

An insight into the lignin peroxidase of *Macrophomina phaseolina*

Mohammed Touaha Akbar, Abdul Musaweer Habib, Dil Umme Salma Chowdhury, Md Iqbal Kaiser Bhuiyan, Kazi Md Golam Mostafa, Sobuj Mondol & Ivan MHAI Mosleh*

Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong - 4331, Bangladesh; Ivan MHAI Mosleh - Email: ivanahmed96@yahoo.com; *Corresponding author

Received July 15, 2013; Accepted July 16, 2013; Published August 07, 2013

Abstract:

Macrophomina phaseolina is one of the deadliest necrotrophic fungal pathogens that infect more than 500 plant species including major food, fiber, and oil crops all throughout the globe. It secretes a cocktail of ligninolytic enzymes along with other hydrolytic enzymes for degrading the woody lignocellulosic plant cell wall and penetrating into the host tissue. Among them, lignin peroxidase has been reported only in *Phanerochaete chrysosporium* so far. But interestingly, a recent study has revealed a second occurrence of lignin peroxidase in *M. phaseolina*. However, lignin peroxidases are of much significance biotechnologically because of their potential applications in bio-remedial waste treatment and in catalyzing difficult chemical transformations. Besides, this enzyme also possesses agricultural and environmental importance on account of their role in lignin biodegradation. In the present work, different properties of the lignin peroxidase of *M. phaseolina* along with predicting the 3-D structure and its active sites were investigated by the use of various computational tools. The data from this study will pave the way for more detailed exploration of this enzyme in wet lab and thereby facilitating the strategies to be designed against such deadly weapons of *Macrophomina phaseolina*. Furthermore, the insight of such a ligninolytic enzyme will contribute to the assessment of its potentiality as a bioremediation tool.

Key words: *Macrophomina phaseolina*, Ligin peroxidase, lignin.

Background:

Plant biomass is primarily consists of cellulose, lignin and hemicellulose. Lignin is the most abundant natural polymer which makes 15 to 30% of the woody cell walls of gymnosperms (softwood) and angiosperms (hardwood). An encrusting matrix is formed by lignin around the cellulose, the most abundant natural polymer. As this matrix significantly prevents the microbial de-polymerization of cellulose, lignin degradation is a crucial step in the global carbon cycle. Furthermore, the presence of this imperturbable polymer is responsible for the inefficient utilization of cellulose in a wide range of industrial processes [1].

Lignin peroxidase (Lip) is a monomeric haemoglycoprotein peroxidase secreted by the white wood-rot fungus *Phanerochaete chrysosporium*. Lignin peroxidases are of much interest biotechnologically because of their potential applications in

bioremedial waste treatment and in catalyzing difficult chemical transformations. Due to their role in lignin biodegradation they are also of agricultural and environmental importance [2].

A recent study revealed the second occurrence of lignin peroxidase in *Macrophomina phaseolina*, which is one of the most devastating necrotrophic fungal pathogens that infect more than 500 plant species throughout the world. It invades the lignocellulosic barrier in fibre crop jute with the help of various ligninolytic and cellulolytic enzymes. The lignin depolymerizing enzymes are laccases, lignin peroxidases, galactose oxidases, and chloroperoxidases, haloperoxidases, and heme peroxidases. Among them, lignin peroxidase is the one, which is main concern of our study. Following invasion, *M. phaseolina* can grow rapidly in infected jute plant and

produces a large amount of sclerotia which plugs the vessels, and then causes wilting of the plant [3].

