IN SILICO CHARACTERIZATION OF SOME FOOD ENZYMES LIKE PROTEASE, CELLULASE AND PECTINASE USING COMPUTATIONAL TOOLS

Megha, S.V., Maragathavalli, S., Brindha, S. and Annadurai, B.  
Research and Development Centre, Bharathiar University, Coimbatore

ABSTRACT
In this study the insilico characterization of some food enzymes like protease, cellulase and pectinase were done. Physicochemical properties of protease, cellulase and pectinase were computed using ExPASy’s protparam tool resulting in primary structure analysis. Protoscale and pI/MW of the enzymes were also computed. Secondary structure predictions were done using GORB IV, SOPMA and transmembrane regions were predicted by TMHMM

KEYWORDS: ProtParam, Prot Scale, GOR IV, SOPMA, TMHMM

INTRODUCTION
Active research on protease, cellulase, pectinase and other related enzymes began in the early 1950s, owing to their enormous potential to convert substrate, the most abundant and renewable source of energy on Earth, to glucose and soluble sugars (Bhat M K, 2000). Proteases are a group of proteolytic enzymes whose catalytic function is to hydrolyse peptide bonds of protein molecules. They are also called proteinase or peptidase. They break the long chain of protein molecule into shorter fragments called peptides and then eventually in to their components called amino acids. Proteolytic enzymes are present in bacteria, archaea, algae, viruses, plants and most abundantly in animals. There are different types of proteolytic enzymes, classified according to the site at which they cleave the protein molecule. Exopeptidase which cleaves at terminal ends of protein and endopeptidase which cleaves the protein molecule at internal regions. Based on protein molecule protease can be classified into seven broad group. The inability of the plants and animals protease to meet the current world demands leads to an increased interest in microbial protease. Among the different protein sources milk is a rich source protein, which can be used for the production of protease. It act as an excellent medium for the growth of microorganism. Proteases are widely used in detergents, food, pharmaceutical and leather tanning industries. Cellulases are inducible enzymes which are synthesized by microorganisms during their growth on cellulose materials. They are studied extensively due to their application in the hydrolysis of cellulose, the most abundant biopolymer and potential source of utilizable sugars, which serves as a raw material in the production of chemicals and fuel (Ali et al., 2011). Cellulase refers to a family of enzymes which act in concert to hydrolyze cellulose. Cellulase is used extensively in the textile and food industries, bioconversion of lignocellulosic wastes to alcohol, animal feed industry as additive, isolation of plant protoplasts, in plant virus studies, metabolic investigations and genetic modification experiments. The extensive use of cellulase in many industries depends on the cost of the enzyme which in turn depends on the method of production. Hence, research all over the world focuses on isolating new, hyper producing microbial strains and also to develop new fermentation processes aimed at reducing the cost of the enzyme with a view to bring down the overall process cost (Suresh et al., 2005). Cellulase production is the most important step in the economical production of ethanol. Pectinase are enzymes that breakdown pectin, a polysaccharide substrate that is found in the cell wall of plants. Pectinases are general name of pectic enzymes which include pectolyase, pectozymes and polygalacturonase. Pectin is jelly like matrix which cement plant cells together in a cell wall. Pectinase is widely used in fruit juice extraction, wine production, paper and pulp industry, textile processing, waste water treatment; animal feed purification of plant viruses. Pectinase is a growing enzyme of biotechnology sector, showing gradual increase of need in market (Garg et al 2016).

Bioinformatics has revolutionized the field of molecular biology. The raw sequence information of proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools. Prediction of protein function is important application of bioinformatics (Prashant et al, 2010). In the present bioinformatics analysis characterization of protease from Pseudomonas aeruginosa, cellulases from Ceratocystis peradoxa and pectinase from Alternaria cepulae were carried out. Protein sequences were retrieved from NCBI and were subjected to ProtParam to analyse various physicochemical properties, secondary structure was predicted by SOPMA, multiple sequence analysis and phylogenetic analysis was carried out by CLC workbench, the protein 3D model and its characteristics were predicted by ESyPred 3D software (Ashokan et al., 2011). These parameters will assist the biochemist and physiologists in extraction, purification, separation and industrial applications of the enzyme.
**SYSTEM & METHODS**

**Protparam Analysis of physicochemical parameters**
The different physicochemical properties of protease, cellulase and pectinase enzymes were computed using ExPASy’s ProtParam tool and these properties can be deduced from a protein sequence which helps in primary structure analysis. The ProtParam includes the following computed parameters: Molecular weight (M.Wt), theoretical pl, instability index (II), alphatic index (AI) and grand average of hydropathicity (GRAVY). The computed isoelectric point (pl) will be useful for developing buffer systems for purification by isoelectric focusing method (Sivakumar et al., 2007).

