INTERNATIONAL JOURNAL OF ADVANCED BIOLOGICAL RESEARCH

www.scienceandnature.org

INSILICO STUDIES AND MOLECULAR MODELLING OF FOOD ENZYMES FROM DIFFERENT SOURCES

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

ABSTRACT
Food enzymes are subjected to *insilico* studies and molecular modelling with a strong basis of biochemical and biophysical knowledge. In recent years, the dramatic development of the genomic and post genomic research has provided this as well as all other fields of life sciences with a massive body of new data, including, but not limited to, protein sequence and structural data. By integrating these new data with the wealth of information available in the literature, it is possible to achieve an unprecedented overview of the properties and functions of Food enzymes in the context of biological systems. To this aim, the role of bioinformatics is essential. In this work, we use bioinformatics tools and databases that we have developed for the study of Food enzymes to gain insights into the functions of components in Food enzymes, its coordination properties, and the usage of Food enzymes in living organisms. The following results like Compute pI/MW, Protscale, Peptide cutter (Primary sequence analysis), GORIV, SOPMA, TmPred, TNHMM (Secondary Structure), Pair wise sequence alignment, Multiple sequence alignment, Wire frame model, Backbone, Sticks, Space fill model, Ball and Stick model, Strands, Cartoons, Molecular surface of Proteinase, Pectinase, Cellulase and Laccase were analysed and presented.

KEY WORDS: PDB, PROSA, Homology modeling, Compute pI, GORIV, TMPred, TNHMM.

INTRODUCTION

Enzymes are in the center of biochemical processes. They catalyze largest part of all chemical reactions in the living organisms (from viruses to human) and are characterized by unique capabilities to accelerate the reaction rates and to catalyze specific or very selective number of chemical transformations. Not surprisingly the enzymes received massive application in biomedicine, pharmacy, biotechnological and chemical industry. The current progress in understanding enzymes underlines the new perspective of their applications and utilization in important areas for us. There is vastly growing amount of novel structures, spectroscopic data about intermediates, novel inhibitors synthesized and even enzymes with novel functions engineered. The current thematic issue of Enzymes studies, its mechanisms, inhibition and dynamics is focused on high quality studies by broad range of experimental and computational methods. Contributions focused on integrated modelling/ experimental or combination between different experimental methods and the multilevel applications of computational methods are investigated. Highly valued will be combined fundamental and innovative contributions focused on the applications of the enzyme mechanisms and in the all areas with impact for the society: industry, health, food etc. Finally it strengthens, develop, demonstrate and facilitate the independence of thinking, creativity, initiative of researchers at all levels.

Proteases execute a large variety of functions and have important biotechnological applications. Proteases represent one of the three largest groups of industrial enzymes and find application in detergents, leather industry, food industry, pharmaceutical industry and bioremediation processes. For an enzyme to be used as an detergent additive it should be stable and active in the presence of typical detergent ingredients, such as surfactants, builders, bleaching agents, bleach activators, fillers, fabric softeners and various other formulation aids.

Cellulase I.U.B.:3.2.1.41, 4-(1, 3; 1, 4)-β-D-Glucan-4-glucanohydrolase, Cellulase refers to a group of enzymes which, acting together, hydrolyze cellulose. It has been reviewed by Whitaker (1971).

In nature cellulose is usually associated with other polysaccharides such as xylan or lignin. It is the skeletal basis of plant cell walls. According to Spano *et al*. (1975) cellulose is the most abundant organic source of food, fuel and chemicals. However, its usefulness is dependent upon its hydrolysis to glucose. Acid and high temperature degradation is unsatisfactory in that the resulting sugars are decomposed; also, waste cellulose contains impurities that generate unwanted by-products under these harsh conditions.

Cellulase is a group of enzymes that catalysescellulolysis. It is mainly produced by fungi, bacteria and some protozoans. The active research of cellulases was started in 1950. After knowing its potentiality to convert lignocellulases. It is studies extensively due to their applications in the hydrolysis of cellulose, the most abundant biopolymer and potential source of utilizable sugar, which serves as a raw material in the production of chemicals and fuel (Ali *et al* 2011, Pradeep *et al*, 2012). Since, Cellulases is used mostly in textiles, food and the bioconversion lignocellulosic waste to alcohol, it becomes industrially important. Because largely is used in the industries, large scale of production (Microbial strains). Isolation and purification, Procedures are required. In

addition to that the computational tools and insilico studies are required to preserve and reduce the cost of cellulase. Bioinformatics revolutionized the field of molecular biology. The raw sequences 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 (Prasanth *et al*., 2010).

Pectinases, or pectinolytic enzymes, are produced by a number of bacteria, yeast, fungi, protozoa, insects, nematodes and plants [23] in order to degrade (to obtain a carbon source) or to modify (in fruit ripening etc.) the heteropolysaccharide pectin. They can be classified, based on the type of linkages they attack, into the esterases, which saponify the substrate, and the depolymerases. The depolymerases can be subdivided based on the bond cleavage mechanism into the class of the hydrolases (hydrolytic cleavage) and the class of the lyases (β-elimination cleavage). Pectinases show different substrate specificity, but basically they can be separated into a group of homogalacturonan and a group of rhamnogalacturonan specific enzymes. Besides the main pectin backbone-degrading enzymes, the 'accessory' enzymes, active towards the side chains of pectin, are needed to fully accomplish pectin degradation.

Laccases (EC 1.10.3.2, benezenediol: oxygen oxidoreductases), first described from the lacquer tree Rhus vernicifera (Yoshida, 1883), are multi-copper oxidases that catabolize a variety of aromatic ring structures, e.g. p-diphenols, but not tyrosine, via reduction of molecular oxygen to water . The general structure of laccases is rather diverse, but the structure of the active site seems to be well conserved in fungal laccases. Laccases usually have three copper ions (T1, T2, and T3) coordinated with histidine residues (Giardina *et al*., 2010, Solomon *et al*., 1996). The T1 copper is also termed the "blue copper", imparting the characteristic blue color. The lack of the T1 copper is a feature of the socalled "yellow" or "white" laccase (Baldrian, 2006). The absence of T1 copper in some laccases has caused some authors to question if such laccases can in fact be termed "true" laccases, although they can oxidize phenols. The term "laccases with unusual spectral properties" has been suggested as more appropriate. Laccases are commonly found in higher plants, fungi, insects, and microorganisms. Plant laccases have been suggested to be involved in lignin polymerization, but experimental proof was missing. Recently, experimental studies in Arabidopsis thaliana and Populus trichocarpa provided evidence that laccases are in fact involved in lignification. Fungal laccases have been more intensively studied, and exhibit various physiological functions, including lignin degradation (Arora & Sharma, 2010; Thurston, 1994), an involvement in virulence, pathogenesis, conidial pigmentation, and morphology.

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 V *et al*, 2010). In the present bioinformatics analysis characterization of Food enzymes from different sources were carried out. Protein sequences were retrieved from NCBI and were subjected to ProtParam to analyze 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. These parameters will assist the biochemist and physiologists in extraction, purification, separation and industrial applications of the enzyme

System (Materials) and Tools

PDB

The PDB is the single worldwide repository for the processing and distribution of 3-D structure data of large molecules of proteins and nucleic acids, as determined by X-ray crystallography or nuclear magnetic resonance (NMR) imaging. The molecules described by the files are usually viewed locally by dedicated software, or can be visualized on the World Wide Web. The number of known protein structures is increasing very rapidly and these are available on the protein Data Bank. There is also a database of structures of 'small' molecules, of interest to biologists concerned with protein-ligand interactions, from the Cambridge Crystallographic Data Centre.

RCSB DATABASE

The World Wide Web site of the protein data bank at the RCSB offers a number of services for submitting and retrieving three-dimensional structure data.The home page of the RCSB site provide links to services for depositing three-dimensional structures, information on how to obtain the status of structures undergoing processing for submission. Ways to download the PDB database and links to other relevant sites and software.

Description of tools used

Protparam

ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in SwissProt or TrEMBL or for a user entered sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated halflife, instability index, aliphatic index and grand average of hydropathicity (GRAVY)

Compute PI/MW

Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of Prot or entries or for user entered sequences

Profilescan

Profile, Scan uses a method called pfscan tofind similarities between a protein or nucleic acid query sequence and a profile library. In this case, three profile libraries are available for searching. First is PROSITE an ExPASy database that catalogs biologically significant sites through the use of motif and sequence profiles and patterns. Second is Pfam. Which is a collection of protein domain families that differ from most such collections in one important aspect the initial alignment of the protein domains is families that differ from most such collections in one important aspect the initial alignment of the protein domains is done by hand? Rather than by depending on automated methods. As such Pfam contains slightly over 500 entries but the entries are potentially of higher quality.

The third profile set is referred to as the Gribskov collection.

SOPMA

The protein Sequence Analysis server at the Centre National de la Recherché Scientifique in Lyons. France takes a unique approach in making secondary Structure predictions: rather than using a single method, it uses five, the predictions from which are subsequently used to come up with a "consensus predictions." The methods used are the GarnierGibrat Robson method the Levin homolog method the double-prediction method the PhD method described above as part of Predict Protein, and the method of CNRS itself, called SOPMA as Briefly, this selfoptimized prediction method builds sub databases is quences with known secondary structure prediction based on sequence similarity. The information from the sub databases is then used to generate a prediction on the query sequence.

SIGNALP

SignalP predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non signal peptide prediction based on a combination of several artificial neural networks and hidden Markov models.

TARGETP

TargetP predicts the sub cellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP).

CHOLOROP

The ChloroP server predicts the presence of chloroplast transit peptides (cTP) in protein sequences and the location of potential cTP cleavage sites. A related service TargetP predicts the sub cellular location of proteins by integrating predictions of chloroplast transit peptides, signal peptides and mitochondrial targeting peptides.

Homology modeling

The amino acid sequence of Food enzyme was obtained from the NCBI protein database (http://www. ncbi.nlm. nih.gov/protein). Crystal structure of Trametes hirsute laccase was taken from the protein data bank (PDB ID: 3FPX) (Berman HM ,2000) and used as the template for building the initial 3D model. The sequence alignment of laccase with the template was accomplished using ClustalW 2.0 (http://www.ebi.ac.uk/ Tools/clustalw2/ index.html). The Modeller 9v7 program (Sali A, Blundell TL (1993) was employed to generate the initial 3D models of laccase. Modeller generates the 3D models by optimization of molecular probability density functions. The optimization process consists of applying the variable target function as well as conjugated gradients and molecular dynamics with simulated annealing. A set of 20 models of Food enzymes were produced based on the resulting alignment obtained above. The outcomes were ranked based on the internal scoring function of Modeller.

RESULTS & DISCUSSION

I. Proteinase

The coils output for proteinase obtained were shown in Figure 1. Coils is program that compares a sequence to a database of known parallel to standard coiled - coils and derives a similarity score. By comparing this score to distribution of scores in globular and coiled-coil proteins, the program then calculates the probability that the sequence will adopt a coiled-coil conformation.

Coils output for proteinase

coils-def-in=../wwwtmp/.COILS.27269.5764.seq -out=../wwwtmp/.COILS.27269.5764.out -mat=2
COILS version 2.1
using MTIDK matrix
no weights
Input file is ../wwwtmp/.COILS.27269.5764.seq
#>proteinase, 466 bases, 7FE6643A checksum.

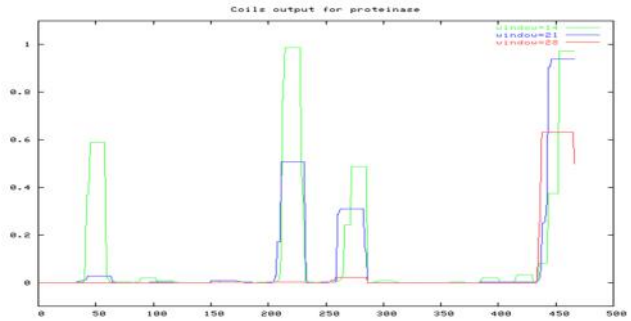


Figure 1 shows the theoretical isoelectric point and molecular weight of the enzyme proteinase from this program the molecular weight of the proteinase is conformed as 43387.78 and the isoelectric point of the protenase is 7.82.

AF015775. Bacillus subtilis...[gi:2415383]
>gi|2415383|gb| AF015775.1|AF015775 Bacillus subtilis YodA (yodA), YodB (yodB), YodC (yodC), YodD (yodD), ABC-transporter (yodE), permease (yodF), proteinase (ctpA), YodH (yodH), YodI (yodI), carboxypeptidase (yodJ), purine nucleoside phosphorylase (deoD), YodL (yodL), YodM (yodM), YodN (yodN), YodO (yodO), YodP (yodP), acetylornitine deacetylase (argE), butirate-acetoacetate CoA transferase (yodR), butyrate acetoacetate-CoA transferase (yodS), YodT (yodT), CgeE (cgeE), CgeD (cgeD), CgeC (cgeC), CgeA (cgeA), CgeB (cgeB), YzxA (yzxA), UDP-glucose epimerase (yodU), YodV (yodV), and YodW (yodW) genes, complete cds; and YodZ (yodZ) gene, partial cds