So, identifying different properties of this enzyme from *M. phaseolina* along with predicting the 3-D structure and its active sites could be helpful in designing strategies against this catastrophic fungus. Moreover, pollutions owing to olive mill wastewaters (OMW) is creating a serious ecological problem in the Mediterranean area, particularly in Italy, Spain, Greece, and Tunisia, which are the main producers of olive oil. These countries produce more than 3×10^7 m³ of OMW per year. Due to high concentration of aromatic compounds, this effluent is black & highly noxious [4]. This wastewater includes the black high-molecular-weight polyphenols such as tannins, anthocyanins, and catechins [5]. The phytotoxic and antibacterial effects of the OMW have been attributed to the phenolic content of OMW [6]. Conventional bacterial water treatment processes are relatively ineffective at removing these pollutants. OMW must be diluted more than 10-fold prior to treatment with aerobic activated sludge or by anaerobic digestion. Hence, it requires more effective treatment.

The irregular and recalcitrant nature of lignin and the fact that the substructures contained in lignin are also found in primary pollutants such as phenols, anisoles, biphenyls, and diarylethers led researchers to postulate that the nonspecific ligninolytic enzyme produced by white rot fungi might be able to degrade such persistent aromatic pollutants. Moreover, a wide variety of phenolic compounds and environmentally persistent xenobiotics and chlorinated hydrocarbons, including DDT [1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane], TNT (2,4,6-trinitrotoluene), alkyl halide insecticides, chloroanilines, polychlorinated phenols, and polychlorinated biphenyls are degradable by *P. chrysosporium*. Thus, characterization of lignin peroxidase enzyme from *M. phaseolina* will be an added benefit in accordance with *P. chrysosporium* in the field of bioremediation waste treatment [7].

Until now, purification from the native fungus (*Phanerochaete chrysosporium*) has been the only method of producing the active enzyme [8]. Heterologous expression in *E. coli* genes, encoding the mature sequence of Lip from *P. chrysosporium* has been done successfully. Moreover, the first successful homologous expression of rLiPH8 in *P. chrysosporium* and the characterization of the recombinant enzyme has also been reported as well as enlightened the fact that increased expression of Lignin peroxidase in *M. phaseolina* can also be achieved [9]. In this present study, several features of lignin peroxidase of *M. phaseolina* have been revealed using different web based tools.

Methodology:

Several characteristics of *M. phaseolina* have been identified along with its 3D structure and active site prediction by using different databases and software's maintaining as shown (Figure 1).

Sequence retrieval and physicochemical characterization

At the beginning, protein sequence of lignin peroxidase of *M. phaseolina* was retrieved in FASTA format from NCBI [10] and then physicochemical properties of the enzyme were obtained by using ProtParam [11].

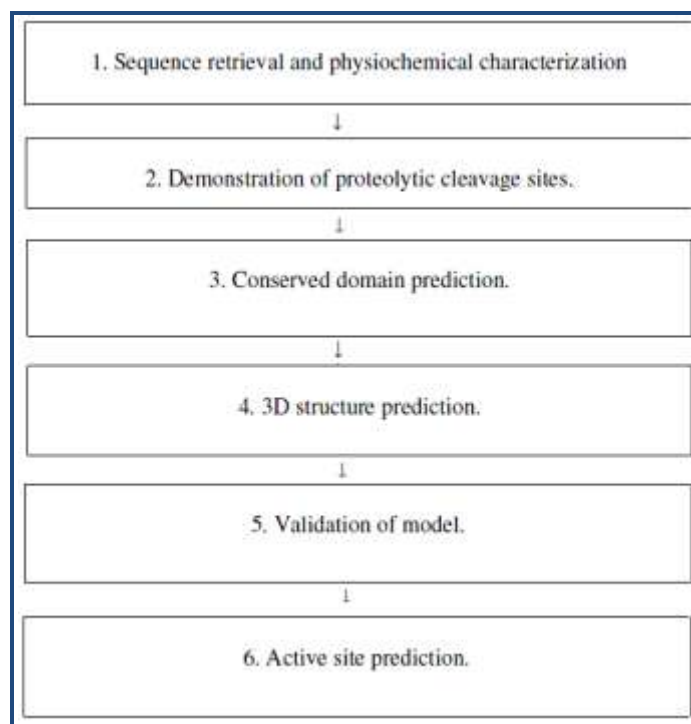


Figure 1: Methodology involved in characterisation of *M. phaseolina*.

