International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244

EXPLORATION OF THERMOLYSIN FROM BACILLUS THURINGIENSIS BT407

Maheswara Reddy Mallu*

Srinivasa Reddy Ronda*

Sandeep Vemula*

Abstract: *Thermolysin (EC 3.4.24.27) is a crucial enzyme which involves in hydrolysis of peptide bonds and peptide bond formation through the reverse reaction of hydrolysis. The exploration Thermolysin producing gene from Bacillus thuringiensis Bt407 with genomic and proteomics studies states that, the gens has pyramidine rich nucleotide sequence with 2682 bases and 1623334.00 Daltons (dsDNA) of molecular weight. The gene has around thirteen specific sites for the restriction sites. The gene contains both left & right primers. It codes for 879 amino acid peptide with the molecular weight of 96.3 KD. The primary structure of the protein produced by this gene reveals that it is an acidic, stable protein. Random coils, alpha helix, beta sheets and beta turns forms the secondary structure. 83 phosphorylation sites (Serine, Threonine & Tyrosine) are present in the protein along with ten N-Glycosylation site.*

Keywords*: Thermolysin, Bacillus thuringiensis, hydrolysis, Glycosylation, Peptide bonds.*

*Department of Biotechnology, KL University, Vaddeswaram-522502, Andhra Pradesh, INDIA.

INTRODUCTION

Thermolysin is a specific enzyme which catalyses the hydrolysis of peptide bonds containing hydrophobic amino acids. The same enzyme is often used for peptide bond formation through the reverse reaction of hydrolysis **[1]**. Thermolysin (EC 3.4.24.27) is a thermo stable neutral metalloproteinase enzyme produced by the many types of Gram-positive bacteria. Bacillus thermoproteolyticus is one of the thermolysin producing Gram-positive bacterium **[2]**. Thermolysin is the most stable member of a family of metalloproteinases produced by various Bacillus species. These enzymes are also termed 'neutral' proteinases or thermolysin -like proteinases (TLPs). Like all bacterial extracellular proteases thermolysin is first synthesised by the bacterium as a pre-proenzyme **[3]**. In the periplasm pre-prothermolysin is then processed into prothermolysin by a signal peptidase. The prosequence then acts as a molecular chaperone and leads to auto cleavage of the peptide bond linking pro and mature sequences. The mature protein is then secreted into the extracellular medium **[4]**. In contrast to many proteins that undergo conformational changes upon heating and denaturation, thermolysin does not undergo any major conformational changes until at least 70 °C **[5].** Thermolysin has a T50 value of 86.9 °C, making it the most thermo stable member of the TLP family **[6]**. Applications of thermolysin includes synthesis of aspartame , less bitter-tasting by-product is produced when the reaction is catalysed by thermolysin **[7]** & determining protein stability in cell lysate using the fast parallel proteolysis (FASTpp) assay **[8].**

METHODOLOGY

Various genomic & proteomic online and offline tools were used to study Thermolysin producing gene. The nucleic acid sequence of gene of *Bacillus thuringiensis Bt407* was retrieved from NCBI (Accession: NC_018877 REGION: 2545020-2547701). Genomic studies were carried out by Bioedit, ORF finder, Primer 3.0 and Genscan followed by proteomic studies to predict primary structure with Protparam & secondary structure with SOPMA **[9]**. Post translational modifications are studied with NetOGlyc, NetNGlyc and NetPhos 2.0 tools.

RESULTS AND DISCUSSION

GENOMICS

Bioedit

DNA molecule: gi|409187965:2545020-2547701 Bacillus thuringiensis Bt407 chromosome, complete genome

Length = 2682 base pairs

Molecular Weight = 812382.00 Da (ssDNA) & 1623334.00 Da (dsDNA)

G+C content = 34.45% A+T content = 65.55%

Nucleotide	Number	Mol%
A	1042	38.85
C	378	14.09
G	546	20.36
T	716	26.70

Nucleotide composition of Thermolysin producing gene from *Bacillus thuringiensis* Bt407 has been predicted by Bioedit & the results reveals that the DNA sequence is pyramidine rich with the molecular weight of more than 1623334Da (dsDNA) containing 2682 base pairs. The CG & AT content was found to be **34.45%** and **65.55%** respectively.

ORF Finder

Open Reading Frame Finder predicts the presence of the possible protein coding region sequence.

