GLOBAL JOURNAL OF BIO-SCIENCE AND BIOTECHNOLOGY

www.scienceandnature.org

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. Protscale 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 microorganisms. 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 cellulosic 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 poly galacturonase. 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 pI, instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY). The computed isolelectric point (pI) will be useful for developing buffer systems for purification by isoelectric focusing method (Sivakumar *et al.*, 2007).

Extinction Coefficient: It gives how much light a protein absorb at a certain wavelength.

pI : pI is the pH at which the protein has no net charge. Protein pI is calculated using pKa value of each amino acid. pKa of amino acid depends on the side chains present in each amino acid. It has an important role in defining pH dependent character of protein.

Half Life: It is a prediction of time it takes for half of the protein in a cell to disappear after its synthesis in a cell

GRAVY(Grand Average of Hydropathy) : It is calculated by adding the hydropathy values of each amino acid residue and dividing the number of residues in the sequence or length of sequence. Increasing the positive score indicates greater hydrophobicity.

Aliphatic index: It is described as the relative volume occupied by the aminoacid (ala, val, iso leu and leu) which have an aliphatic chain on the strand. Increase in the aliphatic value implies increase in the stability of protein. (Walker, 2005).

Instability index: It provides an estimate of the stability of sample protein. The protein whose instability index is smaller than 40 is predicted as stable, a value above 40 may be unstable. (Guruprasad et al, 1990).

Method

To know physicochemical properties of proteinase, cellulase and pectinase first we go to proteomic expasy tools by the help of http://www.expasy.ch/tools. The procedure was explained in general system and method.

Compute pI/Mw

Compute pi/Mw is one of the proteomics ExPASy tools, specifically it is a primary structure analysis tools. This tool calculates the estimated isoelectric point and molecular weight of the protein sequence. Their parameters are useful if we want to know the approximate region of a 2-D gel where a protein may be found.

Method

To know isoelectric point and molecular weight of enzymes first we go to proteomics ExPASy tools through http: www.expasy.ch/tools. The procedure was explained in general system and method.

Prot Scale

Prot Scale is one of the proteomics ExPASy tools, specifically. It is primary structure of protein analysis tool. This tools hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales (Bjellquish et al 1993,Bjellquish et al 2005)

Method

To know hydrophobic 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 bassed on probability parameters derived from the studies of protein teritary strucrure 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 system and method.

TMHMM

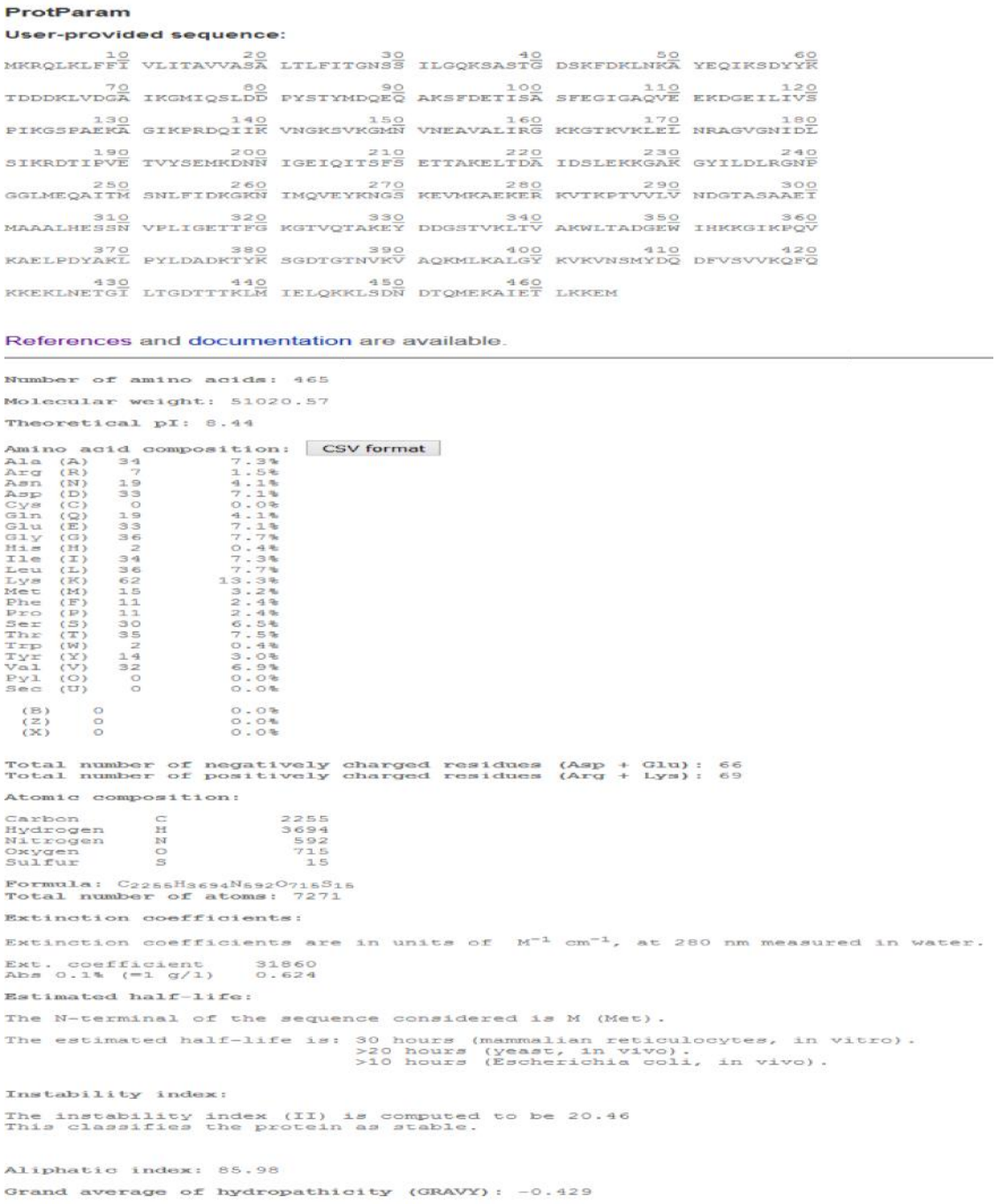
TMHMM is a method for prediction of transmembrane helices based on a hidden Markov model developed by Anders Krogh and Erik Sonnhammer. It predict transmembrane helices and discriminate between soluble and membrane proteins with high degree of accuracy.At a time a user can submit as many as 4000 protein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



ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MKRQLKLFFI VLITAVVASA LTLFITGNSS ILGQKSASTG DSKFDKLNKA YEQIKSDYYK

        70         80         90        100        110        120
TDDDKLVDGA IKGMIQSLDD PYSTYMDQEQ AKSFDETISA SFEGIGAQVE EKDGEILIVS

       130        140        150        160        170        180
PIKGSPAEKA GIKPRDQIIK VNGKSVKGMN VNEAVALIRG KKGTKVKLEL NRAGVGNIDL

       190        200        210        220        230        240
SIKRDTIPVE TVYSEMKDNN IGEIQITSFS ETTAKELTDA IDSLEKKGAK GYILDLRGNP

       250        260        270        280        290        300
GGLMEQAITM SNLFIDKGKN IMQVEYKNGS KEVMKAEKER KVTKPTVVLV NDGTASAAEI

       310        320        330        340        350        360
MAAALHESSN VPLIGETTFG KGTVQTAKEY DDGSTVKLTV AKWLTADGEW IHKKGIKPQV

       370        380        390        400        410        420
KAELPDYAKL PYLDADKTYK SGDTGTNVKV AQKMLKALGY KVKVNSMYDQ DFVSVVKQFQ

       430        440        450        460
KKEKLNETGI LTGDTTTKLM IELQKKLSDN DTQMEKAIET LKKEM
```

References and documentation are available.

Number of amino acids: 465

Molecular weight: 51020.57

Theoretical pI: 8.44

Amino acid composition: CSV format

Ala	(A)	34	7.3%
Arg	(R)	7	1.5%
Asn	(N)	19	4.1%
Asp	(D)	33	7.1%
Cys	(C)	0	0.0%
Gln	(Q)	19	4.1%
Glu	(E)	33	7.1%
Gly	(G)	36	7.7%
His	(H)	2	0.4%
Ile	(I)	34	7.3%
Leu	(L)	36	7.7%
Lys	(K)	62	13.3%
Met	(M)	15	3.2%
Phe	(F)	11	2.4%
Pro	(P)	11	2.4%
Ser	(S)	30	6.5%
Thr	(T)	35	7.5%
Trp	(W)	2	0.4%
Tyr	(Y)	14	3.0%
Val	(V)	32	6.9%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%
(B)		0	0.0%
(Z)		0	0.0%
(X)		0	0.0%

Total number of negatively charged residues (Asp + Glu): 66
Total number of positively charged residues (Arg + Lys): 69

Atomic composition:

Carbon	C	2255
Hydrogen	H	3694
Nitrogen	N	592
Oxygen	O	715
Sulfur	S	15

Formula: $C_{2255}H_{3694}N_{592}O_{715}S_{15}$
Total number of atoms: 7271

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 31860
Abs 0.1% (=1 g/l) 0.624

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 20.46
This classifies the protein as stable.

Aliphatic index: 85.98

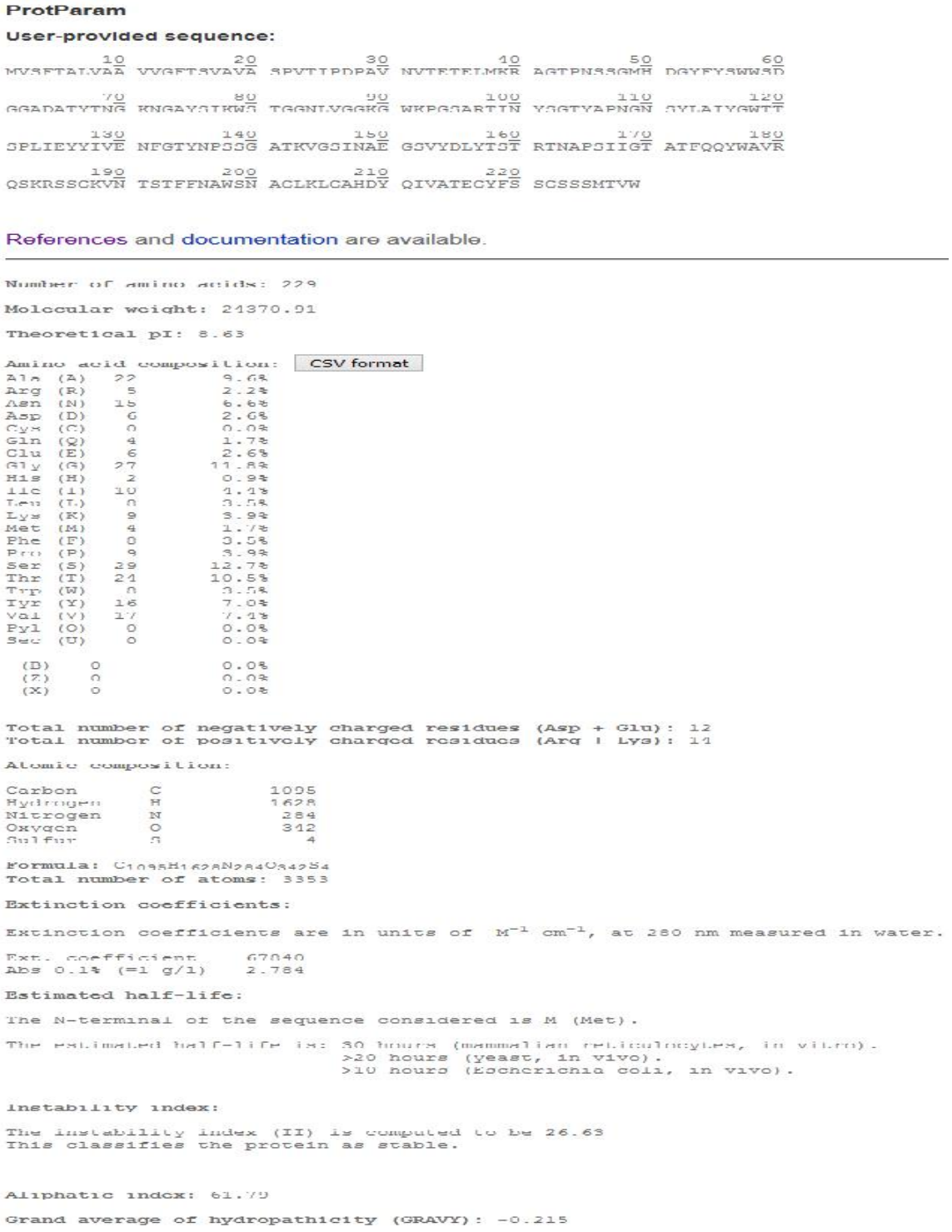
Grand average of hydropathicity (GRAVY): -0.429

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 Hydropathicity (GRAVY). The physicochemical parameters show that the molecular weight of protease is around 51020.57 Da. The instability index is used to measure the invivo 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



ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MVSFTALVAA VVGFTSVAVA SPVTTPDPAV NVTETELMKR AGTPNSSGMH DGYFYSWWSD

        70         80         90        100        110        120
GGADATYTNG KNGAYSIKWS TGGNLVGGKG WKPGSARTIN YSGTYAPNGN SYLAIYGWTT

       130        140        150        160        170        180
SPLIEYYIVE NFGTYNPSSG ATKVGSINAE GSVYDLYTST RTNAPSIIGT ATFQQYWAVR

       190        200        210        220
QSKRSSCKVN TSTFFNAWSN ACLKLGAHDY QIVATECYFS SGSSSMTVW
```