Identification of proteolytic cleavage site

Proteolytic cleavage site was identified by using a web based tool, peptide cutter [12] which predicts protease cleavage sites and sites cleaved by chemicals in a given protein sequence.

Identification of conserved domain

To search the conserved domains present in the Lignin Peroxidase of *Macrophomina phaseolina*, we utilized Conserved Domain Search tool which can be found in NCBI [13]. The sequence can be added to this tool by the use of accession number, gi, or sequence in FASTA format. In this study, we added sequence in FASTA format.

Prediction of three-dimensional structure

To predict the three dimensional structure of Lignin peroxidase, the most popular tool, Phyre2 (Protein Homology/Analogy Recognition Engine) [14] an online protein fold recognition server, was used in this study. The retrieved amino acid sequence in FASTA format was used as input data in Phyre2. In this respect, the intensive mode of protein modeling was selected in order to get an accurate model.

Validation of lignin peroxidase model

To check the accuracy of the predicted three dimensional structure, PDBsum was used. PDBsum is a database that provides an overview of the protein structures stored in the Protein Data Bank for the protein structures that have already been generated. PDBsum has a page for each structure denoted by a specific PDB id. But for lignin peroxidase, there is no PDBsum page. So a customized page for lignin peroxidase was generated by using PDBsum 'generate' option [15]. In this case the PDB file obtained by using Phyre2 server for lignin peroxidase was used as input data. The PDBsum generated result was further subjected to documentation analysis by PROCHECK. To obtain an idea about the protein chains and

their stereochemical quality, like Ramachandran plot quality, peptide bond planarity, Bad non-bonded interactions, main chain hydrogen bond energy, C alpha chirality and over-all G factor and the side chain parameters like standard deviations of chi1 gauche minus [16] PROCHECK analysis was done.

Active site prediction

To predict the active site of the sequence of interest a web based tool named CastP was used [17]. The weighted Delaunay triangulation and the alpha complex for shape measurements

are used by CastP which provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules. It assesses analytically the volume as well as the area of each cavity and pocket, both in molecular surface (MS, Connolly's surface) and solvent accessible surface (SA, Richards' surface). It also measures the number of mouth openings, circumference of mouth lips, and area of the openings, in both SA and MS surfaces for each pocket [18].

