

Original Research Article

In Silico Structural Analysis for Exo-Inulinases in Proteomes of *Streptomyces* sp. using PDB Structures as Templates

K. Mohan Reddy¹, Seelam Siva Deepthi², Jayachandra S. Y.¹, Parameshwar A. B.¹,
Dayanand Agsar², Bikshapathi E³ and M. B. Sulochana^{1*}

¹Department of Biotechnology, Gulbarga University, Kalaburagi-585106, India

²Department of Microbiology, Gulbarga University, Kalaburagi-585106, India

³Proteomics Lab, Centre for Cellular and Molecular Biology (CCMB), Hyderabad-500007, India

*Corresponding author

ABSTRACT

Exo-inulinases are industrial important enzymes that are used in preparations of fructose syrup, an important component of the biopharmaceuticals and food industries. On the other hand, thermo-stable exo-inulinases have great demand in both of these industries. In the present investigation, the structural studies of exo-inulinases of thermo tolerant *Streptomyces* sp. reported in the proteome databases, which has no reports in PDB database compare to fungal and bacterial origin. Through bioinformatics approach, template structure based sequence similarity and comparative modelling are the well-known methods for predicting the protein three dimensional structures. These structural integrity studies are achieved by the selection of query sequences having highest similarity (>40%) and high query coverage (>95%) using the reciprocal BLAST approach, template sequence is used as query retrieved from PDB database. Enzymes undergo many changes when they reacts with different metal salts during fermentation, also their structural integrity will be modified when they interact with the specific substrate, these are studied by finding the active site amino acids and their action using SCFBio (Supercomputing Facility for Bioinformatics and Computational Biology, IIT Delhi) and Dog Site Scorer, a pocket detection and analysis tool.

Keywords

Exo-inulinase,
In silico,
Streptomyces,
BLAST,
PDB, etc

Introduction

Extracellular inulinases (Exo-inulinases) are the class of glycosyl hydrolases that catalyses the polysaccharide inulin, based on the site of catalysis their nomenclature could

be decided by the enzyme commission. X-ray crystallographic structures of Exo-inulinases from the genus *Streptomyces* had least reports, but well reported by *Fungi*.

Inulinases has gained much attention in the Biopharmaceuticals involved in preparation of medical syrups as an ingredient fructose syrup from hydrolysed inulin. Also their importance was much crawled in food industries, using this fructose syrup as sweetener, preparation of sugar jellies, juices, health drinks etc., it has applications in bioenergy industries too [1, 2, 3].

Inulin contains a chains of linear β -2-1-linked D-Fructofuranose monomers terminated with one glucose residue at the reducing end. Bonding nature of this linear polymer contains the chain such as β -(2-1) linked polyfructose with terminal glucose which is again linked to fructose by α -(1-2) bond [4]. Long back, around 1970s the acid hydrolysis of inulin practiced was much famous, but with the advent of new enzyme hydrolysis got lead in production of pure fructose syrups in pharmaceuticals [5].

The scientific slogan such as “All enzymes are proteins but all proteins are not enzymes” depicts us all enzymes in the nature will be in proteinaceous form, thus the enzyme structures are studied through NMR spectroscopy, X-ray crystallography or Circular Dichroism. The advancement in the computational applications over biological data and their fundamental principles expressed through mathematical calculations having much attention in developing the protein three dimensional structures using Bioinformatics (*Insilico*) tools and servers. The present study involves in using developed programmes and softwares such as NCBI BLAST, protein data bank (PDB), EasyModeller 4.0 etc. for developing exo-inulinase structures of glycosyl hydrolase family-32 (Exo-inulinases) of *Streptomyces* (taxid:1883) with highest BLAST hit sequence identities to the template sequence retrieved from PDB database.

With the development of these Exo-inulinase 3D structures, it is possible to understand the structural and functional behaviours of protein sequences which are identified as hypothetical, putative, or uncharacterized. Modeller is a famous tool developed by Andrej Sali and Ben Webb [6, 7] used for the development of the protein 3D structures offline. EasyModeller 4.0 used in the present study, is a GUI (Graphical User Interface) version operated on the baselines of Andrej Sali’s Modeller. Based on the conditions and fundamental regulations required for developing the protein 3D structures are accordingly followed.

Materials and Method

Retrieval of Template Structure and Sequence of Exo-inulinase

Template structure 1Y9M [8] sequence of exo-inulinase of *Aspergillus awamori* (Hydrolase family) was retrieved from PDB (protein data bank) database to explore it among the proteomes of *Streptomyces* sp. using NCBI-BLAST programme [9].

pBLAST Analysis for Exo-inulinases in *Streptomyces* sp.

The retrieved query sequence of *Aspergillus awamori* exo-inulinase from PDB, is used for scouting the entire proteome sets of *Streptomyces* (taxid: 1883) as a selective organism against the non-redundant protein sequence database accessed through pBLAST programme, this can be termed as reciprocal BLAST analysis.

Selection of *Streptomyces* Exo-inulinase Sequences

Highly similar or most conserved exo-inulinase protein sequences were selected

from the pBLAST hits of *Streptomyces* sp. The conservity in between target exo-inulinase sequence and resultant sequences is dependent upon the percentage of query coverage. Later with highest sequence identity and query coverage sequences are used for comparative modelling using EasyModeller 4.0 [10].

Comparative Modelling

EasyModeller is a graphical user interface software which works under the python script launcher available for all operating systems. Python 2.4.7 software is used as a script launcher for EasyModeller in our 32bit based Windows 8.1 operating system. An academic Modeller licence key is used for the external internet based automated Modelling servers such as Swiss Modeller, I-Tasser etc. for structural comparison studies with the developed offline structures. Three sequences from BLAST hit of template sequence as query got the 98% query coverage and 52% highest sequence similarity were retrieved and subjected for homology modelling.

