Advances in Life Science and Technology
ISSN 2224-7181 (Paper) ISSN 2225-062X (Online)
Vol 4, 2012

In silico Characterization of Industrial Important Cellulases using Computational Tools

Pradeep N V[1*], Anupama[1], Vidyashree K G[2], Lakshmi P[2].

1. Assistant Professor, Dept of Biotechnology, Ballari Institute of Technology and Management, Bellary, India.
2. B.E Scholar, Dept of Biotechnology, Ballari Institute of Technology and Management, Bellary, India.

* E-mail of the corresponding author: nagamallivpradeep@gmail.com

Abstract

Cellulases refer to a class of enzymes produced majorly by fungi, bacteria and protozoans that catalyze cellulolysis. Cellulase enzyme is used extensively in various industries, especially in textile, food and in the bioconversion of lignocellulosic wastes to alcohol. The extensive use of cellulase 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. Cellulases from different strains of *Pseudomonas* species were analyzed using computational tools. The physicochemical properties of the selected cellulases were analyzed by using ExPASy's ProtParam tool and it was found that the molecular weight (M.Wt) ranges between 40927.4-100058.7 Da. Isoelectric Points (pI) of all the organisms were found to be acidic in nature. The aliphatic index infers that all the cellulases are stable. The negative value of GRAVY indicates that there will be better interaction with water. The secondary structure prediction was done by SOPMA which showed that random coils dominated all the other conformations. Multiple sequence analysis and evolutionary analysis of cellulases were carried out by CLC workbench. The Phylogenetic analysis was done using Neighbour joining method. The 3D structures of cellualses were obtained by ESyPred 3D server.

Keywords: Cellulases, Enzymes, ProtParam, SOPMA, ESyPred 3D.

1. Introduction

Active research on cellulases and related polysaccharidases began in the early 1950s, owing to their enormous potential to convert lignocellulose, the most abundant and renewable source of energy on Earth, to glucose and soluble sugars (Bhat M K, 2000).

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, single cell protein and other chemicals from renewable cellulosic materials. To date, production of cellulase has been widely studied in submerged culture processes, but relatively high cost of enzyme production has hindered the

industrial application of cellulose bioconversion. It has been reported that solid-state fermentation is an attractive process to produce cellulose economically due to its lower capital investment and lower operating expenses. Another approach to reduce the cost of cellulose production is the use of lignocellulosic materials as substrates rather than expensive pure cellulose. Abundant agricultural residues such as corn stover, wheat straw, rice straw, bagasse, etc. were used in cellulase production. Although these raw materials are cheaper, pretreatment is generally required to improve the utilization ratio of lignocellulosic materials (Jian and Jichu, 2007). Microbial digestion of lignocellulosic waste and the downstream products resulting from it is accomplished by a concerted action of several enzymes, the most prominent of which are the cellulases, which are produced by a number of microorganisms and comprise several different classifications.

There are three major types of cellulase enzymes:

1. Cellobiohydrolase (CBH or 1, 4-β-D-glucan cellobiohydrolase, EC 3.2.1.91)

2. Endo-β-1, 4-glucanase (EG or Endo-1, 4-β-D-glucan 4-glucanohydrolase, EC 3.2.14)

3. β-glucosidase (BG-EC 3.2.1.21) (Rajeev *et al.*, 2005).

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 cellulases from different *Pseudomonas* species 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 (Ashokan *et al*., 2011). These parameters will assist the biochemist and physiologists in extraction, purification, separation and industrial applications of the enzyme.

2. Materials and methods

2.1 Sequence retrieval

The sequences of cellulase were retrieved from NCBI (National Center for Biotechnology Information). Sequences retrieved belong to the *Pseudomonas* species (Accession numbers: ACX31080.1, ADK13057.1, ABY99087.1, EGH91691.1, and ZP_07003414.1).

2.2 Analysis of physicochemical parameters

The different physicochemical properties of cellulase enzyme were computed using ExPASy's ProtParam tool and these properties can be deduced from a protein sequence. 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). The instability index provides an estimate of the stability of our protein. A protein whose instability index is smaller than 40 is predicted as stable; a value above 40 predicts that the protein may be unstable (Guruprasad *et al*, 1990). The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermo stability of globular proteins (Walker, 2005).

2.3 Secondary structure prediction

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 the selected 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. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method (Prashant *et al.,* 2010).

2.4 CLC analysis

The CLC free workbench was used to perform multiple sequence alignment with gap open cost 10.0 and gap open extension cost 1.0. End gap cost was not considered. The resulting alignment was used for the construction of evolutionary tree with neighbour joining algorithm (Ashokan *et al*., 2011).