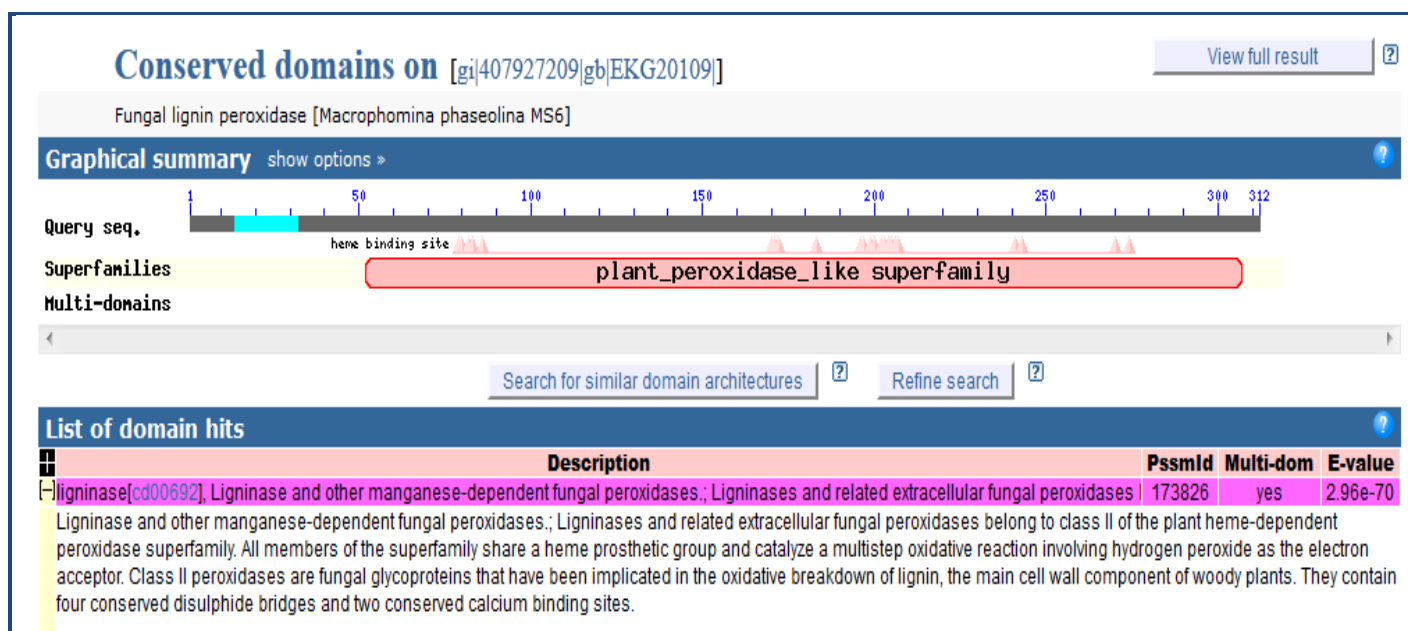


Figure 2: Conserved domain over lignin peroxidase

Results & Discussion:

Physicochemical characterization

Different physicochemical properties of Lignin peroxidase has been characterized. Computation of Isoelectric point (theoretical pI) and molecular weight (Mw) of an amino acid sequence is important because these data dictate the approximate condition of a 2D-gel in which a protein of interest can be detected [19]. Lignin peroxidase is moderately acidic (5.22) in nature. The instability index (II) for the respective enzyme is computed to be 29.87 which classifies the protein as stable. A protein whose instability index is smaller than 40 is predicted as stable; a value above 40 predicts that the protein may be unstable [20].

The half-life is the time which is taken by half of the amount of protein in a cell to disappear after its synthesis in the cell. The half-life is given for three organisms (human, yeast, and E. coli). The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro); >20 hours (yeast, in vivo); >10 hours (*Escherichia coli*, in vivo) [20].

The relative volume of a protein occupied by its aliphatic side chains is termed as Aliphatic index (AI). Aliphatic index plays role in protein thermal stability. The higher the Aliphatic index of proteins, the more thermally stable the proteins. Aliphatic amino acids also are hydrophobic in nature [20]. The aliphatic index of lignin peroxidase is 86.44 which indicated this enzyme is thermally stable as well as it have high amount of

hydrophobic amino acids. The protein is hydrophobic (due to positive GRAVY value: 0.022).