**Method**
To know hydropathic value for individual amino acid present in protein sequence. First we go proteomics ExPASy tools through http://www.expasy.ch/tools. The procedure was explained in general system and method.

**Secondary structure prediction**

**GOR IV**
GOR IV (Garnier Osguthorpe and Robson) method is for the prediction of secondary structure in proteins. GOR method is based on probability parameters derived from the studies of protein tertiary structure solved by X ray crystallography. The GOR method analyses sequences to predict alpha helix, beta sheet, random coil etc. The predicted secondary structure is one with the highest compatible structure with a predicted helix segment of atleast four residues and a predicted extended segment of at least two residues. (Garnier et al 1996).

**Method**
To know the secondary structure of the enzyme sequence. First we go proteomics ExPASy tools through http://www.expasy.ch/tools. The procedure was explained in general system and methods.

**SOPMA**
The secondary structure was predicted by self-optimized prediction method with alignment (SOPMA) (Ashokan et al., 2011). SOPMA was employed for calculating the secondary structural features of theselected protein sequences considered in this study (Neelima et al., 2009). This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. (Altshul et al 1997, Geourjon 1995) SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method (Prashant et al, 2010).

**Method**
To know the secondary structure of the enzyme sequence. First we go proteomics ExPASy tools through http://www.expasy.ch/tools. The procedure was explained in general system and method.

**TMHMM**
TMHMM is a method for prediction of transmembrane helices based on a hidden Markov model developed by Anders Krogh and Erik Sonnhammer. It predict transmembrane helices and discriminate between soluble and membrane proteins with high degree of accuracy. At a time a user can submit as many as 4000 protein sequences in FASTA format each time.

**Method**
First we go proteomics ExPASy tools through http://www.expasy.ch/tools. The home page of proteomics ExPASy tools will appear. The procedure was explained in general system and method.
RESULT

Protparam Protease

FIGURE 1: The ProtParam result of Protease

Inference

The physicochemical properties of protease were predicted by using ProtParam tool (Fig 4.1). The Prot Param includes the following computed parameters: Molecular weight (M Wt), theoretical pI, Instability Index, Aliphatic Index, The Grand Average of Hydrophaticity (GRAVY). The physicochemical parameters show that the molecular weight of protease is around 51020.57 Da. The instability index is used to measure the in vivo half life of a protein (Laskowski et al., 1993). The instability index of 20.46 showed that most of the proteases are stable since their index showed a value less than 40. Isoelectric point (pI) is the pH at which the surface of the protein is covered with charge but the net is zero. The computed pI value of 8.44 shows that proteases are alkaline in nature (pH > 7). The aliphatic index implies on the stability of the protein when its value is high. Here the Aliphatic index value of 85.98 is high showing that the proteases are stable. The GRAVY value of the protease is -0.429 which is lower showed the better interaction of protease with water.
Protparam Cellulase

**FIGURE 2:** The ProtParam result of Cellulase

Inference

The physicochemical properties of cellulase were predicted by using ProtParam tool (Fig 4.2). The ProtParam includes the following computed parameters: Molecular weight (M Wt), theoretical pI, Instability Index, Aliphatic Index, The Grand Average of Hydrophobicity (GRAVY). The physicochemical parameters show that the molecular weight of cellulase is around 24370.91 Da. The instability index is used to measure the invivo half life of a protein (Laskowski et al., 1993). The instability index of 26.63 showed that most of the cellulase are stable since their index showed a value...
The Protparam result of Pectinase protease with water. The physicochemical properties of pectinase were predicted by using ProtParam tool (Fig 4.3). The ProtParam includes the following computed parameters: Molecular weight (M Wt), theoretical pI, Instability Index, Aliphatic Index, The Grand Average of Hydropathicity (GRAVY). The physicochemical parameters shows that the molecular weight of pectinase is around 38816.10 Da. The instability index is used to measure the in vivo half-life of a protein (Laskowski et al., 1993). The instability index of 21.86 showed that most of the pectinase are stable since their index showed a value less than 40. Isoelectric point (pI) is the pH at which the surface of the protein is covered with charge but the net is zero. The computed pI value of 4.85 shows that proteases are acidic in nature (pH < 7). The aliphatic index implies on the stability of the protein when its value is high. Here the Aliphatic index value of 61.79 is high showing that the cellulase are stable. The GRAVY value of the cellulase is 0.215 which is lower showed the better interaction of protease with water.