TCTTTTTCATCGAGCGCACTTCAGTACCGATTTCTTCACAAATCGCAATCGGGATACTGGCTTTGTTTACCAAGAGCAGCTTTGGCTGAATAGGGATGTGCACTATATGTGTGACCGCTCATAATCACACCTGACCCTTGTTTTATG
GTTTCAATAATAGAATCAGATACGACGGCAGCAGCAATAGGTGCATACCCCGCACCGAGTCCCTTCCCCAGTACAGCAATATCAGGTACGGTATCCCAATGCTCTGTCGCGAGCATCCTTCCTGTTCTCCCAAGCCCCGTCATCA
CTTCATCTGCAATAAAAAGCACATCGTGTGTGCGGCATACCTCACTTAATCTCTCATAATATCCCGGAGGCGGAGTAATCGCTGCTCCTGCAGCACCAATAATAGGCTCAGCCACAAAGCCGGCGATGAATTGGCTTCCGATTCT
TTTAATCATGGTGTCCAGTTCATCAGCTGCAGTCTGAACAAAGTCTTCTTCCGTCTCGTGATTCAGCCGATAAATATGTGGAGCTGAGATAGCTGGATACCGCTCAATGAGATGGGTGAACCGGTATCTCCTTTCATAAAAACCA
GACAATGAAAGCGCTCCCAAAGTTATTCCGTGGTAACTGCTCCATCGAGACAAAAAGATGGATTTTTGTGTTTGCTTTTTTTCCTGCCAATACTGAATGGCGATTTTCATAGCTGTTTCTATCGCTTCTGATCCGCTGTTGACAAAA
AAAGACCAATTCACATCTCCGGGCAGCTCCTGTGCCAAGAGAGCGGCTAATTGCTCGGCGGGCTCACTCGTAAACTGTGAGCGGTAAGCAAAAGACACCTGATCAAGCTGTTCTTTCAGCTTCTCAGTCACATCACGAACTCCGT
GGCCGATATTACATGTCACCGCACCTGACGAGCCGTCGAGATATTTTTTTCCGGTCTGATCATAAACATATGAACCCTTCGCATAACTGACAACCGGATAGGCCGAGCTAAGCTCTGGCTTAATCAAATAACTGCTCATGCTCCC
ACCTCAAAACCCCTTTTACCATATTGTATGGTGTCCAAATGGCGGCATGACGTTCCCCGCAAAAACTTTTTCACATAAAAAGAGCCTTCCTAATAAGGAAGGCTCTTTTGAAATTTCAGCATCTAGCAAATAAAAGCGGCTCCAA
CAATAATTAATAAAATGAACAAGACAACAATTAACACAAATGTATTTGTCCCGCAGCCAAATCCGTAACCGTACCCAGAGTATCCATACATGCTG
TTTTCATCCTTTCTTATCACGATATACGACATTTTATGCGGGCCAAGTCCCAAATGGTTTGTGCAAACGCCTATTTCATGTAAGGAAATAATGAAATACAGTATGCACCCGCTCATATTTTTTAAACGTCAGCTTTTCGTAAAACTG
AGTTGCACCGTCAGTGCTTGCAGCCACTAAGTAGGCTGCTTCCATTCCTAGTGATTTGGCGATGGACAGCATTTCTTTCACTAAGATACTGCCGTAGCCCTTCCTTTGAAACTGATCTAAAATTGCCACTTCTTCGATTTTAGCCGT
TTTCTCTTGATGATCGATAAACAGCTCGCCGCACCCAATGGGAACTTTCTGTGCCTGGTCGCTGTAACAAACAACAAGAATGATTATATCTGCTTCATAGAAAGGATATTTTCTTTTTAATTTTTGTTCAGCGATTGCTTCTCCAAT
ATAAATGGAATCGTAAATTTTCATAATAGAAGATCCGTCGCTAAGACTTTTTGCCGTCCCCCATGCGGTTAAAGGATGTCCCATCTGCCCCTTCCAGTCACTAAGCTTCATGTCATATAGTAGCTCATCCTCGACGACAAAACCGC
TTTTTACAAGTGACTGCTTCAAACCGAGCGGAAAAACATGCTGCGGCGACGCTTTTAAATGGAGATAATTCGCATCAAGCAGCTTTGGTACGGAGGGTAAAAAGTTAAGTAACCGCTCCAATGGAAACGTATCATGGAGCTGAA

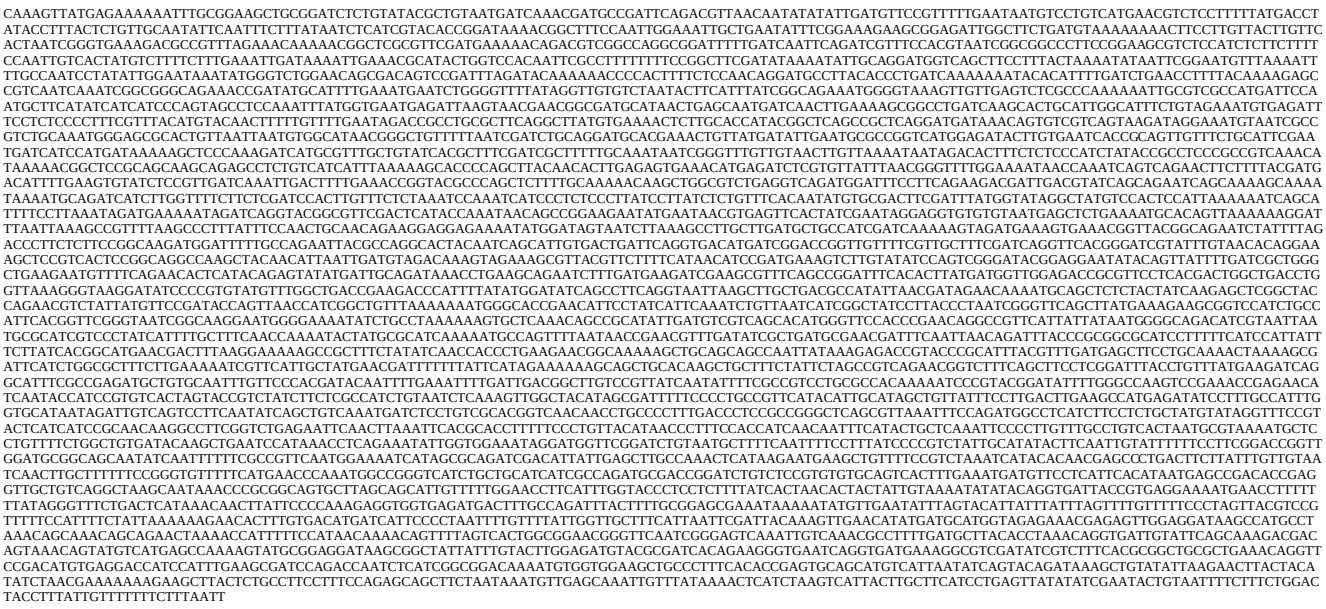


Figure 2 shows the gene sequence of proteinase obtained peptide cutter searches a protein sequence from the SWISS-PROT and / or TrEMBL databases are a user entered protein sequence for protease sequence site. Single proteases and chemicals, a selection are the whole list of proteases and chemicals can be used. Different forms of the output of the result of available. The sequence map is displayed in portions of 10 to 60 aminoacids. when the results are displayed in form of map, the user as the possibility to select one enzymes is choice by mouse clicking. The sites that are potentially cleaved by this enzymes are then displayed in a separate window.

PeptideCutter

The sequence to investingate:

10 20 30 40 50 60 70 80 90 100 110 120

MKRQLKLFFI VLITAVVASA LTLFITGNSS ILGQKSASTG DSKFDKLNKA YEQIKSDYY KTDDDKLVDG AIKGMIQSLD DPYSTYMDQ EQAKSFDETIS ASFEGIGAQV EEKDGEILIV SPIKGSPAEK AGIKPRDQII KVNGKSVKGM NVNEAVALIR GKKGTKVKLE LNRAGVGNIDLSIKRDTIPV ETVYSEMKDN NIGEIQITSF SETTAKELTD AIDSLEKKGA KGYILDLRGN PGGLMEQAIT MSNLFIDKGK NIMQVEYKNG SKEVMKAEKE RKVTKPTVVL VNDGTASAAE IMAAALHESS NVPLIGETTF GKGTVQTAKEY DDGSTVKLT VAKWLTADGE WIHKKGIKPQVKAELPDYAK LPYLDADKTY KSGDTGTNVK VAQKMLKALG YKVKVNSMYD QDFVSVVKQF QKKEKLNETG ILTGDTTTKL MIELQKKLSD NDTQMEKAIE TLKKEM

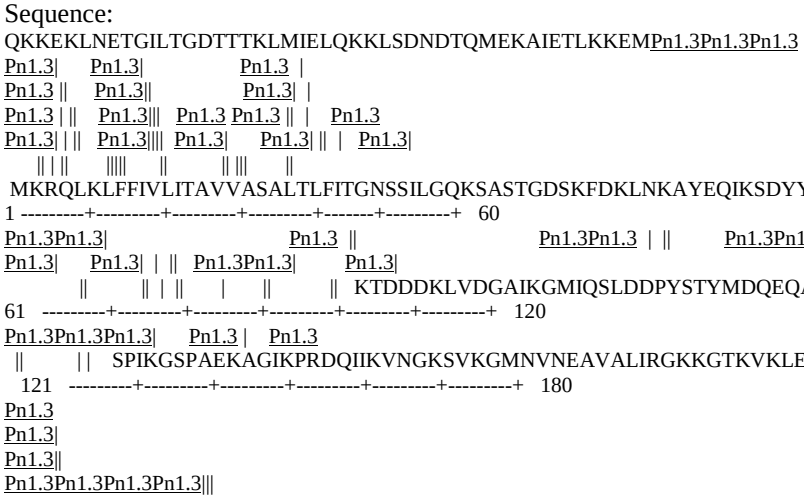
The sequence is 466 amino acids long.

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Pepsin (pH1.3)	99	6 7 7 9 11 12 20 21 22 23 23 24 31 32 43 44 46 47 50 58 58 59 66 67 78 79 82 85 86 94 102 103 117 118 157 158 169 171 180 181 193 194 209 210 217 224 225 232 234 235 236 237 243 244 253 254 254 255 266 267 289 290 305 306 313 319 320 330 338 339 343 344 344 350 351 365 368 371 374 379 395 398 400 401 408 409 412 413 419 426 431 432 439 440 443 444 447 461 462

Figure 3. shows the peptide cutter of proteinase using the pepsin as a cutter

These are the cleavage sites of the chosen enzymes and chemicals mapped onto the entered protein Sequence:



```
|       ||          ||    |    ||    | ||| LSIKRDTIPVETVYSEMKDNNIGEIQITSFSETTAKELTDAIDSLEKKGAKGYILDLRGN
  181  ---------+---------+---------+---------+---------+---------+  240

Pn1.3
Pn1.3Pn1.3|      Pn1.3Pn1.3
Pn1.3|   Pn1.3||     Pn1.3|             Pn1.3|         ||      |||       ||               ||
      PGGLMEQAITMSNLFIDKGKNIMQVEYKNGSKEVMKAEKERKVTKPTVVLVNDGTASAAE
  241  ---------+---------+---------+---------+---------+---------+  300
Pn1.3
Pn1.3|
Pn1.3Pn1.3Pn1.3   ||  Pn1.3
Pn1.3|  Pn1.3Pn1.3|     Pn1.3Pn1.3|   || Pn1.3|
 ||    |    ||       |     ||  ||    ||  IMAAALHESSNVPLIGETTFGKGTVQTAKEYDDGSTVKLTVAKWLTADGEWIHKKGIKPQ
  301  ---------+---------+---------+---------+---------+---------+  360
Pn1.3
Pn1.3    Pn1.3Pn1.3Pn1.3   | Pn1.3|     Pn1.3| vPn1.3   Pn1.3 ||   Pn1.3  ||
Pn1.3 | | |   |          Pn1.3  | ||  Pn1.3|   || Pn1 VKAELPDYAKLPYLDADKTYKSGDTGTNVKVAQKMLKALGYKVKVNSMYDQDFVSVVKQF
  361  ---------+---------+---------+---------+---------+---------+  420
Pn1.3   Pn1.3  |Pn1.3Pn1.3|  |Pn1.3|   Pn1.3  Pn1.3Pn1.3    ||  Pn1.3|  || |        Pn1.3|
     |   ||     || || |            ||

  421  ---------+---------+---------+---------+------  466
```

Figure 4 shows the Prot param it is the user provided sequence of proteinase. It shows a number of amnio acid in proteinase, molecular weight, theoretical pI, amino acid composition, total number of negatively charged residues (Asptculor)., total number of positively charge residues, atomic composition formula, total number atom present in the protenase.Extinct air co-efficient, estimated of life instability indere (II), and grand average hydropathicity (gravy) of proteinase. By this, the number amino acid in proteinase found to be 466. The molecular weight of proteinase is 51148.7. The theoretical isoelectric point of proteinase is 8.44. Total number of negatively charged residues (Asptalu) east 66. Total number positively charged residues (Arg + lys) is 69. The atomic composition of proteins are carbon C: 2260, hydrogen H : 3702, nitrogen N: 594, Oxygen O: 717, Sulphur S: 15. the molecular formula is C 2260 H 3702 N594 O 717 S 15. extinction co-efficient (EC is 31860. Total number atoms present in proteinase is 7288. The estimated of life of proteinase is 30 hours. Instability index of the proteinease is 20.43. And the grand average of hydropathcity is (gravy) is 0.436.

```
10          20          30          40          50          60          70          80          90          100  110       120
      AIKGMIQSLD DPYSTYMDQE QAKSFDETIS ASFEGIGAQV EEKDGEILIV SPIKGSPAEK AGIKPRDQII KVNGKSVKGM NVNEAVA
      LIR GKKGTKVKLE LNRAGVGNID LSIKRDTIPV ETVYSEMKDN NIGEIQITSF SETTAKELTD AIDSLEKKGA KGYILDLRGN PGGLME
      QAITMSNLFIDKGK NIMQVEYKNG SKEVMKAEKE RKVTKPTVVL VNDGTASAAE IMAAALHESSNVPLIGETTF GKGTVQTAKE YDDGSTV
     KLT VAKWLTADGEWIHKKGIKPQ VKAELPDYAKLPYLDADKTY KSGDTGTNVK VAQKMLKALG YKVKVNSMYD QDFVSVVKQF QKKEKLNET
      GILTGDTTTKL MIELQKKLSD NDTQMEKAIE TLKKEM
```

Theoretical pI/Mw: 7.82 / 43387.78

Figure 5 shows the Prot scale user provided sequence of proteinase. Prot scale allows to compute and represent the profile produced by any amino acid scale on a selected protein. An amino acid scale is defined by a numerical value assigned to each type of amino acid and most frequently used scales are hydropobicity scales, and secondary structure conformational parameter scales.

```
10        20        30        40        50  60    70  80        90        100       110       120
MKRQLKLFFI VLITAVVASA LTLFITGNSS ILGQKSASTG DSKFDKLNKA
YEQIKSDYYQ KTDDDKLVDG AIKGMIQSLD DPYSTYMDQE QAKSFDETIS
ASFEGIGAQV EEKDGEILIV SPIKGSPAEK AGIKPRDQII KVNGKSVKGM NVNEAVALIR
GKKGTKVKLE LNRAGVGNID LSIKRDTIPV ETVYSEMKDN NIGEIQITSF
SETTAKELTD AIDSLEKKGA KGYILDLRGN
PGGLMEQAIT MSNLFIDKGK NIMQVEYKNG SKEVMKAEKE RKVTKPTVVL VNDGTA
IMAAALHESS NVPLIGETTF GKGTVQTAKE YDDGSTVKLT VAKWLTADGE
WIHKKGIKPQ VKAELPDYAK LPYLDADKTY KSGDTGTNVK VAQKMLKALG
YKVKVNSMYD QDFVSVVKQF QKKEKLNETG ILTGDTTTKL MIELQKKLSD
NDTQMEKAIE TLKKEM
SEQUENCE LENGTH: 466
```