1. It was been identified that the gene codes 13 ORFs in it.
2. The largest protein-coding region (exon) was identified in the 1st frame of the direct strand from the position 1 to 2681 of length 2681 bases.

Primer 3.0

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	1319	20	60.07	50.00	4.00	2.00	TAGGAGAACAAACGGGATCG
RIGHT PRIMER	1516	20	60.02	55.00	4.00	0.00	CGCCGTTATGAGTACCACCT

SEQUENCE SIZE: 2682 INCLUDED REGION SIZE: 2682

PRIMER3 predicts the presence of the left and the right primers of length 20 residues in the oligonucleotide query. It gives the presence of the left primer starting from 1319th position

with GC content of 50% and the right primer at starting from 1516th position with GC content of 55%.

Genscan

Predicted genes/exons:

Gn.Ex	Type	S	Begin	...End	Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr.
1.01	Init	+	50	2684	2635	1	1	61	92	2046	0.556	189.85

GenScan predicts the presence of the exon in the direct strand from the position 50 to 2684.

1. The probability of the predicted output for the exon is **0.556.**
2. The predicted exon is categorized as a strong exon as the exon score is greater than 100 i.e., **189.85.**
3. The predicted peptide of the above gene is of about 879 amino acids in length.

PROTEOMICS

Primary Structure Prediction

Bioedit

Length = 879 amino acids Molecular Weight = 96336.06 Daltons

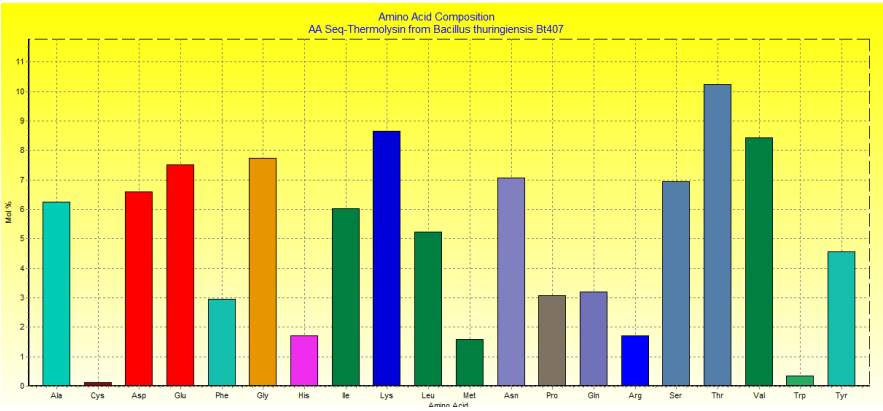

Figure 1 : Aminoacid composition of *protein*

Amino acid composition of protein has been predicted by Bioedit & the results reveal that the protein is rich in Thr, Lys and Val. The protein has a molecular weight protein of 96336.06 Daltons containing 879 amino acids.

PROTPARAM

Number of amino acids: 879 Theoretical pI: 5.02

Total number of negatively charged residues (Asp + Glu): 162

Total number of positively charged residues (Arg + Lys): 130

Amino acid composition:

Ala (A) 55	6.3%	Phe (F) 26	3.0%
Arg (R) 15	1.7%	Pro (P) 27	3.1%
Asn (N) 62	7.1%	Ser (S) 61	6.9%
Asp (D) 58	6.6%	Thr (T) 90	10.2%
Cys (C) 1	0.1%	Trp (W) 3	0.3%
Gln (Q) 28	3.2%	Tyr (Y) 40	4.6%
Glu (E) 66	7.5%	Val (V) 74	8.4%
Gly (G) 68	7.7%	Pyl (O) 0	0.0%
His (H) 15	1.7%	Sec (U) 0	0.0%
Ile (I) 53	6.0%	(B) 0	0.0%
Leu (L) 46	5.2%	(Z) 0	0.0%
Lys (K) 76	8.6%	(X) 1	0.1%
Met (M) 14	1.6%		

Total number of negatively charged residues (Asp + Glu): 124

Total number of positively charged residues (Arg + Lys): 91

Extinction coefficients:

Ext. coefficient 76100

Abs 0.1% (=1 g/l) 0.790, assuming all pairs of Cys residues form cystines

Ext. coefficient 76100

Abs 0.1% (=1 g/l) 0.790, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 21.63

This classifies the protein as stable.