References and documentation are available.

```
Number of amino acids: 229

Molecular weight: 24370.91

Theoretical pI: 8.63

Amino acid composition:   CSV format
Ala (A)   22          9.6%
Arg (R)    5          2.2%
Asn (N)   15          6.6%
Asp (D)    6          2.6%
Cys (C)    0          0.0%
Gln (Q)    4          1.7%
Glu (E)    6          2.6%
Gly (G)   27         11.8%
His (H)    2          0.9%
Ile (I)   10          4.4%
Leu (L)    8          3.5%
Lys (K)    9          3.9%
Met (M)    4          1.7%
Phe (F)    8          3.5%
Pro (P)    9          3.9%
Ser (S)   29         12.7%
Thr (T)   24         10.5%
Trp (W)    8          3.5%
Tyr (Y)   16          7.0%
Val (V)   17          7.4%
Pyl (O)    0          0.0%
Sec (U)    0          0.0%

  (B)     0          0.0%
  (Z)     0          0.0%
  (X)     0          0.0%

Total number of negatively charged residues (Asp + Glu): 12
Total number of positively charged residues (Arg + Lys): 14

Atomic composition:

Carbon      C          1095
Hydrogen    H          1628
Nitrogen    N           284
Oxygen      O           342
Sulfur      S             4

Formula: C1095H1628N284O342S4
Total number of atoms: 3353

Extinction coefficients:

Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.

Ext. coefficient    67040
Abs 0.1% (=1 g/l)   2.784

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 26.63
This classifies the protein as stable.


Aliphatic index: 81.79

Grand average of hydropathicity (GRAVY): -0.215
```

FIGURE 2: The ProtParam result of Cellulase

Inference

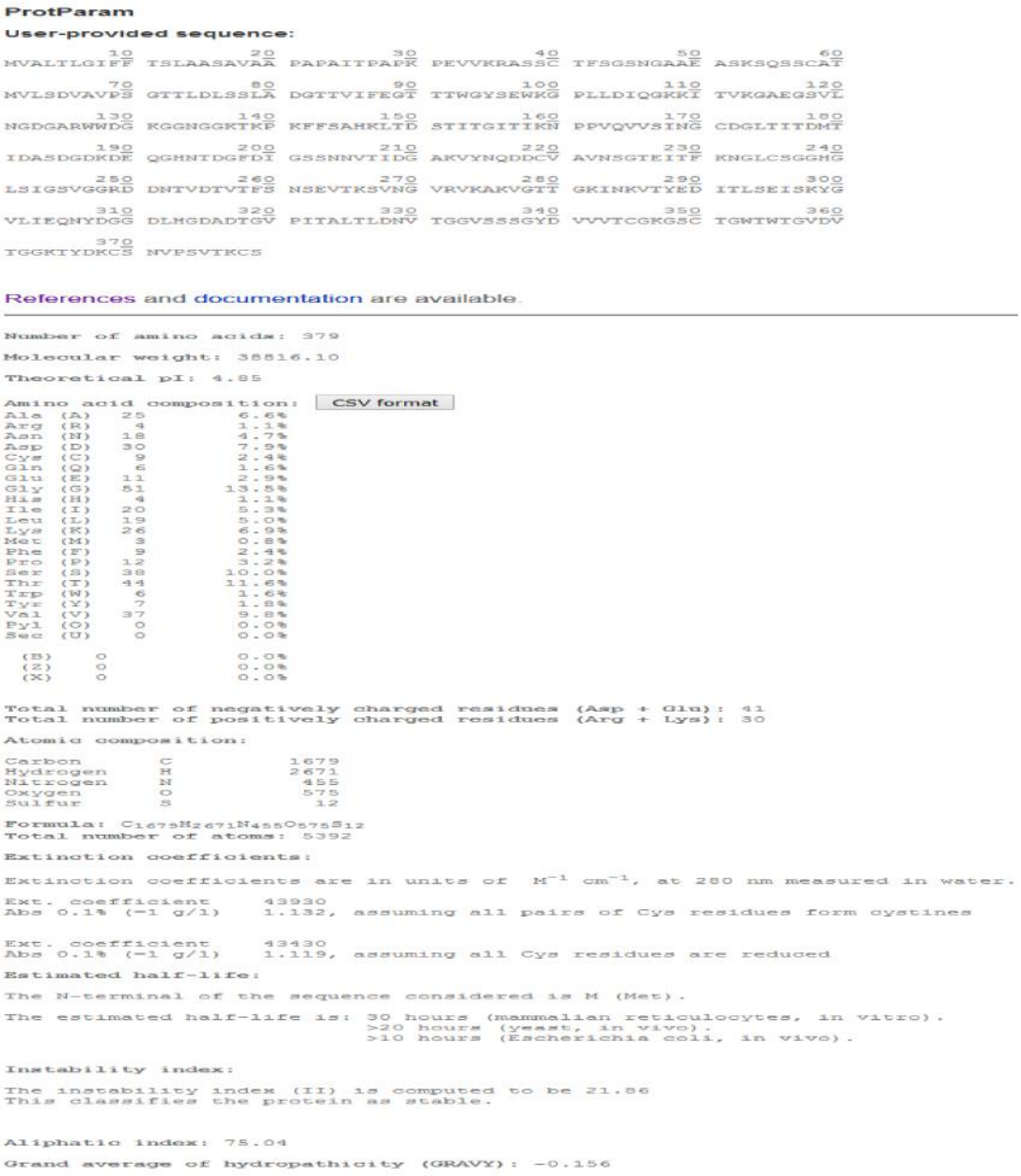
The physicochemical properties of cellulase were predicted by using ProtParam tool (Fig 4.2). The Prot Param includes the following computed parameters: Molecular weight(M Wt), theoretical pI, Instability Index, Aliphatic Index, The Grand Average of Hydropathicity(GRAVY). The physicochemical parameters 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