Using the scale **Hphob. / Kyte & Doolittle**, the individual values for the 20 amino acids are:
Ala: 1.800 Arg: -4.500 Asn: -3.500 Asp: -3.500 Cys: 2.500 Gln: -3.500
Glu: -3.500 Gly: -0.400 His: -3.200 Ile: 4.500 Leu: 3.800 Lys: -3.900
Met: 1.900 Phe: 2.800 Pro: -1.600 Ser: -0.800 Thr: -0.700 Trp: -0.900
Tyr: -1.300 Val: 4.200 Asx: -3.500 Glx: -3.500 Xaa: -0.490

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
```

MIN: -2.789
MAX: 3.256

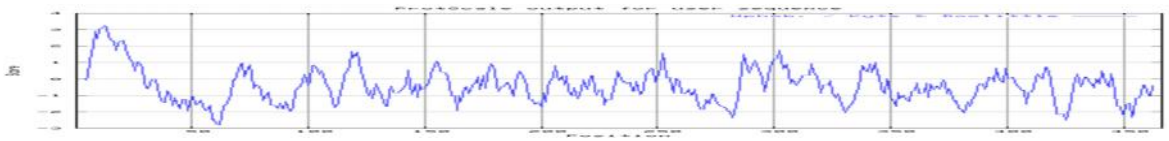
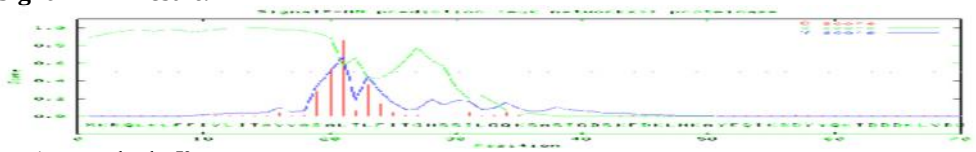


Figure 6 shows the prot scale output for user sequence by this it is known individual hydrophobic values for 20 amino acid is Ala: 1.800, Arg: 4.500, Asn:3.500, Asp:3.500, Cys:2.500, Gln:3.500, Glu:3.500, Gly: 0.400, His:3.200, Ele: 4.500, Leu:3,800, Lys:3.900, Met:1.900, Phe:2.800, Pro:1.600, Ses: 0.800, Thr: 0.700, Trp: 0.900, Tyr:1.300, Val: 4.200, Glx: 3.500, Xaa: 0.490 from the

above graph the maximum hydrophobic value is 3.256, the minimum hydrophobic value: 2.789.

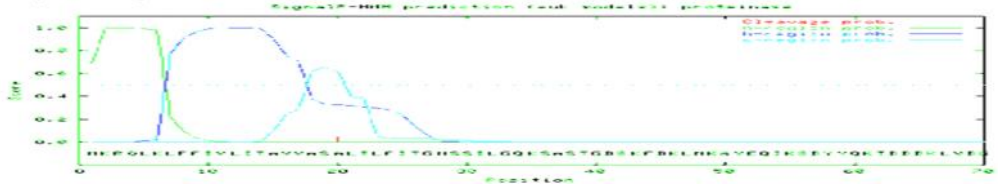
>proteinase

SignalP-NN result:



>proteinase length = 70
Measure Position Value Cutoff signal peptide?
max. C 21 0.860 0.32 YES max. Y 21 0.674 0.33 YES
max. S 14 0.997 0.87 YES mean S 1-20 0.963 0.48 YES
D 1-20 0.819 0.43 YES
Most likely cleavage site between pos. 20 and 21: ASA-LT

Figure 7 shows the signal P using the neural networks by this it is understood the proteinase length 70, measure position value cut of is a 0.860 to 0.819 and minimum C is 0.32 to 0.43, the most likely cleavage site between position 20 and 21: ASA - L& T



SignalP-HMM result

>proteinase

Prediction: Signal peptide Signal peptide probability: 0.685

Signal anchor probability: 0.315 Max cleavage site probabilities: 0.341 between pos. 22 and 23

Figure 8 shows the signal P HMM result by this it is understood the proteinase prediction is a signal peptide. Signal peptide probability is 0.685 and signal anchor probability is 0.315. The cleavage site probability is a 0.341 between position 22 and 23 by this it is understood the signal peptide present is only as secretary protein type..

SOSUI

Query title : proteinase Total length : 466 A. A. Average of hydrophobicity : -0.435622

This amino acid sequence is of a MEMBRANE PROTEIN which have 1 transmembrane helix.

No.	N terminal	transmembrane region	C terminal	Type	Length
1	7	LFFIVLITAVVASALTLFITGNS	29	PRIMARY	23

Figure 9 shows SOSUI result for proteinase this stable shows the transmembrane helix region of proteinase and the type of protein and length of the transmembrane region. By this it is known transmembrane region of the proteinase is LEFILITAVVASALTLITNGS. The N-terminal end of the transmembrane regions 7th position, C-terminal end of the transmembrane region 29th position, type of proteinase is primary type. The length of the transmembrane region in proteinase is 23. from the above results it is concluded that there are 23 amino acids are present in the transmembrane regions. And that too from the 7th position of N-terminal to 29th position of C-terminal of the proteins. Hence, it is concluded the protease is the membrane protein

max=33 | html | proteinase | plain_text |
MKRQLKLFFIVLITAVVASALTLFITGNSSILGQKSASTGDSKFDKLNKAYEQIKSDYYQKTDDDKLVDG%0D%0AAIKGMIQSLDDPYSTYMDQEQAKSFDETISASFEGIGAQVEEKDGEILIVSP
IKGSPAEKAGIKPRDQII%0D%0AKVNGKSVKGMNVNEAVALIRGKKGTKVKLELNRAGVGNIDLSIKRDTIPVETVYSEMKDNNIGEIQITSF%0D%0ASETTAKELTDAIDSLEKKGAKGYILDL
RGNPGGLMEQAITMSNLFIDKGKNIMQVEYKNGSKEVMKAEKE%0D%0ARKVTKPTVVLVNDGTASAAEIMAAALHESSNVPLIGETTFGKGTVQTAKEYDDGSTVKLTVAKWLTADGE%0D
%0AWIHKKGIKPQVKAELPDYAKLPYLDADKTYKSGDTGTNVKVAQKMLKALGYKVKVNSMYDQDFVSVVKQF%0D%0AQKKEKLNETGILTGDTTTKLMIELQKKLSDNDTQMEKAIETLK
KEM%0D%0A

TMpred output for proteinase

Sequence: MKR...KEM, length: 466

Prediction parameters: TM-helix length between 17 and 33

1.) Possible transmembrane helices

The sequence positions in brackets denominate the core region.Only scores above 500 are considered significant.

Inside to outside helices : 2 found from to score center

7 (9) 27 (25) 2524 17 286 (286) 310 (305) 67 297

Outside to inside helices : 1 found from to score center

7 (7) 25 (25) 2463 17

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.Helices shown in brackets are considered insignificant.A "+"-symbol indicates a preference of this orientation.A "++"-symbol indicates a strong preference of this orientation.

inside->outside | outside->inside

7- 27 (21) 2524 | 7- 25 (19) 2463 (286- 310 (25) 67 ++) |

2.) Table of correspondences

3.) Suggested models for transmembrane topology

N-terminus inside 1 strong transmembrane helices, total score : 2524 from to length score orientation 1 7 27 (21) 2524 i-o alternative model 1 strong transmembrane helices, total score : 246 from to length score orientation 1 7 25 (19) 2463 o-.

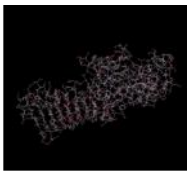


Fig. 11 shows the Wire frame model from Pseudomonas aeruginosa

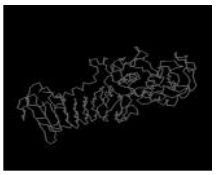


Fig. 12 shows backbone of proteinase from Psuedomonas

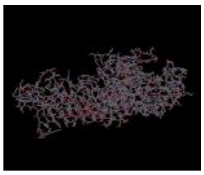


Fig. 13 shows the Stick model of Proteinase of Pseudomonas

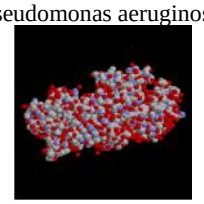


Fig 14. shows the Spacefill model of Pseudomonas aeruginosa

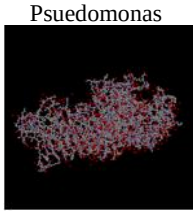


Fig 15. shows the .Ball ad Stickof protienase

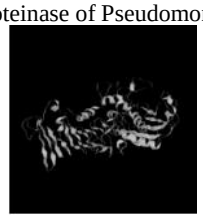


Fig 16. shows the ribbon model proteinase

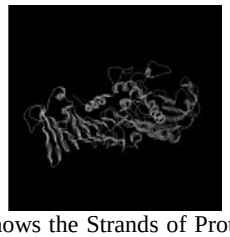


Fig 17. shows the Strands of Proteinase of Pseudomonas aeruginosa Proteinase

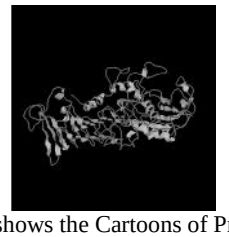


Fig 18. shows the Cartoons of Proteinas from Pseudomonas aeruginosa

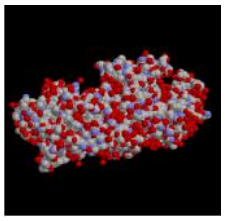


Fig 19. shows Molecular surface of from Pseudomonas aeruginosa

DISCUSSION

From the above picture the Proteinase produced from the organism have 2 number of chains,470(340) number of groups ,3505 number of atoms,56 bonds,12 helices,21 strands,zero number of turns and 3581 number of bonds.

II Cellulase.

ProtParam Result:

User-provided sequence:

10 20 30 40 50 60

MNCRKYLLSG LAVFGLAATS AVAALSTDDY VEAAWMTTRF FGAQRSGQGP NWILDGTSNPTSFTKDSYNG KDVSGGWFDC GDHVMYGQSQ GYASYVLALA YAEFTEVSTT FILVTTPTTR KPTTTPMKSG KPNKVRDLLE ELRYEADFWV KAAIDGNNFV TVKGDGNADH QKWVTAGKLGSGEGGEP RCITGNANDG FTSGLAA AML AVMARVDPDT ANQAKYLKAA KTAYSYAKSH KGVTNSQGFY ESSWWDGRWE DGPFLAELEL YRTTGENSYK TAAIDRYDNL KFSLGEMYSNVVPLSA VMAEAVFEET PHGMRKEAIG VLDLIYEEKA KDKIFQNPNG MGSGKFPVRV PSGGAFLYAL SDKFNNTNEH MEMIEKNVSY LLGDNGSKKS YVVGFSKNGA NAPSRPYYANEKRWRR SRRCSESSRK EQALGRYDCW RLY .

Number of amino acids: 453 **Molecular weight:** 50042.0 **Theoretical pI:** 8.35

Amino acid composition:

Ala (A) 45	9.9%	Arg (R) 23	5.1%	Asn (N) 25	5.5%	Asp (D) 24	5.3%	Cys (C) 5	1.1%
Gln (Q) 9	2.0%	Glu (E) 26	5.7%	Gly (G) 44	9.7%	His (H) 8	1.8%	Ile (I) 9	2.0%
Leu (L) 30	6.6%	Lys (K) 30	6.6%	Met (M) 13	2.9%	Phe (F) 20	4.4%	Pro (P) 16	3.5%
Ser (S) 36	7.9%	Thr (T) 30	6.6%	Trp (W) 10	2.2%	Tyr (Y) 24	5.3%	Val (V) 26	5.7%

Total number of negatively charged residues (Asp + Glu): 50 Total number of positively charged residues (Arg + Lys): 53

Atomic composition: Carbon 2218 Hydrogen 3378 Nitrogen N 612 Oxygen O 678 Sulfur S 18

Formula: $C_{2218}H_{3378}N_{612}O_{678}S_{18}$ **Total number of atoms:** 6904

Extinction coefficients: Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water. Ext. coefficient 91010 Abs 0.1% (=1 g/l) 1.819, assuming ALL Cys residues appear as half cystines Ext. coefficient 90760 Abs 0.1% (=1 g/l) 1.814, assuming NO Cys residues appear as half cystines

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 29.52

This classifies the protein as stable.

RESULT

The given protein is calculated with various parameters to give good results.There are number of parameters are used to analyse the given protein sequence.

1. Calculated number of amino acid present in the given protein sequence is 453.
2. Calculated the molecular weight of protein sequence is 50042.0
3. Calculated the theoretical isoelectric point of protein sequence is 8.35

4. Calculated each amino acid sequence molecular weight represents in a molecular percent.
5. Calculated the number of positive charges amino acids is 53
6. Calculated the number of negative charge amino acids is 50
7. Calculated the number of atoms of amino acid of protein is 6904
8. Give the molecular formula of given protein sequence is C_{2218} H_{3378} N_{612} O_{678} S_{18}
9. The extinction co-efficient of protein sequence is calculated.
10. Computed the half-life of given protein is 30 (hours)
11. Computed the instability of given protein sequence is 29.52

Inference

From this result I have got protein molecular weight, isoelectric point, each aminoacids composition with molar percent values, positive charge residues and negative charge residues, number of atoms of amino acids , formula of proteins, total number of atoms, stability of proteins, half of proteins and extinction co-efficient in the protein sequence.

Result

The given protein is calculated with various parameters to give good results. There are number of parameters are used to analyse the given protein sequence.

Inference

In this site calculate the molecular weight of the 1ut9A sequence is 50042.0 and isoelectric point is 8.35.

3. Reverse Translate Results:

Results for 394 residues sequence "P23665|GUNA_FIBSU Endoglucanase A - Fibrobacter MNCRKYLLSGLAVFGLAATSAVAALSTDDYVEAAWMTTRFFGAQRSGQGPNWILDGTSNP" starting "TSFTKDSYNG"

>reverse translation of P23665|GUNA_FIBSU Endoglucanase A - Fibrobacter MNCRKYLLSGLAV FGLAATSA VAALST DDYV EAAWMTTRFFGAQRSGQGPNWILDGTSNP to a 1182 base sequence of most likely codons.

accagctttaccaaagatagctataacggcaaagatgtgagcggcggctggtttgattgcggcgatcatgtgatgtatggccagagccagggctatgcgagctatgtgctggcgctggcgtatgcggaa
tttaccgaagtgagcaccacctttattctggtgaccaccccgaccacccgcaaaccgaccaccaccccgatgaaaagcggcaaaccgaacaaagtgcgcgatctgctggaagaactgcgctatgaag
cggatttttgggtgaaagcggcgattgatggcaacaactttgtgaccgtgaaaggcgatggcaacgcggatcatcagaaatgggtgaccgcgggcgcgatgagcaaactgggcagcggcgaaggc
ggcgaaccgcgctgcattaccggcaacgcgaacgatggctttaccagcggcctggcggcggcgatgctggcggtgatggcgcgcgtggatccggataccgcgaaccaggcgaaatatctgaaag
cggcgaaaaccgcgtatagctatgcgaaaagccataaaggcgtgaccaacagccagggcttttatgaaagcagctggtgggatggccgctgggaagatggcccgtttctggcggaactggaactgta
tcgcaccaccggcgaaaacagctataaaaccgcggcgattgatcgctatgataacctgaaatttagcctgggcgaaggcacccattttatgtatagcaacgtggtgccgctgagcgcggtgatggcgg
aagcggtgtttgaagaaaccccgcatggcatgcgcaaagaagcgattggcgtgctggatctgatttatgaagaaaaagcgaaagataaaatttttcagaacccgaacggcatgggcagcggcaaattt
ccggtgcgcgtg

ccccgagcggcggcgcgtttctgtatgcgctgagcgataaatttaacaacaccaacgaacatatggaaatgattgaaaaaaacgtgagctatctgctgggcgataacggcagcaaaaaaagctatgtg
gtgggctttagcaaaaacggcgcgaacgcgccgagccgcccgcatcatcgcggctattatgcgaacgaaaaacgctggcgccgcagccgccgctgcagcgaaagcagccgcaaagaacaggcg
ctgggccgctatgattgctggcgcctgtattaa

Result: The given protein sequence is converted to DNA by using Reverse Translate.

4. ScanProsite Results Viewer:

This view shows ScanProsite results together with ProRule-based predicted intra-domain features

MNCRKYLLSGLAVFGLAATSAVAALSTDDYVEAAWMTTRFFGAQRSGQGPNWILDGTSNPTSFTKDSYNGKDVSGGWFDCGDHVMYGQSQ
GYASYVLALAYAEFTEVSTTFILVTTPTTRKPTTTPMKSGKP
NKVRDLLEELRYEADFWVKAAIDGNNFVTVKGDGNADHQKWVTAGAMSKLGSGEGGEPRCITGNANDGFTSGLAAAMLAVMARVDPDTAN
QAKYLKAAKTAYSYAKSHKGVTNSQGFYESSWWDGRWEDGPFLAELELYRTTGENSYKTAAIDRYDNLKFSLGEGTHFMYSNVVPLSAVMA
EAVFEETPHGMRKEAIGVLDLIYEEKAKDKIFQNPNGMGSGKFPVRVPSGGAFLYALSDKFNNTNEHMEMIEKNVSYLLGDNG
SKKSYVVGFSKNGANAPSRPHHRGYYANEKRWRRSRRCSESSRKEQALGRYDCW

Inference

Scanprosite search a given protein against prosite database to occurrence of pattern and profile.GLYCOSYL_HYDROL_F9_1 active site is found between 403 -419 in the sequence.

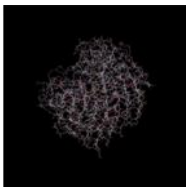


Fig 20.Wire frame model of Cellulase Ceratocystis paradoxa paradoxa

Fig 21 Backbone of Cellulase

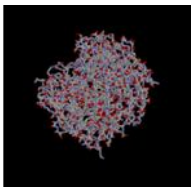


Fig 22.Sticks of Cellulase from from Ceratocystis

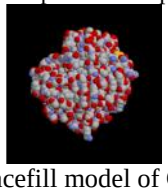


Fig23.spacefill model of Cellulase Ceratocystis paradoxa

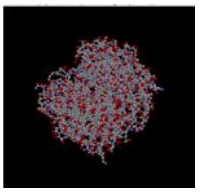


Fig 24 Ball and Stick model of Cellulase

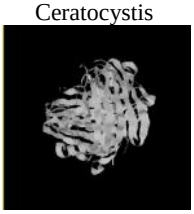


Fig 25 shows Ribbon model of Cellulase

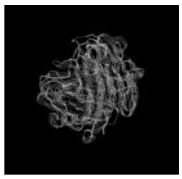


Fig 24 Strand model of Cellulase from Ceratocystis paradoxa Cellulase

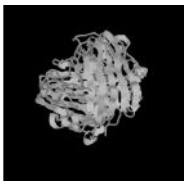


Fig 25 .Cartoon model of Cellulase Cellulase

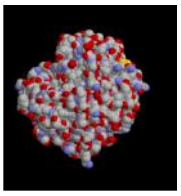


Fig 26 shows Molecular surface of Ceratocystis paradoxa

Discussion

From the above picture the Cellulase produced from the organism have 6 number of chains,440 number of groups 3331 number of atoms,10 bonds,6 helices,32 strands,zero number of turns and 3429 number of bonds

III Pectinase

Primary sequence analysis

Compute pI/MwTheoretical pI/Mw (average) for the use r-entered sequence:

10 20 30 40 50 60 70 80 90 100 110 120
MVALTLGIFF TSLAASAVAA PAPAITPAPK PEVVKRASSC TFSGSNGAAE ASKSQSSCAT MVLSDVAVPS GTTLDLSSLA DGTTVIFEGT TTWGYSEWKG PLLDIQGKKI TVKGAEGSVL
NGDGARWWDG KGGNGGKTKP KFFSAHKLTD STITGITIKN PPVQVVSING CDGLIDASDGDKDE QGHNTDGFDI GSSNNVTIDG AKVYNQDDCV AVNSGTEITF
LSIGSVGGRD DNTVDTVTFS NSEVTKSVNG VRVKAKVGTT GKINKVTYED ITLSEISKYG VLIEQNYDGG DLHGDADTGV PITALTLDNV TGGVSSSGYD VVVTCGKGSC TGWTWTGV
TGGKTYDKCS NVPSVTKCS

Theoretical pI/Mw: 4.85 / 38816.1 **protparam:**

Number of amino acids: 379

Molecular weight: 38816.1

Theoretical pI: 4.85

Amino acid sequence:

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 2671 Nitrogen 455 Oxygen 575Sulfur S12

Formula: $C_{1679}H_{2671}N_{455}O_{575}S_{12}$

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

There are number of parameters are used to analyze the given protein seuquence.

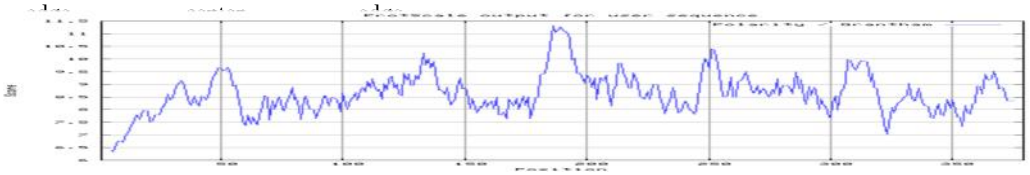
Protscale

Using the scale Polarity/Grantham,the individual values for the 20 amino acids are: Ala: 8.100 Arg: 10.500 Asn: 11.600 Asp: 13.000 Cys: 5.500 Gln: 10.500 Glu: 12.300 Gly: 9.000 His: 10.400 Ile: 5.200 Leu: 4.900 Lys: 11.300 Met: 5.700 Phe: 5.200 Pro: 8.000 Ser: 9.200 Thr: 8.600 Trp: 5.400
Tyr: 6.200 Val: 5.900 : 12.300 : 11.400 : 8.325

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

1 Ala: 89.000 Arg: 174.000 Asn: 132.000 Asp: 133.000 Cys: 121.000 Gln: 146.000 Glu: 147.000 Gly: 75.000 His: 155.000 Ile: 131.000 Leu: 131.000 Lys: 146.000 Met: 149.000 Phe: 165.000 Pro: 115.000 Ser: 105.000 Thr: 119.000 Trp: 204.000 Tyr: 181.000 Val: 117.000 : 132.500 : 146.500 : 136.750

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



SEQUENCE LENGTH: 379

Using the scale Molecular weight, the individual values for the 20 amino acids are:
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

Peptide cutter

The sequence is 379 amino acids long. Trypsin enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	27	35 36 53 99 108 109 113 126 131 137 141 147 159 212 231 249 266 272 274 276 282 285 298 347 364 368 377

Tryps