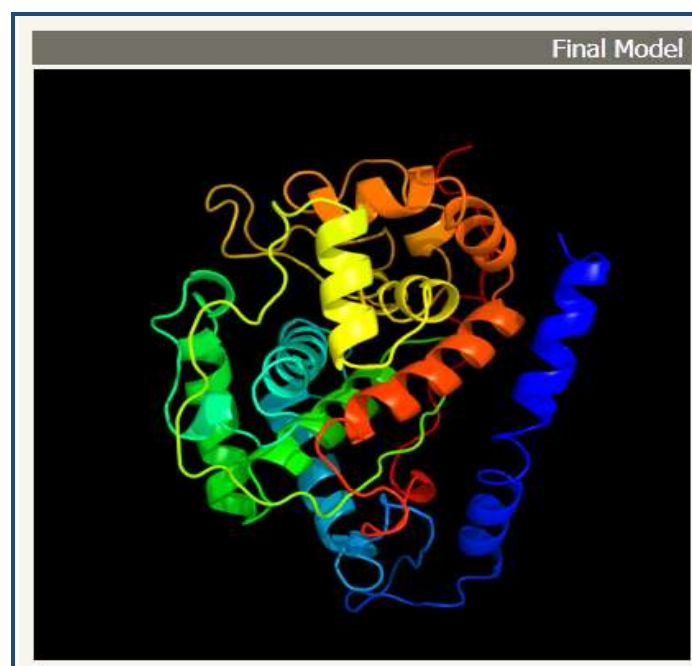


Figure 3: The final 3D structure of lignin peroxidase model from *Macrophomina phaseolina*. The α -helices and β -sheets are

represented by red helices and yellow ribbons respectively. This model was conducted by Phyre2 MODELLER program.

Identification of proteolytic cleavage site

Protease digestion can be useful if one wants to carry out experiments on a portion of a protein, separate the domains in a protein, remove a tag protein when expressing a fusion protein, or make sure that the protein under investigation is not sensitive to endogenous proteases. One or several reagents can be selected from a list of (currently) 33 proteases and chemicals [20]. The proteolytic cleavage sites found in lignin peroxidase enzyme of *Macrophomina phaseolina* is given in Table 1 (see the supplementary).

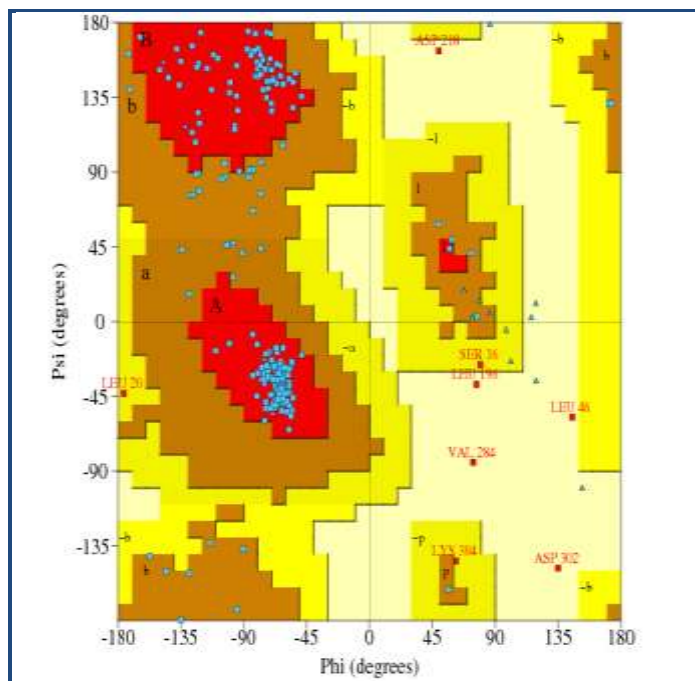


Figure 4: Ramachandran's Map of lignin peroxidase model from *Macrophomina phaseolina*. The Plot calculation was done with PROCHECK program

Identification of conserved domain

The fundamental units of protein structure, function, and evolution are known as protein domain. As a result, domain

prediction is critical step in the annotation of protein sequences [21]. Domain predictions are also the starting point for a range of more sophisticated analyses, including comparative genomics of domain families in diverse organisms studies of the evolution of protein and domain structure and function, prediction of protein-protein interactions and identification of complex evolutionary relationships [22].

The conserved domain of lignin peroxidase lies between 30-300 amino acids of the enzyme (Figure 2). Within the conserved domain, there are several heme-binding sites. This enzyme belongs to a group of peroxidases or a superfamily named plant peroxidase-like superfamily. The enzyme, lignin peroxidase contain a heme prosthetic group (ferriprotoporphyrin IX), which catalyzes a multistep oxidative reaction involving hydrogen peroxide as the electron acceptor.