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 8.63 shows that cellulase 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 61.79 is high showing that the cellulase are stable. The GRAVY value of the cellulasee is -0.215 which is lower showed the better interaction of protease with water

Protparam Pectinase

Inference

The physicochemical properties of pectinase were predicted by using ProtParam tool (Fig 4.3). The Prot Param 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 invivo 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 75.04 is high showing that the pectinase are stable. The GRAVY value of the pectinase is -0.156 which is lower showed the better interaction of pectinase with water.



ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MVALTLGIFF TSLAASAVAA PAPAITPAPK PEVVKRASSC TFSGSNGAAE ASKSQSSCAT

        70         80         90        100        110        120
MVLSDVAVPS GTTLDLSSLA DGTTVIFEGT TTWGYSEWKG PLLDIQGKKI TVKGAEGSVL

       130        140        150        160        170        180
NGDGARWWDG KGGNGGKTKP KFFSAHKLTD STITGITIKN PPVQVVSING CDGLTITDMT

       190        200        210        220        230        240
IDASDGDKDE QGHNTDGFDI GSSNNVTIDG AKVYNQDDCV AVNSGTEITF KNGLCSGGHG

       250        260        270        280        290        300
LSIGSVGGRD DNTVDTVTFS NSEVTKSVNG VRVKAKVGTT GKINKVTYED ITLSEISKYG

       310        320        330        340        350        360
VLIEQNYDGG DLHGDADTGV PITALTLDNV TGGVSSSGYD VVVTCGKGSC TGWTWTGVDV

       370
TGGKIYDKCS NVPSVTKCS
```

References and documentation are available.

```
Number of amino acids: 379

Molecular weight: 38816.10

Theoretical pI: 4.85

Amino acid composition:   CSV format

Ala (A)   25          6.6%
Arg (R)    4          1.1%
Asn (N)   18          4.7%
Asp (D)   30          7.9%
Cys (C)    9          2.4%
Gln (Q)    6          1.6%
Glu (E)   11          2.9%
Gly (G)   51         13.5%
His (H)    4          1.1%
Ile (I)   20          5.3%
Leu (L)   19          5.0%
Lys (K)   26          6.9%
Met (M)    3          0.8%
Phe (F)    9          2.4%
Pro (P)   12          3.2%
Ser (S)   38         10.0%
Thr (T)   44         11.6%
Trp (W)    6          1.6%
Tyr (Y)    7          1.8%
Val (V)   37          9.8%
Pyl (O)    0          0.0%
Sec (U)    0          0.0%

  (B)     0          0.0%
  (Z)     0          0.0%
  (X)     0          0.0%

Total number of negatively charged residues (Asp + Glu): 41
Total number of positively charged residues (Arg + Lys): 30

Atomic composition:

Carbon      C          1679
Hydrogen    H          2671
Nitrogen    N           455
Oxygen      O           575
Sulfur      S            12

Formula: C1679H2671N455O575S12
Total number of atoms: 5392

Extinction coefficients:

Extinction coefficients are in units of  M-1 cm-1, at 280 nm measured in water.

Ext. coefficient      43930
Abs 0.1% (=1 g/l)     1.132, assuming all pairs of Cys residues form cystines

Ext. coefficient      43430
Abs 0.1% (=1 g/l)     1.119, 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 21.86
This classifies the protein as stable.

Aliphatic index: 75.04

Grand average of hydropathicity (GRAVY): -0.156
```

FIGURE 3: The Protparam result of Pectinase

Protscale of Protease

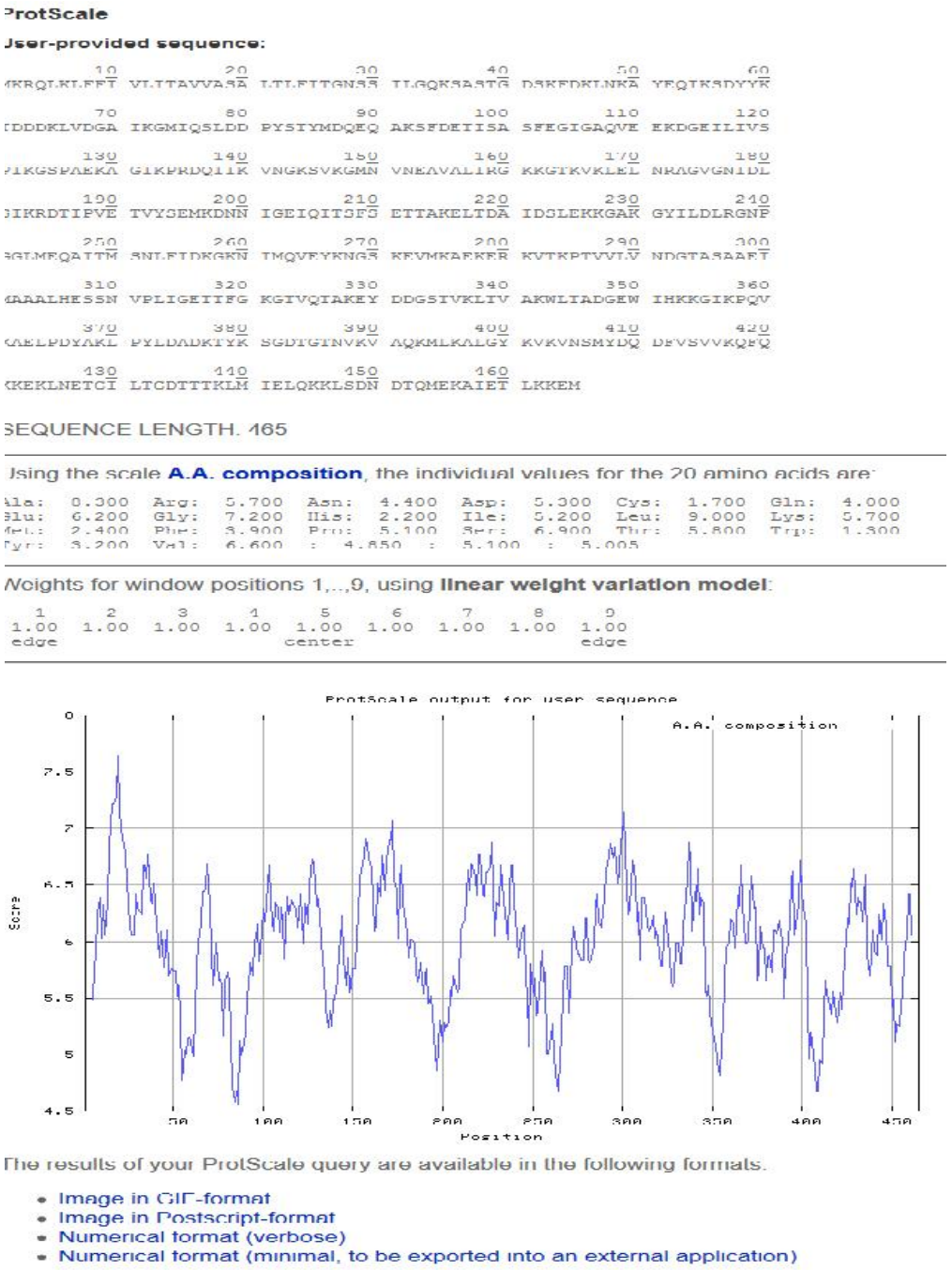


FIGURE 4. The ProtScale of protease

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. (Bjellquish et al 1993,Bjellquish et al 2005).

Protscale cellulase

ProtScale

User-provided sequence:

```
        10         20         30         40         50         60
MVSFTALVAA VVGFTSVAVA SPVTIPDPAV NVTETELMKR AGTPNSSGMH DGYFYSWWSD

        70         80         90        100        110        120
GGADATYTNG KNGAYSIKWS TGGNLVGGKG WKPGSARTIN YSGTYAPNGN SYLAIYGWTT

       130        140        150        160        170        180
SPLIEYYIVE NFGTYNPSSG ATKVGSINAE GSVYDLYTST RTNAPSIIGT ATFQQYWAVR

       190        200        210        220
QSKRSSGKVN TSTFFNAWSN AGLKLGAHDY QIVATEGYFS SGSSSMTVW
```

SEQUENCE LENGTH: 229

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:   7.200  His:   2.200  Ile:   5.200  Leu:   9.000  Lys:   5.700
Met:   2.400  Phe:   3.900  Pro:   5.100  Ser:   6.900  Thr:   5.800  Trp:   1.300
Tyr:   3.200  Val:   6.600  :  4.850  :  5.100  :  5.005
```

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

```
   1     2     3     4     5     6     7     8     9
1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
edge                    center                edge
```

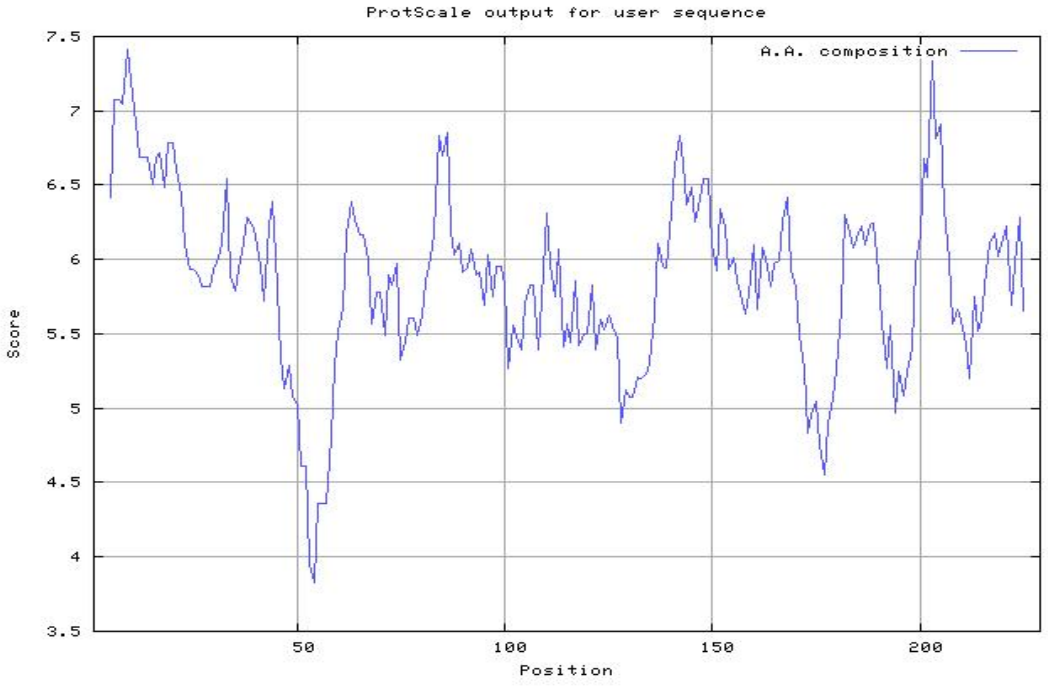


FIGURE 5: ProtScale of Cellulase

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).

Protscale Pectinase

ProtScale

User-provided sequence:

```
        10         20         30         40         50         60
MVALTLGIFF TSLAASAVAA PAPAITPAPK PEVVKKASSC TFSGSNGAAE ASKSQSSCAT

        70         80         90        100        110        120
MVLSDVAVPS GTTLDLSSLA DGTTVIFEGT TTWGYSEWKG PLLDIQGKKI TVKGAEGSVL

       130        140        150        160        170        180
NGDGARWWDG KGGNGGKTKP KFFSAHKLTD STITGITIKN PPVQVVSING CDGLTITDMT

       190        200        210        220        230        240
IDASDGDKDE QGHNTDGFDI GSSNNVTIDG AKVYNQDDCV AVNSGTEITF KNGLCSGGHG

       250        260        270        280        290        300
LSIGSVGGRD DNTVDTVTFS NSEVTKSVNG VRVKAKVGTT GKINKVTYED ITLSEISKYG

       310        320        330        340        350        360
VLIEQNYDGG DLHGDADTGV PITALTLDNV TGGVSSSGYD VVVTCGKGSC TGWTWTGVDV

       370
TGGKTYDKCS NVPSVTKCS
```