```
MVALTLGIFFTSLAASAVAAPAPAITPAPKPEVVKRASSCTFSGSNGAAEASKSQSSCAT
  1  ---------+---------+---------+---------+---------+---------+  60
   MVLSDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVL
 61  ---------+---------+---------+---------+---------+---------+  120
NGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMT
 121  ---------+---------+---------+---------+---------+---------+  180
IDASDGDKDEQGHNTDGFDIGSSNNVTIDGAKVYNQDDCVAVNSGTEITFKNGLCSGGHG
 181  ---------+---------+---------+---------+---------+---------+  240
LSIGSVGGRDDNTVDTVTFSNSEVTKSVNGVRVKAKVGTTGKINKVTYEDITLSEISKYG
 241  ---------+---------+---------+---------+---------+---------+  300
VLIEQNYDGGDLHGDADTGVPITALTLDNVTGGVSSSGYDVVVTCGKGSCTGWTWTGVDV
 301  ---------+---------+---------+---------+---------+---------+  360
TGGKTYDKCSNVPSVTKCS
 361  ---------+---------  379
```

Peptide mass

The selected enzyme is: Trypsin
Maximum number of missed cleavages (MC): 0
All cysteines in reduced form.
Methionines have not been oxidized.
Displaying peptides with a mass bigger than 500 Dalton.
Using monoisotopic masses of the occurring amino acid residues and giving peptide masses as $[M+H]^+$.

The peptide masses from your sequence are:
[Theoretical pI: 4.85 / Mw (average mass): 38816.10 / Mw (monoisotopic mass): 38792.04]
93.4% of sequence covered (you may modify the input parameters to display also peptides < 500 Da or > 100000000000 Da):

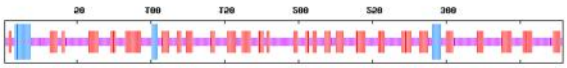
Secondary structure prediction

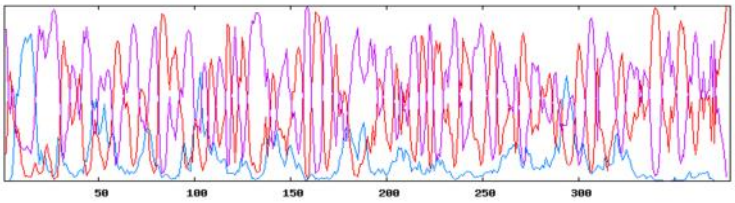
GOR-IV results for endopolygalacturonase of alternaria cepula

```
 10        20        30        40        50        60        70
 MVALTLGIFFTSLAASAVAAPAPAITPAPKPEVVKRASSCTFSGSNGAAEASKSQSSCATMVLSDVAVPS
 ccccceecchhhhhhhhhhhccccccccccccceeeeeccceeccccccccccccccccceeeeeeeccccc
 GTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLNGDGARWWDGKGGNGGKTKP
 ccceeeeccccccceeeeeeeeeeccccccchhhhccceeeeeccccceeeccceeeeeccccccccccc
 KFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMTIDASDGDKDEQGHNTDGFDIGSSNNVTIDG
 ceeeccccccccceeeeeccccceeeeeeccccceeeeeccceeccccccccccccccccceeeccccceeecc
 AKVYNQDDCVAVNSGTEITFKNGLCSGGHGLSIGSVGGRDDNTVDTVTFSNSEVTKSVNGVRVKAKVGTT
 eeecccccceeeeccccceeeeeccccccccceeeeeeeccccccccceeeeeccccccccccceeeeeeeccc
 GKINKVTYEDITLSEISKYGVLIEQNYDGGDLHGDADTGVPITALTLDNVTGGVSSSGYDVVVTCGKGSC
 cceeeeeeccchhhhhhccceeeeeccccccccccccccceeeeeccccccccccccceeeeeeeccccc
 TGWTWTGVDVTGGKTYDKCSNVPSVTKCS
 cceeeeeeeccccccccccccceeeeec
```

Sequence length : 379
GOR4 :
Alpha helix (Hh) : 21 is 5.54% 3_{10} helix (Gg) : 0 is 0.00% Pi helix (Ii) : 0 is 0.00%
Beta bridge (Bb) : 0 is 0.00% Extended strand (Ee) : 137 is 36.15%

Beta turn (Tt) : 0 is 0.00% Bend region (Ss) : 0 is 0.00% Random coil (Cc) : 221 is 58.31%
Ambigous states (?) : 0 is 0.00% Other states : 0 is 0.00%





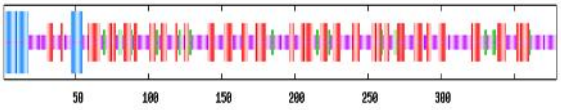
SOPMA results for endopolygalacturonase of Alternaria cepulae

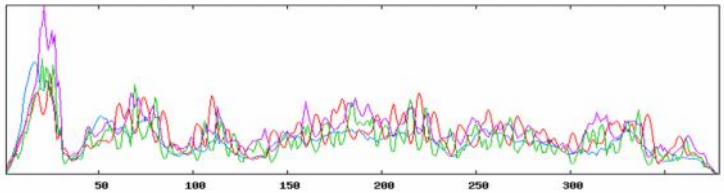
```
        10        20        30        40        50        60        70
         |        |         |         |         |         |         |
MVALTLGIFFTSLAASAVAAPAPAITPAPKPEVVKRASSCTFSGSNGAAEASKSQSSCATMVLSDVAVPS
eehhhhhhhhhhhhhhhhcccccccccccceeeeccccceeccccchhhhhhhhcccceeeeeeeecct
GTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLNGDGARWWDGKGGNGGKTKP
tceeeeeeccttceeeeetceeeccccccccceeeettcceeeeccccceecttcceeettccccccccc
KFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMTIDASDGDKDEQGHNTDGFDIGSSNNVTIDG
ceeeeeccccceeeeeeeccccceeeeeccccceeeeeeeeeccttccccccccccceeeeccccceeeee
AKVYNQDDCVAVNSGTEITFKNGLCSGGHGLSIGSVGGRDDNTVDTVTFSNSEVTKSVNGVRVKAKVGTT
eeeccttceeeeettcceeeeeccccccccceeeeeeccccccccceeeeeeetceeeccctteeeeeeeccc
GKINKVTYEDITLSEISKYGVLIEQNYDGGDLHGDADTGVPITALTLDNVTGGVSSSGYDVVVTCGKGSC
ccceeeeeccceeeccccceeeeeccccccccccccccccceeeeeeeeccttcccttteeeeeeeccccc
TGWTWTGVDVTGGKTYDKCSNVPSVTKCS
cceeeeeeeetccccccccccccccc
```

Sequence length : 379

SOPMA :

Alpha helix (Hh) : 24 is 6.33% 3_{10} helix (Gg) : 0 is 0.00% Pi helix (Ii) : 0 is 0.00%
Beta bridge (Bb) : 0 is 0.00% Extended strand (Ee) : 145 is 38.26 % Beta turn (Tt) : 26 is 6.86%
Bend region (Ss) : 0 is 0.00%Random coil (Cc) : 184 is 48.55% Ambigous states (?) : 0 is 0.00%
Other states : 0 is 0.00%





SUMO plot result for EPG

Protein :	ENDOPOLYGLACTURONASE
Length:	379 aa