3-D structure determination

Three-dimensional (3D) protein structures provide valuable insights into the molecular basis of protein function, allowing an effective design of experiments, such as site-directed mutagenesis, studies of disease-related mutations or the structure based design of specific inhibitors [23]. Therefore, the high resolution 3D structure of a protein is the key to the understanding and manipulating of its biochemical and cellular functions [24].

The theoretical structure (Figure 3) of lignin peroxidase from *M phaseolina* is generated using Phyre2. Greater than 90% confidence match was obtained, which is a high confidence match. This implies that, overall fold of the model was almost certainly correct and the central core of the model was tend to be accurate [25].

Validation of Lignin peroxidase Model

Accuracy of the protein model was measured by PROCHECK. Parameter comparisons of these proteins were made with well-refined structures that have similar resolution. Ramachandran plot and Ramachandran plot statistics is shown in (Figure 4) and Table 2 (see the supplementary). Thus, analysis of PROCHECK reveals that all residues are within the limits of the Ramachandran plot and hence it can be considered as a good model.

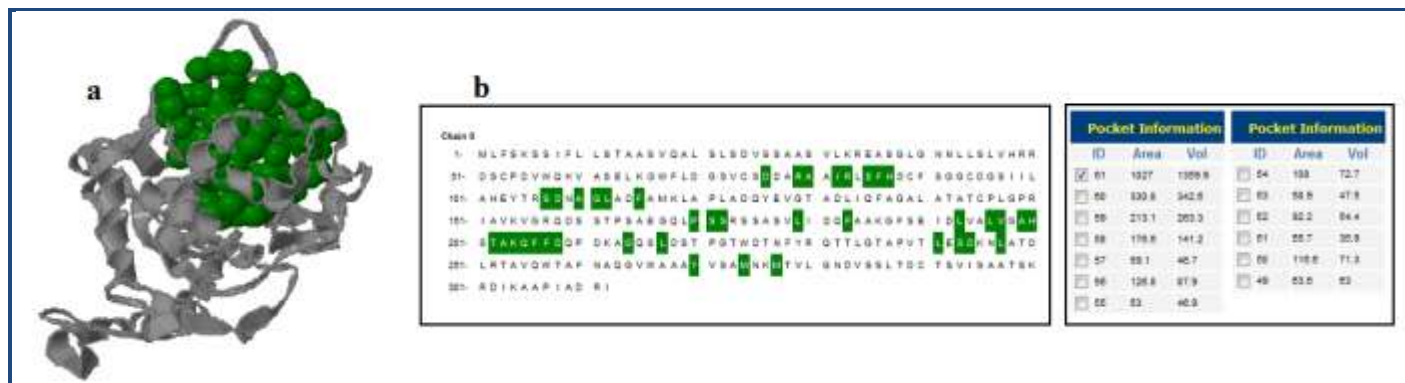


Figure 5: a) Active site information by CASTp. Green color shows the active site position from 76 to 287 with the β -sheet in between them and 3D representation of the best active site of lignin peroxidase generated by using CASTp; b) The table shows the area and the volume for different active sites of lignin peroxidase and the best active site remains in an area of 1027 and a volume of 1359.9 amino acid.

Active site prediction

The identification and characterization of functional sites on proteins has increasingly become an area of interest. Because analysis of the active site residues for the binding of ligands provides insight towards the design of inhibitors of an enzyme [26]. In this study, we have also reported the best active site area of the experimental enzyme as well as the number of amino acid involved in it. (Figure 5a & 5b).

Conclusion:

Lignin peroxidase of *M. phaseolina*, an enzyme having undoubted biotechnological prospect, was terra incognita. This work has unearthed its several crucial aspects, on the basis of which, it could be concluded that further inquiry into such a deadly weapon of *M. phaseolina* might be of significant value in the regulation of the diseases caused by this phytopathogen. The model of the enzyme can be effectively used to design mutagenesis experiments and broad screening on inhibitors of the protein could be done as well using knowledge obtained from this study. Last but not least, the study paves the way to assess *M. phaseolina* as a potential bioremedial agent such as in phenolic-rich effluent olive oil waste water.