2.5 ESyPred 3D server

The 3D structures of cellulases were generated by homology modeling using ESyPred 3D server. ESyPred 3D models query proteins using a 3D structure template present in the database using ALIGN program (Christophe *et al*., 2002; Sivakumar *et al.,* 2007).

3. Results and discussion

3.1 Prot Param

The physicochemical properties of cellulases were predicted by using ProtParam tool. The physicochemical properties (table1) show that molecular weight is highest in *Pseudomonas aeruginosa* (100058.7 Da) and lowest in *Pseudomonas putida* (40927.4 Da). The instability index showed that cellulase of *Pseudomonas fluorescens* (45.28) is more stable than other species studied. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of the protein is zero. The computed pI value of ACX31080.1, ADK13057.1, ABY99087.1, EGH91691.1, ZP_07003414.1 (pI<7) indicates that these cellulases are acidic in nature. On the basis of instability index (II) ExPASy's protparam classifies the ACX31080.1 (*Pseudomonas aeruginosa*), ADK13057.1 (*Pseudomonas fluorescens*), ABY99087.1 (*Pseudomonas putida*) cellulases as unstable (II>40) and other cellulases EGH91691.1 (*Pseudomonas syringae*) ZP_07003414.1 (*Pseudomonas savastanoi*) as stable (II<40). The very high aliphatic index of all cellulases infers that they may be stable for a wide range of temperature. The very low GRAVY index of cellulases ADK13057.1, EGH91691.1 infers that these cellulases could result in a better interaction with water. Similar trend was generated for Protein Feature Based Identification of Cell Cycle Regulated Proteins in Yeast (Ulrik *et al*., 2003; Jianwen *et al*., 2009).

3.2 SOPMA

The secondary structure prediction (Table 2) shows that random coil predominates the other structures whereas β-turn being the least conformational structure. Random coil is dominant in the cellulases of ACX31080.1, ADK13057.1, EGH91691.1, ZP_07003414.1 except ABY99087.1. In case of ABY99087.1 (*Pseudomonas putida*) α-helix being 40.97% and random coil 39.62% showing nearly equal proportions. In all the cellulases analyzed, it was clearly noticed that β-turns showing very less percentage of conformation (below 10%). In all the cellulases extended strands were ranging from 10-20%. The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. (Jyotsna *et al.,* 2010; Ojeiru *et al.,* 2010). Similarly SOPMA was used for structure prediction of thioredoxin protein in wheat by Prabhavathi *et.al* (2011).

3.3 Multiple sequence analysis

Cellulases from different organisms were subjected to multiple sequence analysis using CLC workbench. Similarly phylogenetic relationships between the catalase of the different taxonomic groups were compared using multiple sequence analysis by Ashokan *et.al* (2011).

3.4 Phylogenetic tree

In order to find out any taxonomic variation in the cellulase, multiple sequence alignment has been performed. The MSA (Figure 1) followed by phylogenic tree construction (Figure 2) using Neighbour joining method showed a clear phylogenetic relationship between the cellulase of the different taxonomic groups. Similar method was carried out for Biochemical and Genetic Analyses of a Catalase from the Anaerobic Bacterium *Bacteroides fragilis* (Edson *et al.*, 1995).

3.5 ESyPred 3D Analysis

The 3D structure analyses of cellulases were done by using ESyPred 3D tool (figure 3). Predicting the protein 3D structure by this method is used which implements the four steps of the homology modeling approach:

1. Databanks searching to identify the structure homology,
2. Target-template alignment,
3. Model building and optimization and
4. Model evaluation (Christophe *et al*., 2002).

4. Conclusion

For obtaining desirable results in industrial application, it is essential to manipulate the characteristic properties of enzyme which is a tedious task. Protein engineering techniques used to achieve this goal require a sound knowledge about the protein both at sequence and structure level. In this study, 5 Cellulases sequences were selected to acquire an understanding about their physico-chemical properties and various protein structure levels by using in silico techniques. Physicochemical properties of cellulases were analyzed by using ProtParam. Physicochemical characterization studies give more insight about the properties such as M.Wt, pI, AI, GRAVY and Instability Index that are essential and vital in providing data about the proteins and their properties. Secondary structure prediction was carried out using SOPMA. SOPMA predicted that all the cellulases contain large percentage of random coils and the least conformation was of β-turns. Multiple sequence alignment and evolutionary analysis were done by using CLC workbench, construction of evolutionary tree was done by using neighbour joining algorithm. Multiple sequence analysis and the evolutionary tree showed that the cellualses from organisms bearing the accession number ACX31080.1 and ADK13057.1 are closely related. The 3D structures of cellulases were generated by homology modeling using ESyPred 3D server. This study will provide an insight about the physiochemical properties and function of cellualses and thus aid in formulating their uses in industries.