```
  1 MVALTLGIFF TSLAASAVAA PAPAITPAPK PEVVKRASSC TFSGSNGAAE
 51 ASKSQSSCAT MVLSDVAVPS GTTLDLSSLA DGTTVIFEGT TTWGYSEWKG
101 PLLDIQGKKI TVKGAEGSVL NGDGARWWDG KGGNGGKTKP KFFSAHKLTD
151 STITGITIKNPPVQVVSING CDGLTITDMT IDASDGDKDE QGHNTDGFDI
201 GSSNNVTIDG AKVYNQDDCV AVNSGTEITFKNGLCSGGHG LSIGSVGGRD
251 DNTVDTVTFS NSEVTKSVNG VRVKAKVGTT GKINKVTYED ITLSEISKYG
301 VLIEQNYDGG DLHGDADTGV PITALTLDNV TGGVSSSGYD VVVTCGKGSC
351 TGWTWTGVDV TGGKTYDKCS NVPSVTKCS
```

Motifs with high probability
Motifs with low probability
Overlapping Motifs

No.	Pos.	Group	Score	No.	Pos.	Group	Score
1	K159	ITGIT **IKNP** PVQVV	0.84	5	K99	WGYSE **WKGP** LLDIQ	0.54
2	K231	GTEIT **FKNG** LCSGG	0.68	6	K131	ARWWD **GKGG** NGGKT	0.50
3	K276	GVRVK **AKVG** TTGKI	0.62	7	K188	DASDG **DKDE** QGHNT	0.50
4	K30	AITPA **PKPE** VVKRA	0.61	8	K141	GGKTK **PKFF** SAHKL	0.26

Tmpred results

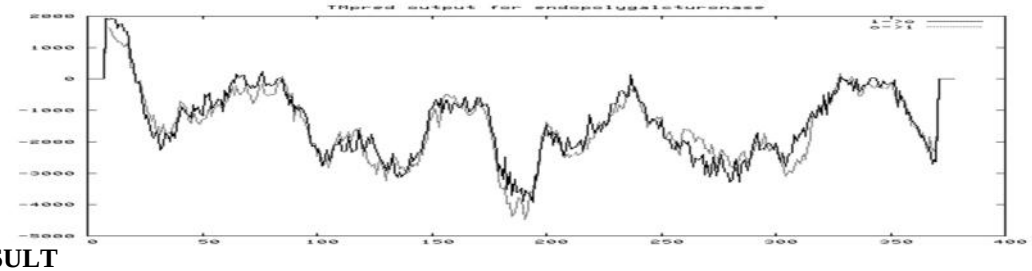
Possible transmembrane helices

The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.
Inside to outside helices: 5 found from to score center
1 (1) 21 (18) 1934 9 61 (67) 93 (87) 234 77 230 (230) 248 (248) 139 238 320 (320) 339 (337) 65 329 330 (335) 353 (353) 42 343
Outside to inside helices : 3 found from to score center
1 (1) 17 (17) 1624 9 76 (76) 98 (95) 74 86 319 (319) 345 (337) 167 329

2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.
Helices shown in brackets are considered insignificant.
A "+"-symbol indicates a preference of this orientation.
A "++"-symbol indicates a strong preference of this orientation.
inside->outside | outside->inside
1- 21 (21) 1934 ++ | 1- 17 (17) 1624 (61- 93 (33) 234 +) |(76- 98 (23) 74)
(230- 248 (19) 139 ++) | (320- 339 (20) 65) |(319- 345 (27) 167 +) (330- 353 (24) 42 ++) |

2 possible models considered, only significant TM-segments used
-----> STRONGLY prefered model: N-terminus inside
1 strong transmembrane helices, total score : 1934
from to length score orientation
1 1 21 (21) 1934 i-o
------>alternative model
1 strong transmembrane helices, total score : 1624
from to length score orientation
1 1 17 (17) 1624 o-i

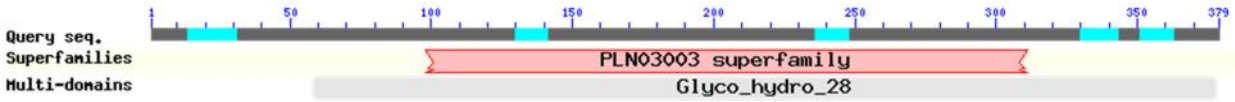


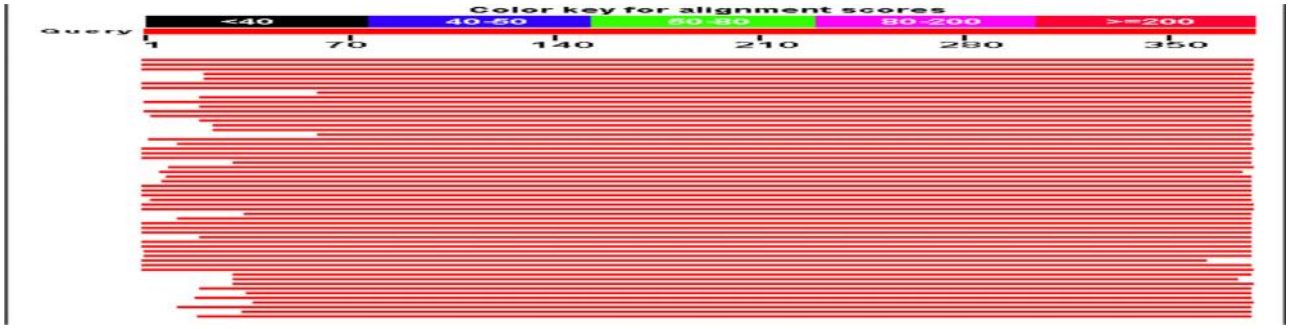
TMHMM RESULT
gi|13160919|dbj|BAB32924.1| Length: 379
gi|13160919|dbj|BAB32924.1| Number of predicted TMHs: 0
gi|13160919|dbj|BAB32924.1| Exp number of AAs in TMHs: 3.20296
gi|13160919|dbj|BAB32924.1| Exp number, first 60 AAs: 3.20219
gi|13160919|dbj|BAB32924.1| Total prob of N-in: 0.14545
gi|13160919|dbj|BAB32924.1|TMHMM2.0 outside 1 379
ppmtogif: computing colormap... ppmtogif: 5 colors found

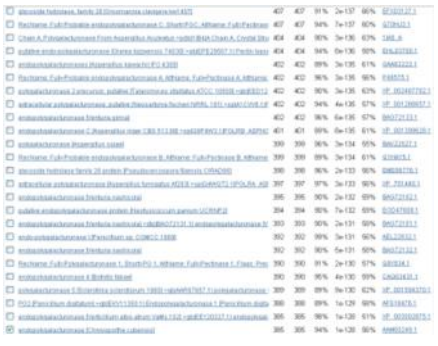
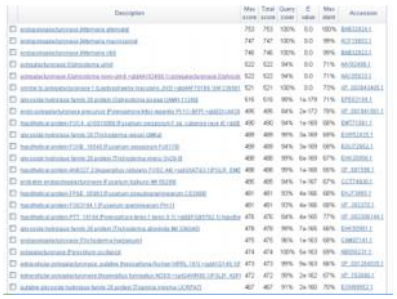
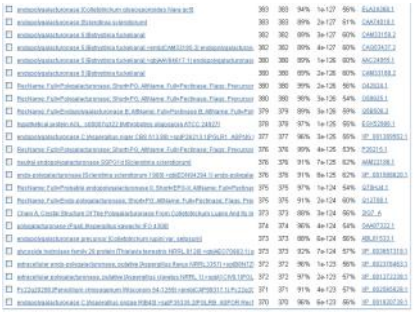
SOUI
Total length : 379 A. A. Average of hydrophobicity : -0.155673
This amino acid sequence is of a MEMBRANE PROTEIN which have 1 transmembrane helix.

No.	N terminal	transmembrane region	C terminal	type	length
1	3	ALTLGIFFTSLAASAVAAPAPAI	25	PRIMARY	23

PAIRWISE SQUENCE ALIGNMENT
BLAST results
Putative conserved domains have been detected, click on the image below for detailed results.





Multiple sequence alignment and phylogenetic analysis

Scores Table

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	cepulae	145	2	alternata	140	100
1	cepulae	145	3	citriarbusti	154	100
1	cepulae	145	4	perangusta	159	99
2	alternata	140	3	citriarbusti	154	100
2	alternata	140	4	perangusta	159	99
3	citriarbusti	154	4	perangusta	159	99

Alignment

CLUSTAL W (1.82) multiple sequence alignment

```
cepulae         VLSDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLN 60
alternata       VLSDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLN 60
citriarbusti    VLSDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLN 60
perangusta      VLSDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVKGAEGSVLN 60
                ************************************************************

cepulae         GDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMTI 120
alternata       GDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMTI 120
citriarbusti    GDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPVQVVSINGCDGLTITDMTI 120
perangusta      GDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGIAIKNPPVQVVSINGCDGLTITDMTI 120
                ************************************:***********************

cepulae         DASDGDKDEQGHNTDGFDIGSSNNV---------- 145
alternata       DASDGDKDEQGHNTDGFDIG---------- 140
citriarbusti    DASDGDKDEQGHNTDGFDIGSSNNVIIDGAKVYN----- 154
perangusta      DASDGDKDEQGHNTDGFDIGSSNNVIIDGAKVYNSSNNV 159
                **********
```

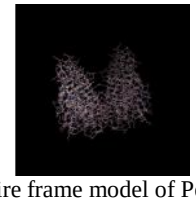


Fig 27.Wire frame model of Pectinase in Alternaria cepulae

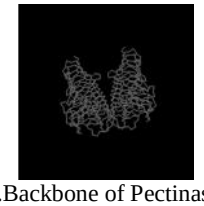


Fig 28.Backbone of Pectinase from Alternaria cepulae

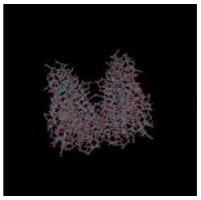


Fig 28. Backbone of Pectinase Alternaria cepulae

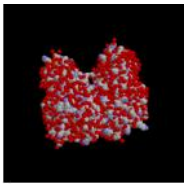


Fig 30.Spacefill model of Pctinase fromAlternaria cepulae

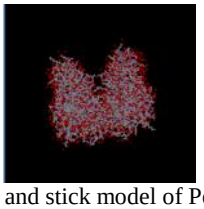


Fig 31.Ball and stick model of Pecinase in Alternaria cepulae

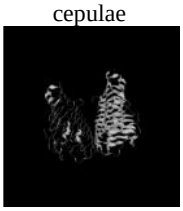


Fig 32.Ribbon model of Pectinase inAlternaria cepulae

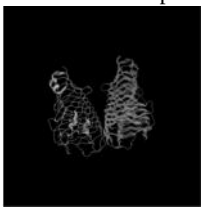


Fig 33.Strands of Pectinase from Alternaria cepulae

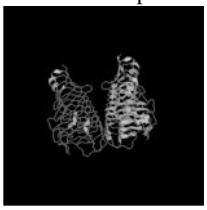


Fig 34.Cartoons of Pectinase in Alternaria cepulae

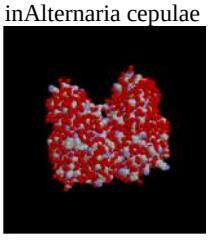


Fig 35 Molecular surface of Alternaria cepulae

Discussion

From the above picture the Pectinase produced from the Alternaria cepulae have 6 number of chains,670 number of groups 4918 number of atoms,64 bonds,8 helices,43 strands,zero number of turns and 5036 number of bonds

The Local Pairwise Alignment of Two Sequences

Here below, the classical text representation of a pairwise alignment of two sequences (THIO_ECOLI and PDI_ASPNG). This alignment was obtained with the Smith-Waterman algorithm, a BLOSUM62 similarity matrix and (-11/-1) for gap opening and

37.5% identity in 80 aa overlap; score: 122