### Table: Protparam Pectinase

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight (M Wt)</td>
<td>38816.10 Da</td>
</tr>
<tr>
<td>Isoelectric point (pI)</td>
<td>4.85</td>
</tr>
<tr>
<td>Instability Index</td>
<td>21.86</td>
</tr>
<tr>
<td>Aliphatic Index</td>
<td>61.79</td>
</tr>
<tr>
<td>Grand Average of Hydropathicity (GRAVY)</td>
<td>-0.156</td>
</tr>
</tbody>
</table>

**FIGURE 3:** The Protparam result of Pectinase
In silico characterization of some food enzymes Using computational tools

Inference

The ProtScale parameter many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales. Here the Prot Scale parameter (Fig 4.4) of amino acid composition was conducted and analysed. The ProtScale parameter is mainly used for the construction of Kyte and Doolittle hydropathy plot. It has shown the composition of all the amino acids present in the protease sequence. (Bjellquist et al 1993, Bjellquist et al 2005).
Protscale cellulase

ProtScale

User-provided sequence:

10  20  30  40  50  60
MVSSFLVAA VQGFTSVAVV SFPVITPDAVP NVTETELMEQ AGTPHSSGMN DGYPFNSWAVD
70  80  90  100  110  120
GGDARYTNG KHGAYSGIKS TGGNLYOOGH WRPSGSARTIN YSGTYAPNQ SYLAIYGMTT
130 140 150 160 170 180
SPLTYYVE NFOTYNFSSG ATVQGSKHAQ SVYDLITST RINAPSQITG ATIFQQHAVVR
190 200 210 220
QSKRESGKVN TSTPPNAWSN ASLKLGAHDY QIVATEGYFQS SGSSSMTW

SEQUENCE LENGTH: 229

Using the scale A.A. composition, the individual values for the 20 amino acids are:

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>8.900</td>
</tr>
<tr>
<td>Arg</td>
<td>5.700</td>
</tr>
<tr>
<td>Asn</td>
<td>6.400</td>
</tr>
<tr>
<td>Asp</td>
<td>5.200</td>
</tr>
<tr>
<td>Cys</td>
<td>1.700</td>
</tr>
<tr>
<td>Gln</td>
<td>4.000</td>
</tr>
<tr>
<td>Glu</td>
<td>6.200</td>
</tr>
<tr>
<td>Gly</td>
<td>7.200</td>
</tr>
<tr>
<td>His</td>
<td>2.200</td>
</tr>
<tr>
<td>Ile</td>
<td>5.200</td>
</tr>
<tr>
<td>Ile</td>
<td>5.200</td>
</tr>
<tr>
<td>Leu</td>
<td>9.000</td>
</tr>
<tr>
<td>Lys</td>
<td>5.700</td>
</tr>
<tr>
<td>Met</td>
<td>2.400</td>
</tr>
<tr>
<td>Phe</td>
<td>5.900</td>
</tr>
<tr>
<td>Pro</td>
<td>5.100</td>
</tr>
<tr>
<td>Ser</td>
<td>6.900</td>
</tr>
<tr>
<td>Thr</td>
<td>5.800</td>
</tr>
<tr>
<td>Trp</td>
<td>1.500</td>
</tr>
<tr>
<td>Tyr</td>
<td>3.200</td>
</tr>
<tr>
<td>Val</td>
<td>6.600</td>
</tr>
</tbody>
</table>

Weights for window positions 1...9, using linear weight variation model:

<table>
<thead>
<tr>
<th>Window</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Inference

The ProtScale parameter many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales. Here the Prot Scale parameter (Fig 4.5) of cellulase amino acid composition was conducted and analysed. The ProtScale parameter is mainly used for the construction of Kyte and Doolittle hydropathy plot. It has shown the composition of all the amino acids present in the cellulase sequence. (Bjellquish et al 1993, Bjellquish et al 2005).
In silico characterization of some food enzymes Using computational tools