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.

Ashokan K V, Mundaganur D S and Mundaganur Y D. (2010). 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.

Christophe L, Nadia L, Xavier D B and Eric D, (2002). ESyPred3D: Prediction of Proteins 3D Structures. *Bioinformatics*. 18(9), 1250-1256.

Edson R R and Smith C J, (1995). Biochemical and Genetic Analyses of a Catalase from the Anaerobic Bacterium Bacteroides fragilis. *Journal of Bacteriology*, 177(11), 3111–3119.

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

Jian L and Jichu Y, (2007). Cellulase Production by *Trichoderma koningii* AS3.4262 in Solid-State Fermentation Using Lignocellulosic Waste from the Vinegar Industry, *Food Technol. Biotechnol*, 45 (4), 420–425.

Jianwen F, Yakov M K and Robert P H, (2009). Bioinformatic Analysis of Xenobiotic Reactive Metabolite Target Proteins and their Interacting Partners, *BMC Chemical Biology*, 9(5), 1-13.

Jyotsna C, Ashish P, Shailendra G, Verma M K. (2010). Homology Modeling and Binding Site Identification of 1 deoxy d- xylulose 5 phosphate Reductoisomerase of Plasmodium Falciparum: New drug target for Plasmodium falciparum. *International Journal of Engineering Science and Technology*, 2 (8), 3468-3472.

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

Ojeiru F E, Kazuya T, Yuki H, Mohammed S. M and Shunsuke M, (2010). Circular Dichroism Studies on C-terminal Zinc Finger Domain of Transcription Factor GATA-2. *Yonago Acta medica*, 53, 25–28.

Prabhavathi M, Ashokkumar K, Geetha N, Saradha Devi K M, (2011). Homology modeling and structure prediction of thioredoxin (TRX) protein in wheat (Triticum aestivum L). *International Journal of Biosciences*, 1(1), 20-32.

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.

Rajeev K S, Reeta R S and Ashok P, (2005). Microbial cellulases- production, applications and challenges, *Scientific and IndustrialResearch*, 64, 832-844.

Sivakumar K, Balaji S and Gangaradhakrishnan, (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

Ulrik de L, Thomas S J, Lars J J and Soren B, (2003). Protein Feature Based Identification of Cell Cycle Regulated Proteins in Yeast. *Journal of Molecular Biology*, 329, 663-674.

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

Table 1: Parameters computed using ExPASy's ProtParam tool.

S.No	Organism name	Accession number	No. of amino acids	M. wt	pI	II (%)	AI	GRAVY
1	*Pseudomonas aeruginosa*	ACX31080.1	960	100058.7	5.86	44.70	71.10	-0.165
2	*Pseudomonas fluorescens*	ADK13057.1	748	80150.0	5.33	45.28	64.65	-0.395
3	*Pseudomonas Putida*	ABY99087.1	371	40927.4	5.63	41.82	88.03	-0.204
4	*Pseudomonas syringae*	EGH91691.1	390	42347.8	6.67	31.92	84.87	-0.237
5	*Pseudomonas savastanoi*	ZP_07003414.1	390	42383.9	6.61	31.36	84.62	-0.236

Table 2: Percentage of amino acids sequence forming secondary structure in SOPMA prediction.

S.NO	Organism name	Accession number	α-helix (Hh) (%)	β-turns (Tt) (%)	Extended strands (Ee) (%)	Random coils (%)
1	*Pseudomonas Aeruginosa*	ACX31080.1	21.04	5.10	20.52	53.33
2	*Pseudomonas Fluorescens*	ADK13057.1	18.72	6.82	17.91	56.55
3	*Pseudomonas Putida*	ABY99087.1	40.97	8.63	10.78	39.62
4	*Pseudomonas Syringae*	EGH91691.1	30.77	7.44	17.44	44.36
5	*Pseudomonas Savastanoi*	ZP_07003414.1	35.13	7.18	17.44	40.26