```
              20        30        40        50            60
THIO_ECOLI SFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQ------GKLTVAKLNIDQNP
           :.  :. :  .:..:.: ::: :: .::  ::.: :     .:.:.::..  :
PDI_ASPNG  SYKDLVIDNDKDVLLEFYAPWCGHCKALAPKYDELAALYADHPDLAAKVTIAKIDATAND
          370       380       390       400       410       420
```

The amino acids of the query sequence (THIO_ECOLI) are represented using the grayed residues at the top of the grayed background histogram. Hence the full length of the query sequence is shown.

The local alignment of PDI_ASPNG on the query is represented by the sequence in black. The "+" signs at both ends of the aligned sub-sequence indicate that the alignment is local on PDI_ASPNG (the symbols "<" and ">" can be used to tag sequence extremities).

The Smith-Waterman score (122) is proportional to the sum of the areas of the red, blue and orange rectangles. The areas of the rectangles located below the aligned sequence are negative.

The area of every red rectangle corresponds to the score attributed by the similarity matrix to an observed pair of amino acids. The underlying gray rectangles represent the maximal score possible at every position of the query, which correspond to the diagonal elements of the similarity matrix in this example.

Two gaps are present in this example. The first one is an insertion (relative to the query) and is represented with lowercase letters. The second one is a deletion (relative to the query) and is represented with "-".

The cost of a gap is proportional to the sum of the areas of the adjacent blue and yellow rectangles. The area of the two blue rectangles represent the "gap existence" cost which is equally divided into an opening and a closing penalty. The orange rectangles represent the costs for extending the gap.

Alignment of a Sequence on a Profile

The pairwise alignment below corresponds to the one obtained when the PDI_ASPNG sequence is searched with the THIOREDOXIN_2 profile. For the sake of the textual representation, the profile positions were symbolized by the residues of the "consensus" sequence of the multiple sequence alignment from which the profile was derived. This alignment is not fundamentally different from the one considered before

```
consensus   1 XVXVLSDENFDEXVXDSDKPVLVDFYAPWCGHCRALAPVFEELAEEYK----DBVKFVKV  -48
                                  : :       : : :  ::  : : ::::: : :    : :  :       :
PDI_ASPNG 360 PVTVVVAHSYKDLVIDNDKDVLLEFYAPWCGHCKALAPKYDELAALYAdhpdLAAKVTIA  -97

consensus  57 DVDENXELAEEYGVRGFPTIMFF--KBGEXVERYSGARBKEDLXEFIEK           -1
                                    :         : ::           : : :  : :  :
PDI_ASPNG 420 KID-ATANDVPDPITGFPTLRLYpaGAKDSPIEYSGSRTVEDLANFVKE           -49
```

but the textual representation does not reveal the additional information carried on by the profile scoring system, that eventually makes the identification by the profile so "informative". The alternative graphical representation of this alignment reveals much of this extra information.

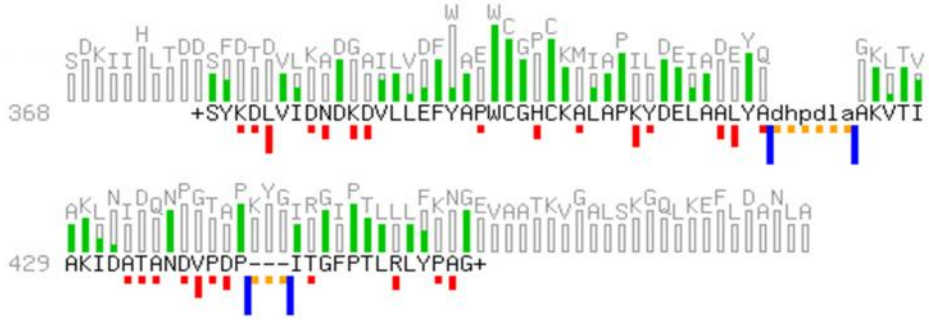
In strong contrast to the previous example, the scoring system is heavily position-dependent: The area of every red rectangle corresponds to the score attributed by the profile for the presence of a particular residue at a particular position. The underlying gray rectangle represents the maximal score possible at that position. The amino acids of the profile consensus that might contribute the most to the profile score are represented in gray at the top of the background histogram.Three gaps are presented in this example. They score differently as the system of gap penalties is also position dependent in a profile.Two cysteines are found among the highest scoring residues of the above example. Actually they form the active site of thioredoxins. A proline residue, which is quite distant on the sequence, also rewards a particularly high score. Actually, this proline is spatially located close to the active site as shown on the figure below. Obviously, this is a case where the alignment of a sequence on a profile can provides indication for the possible function of selected residues.

Description	N-glycosylation site.
myhits	query by motif
MyHits synonyms	ASN_GLYCOSYLATION , PS00001

```
ID   ASN_GLYCOSYLATION; PATTERN.
AC   PS00001;
DT   APR-1990 (CREATED); APR-1990 (DATA
UPDATE); APR-1990 (INFO UPDATE).
DE   N-glycosylation site.
PA   N-{P}-[ST]-{P}.
CC   /SITE=1,carbohydrate;
CC   /SKIP-FLAG=TRUE;
```

CC /VERSION=1; PR PRU00498; DO PDOC00001; //	
Description	cAMP- and cGMP-dependent protein kinase phosphorylation site.
myhits	query by motif
MyHits synonyms	CAMP_PHOSPHO_SITE , PS00004
ID CAMP_PHOSPHO_SITE; PATTERN. AC PS00004; DT APR-1990 (CREATED); APR-1990 (DATA UPDATE); APR-1990 (INFO UPDATE). DE cAMP- and cGMP-dependent protein kinase phosphorylation site. PA [RK](2)-x-[ST]. CC /SITE=3,phosphorylation; CC /SKIP-FLAG=TRUE; CC /VERSION=1; DO PDOC00004; //	