SEQUENCE LENGTH: 379

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:   7.200   His:   2.200   Ile:   5.200   Leu:   9.000   Lys:   5.700
Met:   2.400   Phe:   3.900   Pro:   5.100   Ser:   6.900   Thr:   5.800   Trp:   1.300
Tyr:   3.200   Val:   6.600   :   4.050   :   5.100   :   5.005
```

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

```
   1     2     3     4     5     6     7     8     9
1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
edge                     center                edge
```

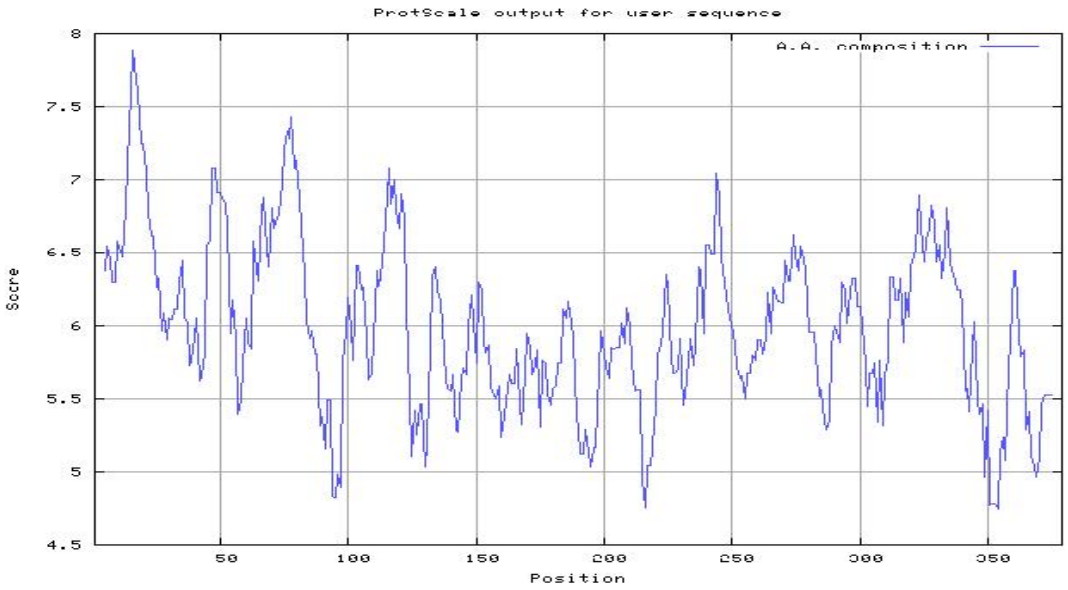


FIGURE 6: ProtScale of Pectinase

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

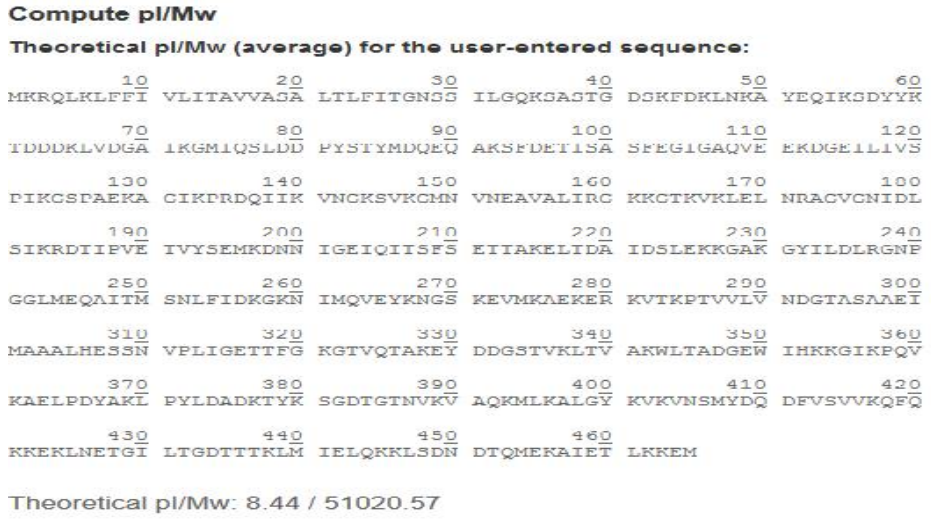


FIGURE 7: Theoretical pI /MW of protease

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

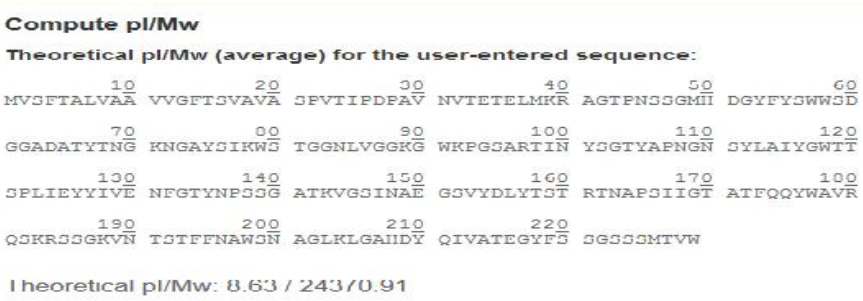


FIGURE 8: Theoretical pI /MW of Cellulase
Inference: The theoretical pI and Molecular weight of cellulase (Fig 4.8)was found to be 8.35 and 50042 respectively
Theoretical pI /MW of Pectinase

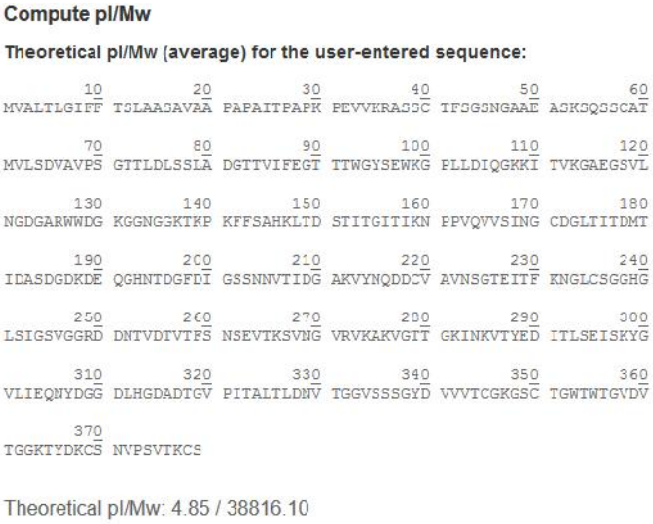


FIGURE 9: Theoretical pI /MW of Pectinase
Inference: The theoretical pI and Molecular weight of pectinase (Fig4. 9)was found to be 4.85 and 38816.10 respectively

GOR IV
Protease

FIGURE 10: GOR IV of protease

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

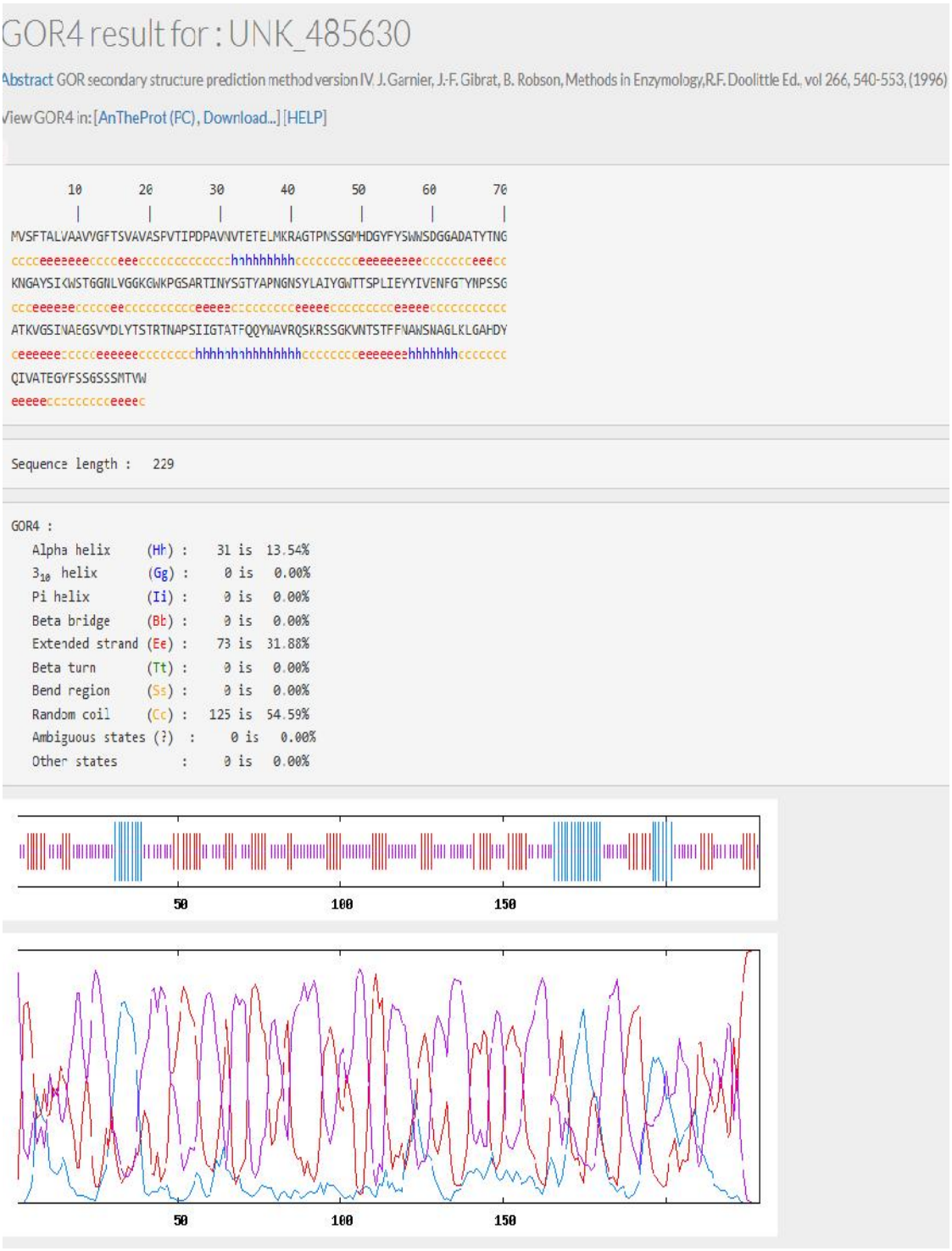


FIGURE 11: GOR IV of Cellulase

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.

GOR IV of Pectinase

FIGURE 12: GOR IV of Pectinase

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

Protease

FIGURE 13: SOPMA of protease

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

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%.

TMHMM
Protease

TMHMM result

HELP with output formats