Aliphatic index: 74.60

Grand average of hydropathicity (GRAVY): -0.538

The protein was acidic in nature, as it has more negatively charged residues than the positively charged ones. The protein is classified as stable, as the instability index is computed to be 21.63. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 74.6. The protein is highly hydrophilic as the Grand Average of Hydropathicity value is –0.204, which is very much lesser than 0.05.

SECONDARY STRUCTURE PREDICTION

SOPMA

Sequence length : 879

Alpha helix (Hh) : 203 is 23.09%	Beta turn (Tt) : 45 is 5.12%
3_{10} helix (Gg) : 0 is 0.00%	Bend region (Ss) : 0 is 0.00%
Pi helix (Ii) : 0 is 0.00%	Random coil (Cc) : 438 is 49.83%
Beta bridge (Bb) : 0 is 0.00%	Ambigous states (?) : 0 is 0.00%
Extended strand (Ee) : 193 is 21.96%	Other states : 0 is 0.00%

SOPMA (Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments) method was used to predict the secondary structure of protein. It was found that 23.09% of amino acids fall on Alpha helix region, 21.96% of amino acid was found to be lays in beta sheet, 5.12% in Beta turn and the remaining 49.83% tends to form random coil.

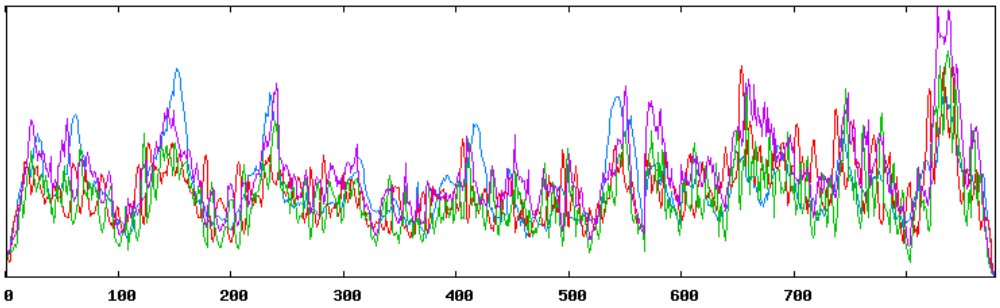

Figure 2 : Secondary structure of protein

POST TRANSLATIONAL MODIFICATIONS

NetNGlyc

Name: Sequence Length: 879

SeqName	Position	Potential	Jury agreement	N-Glyc result
Sequence	137 NLTQ	0.8014	(9/9)	+++
Sequence	148 NLSK	0.7473	(9/9)	++
Sequence	164 NDTK	0.7747	(9/9)	+++
Sequence	239 NKTS	0.6940	(9/9)	++
Sequence	243 NFTS	0.7143	(9/9)	++
Sequence	374 NISM	0.4520	(6/9)	-
Sequence	437 NWTI	0.5221	(7/9)	+
Sequence	453 NPTA	0.5611	(7/9)	+ WARNING: PRO-X1.
Sequence	458 NFSE	0.4669	(5/9)	-
Sequence	486 NSSI	0.5084	(4/9)	+
Sequence	529 NMTS	0.6761	(9/9)	++
Sequence	647 NTTA	0.4264	(6/9)	-
Sequence	835 NKTN	0.6856	(9/9)	++
Sequence	841 NISN	0.4581	(6/9)	-
Sequence	844 NNSN	0.3964	(7/9)	-

The protein has ten potential N-Glycosylation sites.