```
                70        80
THIO_ECOLI GTAPKYGIRGIPTLLLFKNG
           :  : :.::: :. :
PDI_ASPNG  VPDP---ITGFPTLRLYPAG
```

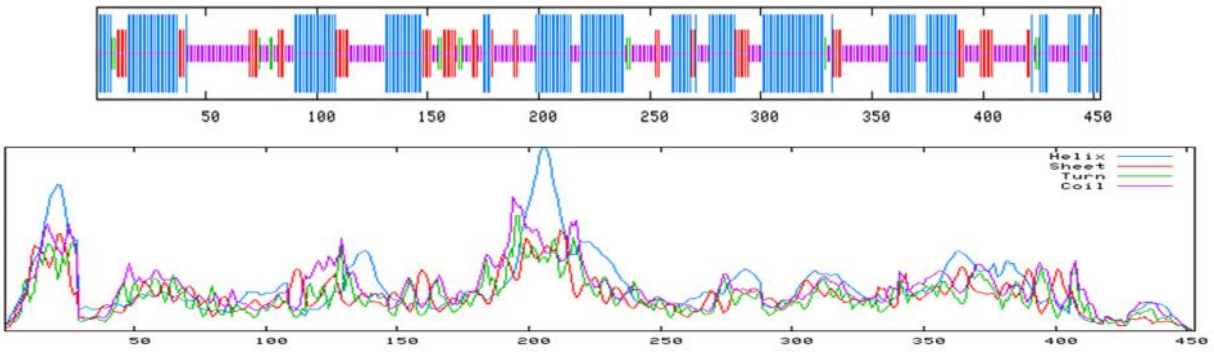


Inference

Motifscan is a program for finding motifs in the given sequence. The above results significantly shows some the important motif, its functions and the family where the motifs belongs to which implies the protein sequence. The result also gave the post translational modification.

6. SOPMA result for: UNK_219630
Sequence length : 453
SOPMA :

Alpha helix	(Hh) : 201 is 44.37%	3_{10} helix	(Gg) : 0 is 0.00%	Pi helix	(Ii) : 0 is 0.00%
Beta bridge	(Bb) : 0 is 0.00%	Extended strand	(Ee) : 62 is 13.69%		
Beta turn	(Tt) : 14 is 3.09%	Bend region	(Ss) : 0 is 0.00%	Random coi	(Cc) : 176 is 38.85%
Ambigous states (?)	: 0 is 0.00%	Other states	: 0 is 0.00%		



Parameters :
Window width : 17 Similarity threshold : 8 Number of states : 4
Top of Form
Result:
The length of the given sequence is 453
The alpha helix of the given sequence in 44.30%
The beta strand of the given sequence in 13.69%
The beta turns of the given sequence in 3.09%

The coils of the given sequence in 38.85%
The output of above parameters values shows in a graphics display.

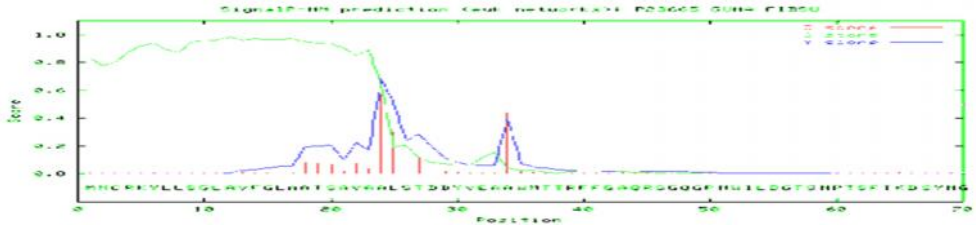
Inference

SOPMA predict secondary structure for 1ut9A and also it gives length, alpha helix, beta strand, beta turn, coils and the output of graphics display.

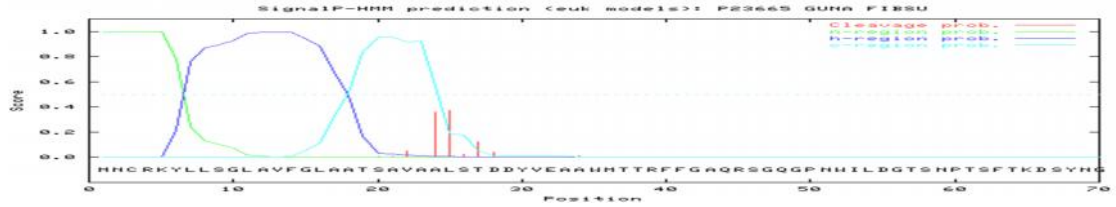
7. SignalP 3.0 Server - prediction results

Using neural networks (NN) and hidden Markov models (HMM) trained on eukaryotes
>P23665_GUNA_FIBSU Endoglucanase A - Fibrobacter succinogenes
SignalP-NN

Result:



>P23665_GUNA_FIBSU length = 70
Measure Position Value Cutoff signal peptide?
max. C 24 0.568 0.32 YES max. Y 24 0.679 0.33 YES
max. S 12 0.985 0.87 YES mean S 1-23 0.916 0.48 YES
D 1-23 0.798 0.43 YES
Most likely cleavage site between pos. 23 and 24: AVA-AL
SignalP-HMM result:



=
>P23665_GUNA_FIBSU
Prediction: Signal peptide
Signal peptide probability: 0.992 Signal anchor probability: 0.007
Max cleavage site probability: 0.375 between pos. 24 and 25

Inference:

The result implies that the given protein sequence contain a signal sequence this gives a clue that the protein is both cytosolic protein by nn and by hmm. The sequence in the aminoterminal 24 or 25 amino acid from this results that our 1ut9A sequence contain any signal sequence which is in first 23 aminoacid by neural network.

TargetP 1.1 Server - prediction results
Number of query sequences: 1
Cleavage site predictions included.
Using NON-PLANT networks.

Name	Len	mTP	SP	other	Loc	RC	TPlen
P23665_GUNA_FIBSU	453	0.069	0.910	0.022	S	1	23
cutoff		0.000	0.000	0.000			

Inference:

The result implies that the given protein contain a signal sequence having a probe of 0.069mTP. From this result is noted that protein contain signal sequence having destination mitochondria.

MOD BASE Result

Cross-references		
Template Structure		
PDB	1ut9	cellulose 1,4-beta-cellobiosidase: catalytic domain, residues 208-816
DBALI	1ut9A	
Jena Image Library	1ut9	
Target Sequence		
SwissProt	P23664	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)(Cellulase).
UniProt	P23664	
InterPro	P23664	
PFAM	P23664	

PRODOM	P23664	
SwissProt	P23665	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)(Cellulase).
UniProt	P23665	
InterPro	P23665	
PFAM	P23665	
PRODOM	P23665	
GenPept	121804	GUNA_FIBSU ENDOGLUCANASE A PRECURSOR (ENDO-1,4-BETA-GLUCANASE) (CELLULASE)

10. PFAM Result:

This is the summary of UniProt entry GUNA_FIBSU (P23665).

Description: Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)(Cellulase).

Source organism: Fibrobacter succinogenes (Bacteroides succinogenes). (NCBI taxonomy ID 833)
ViewPfam genome data.

Length: 453 amino acids

Source	Domain	Start	End
PfamA	Glycohydro9	29	445

11. cholorop Result:

ChloroP 1.1 Server - prediction results
TechnicalUniversity of Denmark

chlorop v1.1 prediction results
##########################
Number of query sequences: 1

Name	Length	Score	cTP	CS-	cTP-score	length
P23665_GUNA_FIBSU	453	0.445	-	-	0.385	65

---Detailed output----------

Residue	--NN-score--	CS-score	Raw	Deriv.
Name: P23665_GUNA_FIBSU				
1 M	0.319	0.000	0.000	
2 N	0.293	0.000	0.000	
3 C	0.261	0.000	-9.099	
4 R	0.204	0.000	-11.993	
5 K	0.239	0.000	-2.983	
6 Y	0.233	0.086	-18.693	
7 L	0.178	0.084	-12.326	
8 L	0.159	0.068	-8.693	
9 S	0.155	0.049	-7.414	
10 G	0.162	0.039	-1.941	
11 L	0.154	0.018	-12.483	
12 A	0.145	-0.020	-12.002	
13 V	0.150	-0.062	-5.568	
14 F	0.160	-0.094	-4.701	
15 G	0.188	-0.097	-4.365	
16 L	0.263	-0.102	-4.106	
17 A	0.326	-0.056	-2.711	
18 A	0.300	0.022	1.457	
19 T	0.178	0.084	0.796	
20 S	0.238	0.097	-4.836	
21 A	0.145	0.139	1.273	
22 V	0.117	0.130	-1.799	
23 A	0.139	0.100	2.960	
24 A	0.130	0.091	-2.509	
25 L	0.078	0.094	-4.243	3.625
230 A	0.190	-0.040	-2.597	
231 K	0.203	-0.042	-4.373	
232 T	0.341	-0.021	-5.670	
233 A	0.208	0.048	-3.222	
234 Y	0.247	0.064	-1.432	
235 S	0.254	0.100	-1.887	
236 Y	0.157	0.145	-9.659	
237 A	0.124	0.133	-1.196	
238 K	0.077	0.090	-5.326	
239 S	0.075	0.051	-3.058	
240 H	0.094	0.000	-5.782	
241 K	0.172	-0.046	-11.904	
242 G	0.122	-0.036	-5.991	
243 V	0.140	-0.038	-7.535	
244 T	0.158	-0.014	-12.991	
245 N	0.167	0.012	-3.089	
246 S	0.137	0.054	-6.739	
247 Q	0.126	0.069	-9.652	
248 G	0.083	0.092	-2.020	
249 F	0.114	0.092	-19.985	
250 Y	0.030	0.103	-16.891	
251 E	0.026	0.079	-4.705	
252 S	0.015	0.059	-13.327	
253 S	0.028	0.035	-15.077	
254 W	0.014	0.028	-16.735	
255 W	0.014	0.007	-8.607	
256 D	0.015	0.005	-17.010	
257 G	0.021	0.003	-14.613	
258 R	0.011	0.007	-17.465	
259 W	0.016	0.005	-12.809	
260 E	0.011	0.007	-19.627	
261 D	0.010	0.008	-15.233	
262 G	0.009	0.008	-16.961	
263 P	0.006	0.006	-20.577	
264 F	0.006	0.006	-14.934	
265 L	0.005	0.004	-9.053	
266 A	0.005	0.003	-2.694	
267 E	0.004	0.002	-12.965	
268 L	0.004	0.001	-4.778	
269 E	0.005	0.000	-12.479	
270 L	0.004	0.000	-9.483	
271 Y	0.005	0.000	-12.199	
272 R	0.004	0.000	-4.310	
273 T	0.005	0.000	-5.509	
274 T	0.004	-0.002	-8.646	
275 G	0.004	-0.002	-6.838	
276 E	0.004	-0.002	-5.341	
277 N	0.007	-0.002	-6.701	

278 S	0.012	-0.001	-10.728
279 Y	0.005	0.002	-11.997
. 300 F	0.191	-0.002	-6.305
301 M	0.065	0.046	-16.878
302 Y	0.120	0.040	-2.646
303 S	0.090	0.045	-5.758
304 N	0.072	0.041	-14.206
305 V	0.084	0.032	-7.193
306 V	0.081	0.021	-11.649
307 P	0.128	0.037	-8.298
308 L	0.044	0.058	3.668
309 S	0.039	0.054	-2.246
310 A	0.032	0.053	5.602
311 V	0.018	0.047	0.172
312 M	0.031	0.037	-3.748
313 A	0.020	0.023	9.437
314 E	0.011	0.021	-4.202
315 A	0.007	0.017	-4.646
316 V	0.006	0.013	-8.311
317 F	0.005	0.010	-4.336
318 E	0.005	0.005	-5.737
319 E	0.004	0.002	-12.829
320 T	0.004	0.001	-12.998
321 P	0.005	0.000	-18.230
322 H	0.006	0.000	-0.877
323 G	0.004	0.001	0.421
324 M	0.004	0.001	-16.010
325 R	0.004	0.001	-11.684
326 K	0.004	0.001	0.024
327 E	0.004	0.000	-10.103
328 A	0.004	0.000	-7.551
329 I	0.004	0.000	-18.596
330 G	0.004	0.000	-17.050
331 V	0.004	0.000	-15.266
332 L	0.004	0.000	-10.781
333 D	0.004	0.000	-5.838
334 L	0.004	0.000	-8.513
335 I	0.004	0.000	-10.881
336 Y	0.004	0.000	-4.989
337 E	0.004	0.000	1.347
338 E	0.004	0.000	-6.093
339 K	0.004	0.000	-12.203
340 A	0.004	0.000	-12.895
341 K	0.004	-0.001	-9.088
342 D	0.004	-0.003	-5.093
343 K	0.004	-0.004	-20.862
344 I	0.005	-0.005	-12.479
345 F	0.008	-0.006	-11.036
346 Q	0.013	-0.005	-9.247
347 N	0.010	-0.005	-9.502
348 P	0.007	-0.008	-13.701
349 N	0.011	-0.015	-5.944
350 G	0.010	-0.026	-1.279
351 M	0.022	-0.034	-15.129
352 G	0.028	-0.046	-6.396
353 S	0.045	-0.063	-8.480
354 G	0.072	-0.085	-13.402
355 K	0.052	-0.100	-12.449
356 F	0.091	-0.110	-13.361
357 P	0.131	-0.100	-12.041
358 V	0.196	-0.080	-6.216
359 R	0.206	-0.045	-6.591
360 V	0.145	-0.057	2.188
361 P	0.112	-0.075	-11.888
362 S	0.130	-0.122	-7.147
363 G	0.172	-0.165	-4.715
364 G	0.403	-0.215	-13.443
365 A	0.326	-0.146	-11.599
366 F	0.368	-0.104	-16.571
367 L	0.346	-0.020	-8.458
368 Y	0.396	0.063	-8.627
369 A	0.255	0.177	-5.931
370 L	0.298	0.191	-8.411
371 S	0.202	0.235	-3.758
372 D	0.149	0.234	-9.527
373 K	0.050	0.213	-11.141
374 F	0.036	0.148	-11.217
375 N	0.051	0.107	-11.016
376 N	0.043	0.062	-12.559
377 T	0.053	0.034	-10.016
378 N	0.030	0.023	-14.476
379 E	0.024	0.023	-6.868
380 H	0.029	0.023	-12.811
381 M	0.024	0.022	-7.634
382 E	0.012	0.020	-2.133
383 M	0.011	0.012	-8.257
384 I	0.012	0.009	-7.508
385 E	0.011	0.008	-5.092
386 K	0.012	0.004	-13.414
387 N	0.011	0.002	-19.305
388 V	0.007	0.002	-5.592
389 S	0.008	0.001	-8.750
390 Y	0.011	-0.001	-6.387
391 L	0.012	-0.001	-11.895
392 L	0.008	-0.001	-10.552
393 G	0.009	-0.003	-1.033
394 D	0.012	-0.003	-7.692
395 N	0.013	-0.002	-9.534
396 G	0.014	-0.002	-3.973
397 S	0.013	0.000	-15.754
398 K	0.011	0.000	-10.130
399 K	0.012	-0.001	-1.166
400 S	0.014	-0.004	-9.159
401 Y	0.008	-0.011	-8.042
402 V	0.018	-0.017	-7.423
403 V	0.016	-0.026	-8.978
404 G	0.027	-0.074	2.122
405 F	0.050	-0.128	-3.273
406 S	0.033	-0.210	-9.450
407 K	0.066	-0.331	-6.144
408 N	0.261	-0.443	-6.554
409 G	0.314	-0.493	-2.113
410 A	0.496	-0.509	-5.342
411 N	0.661	-0.443	-13.495
412 A	0.677	-0.351	-3.044
413 P	0.754	-0.253	-15.521
414 S	0.683	-0.173	-13.193
415 R	0.611	-0.124	-6.747
416 P	0.828	-0.148	-3.655
417 H	0.797	-0.116	-15.866
418 H	0.848	-0.090	-13.042
419 R	0.807	-0.028	-10.058
420 G	0.847	0.018	-5.272
421 Y	0.832	0.126	-15.462
422 Y	0.787	0.218	-8.415
423 A	0.633	0.304	-3.708
424 N	0.701	0.332	-7.305
425 E	0.543	0.408	-11.534
426 K	0.376	0.443	-15.151
427 R	0.349	0.420	-14.359
428 W	0.279	0.396	-9.030
429 R	0.211	0.370	-14.699
430 R	0.066	0.311	-7.581
431 S	0.035	0.226	-7.550
432 R	0.030	0.163	-14.785
433 R	0.054	0.104	-4.654
434 C	0.019	0.069	2.204
435 S	0.012	0.034	-4.641

436	E	0.010 0.025	3.594
437	S	0.005 0.021	-1.691
438	S	0.004 0.016	-5.208
439	R	0.004 0.006	-11.112
440	K	0.004 0.003	-5.325
441	E	0.004 0.001	-5.809
442	Q	0.004 0.000	-2.842
443	A	0.004 0.000	-10.970

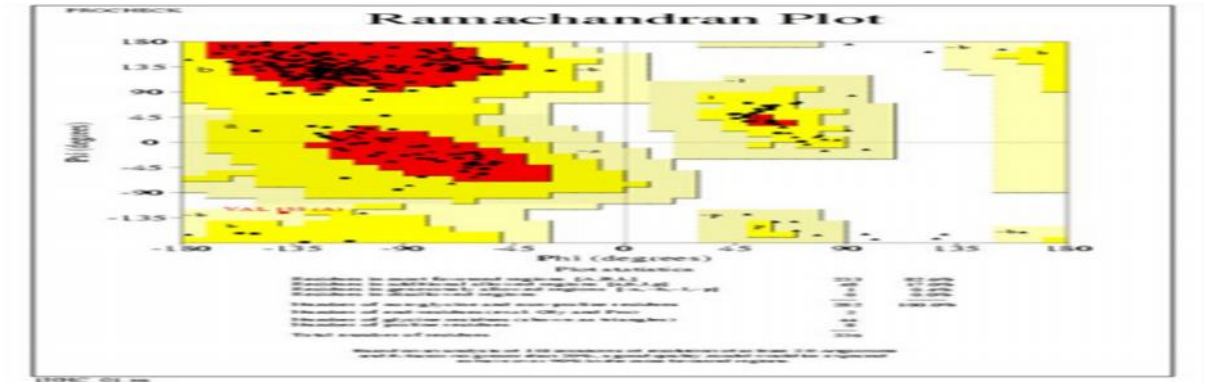
---End----------

Molecular modelling of laccase

The sequence of Laccase was retrieved from Universal ProteinResource (UniProt) and its corresponding sequence id was P51589. Itconsists of 529 amino acids. This sequence was subjected to similaritysearch against Protein Data Bank, using the BLAST tool offered byNCBI. Later, the templates were selected on the basis of structural hitsand its alignment pattern against the query sequence. The selected templateswere as follows: chain A of 1GYC, chain A of 3KW7 and chainA of to 1V10.

The advanced modelling tutorial package offered in MODELLERwas utilized for comparative molecular modelling. The DOPE scorebelonging to the best modeled structure was -60304.7734. The stereo chemistry qualities of the structures were validated with PROCHECK[37] structural validation tool. PROCHECK results clearly indicatedthe higher fidelity of modeled Laccase structure .

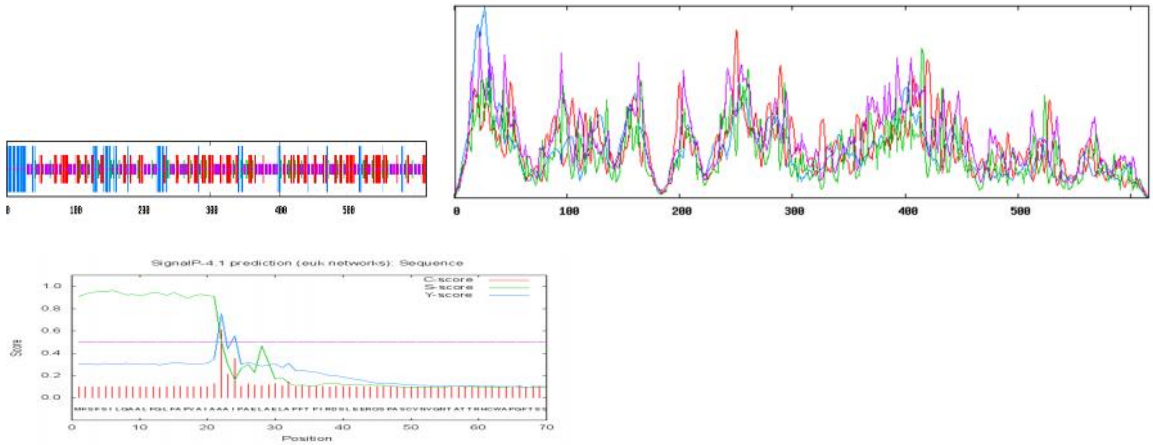
PROCHECK RESULT



Active site analysis

407 O TRP A 90 , 418 N LYS A 91 669 C GLY A 124 , 671 N THR A 125 678 N ASN A 126 , 686 N GLY A 127 690 N GLY A 128 , 694 N LYS A 129 703 N THR A 130 , 710 N LYS A 131 719 N PRO A 132 , 726 N LYS A 133 912 CB VAL A 155

Laccase



***Alternaria alternata* [gbpln]: 56 CDS's (24785 codons)**

AmAcid	Codon	Number	/1000	Fraction	..				
Gly	GGG	215.00	8.67	0.00	Gly	GGA	424.00	17.11	0.00
Gly	GGT	483.00	19.49	0.00	Gly	GGC	630.00	25.42	0.00
Glu	GAG	890.00	35.91	0.00	Glu	GAA	556.00	22.43	0.00
Asp	GAT	565.00	22.80	0.00	Asp	GAC	877.00	35.38	0.00
Val	GTG	267.00	10.77	0.00	Val	GTA	230.00	9.28	0.00
Val	GTT	486.00	19.61	0.00	Val	GTC	675.00	27.23	0.00
Ala	GCG	358.00	14.44	0.00	Ala	GCA	482.00	19.45	0.00
Ala	GCT	701.00	28.28	0.00	Ala	GCC	755.00	30.46	0.00
Arg	AGG	166.00	6.70	0.00	Arg	AGA	194.00	7.83	0.00
Ser	AGT	181.00	7.30	0.00	Ser	AGC	398.00	16.06	0.00
Lys	AAG	920.00	37.12	0.00	Lys	AAA	325.00	13.11	0.00
Asn	AAT	289.00	11.66	0.00	Asn	AAC	595.00	24.01	0.00
Met	ATG	530.00	21.38	0.00	Ile	ATA	237.00	9.56	0.00
Ile	ATT	383.00	15.45	0.00	Ile	ATC	676.00	27.27	0.00

Thr	ACG	273.00	11.01	0.00	Thr	ACA	343.00	13.84	0.00
Thr	ACT	370.00	14.93	0.00	Thr	ACC	514.00	20.74	0.00
Trp	TGG	302.00	12.18	0.00	End	TGA	14.00	0.56	0.00
Cys	TGT	184.00	7.42	0.00	Cys	TGC	304.00	12.27	0.00
End	TAG	16.00	0.65	0.00	End	TAA	26.00	1.05	0.00
Tyr	TAT	229.00	9.24	0.00	Tyr	TAC	393.00	15.86	0.00
Leu	TTG	356.00	14.36	0.00	Leu	TTA	109.00	4.40	0.00
Phe	TTT	308.00	12.43	0.00	Phe	TTC	634.00	25.58	0.00
Ser	TCG	330.00	13.31	0.00	Ser	TCA	289.00	11.66	0.00
Ser	TCT	433.00	17.47	0.00	Ser	TCC	425.00	17.15	0.00
Arg	CGG	156.00	6.29	0.00	Arg	CGA	239.00	9.64	0.00
Arg	CGT	277.00	11.18	0.00	Arg	CGC	289.00	11.66	0.00
Gln	CAG	577.00	23.28	0.00	Gln	CAA	387.00	15.61	0.00
His	CAT	258.00	10.41	0.00	His	CAC	344.00	13.88	0.00
Leu	CTG	418.00	16.87	0.00	Leu	CTA	244.00	9.84	0.00
Leu	CTT	496.00	20.01	0.00	Leu	CTC	585.00	23.60	0.00
Pro	CCG	249.00	10.05	0.00	Pro	CCA	274.00	11.06	0.00
Pro	CCT	348.00	14.04	0.00	Pro	CCC	304.00	12.27	0.00

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

10 20 30 40 50 60 70 80 90 100 110 120
MKSFSILGAA LFGLFAPVAI AAAIPAELAE LAPFTPIRDS LEERQSPASC VNVGNTATTR HCWAPGFTSS TDMYTSWPNT
GVVRSYNLRI ENTTCNPDGA GSRVCMLING RYPGPTIVANWGDTIRVTVR NLLQANGTSI HWHGFRMLNK NIQDGVNGIT
ECALAPNDVK TYEFQATEYGTTWYHSHFSH QYGDGVVGTV IVNGPATANY DEDLGVMPIT DWYYQTAYQA
ASIAFQNGQGLGPPVGDNI LINGTAKNAA GGGAWNNVKI QAGKRYRLRL VNTAVDTNMV VNLDGHPFQV IATDFVPINP
YNTSHLQIGI GQRYDVIITA NQTAGNYWFR AVADGLCQSR NTREGRAVFTYQGQTVADPT SNSTAIPFTE CVDPVTSPKI
AKNVPSTTFA AQAKSLPVAF GPVAANGNTV LWTINGTSMI IDPGKPTIKY VAETNNSFPQ SYNVVEVPST SASTWSYWVV
QQAVGAPPLAHPIHLHGHDS YVLGAGDGQF NVSTHFSQLR FTNPPRRDVT QLKKNGWLVL AYPTDNPGAW
LMHCHIAFHV GMGLSVQFLE RKQSINLPAP GSEWYGNCNK WASYKAGTTD IWPQDDSGLK KRWPPLIEGG STFRLD

Theoretical pI/Mw: 6.96 / 66922.21

ProtParam

Laccase {ECO:0000313|EMBL:OAG13331.1}
Alternaria alternata (Alternaria rot fungus) (Torula alternata).
The computation has been carried out on the complete sequence (**616** amino acids).
Number of amino acids: 616 **Molecular weight:** 66922.21

Theoretical pI: 6.96

Amino acid composition

Ala (A)	59	9.6%	Arg (R) 24	3.9%	Asn (N) 45	7.3%
Asp (D)	27	4.4%	Cys (C) 9	1.5%	Gln (Q) 27	4.4%
Glu (E)	16	2.6%	Gly (G) 55	8.9%	His (H) 16	2.6%
Ile (I)	34	5.5%	Leu (L) 36	5.8%	Lys (K) 18	2.9%
Met (M)	9	1.5%	Phe (F) 23	3.7%	Pro (P) 41	6.7%
Ser (S)	37	6.0%	Thr (T) 53	8.6%	Trp (W) 17	2.8%
Tyr (Y)	23	3.7%	Val (V) 47	7.6%	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): 43
Total number of positively charged residues (Arg + Lys): 42
Atomic composition:

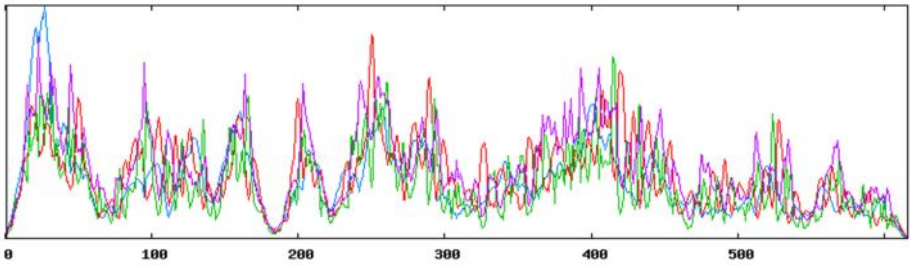
Carbon	C	2994	Hydrogen	H	4557
Nitrogen	N	827	Oxygen	O	888
Sulfur	S	18			

Formula: $C_{2994}H_{4557}N_{827}O_{888}S_{18}$
Total number of atoms: 9284
Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 128270
Abs 0.1% (=1 g/l) 1.917, assuming all pairs of Cys residues form cystines
Ext. coefficient 127770
Abs 0.1% (=1 g/l) 1.909, 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 34.80
This classifies the protein as stable.
Aliphatic index: 76.02

Grand average of hydropathicity (GRAVY): -0.218
SOPMA result for : UNK_347730