```
# WEBSEQUENCE Length: 465
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 21.01813
# WEBSEQUENCE Exp number, first 60 AAs:  21.01794
# WEBSEQUENCE Total prob of N-in:        0.99922
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE     TMHMM2.0        inside      1     6
WEBSEQUENCE     TMHMM2.0        TMhelix     7    26
WEBSEQUENCE     TMHMM2.0        outside    27   465
```

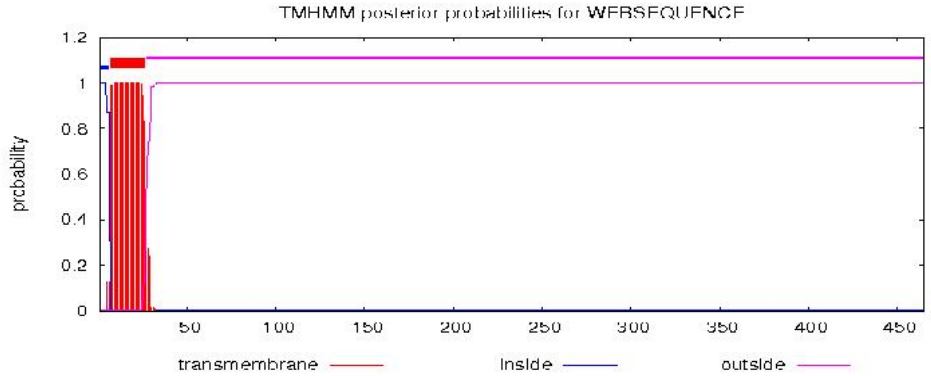


FIGURE 16: TMHMM OF Protease

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

HELP with output formats