References:

- [1] Gold MH & Alic M, *Microbiological Reviews*. 1993 **57**: 605 [PMID: 8246842]
- [2] Doyle WA & Smith AT, *Biochem J*. 1996 **315**: 15 [PMID: 8670100]
- [3] Islam MS *et al.* *BMC Genomics*. 2012 **13**: 493 [PMID: 22992219]
- [4] Sayadi S & Ellouz R, *J Chem Technol Biotechnol*. 1993 **57**: 141
- [5] Hamdi M *et al.* *Appl Microbiol Biotechnol*. 1991 **34**: 829
- [6] Moreno E *et al.* *Microbios*. 1987 **51**: 169
- [7] Sayadi S & Ellouz R, *Appl Environ Microbiol*. 1995 **61**: 1098 [PMID: 16534959]
- [8] Tien M & Kirk TK, *Proc Natl Acad Sci U S A*. 1984 **81**: 2280 [PMID: 16593451]
- [9] Reiser J *et al.* *Appl Environ Microbiol*. 1993 **59**: 2897 [PMID: 8215362]
- [10] <http://www.ncbi.nlm.nih.gov/>
- [11] <http://web.expasy.org/protparam/>
- [12] <http://web.expasy.org/peptidecutter/>
- [13] <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>
- [14] <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>
- [15] <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>
- [16] Morris AL *et al.* *Proteins*. 1992 **12**: 345 [PMID: 1579569]
- [17] Dundas J *et al.* *Nucl Acids Res*. 2006 **34**: W116. [PMID: 16844972]
- [18] Panda S & Chandra G, *Bioinformation*. 2012 **8**: 891 [PMID: 23144546]
- [19] http://web.expasy.org/docs/expasy_tools05.pdf
- [20] Stein L, *Nat Rev Genet*. 2001 **2**: 493 [PMID: 11433356]
- [21] Ochoa A *et al.* *BMC Bioinformatics*. 2011 **12**: 90 [PMID: 21453511]
- [22] Schwede T *et al.* *Nucleic Acids Res*. 2003 **31**: 3381 [PMID: 12824332]
- [23] Fidanova S & Lirkov I, *Seria Matematica-Informatica*. 2009 **17**: 33
- [24] Kelley LA & Sternberg MJ, *Nat Protoc*. 2009 **4**: 363 [PMID: 19247286]
- [25] Laurie AT & Jackson RM, *Curr Protein Pept Sci*. 2006 **7**: 395 [PMID: 17073692]
- [26] Huang B & Schroeder M. *BMC Structural Biology* 2006 **6**:19 [PMID:16995956]

Edited by P Kanguane

Citation: Akbar *et al.* *Bioinformation* 9(14): 730-735 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Result from PeptideCutter