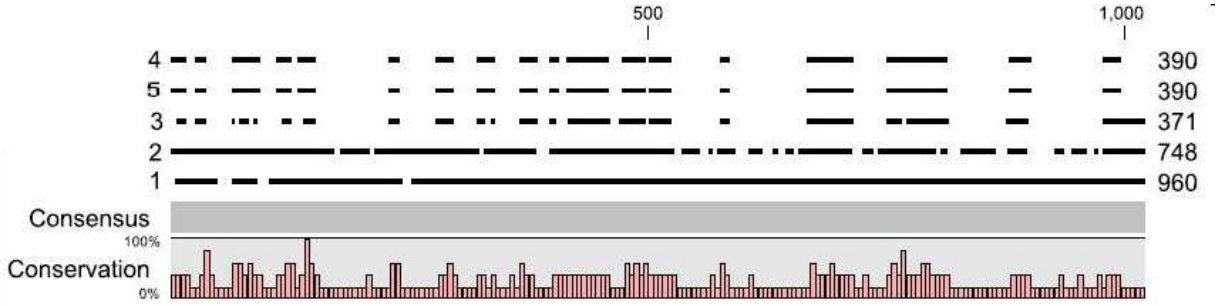

Figure 1: Results of multiple sequence analysis using CLC work bench (1- *Pseudomonas aeruginosa*, 2- *Pseudomonas fluorescens*, 3- *Pseudomonas putida*, 4- *Pseudomonas syringae*, 5- *Pseudomonas savastanoi*).

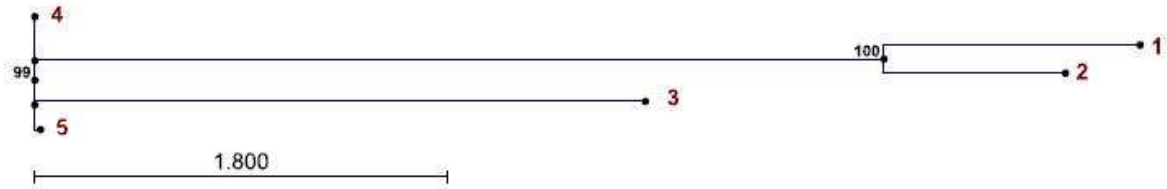

Figure 2: Results of phylogenetic analysis using CLC workbench (1- *Pseudomonas aeruginosa*, 2- *Pseudomonas fluorescens*, 3- *Pseudomonas putida*, 4- *Pseudomonas syringae*, 5- *Pseudomonas savastanoi*).

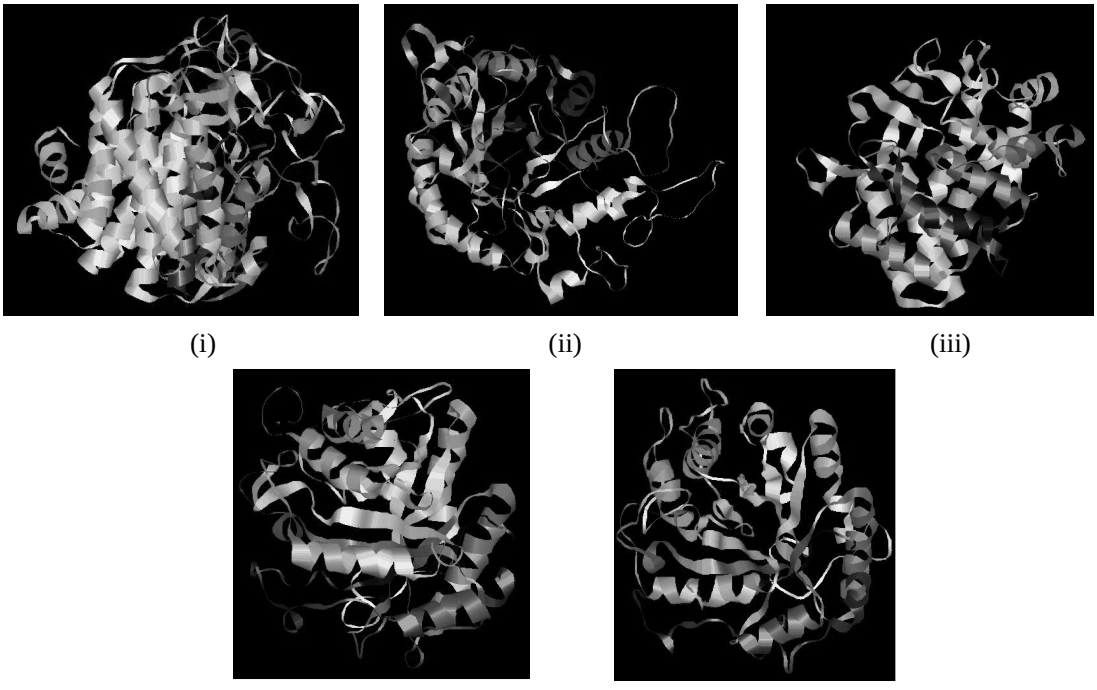

Figure 3: 3D structures of cellulases using ESyPred 3D (i- *Pseudomonas aeruginosa*, ii- *Pseudomonas fluorescens*, iii- *Pseudomonas putida*, iv- *Pseudomonas syringae*, v- *Pseudomonas savastanoi*).