```
# WEBSEQUENCE Length: 229
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 9.97104999999999998
# WEBSEQUENCE Exp number, first 60 AAs:  9.50842
# WEBSEQUENCE Total prob of N-in:        0.38832
WEBSEQUENCE     TMHMM2.0        outside     1   229
```

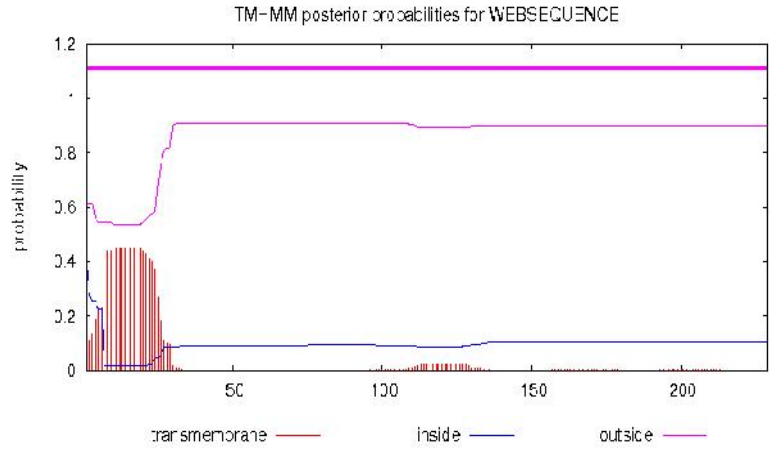


FIGURE 17: TMHMM OF Cellulase

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

HELP with output formats

```
# WEBSEQUENCE Length: 379
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 3.20296
# WEBSEQUENCE Exp number, first 60 AAs:  3.20219
# WEBSEQUENCE Total prob of N-in:        0.14545
WEBSEQUENCE     TMHMM2.0        outside      1   379
```

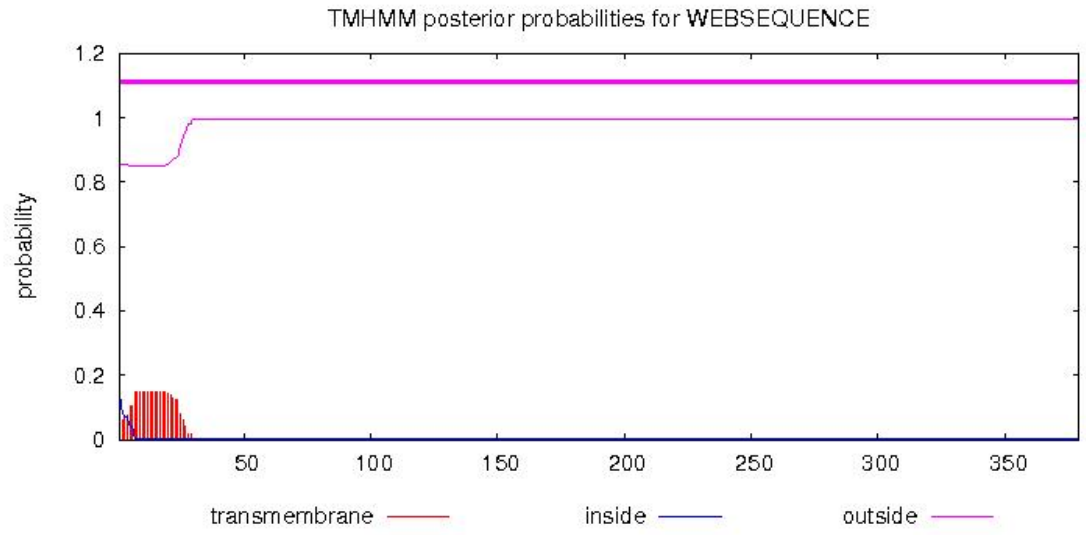


FIGURE 18: TMHMM OF Pectinase

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 have transmembrane regions but for pectinase and cellulase there was no transmembrane region. When the sequences of each enzymes were compaired 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 extend 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

Ali, A.J., Ahmed, A.H., Zahraa, A. and Umar, A. (2011) Optimization of Cellulase Production by *Aspergillus niger*

and *Tricoderma viride* using Sugar Cane Waste, *Journal of Yeast and Fungal Research*, 2(2), 19–23.

Altschul, Stephen, F., John, C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Ashokan, K.V., Mundaganur, D.S. and Mundaganur, Y.D. (2011) Catalase: Phylogenetic Characterization to Explore Protein Cluster, *Journal of Research in Bioinformatics*, 1, 001-008.

Bhat, M.K. (2000) Cellulases and Related Enzymes in Biotechnology, *Biotechnology Advances*, 18, 355–383.

Bjellqvist, B., Hughes, G.J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.Ch., Frutiger, S. & Hochstrasser, D.F. (1993) *The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences.* Electrophoresis, 14, 1023-1031

Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E.,2005, *Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions.* Electrophoresis, 15, 529-539.

Garnier, J., Gibrat, J.F., Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence, Methids in enzymology, vol. 266, 540-553

Garg, G., Singh, A., Kaur, A., Singh, R., Kaur, J. and Mahajan, R. (2016) Microbial pectinase : an ecofriendly tool of nature for industries, 3 biotech, Vol 6(1), 47.

Geourjon, C. & Deléage, G. (1995) SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments., Cabios 11, 681-684.

Guruprasad, K., Bhasker Reddy, B.V. and Madhusudan, W. P. (1990) Correlation between stability of aprotein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from itsprimary sequence, *Protein Engineering* 4(2), 155-161.

Laskowski, R.A., Mac Arthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROTCHEK: A program to check the stereo chemical quality of protein structure, Journal of Applied Crystallography, 26, 283-291.

Neelima, A., Amit, K.B., Srilaxmi, M., Upadhyayula, S M. (2009) Comparative Characterization of Commercially important Xylanase Enzymes, Bioinformation, 3(10), 446-453.

Prashant, V.T., Uddhav, S.C., Madura, S. M., Vishal, P.D. and Renuka, R.K. (2010)Secondary Structure Prediction and Phylogenetic Analysis of Salt Tolerant Proteins, *Global Journal of Molecular Sciences*, 5 (1),30-36

Sivakumar, K., Balaji, S. and Gangaradha krishnan (2007) In-silico Characterization of Antifreeze Proteins using Computational Tools and Servers. *J. Chem. Sci*., 119(5), 571–579.

Suresh, C.K., Snishamol, C. and Nagendra, P.G., (2005) Cellulase Production by Native Bacteria Using Water Hyacinth as Substrate under Solid State Fermentation, *Malaysian Journal of Micorbiology*, 1(2), 25-29.

Walker, J.M. (2005) The Proteomics Protocols Handbook. Humana Press (Chap. 52).