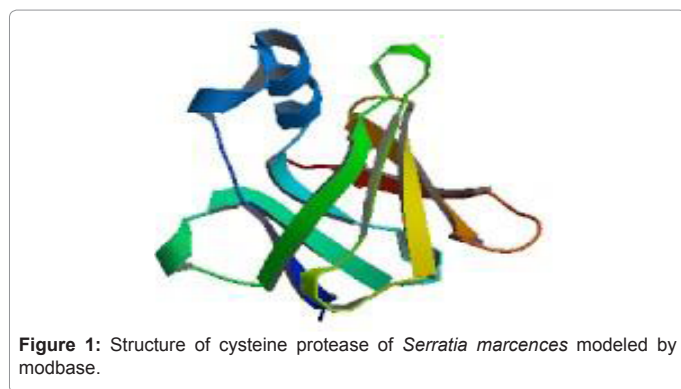


Figure 1: Structure of cysteine protease of *Serratia marcescens* modeled by modbase.

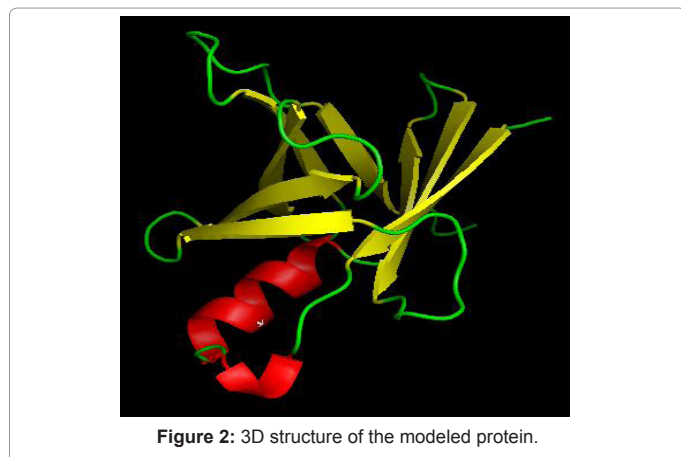


Figure 2: 3D structure of the modeled protein.

Protein structure	No. of residues in favoured region	Additional allowed regions	Generally allowed region	Disallowed region
B99990001	95	4	0	0

Table 4: Analysis of Ramachandran plot statistics of predicted model.

nihserver.mbi.ucla.edu server, 69.37% of the residues have score of 0.2 while errat shows 95.098 as a quality factor (Table 4).

Summary

In this study, 56k protease (Cysteine protease) of *Serratia marcescens* was amplified and sequenced, and the sequence was submitted to GenBank. *In silico* homology modeling of this protein was carried out. The primary structure prediction and physiochemical characterization were performed using ProtParam. The secondary and tertiary structures were modeled and then validated using PROCHECK and PROSA. This structure will provide a good foundation to further analyse this protein, for protein-protein interaction and functional analysis. Still, there is promise that the continued elucidation of specific physiological functions for cysteine proteases, will presage new therapeutic tools.

References

- Grimont F, Grimont PAD (2006) The Genus *Serratia*. Prokaryotes 6: 219-244.
- Quyen DT, Nguyen TT, Le TT, Kim HK, Oh TK, et al. (2004) A novel lipase/

- chaperone pair from *Ralstonia* sp. MI: analysis of the folding interaction and evidence for gene loss in *R. solanacearum*. Mol Genet Genomics 272: 538-549.
- Doerr M, Traub WH (1982) Proteolytic activity of selected clinical isolates of *Serratia marcescens* against human serum proteins. Zentralbl Bakteriol Mikrobiol Hyg A 252: 196-207.
 - Rzychon M, Chmiel D, Stec-Niemczyk J (2004) Modes of inhibition of cysteine proteases. Acta Biochim Pol 51: 861-873.
 - Kreger AS, Lyerly DM, Hazlett LD, Berk RS (1986) Immunization against experimental *Pseudomonas aeruginosa* and *Serratia marcescens* keratitis. Vaccination with lipopolysaccharide endotoxins and proteases. Invest Ophthalmol Vis Sci 27: 932-939.
 - Wilson K (1987) Preparation of genomic DNA from bacteria. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, Current protocols in molecular biology. Wiley Interscience, Brooklyn, NY 1: 2.4.1-2.4.5.
 - Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182: 319-326.
 - 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.
 - Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88: 1895-1898.
 - Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 266: 540-553.
 - Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-684.
 - Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37: D387-D392.
 - Lambert C, Leonard N, De Bolle X, Depiereux E (2002) ESyPred3D: prediction of proteins 3D structures. Bioinformatics 18: 1250-1256.
 - Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8: 477-486.
 - Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.
 - Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56.
 - Guex N, Peitsch MC (1997) SWISSMODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723.
 - Chapman HA, Riese RJ and Shi GP (1997) Emerging roles for cysteine proteases in human biology. Annu Rev Physiol 59: 63-88.
 - Takahashi S, Ushiyama S, Suzuki T, Ogawa K, Oda K (1997) Purification and characterization of cysteine proteinase from a baculovirus gene. Biosci Biotechnol Biochem 61: 1507-1511.