```
      10        20        30        40        50        60        70
       |        |         |         |         |         |         |
MKSFSILGAALFGLFAPVAIAAAIPAELAELAPFTPIRDSLEERQSPASCVNVGNTATTRHCWAPGFTSS
hhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhcccccccchhhhhtcccccceeeetcccccccccctttcccc
TDMYTSWPNTGVVRSYNLRIENTTCNPDGAGSRVCMLINGRYPGPTIVANWGDTIRVTVRNLLQANGTSI
ceeeecccccceeeeeeeeeetccccccccccceeeeeettccccceeeecttcheeeehhhhhhtttcce
HWHGFRMLNKNIQDGVNGITECALAPNDVKTYEFQATEYGTTWYHSHFSHQYGDGVVGTVIVNGPATANY
eeehhhhhhhhhhhttccthhhhhecctttccceeeeeeecchtceeeccccccctttteeeeeeeccccccccc
DEDLGVMPITDWYYQTAYQAASIAFQNGQAGLGPPVGDNILINGTAKNAAGGGAWNNVKIQAGKRYRLRL
ccttttccccchhhhhhhhhhheeehttttttccccccceeeeetcccccttcccceeeecttcceeeee
VNTAVDTNMVVNLDGHPFQVIATDFVPINPYNTSHLQIGIGQRYDVIITANQTAGNYWFRAVADGLCQSR
eehcccteeeeeettcceeeeeeeeccccccccceeeeccccceeeeeeeccttteeeehhhtthhhcc
NTREGRAVFTYQGQTVADPTSNSTAIPFTECVDPVTSPKIAKNVPSTTFAAQAKSLPVAFGPVAANGNTV
Ccttcceeeeettceeccccccccccceeccccccccccccccccchhhhhtcccceeeccccttcee
LWTINGTSMIIDPGKPTIKYVAETNNSFPQSYNVVEVPSTSASTWSYWVVQQAVGAPPLAHPIHLHGHDS
eeeetcceeeecttcccceeeehccccccccceeeeccccccccceeeeehhhtcccccccceeettcce
YVLGAGDGQFNVSTHFSQLRFTNPPRRDVTQLKKNGWLVLAYPTDNPGAWLMHCHIAFHVGMGLSVQFLE
eeeettcccceeeeeeeeeeeccccccchhhhctttteeeeeccccttceeeeeeeeeeechtceeeeec
RKQSINLPAPGSEWYGNCNKWASYKAGTTDIWPQDDSGLKKRWPPLIEGGSTFRLD
Ttccccccccccceeccccchhhhcttceeeccccctttccccccccctteeeec
```

Sequence length : 616 SOPMA :
Alpha helix (Hh) :93 is 15.10% 310helix(Gg) :0 is0.00% Pi helix(Ii):0 is 0.00 Beta bridge (Bb) : 0 is 0.00% Extended strand (Ee) : 187 is 30.36% Beta turn (Tt) 73 is 11.85% Bend region (Ss) : 0 is 0.00% Random coil (Cc) : 263 is 42.69%Ambiguous states (?) : 0 is 0.00 Other states :0 is 0.00%Parameters : Window width : 17 Similarity threshold : 8 Number of states 4



ScanProsite Results Viewer

Ouput format: Graphical view - this view shows ScanProsite results together with ProRule-based predicted intra-domain features.

Hits for all PROSITE (release 20.129) motifs on sequence A0A177D1J9 [UniProtKB/TrEMBLfound: 2 hits in 1 sequence

A0A177D1J9 **A0A177D1J9_ALTAL** (616 aa)
SubName: Full=Laccase {ECO:0000313|EMBL:OAG13331.1};. *Alternaria alternata (Alternaria rot fungus) (Torula alternata)*

```
MKSFSILGAALFGLFAPVAIAAAIPAELAELAPFTPIRDSLEERQSPASCVNVGNTATTRHCWAPGFTSSTDMYTSWPNTGVVRSYNLRIENTTCN
PDGAGSRVCMLINGRYPGPTIVANWGDTIRVTVRNLLQANGTSIHWHGFRMLNKNIGVNGITECALAPNDVKTYEFQATEYGTTWYHSHFSHQ
YGDGVVGTVIVNGPATANYDEDLGVMPITDWYYQTAYQAASIAFQNGQAGLGPPVGDNILINGTAKNAAGGGAWNNVKIQAGKRYRLRLVN
TAVDTNMVVNLDGHPFQVIATDFVPINPYNTSHLQIGIGQRYDVIITANQTAGNYWFRAVADGLCQSRNTREGRAVFTYQGQTVADPTSNSTAI
PFTECVDPVTSPKIAKNVPSTTFAAQAKSLPVAFGPVAANGNTVLWTINGTSMIIDPGKPTIKYVAETNNSFPQSYNVVEVPSTSASTWSYWVVQ
QAVGAPPLAHPIHLHGHDSYVLGAGDGQFNVSTHFSQLRFTNPPRRDVTQLKKNGWLVLAYPTDNPGAWLMHCHIAFHVGMGLSVQFLERK
QSINLPAPGSEWYGNCNKWASYKAGTTDIWPQDDSGLKKRWPPLIEGGSTFRLD
```

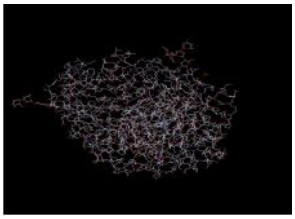


Fig 36:Wire frame model of Laccase Laccase

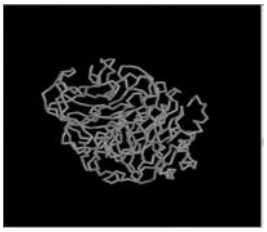


Fig 37 :Backbone model of Laccase

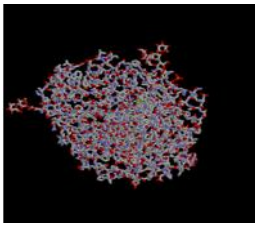


Fig 38 Sticks of Laccase

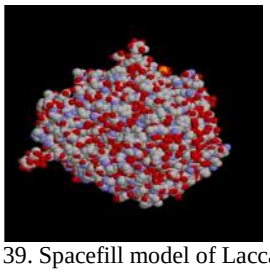


Fig 39. Spacefill model of Laccase

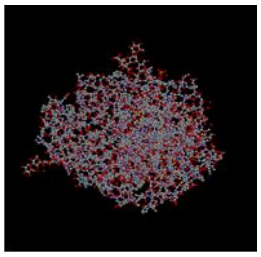


Fig. 40 Ball and Stick model of Laccase

Fig 41 Ribbon moel of Laccase

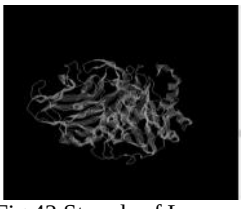


Fig 42:Strands of Laccase

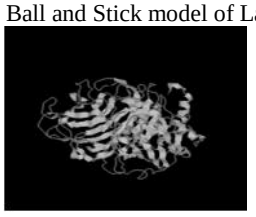


Fig 43:Cartoons of Laccse

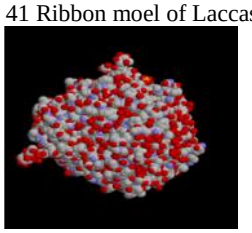


Fig. 44 Molecular Strand of Laccase

DISCUSSION

From the above picture the Laccase produced from the Alternaria cepulae have 2 number of chains,499 number of groups, 3806 number of atoms,197 bonds,13 helices,31 strands,zero number of turns and 4107 number of bonds

REFERENCES

Ali, A.J., Ahmed, A.H., Zahraa, A. & Umar, A. (2011) Optimization of Cellulase Production by *Aspergillus niger* and*Tricoderma viride* using Sugar Cane Waste, *Journal of Yeast and*

Baldrian, P., Snajdr, J. (2006) Production of ligninolytic enzymes by litter-decomposing fungi and their ability to decolorize synthetic dyes. Enzyme Microbiol Technol 39: 1023-1029.*Fungal Research*, 2(2), 19 – 23

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The protein data bank.Nucleic Acids Res 28:235–242

Pradeep, N.V., Anupama, Vidyashree, K.G., Lakshmi, P. (2012) In silico Characterization of Industrial Important Cellulases using Computational tool.*Adv Life Sci Tech.* 4:8-15.

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

Sali, A., Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815

Spano, L., Medeiros, J., Mandels, M. (1975) Enzymatichydrolysis of cellulosic waste to glucose, pollution abatement DIV., Food SVCS. Labs. Us Army Natick, MA, USA, 7,

Whitaker, D.R. (1971) Cellulases. *The enzymes*, *5*, 273-290.

Whitaker, H.A. and McAdam, Dale, W. "Language production: Electroencephalographic localization in the normal human brain." *Science*172.3982 (1971): 499-502.

Yoshida, F. and Mizusawa, K. (1976) Experimental suppliers, 26,61

Zhou, H. and Zhou, Y. (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 11:2714–2726.

.