NetPhos 2.0

Phosphorylation sites predicted: Ser: 34 Thr: 28 Tyr: 21

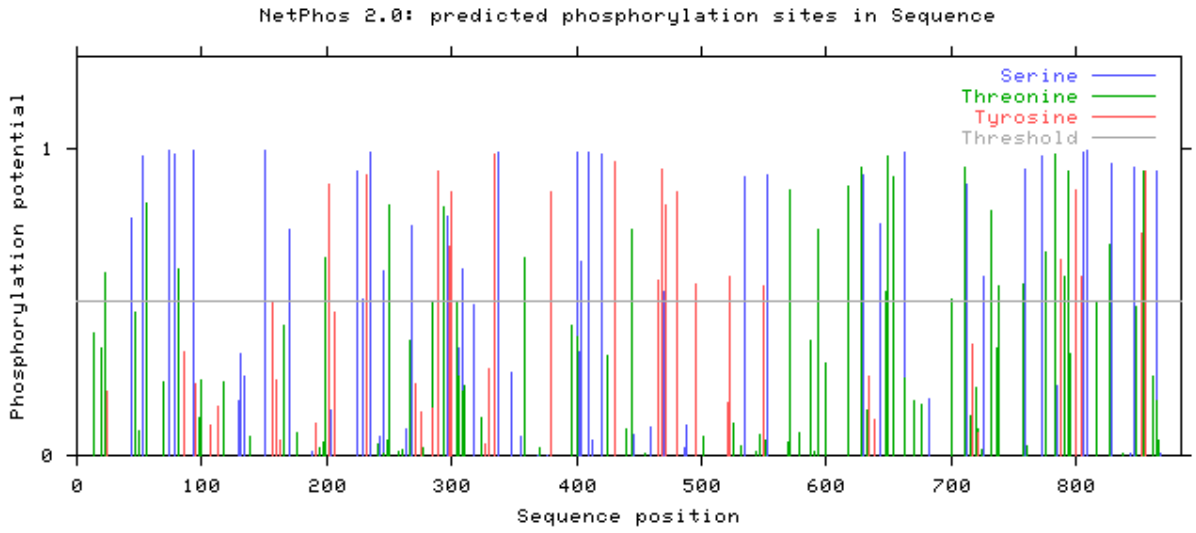

Figure 3 : Net Phosphorylation sites of protein

Above Netphos results confirms that it contains the phosphorylation sites of 34, 28 and 21 for Serine, Threonine & Tyrosine respectively.

CONCLUSION

Thermolysin is an important enzyme that catalyses the hydrolysis of peptide bonds and peptide bond formation through the reverse reaction of hydrolysis. The gene for thermolysin from *Bacillus thuringiensis Bt407* (Accession: NC_018877 REGION: 2545020-2547701) codes for this thermo stable enzyme. Genomics reveals that the pyrimidine rich gene codes for 879 amino acid peptide with the molecular weight of 1623334 Daltons (dsDNA). The gene contains both left & right primers, several restriction sites for restriction sites. In proteomics primary structure says that it is an acidic, stable protein. Secondary structure reveals that it mainly contains the random coils, alpha helix & beta sheets. The protein has a total of 83 phosphorylation sites (Serine-34, Threonine-28 & Tyrosine-21) along with ten O-Glycosylation sites. This insilico study is very much useful for designing the *Bacillus thuringiensis Bt407* for the enhancement of Thermolysin production.

REFERENCES

1. Trusek-Holownia A. (2003): Synthesis of ZAlaPheOMe, the precursor of bitter dipeptide in the two-phase ethyl acetate-water system catalysed by thermolysin, *J. Biotechno.* 102 (2): 153–163.
2. Endo, S. (1962): Studies on protease produced by thermophilic bacteria, *J. Ferment. Techno,* 40: 346–353.
3. Yasukawa K, Kusano M, Inouye K. (2007): A new method for the extracellular production of recombinant thermolysin by co-expressing the mature sequence and pro-sequence in Escherichia coli, *Protein Eng. Des. Se,* 20 (8): 375–383.
4. Inouye K, Kusano M. (2007): Engineering, expression, purification, and production of recombinant thermolysin, *Biotechnol. Annu. Rev,*13: 43–64.
5. Matthews BW, Weaver LH and Kester WR. (1974): The conformation of thermolysin, *J. Biol. Chem.* 249 (24): 8030–8044.
6. Eijsink VG, Veltman OR (1995): Structural determinants of the stability of thermolysin-like proteinases, *Nat. Struct. Biol.* 2 (5): 374–379.

7. Yagasaki, Makoto; Hashimoto, Shin-ichi (2008): Synthesis and application of dipeptides; current status and perspectives, *Applied Microbiology and Biotechnology* 81 (1): 13–22.
8. http://www.plosone.org/article/fetchObjectAttachment.action;jsessionid=CE17B6912F77B4069E4969431710B8A7?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0046147&representation=PDF.
9. C. Geourjon & G. Deleage, (1995): SOPMA : Significant improvement in protein secondary structure prediction by c prediction from alignments and joint prediction, CABIOS, 11, 681-684.