Name of enzyme	No. of cleavages	Positions of cleavage site
Arg-C proteinase	13	34 49 50 79 83 106 150 157 173 230 252 301 311
Asp-N endopeptidase	25	23 50 54 69 75 76 87 94 107 113 123 131 158 180 191 207 210 217 224 243 249 282 288 301 309
Asp-N endopeptidase + N-terminal Glu	32	23 34 50 54 62 69 75 76 87 94 102 107 113 123 126 131 158 165 180 189 191 207 210 217 224 241 243
BNPS-Skatole	5	57 67 224 257 266
CNBr	4	1 117 274 277
Caspase1	1	250
Chymotrypsin-high specificity (C-term to [FYW], not before P)	22	3 9 57 67 68 86 90 104 115 126 136 183 188 206 207 224 228 229 257 260 266 270
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	59	1 2 3 9 10 11 20 22 32 39 43 44 46 48 57 64 67 68 69 84 86 90 100 102 104 112 115 117 119 122 126 133 136 140 147 179 183 188 193 196 200 206 207 217 224 228 229 234 241 247 251 257 260 266 270
Clostripain	13	274 277 280 287
Formic acid	25	34 49 50 79 83 106 150 157 173 230 252 301 311
Glutamyl endopeptidase	7	24 51 55 70 76 77 88 95 108 114 124 132 159 181 192 208 211 218 225 244 250 283 289 302 310
Iodosobenzoic acid	5	35 63 103 127 166 190 242
LysC	5	57 67 224 257 266
LysN	13	5 33 59 65 118 154 186 204 212 245 276 300 304
LysN	13	4 32 58 64 117 153 185 203 211 244 275 299 303
NTCB (2-nitro-5thiocyanobenzoic acid)	6	52 73 88 93 144 289
Pepsin (pH1.3)	92	1 2 2 3 8 9 9 10 10 11 19 20 21 22 31 32 38 39 42 43 43 44 45 46 56 57 63 64 66 68 68 69 84 86 90 99 100 103 111 112 114 115 118 121 125 126 132 133 135 136 139 140 146 169 178 179 182 183 187 192 193 195 196 205 207 216 217 223 224 227 228 228 229 233 234 240 241 246 250 251 256 257 259 260 265 266 269 270 279 280 286 287
Pepsin (pH>2)	76	1 2 2 3 8 9 9 10 10 11 19 20 21 22 31 32 38 39 42 43 43 44 45 46 63 64 68 68 69 84 86 90 99 100 111 112 114 115 118 121 132 133 135 136 139 140 146 169 178 179 182 183 187 192 193 195 196 205 207 216 217 227 228 233 234 240 241 246 250 251 259 260 279 280 286 287
Proteinase K	157	2 3 8 9 10 11 13 14 15 17 19 20 22 25 28 29 31 32 35 36 39 43 44 46 47 56 57 60 61 63 64 67 68 69 73 78 80 81 82 84 86 90 98 99 100 101 103 104 105 110 112 113 115 116 119 120 122 123 126 127 128 130 131 133 134 136 137 139 140 141 142 143 144 147 151 152 153 155 162 165 166 169 176 178 179 180 183 184 185 188 190 191 193 194 195 196 197 199 202 203 206 207 213 217 220 223 224 226 228 229 232 233 234 236 237 239 240 241 242 247 248 249 251 253 254 255 257 258 259 260 262 265 266 267 268 269 270 271 273 278 279 280 284 287 288 291 293 294 296 297 298 298 303 305 306 308 309 312

Name of enzymes	No.of cleavages	Positions of cleavage sites
Staphylococcal peptidase I	7	35 63 103 127 166 190 242
Thermolysin	106	1 2 7 8 9 10 13 14 16 18 19 21 27 28 30 31 38 42 43 45 46 59 60 67 68 72 79 80 81 83 85 89 97 98 99 100 109 111 112 115 116 118 121 122 130 133 135 136 138 139 140 142 146 150 151 152 154 164 175 177 178 179 182 183 184 187 193 194 195 196 198 202 205 206 212 216 227 233 238 240 246 247 253 254 258 259 261 264 266 267 268 270 272 273 276 278 279 286 292 293 295 296 304 307 308 311
Thrombin	1	150
Trypsin	26	5 33 34 49 50 59 65 79 83 106 118 150 154 157 173 186 204 212 230 245 252 276 300 301 304 311

Table 2: Physiochemical characters as predicted by ExPASy's prot-param program

Ramachandran plot statistics	No. of residues	Percentage
Residues in the most favored region(A,B,L)	235	83.9%
Residues in the additional allowed region (a,b,l,p)	37	13.2%
Residues in the generously allowed region (a,b,l,p)	3	1.1%
Residues in the disallowed region	5	1.8%
Number of non-glycine & non-proline residues	280	100.0%
Number of end residues (excl. gly & pro)	2	
Number of glycine residues (shown in triangles)	20	
Number of proline residues	10	
Total number of residues	312	