**ProtScale Pectinase**

**ProtScale**

User-provided sequence:

```
HVALLGDDF TSLASSMMA HAPRITAPR PEVVRASSU TPSGSPMAR ASEKQCSSL
70    80    90   100   110   120
HVLSDVAVPL GYIGLESLA DGSTVFSGE TTWSYKEMG PLEDIQKQY TVKEAGCSTV
130   140   150   160   170   180
NKDGARWMDI KEGNQRKRPI KFPPSAHMLT STITENTH CPFQYVSING QDDLTIHTMT
190   200   210   220   230   240
IDADDDOKE QHRXDQDGQ GDQVTDGQ DQYRHQDV AVHGQRTIFK MNSLCQEPHS
250   260   270   280   290   300
LSIIGSVGGRD DNTVDTVPS NSVRVKAEM RVRKARYST QMKHMKVTV VINLQKSYD
310   320   330   340   350   360
VLERQNYDQG DLHHDATCGV FITALTLDIV TGGVSSGVD VVTGQKGSC TQIRMTGQDV
370
```

SEQUENCE LENGTH: 370

Using the scale **A.A. composition**, the individual values for the 20 amino acids are:

- Ala: 8.300
- Arg: 5.700
- Asn: 4.400
- Asp: 5.300
- Cys: 1.700
- Gln: 4.000
- Glu: 6.200
- Gly: 6.200
- His: 2.200
- Ile: 8.200
- Leu: 9.000
- Lys: 6.700
- Met: 2.400
- Phe: 3.500
- Pro: 5.100
- Ser: 6.800
- Thr: 5.800
- Trp: 1.300
- Tyr: 6.200
- Val: 6.600
- Phe: 4.200
- Ser: 3.000
- Asp: 5.000

Weights for window positions 1, 9, using **linear weight variation model**:

<table>
<thead>
<tr>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>edge</td>
<td>center</td>
<td>edge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 6:** ProtScale of Pectinase

**Inference**

The ProtScale parameter many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales. Here the ProtScale parameter (Fig 4.6) of pectinase amino acid composition was conducted and analysed. The ProtScale parameter is mainly used for the construction of Kyte and Doolittle hydropathy plot. It has shown the composition of all the amino acids present in the pectinase sequence. (Bjellquish et al, 1993, Bjellquish et al, 2005)
Compute pI/Mw of enzymes

Theoretical pI/Mw of protease

![Theoretical pI /MW of protease]

**Inference:** The theoretical pI and Molecular weight of protease (Fig 4.7) was found to be 8.44 and 51020 respectively.

Theoretical pI/Mw of Cellulase

![Theoretical pI /MW of Cellulase]

**Inference:** The theoretical pI and Molecular weight of cellulase (Fig 4.8) was found to be 8.35 and 50042 respectively.

Theoretical pI /MW of Pectinase

![Theoretical pI /MW of Pectinase]

**Inference:** The theoretical pI and Molecular weight of pectinase (Fig 4.9) was found to be 4.85 and 38816.10 respectively.
Inference
The GOR method is for the prediction of secondary structure in proteins. The GOR method analyses sequences to predict alpha helix, beta sheet, random coil etc. In the GOR IV analysis (Fig 4.10) of protease enzyme sequence secondary structure prediction showed that the alpha helix account for 39.78% and extended strands account for 16.56%. There are 43.66% random coils in the secondary structure of protease.
**GOR IV of Cellulase**

Inference

The GOR method is for the prediction of secondary structure in proteins. The GOR method analyses sequences to predict alpha helix, beta sheet, random coil etc. In the GOR IV analysis (Fig 4.11) of cellulase enzyme sequence secondary structure prediction showed that the alpha helix account for 13.54% and extended strands account for 31.88%. There are 54.59% random coils in the secondary structure of cellulose.
Inference
The GOR method is for the prediction of secondary structure in proteins. The GOR method analyses sequences to predict alpha helix, beta sheet, random coil etc. In the GOR IV analysis (Fig 4.12) of pectinase enzyme sequence secondary structure prediction showed that the alpha helix account for 5.54% and extended strands account for 36.15%. There are 58.31% random coils in the secondary structure of pectinase.
SOPMA

Inference

SOPMA is another secondary structure prediction tool available in ExPASy. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. It (Fig 4.13) shows that 43.23% of the protease sequence can attain alpha helix, 18.49% of extended strands, β strand account for 12.26% and random coils were 26.02%.
SOPMA of Cellulase

<table>
<thead>
<tr>
<th>Sequence length</th>
<th>229</th>
</tr>
</thead>
</table>

**SOPMA**

- **Alpha helix (α)**: 34 is 14.95%
- **β-helix (β)**: 0 is 0.00%
- **π-helix (π)**: 0 is 0.00%
- **β sheets**: 326 is 40.17%
- **Extended strand (E)**: 72 is 31.44%
- **Beta turn (T)**: 31 is 13.54%
- **Bend region (S)**: 0 is 0.00%
- **Random coil (C)**: 92 is 40.17%
- **Ambiguous states (?)**: 0 is 0.00%
- **Other states**: 0 is 0.00%

**FIGURE 14**: SOPMA of Cellulase

**Inference**

SOPMA is another secondary structure prediction tool available in ExPASy. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. It Fig 4.14 shows that 14.85% of the cellulase sequence can attain alpha helix, 31.44% of extended strands and random coils were 40.17%.
SOPMA of Pectinase

**FIGURE 15**: SOPMA of Pectinase

**Inference**
SOPMA is another secondary structure prediction tool available in ExPASy. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. It (Fig 4.15) shows that 7.92% of the pectinase sequence can attain alpha helix, 38.26% of extended strands, β strand account for 10.03% and random coils were 43.80%.
**Inference:** The TMHMM result (Fig 4.16) shows that there is one transmembrane helix in the protease sequence and since the expected number of amino acid in the trans membrane region is more than 18 it implies that it can be a transmembrane protein or signal peptide with 21.01813 amino acids.

**TMHMM OF Cellulase**

```
TMHMM result

Number of predicted TMH:  0
Exp number of AAs in TMHs:  9.971049999999339
Total prob of P-site:  0.38082

FIGURE 17: TMHMM OF Cellulase
```
Inference: The TMHMM result (Fig 4.17) shows that there is no transmembrane helix in the cellulase sequence since the expected number of amino acid in the trans membrane region is more than 18.

TMHMM OF Pectinase

**TMHMM result**

<table>
<thead>
<tr>
<th>WEBSEQUENCE</th>
<th>Length: 379</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEBSEQUENCE</td>
<td>Number of predicted TMHs: 0</td>
</tr>
<tr>
<td>WEBSEQUENCE</td>
<td>Exp number of AAs in TMHs: 3.20296</td>
</tr>
<tr>
<td>WEBSEQUENCE</td>
<td>Exp number, first 60 AAs: 3.20219</td>
</tr>
<tr>
<td>WEBSEQUENCE</td>
<td>Total prob of N-in: 0.14545</td>
</tr>
</tbody>
</table>

**FIGURE 18: TMHMM OF Pectinase**

Inference: The TMHMM result (Fig 4.18) shows that there is no transmembrane helix in the pectinase sequence and since the expected number of amino acid in the trans membrane region is more than 18.

**CONCLUSION**

Protease, Cellulases and Pectinase refer to a class of enzymes produced majorly by fungi, bacteria and protozoans that catalyze proteolysis, cellulolysis and pecteolysis. These enzymes used extensively in various industries, especially in textile, food, paper and pulp, wine production, leather, fruit and juice, animal feed, detergent and in the bioconversion of wastes. The extensive use of these enzymes in industries depends on the cost of the enzyme and hence considerable research is being carried out to isolate better microbial strains and also to develop new fermentation processes with the aim to reduce the product cost. Protease from *Pseudomonas* species, Cellulase from *Ceratocystis sp* and Pectinase from *Alternaria* sp were analyzed using computational tools. The physicochemical properties of the selected enzyme proteins were analyzed by using ExPASy’s ProtParam tool and it was found that the molecular weight (M.Wt) ranges between 51020 Da for protease, 24370.91Da for cellulase and 38816Da for pectinase. Isoelectric Points (pI) of the organisms were found to be acidic in nature for pectinase and alkaline in nature for protease and cellulase. The aliphatic index infers that all the three enzymes were stable. The negative value of GRAVY indicates that there will be better interaction with water. The secondary structure prediction was done by SOPMA, GOR IV which showed that random coils dominated all the other conformations in all the three enzymes. The transmembrane protein prediction was conducted using TMHMM which shows that, for protease it has transmembrane regions but for pectinase and cellulase there was no transmembrane region. When the sequences of each enzymes were compared by multiple sequence alignment, protease showed similarity with the carboxyl terminal of protease producing organisms.

**ACKNOWLEDGEMENT**

The author S V Megha is grateful to R & D Bharatiyar university and Biotechnology department. Megha extends her gratitude to Dr. B Annadurai, Dr. R Puvanakrishnan, Scientist at department of Biotechnology CLRI, Chennai, Dr R Easwaramoorthy and all the well wishers for useful suggestions.

**REFERENCES**

In silico characterization of some food enzymes Using computational tools