Exo-inulinase 3D Structure Evaluation

Structure evaluation of modelled proteins were analysed through PDBsum generate [11] server for predicting the model quality with the results of Ramachandran plot and G-score. In this method the modelled PDB file (.pdb format) could be uploaded to the PDBsum generate server including with an email ID for receiving the generated results. The inbox of this email contains entry ID and password, which could be operated on new window to retrieve the results.

Active Site Amino Acids Identification

Active sites are the functional sites that processes the substrates or molecules that

bind to it. Catalytic sites are different from active sites, sometime active sites may also gains the catalytic property. The active cite prediction is performed by SCFBio [12] server provided by IIT, Delhi. The detected cavities list and their sequences are provided in Table. 1.

Structural and Functional Domains Analysis for Selected Sequences

The structural domain analysis of 3 selected Glycosyl hydrolase family 32 (Exo-inulinases) were analysed by PyMOL Viwer [13] and supported by TmHMM [14]. Functional domain analysis is performed by InterPro [15], a protein sequence analysis and classification tool for three selected exo-inulinase sequences of *Streptomyces* sp.

Structural Analog Binding Properties

Analogs are the similar structures with different properties which may bind for one or more reason during the biological metabolic processes. The binding properties were studied by COFACTOR programme provided by ZhangLab [16].

Results and Discussion

Retrieval and pBLAST Analysis of Template Structure Sequence

The basic steps of molecular modelling techniques for unknown query sequences, will be started by searching and selection of the template structures. In the present study, template structure sequence (Fig. 1) were first retrieved from PDB database and used for the reciprocal BLAST analysis for identifying the number of Exo-inulinase hits among *Streptomyces* sp. Three Glycosyl hydrolase family 32 sequences (Fig. 2) of different *Streptomyces* species were retrieved by pBLAST analysis based on the

highest query coverage and sequence similarities. Among 47 sequences only three sequences got 99% query coverage and 52% sequence identity which are taken for further analysis.

For structural studies the query sequences should match with high query coverage and identity might be consider according to the standard percentage (>40%).

Comparative Modelling

Homology modelling (Comparative modelling) for 3 query sequences constructed in Easy Modeller and Swiss Modeller are similar, where both will operate on the same command line developed by Andrej Sali Lab. The constructed 3D protein structure and their template structures are displayed in Fig. 3. The structures modelled has less query coverage and only single chain templates are available in PDB database, thus the structures developed need still more efficient templates.

Structure Evaluation

Every three dimensional structure has to undergo evaluation process to check the stereo chemical quality, which is resulted by Ramachandran plot for the most favoured regions of amino acids in respective three blocks viz. A, B & L. The percentage of most favoured regions of all three models were average at 80%, one them *Streptomyces scabiei* sequence got 83% as shown in Fig. 4. According to Ramachandran plot analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0 and a good quality model would be expected to have over 90% in the most favoured regions [A, B, L]. Thus, there are no suitable NMR/XRC templates in PDB database, a wide scope is available for experimental

structure of *Streptomyces* sp.

Identification of Active sites

Active sites were predicted for the query structure of *Streptomyces scabiei* by SCFBio programme of IIT Delhi. 29 active sites were obtained and results were tabulated in Table 1, also the position of each active site was expressed in the pdb structure by red spheres as shown in the Fig. 5.

Structural and Functional Domain Analysis

Structural domains were analysed by PyMOL are shown in the modelled exo-inulinase protein of *Streptomyces scabiei* which can be observed in structures of modelled proteins or template structures (Fig. 2 or 4) and the domain locations were analysed by TmHMM server, only two proteins have the transmembrane domain outside as shown in the Fig. 6. Using InterPro programme server the functional domains were analysed and each domain expressed with function as shown in the Fig. 7.

Structural Analog Studies

Five structural analogs were found for the modelled structure of *Streptomyces scabiei*, a single analog binding structure for the modelled structure is shown in the Fig. 8. The list of enzyme structural analogs were tabulated in Table 2.

Glycosyl hydrolase family 32 (EC 3.2.1.80) are the exo-inulinases that hydrolyse polyfructan carbohydrates. *Streptomyces* sp. has less reports in production of exo-inulinases, also the structural studies like NMR, X-Ray Crystallography (XRC) are not available in PDB database. Thus *In silico* prediction of three dimensional structures were conducted in this study. As

modelling of protein 3D structures require perfect matching template structures in PDB database, only three exo-inulinase N-terminal domain structures of *Aspergillus awamori* are available. Based on the

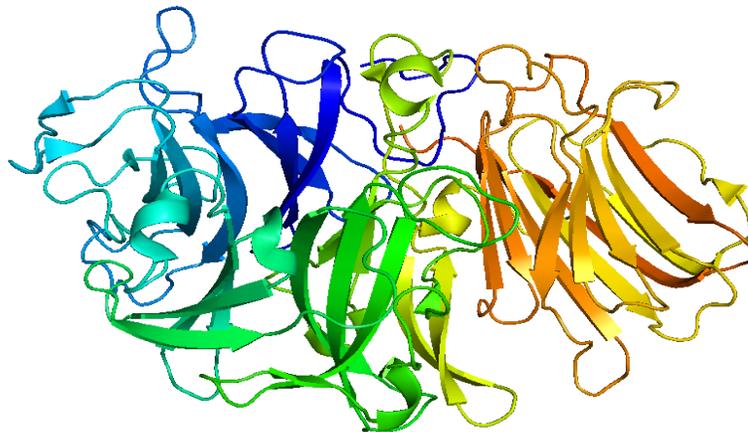
template search in Swiss Modeller server, there are no structures matching with high query coverage and perfect sequence similarity, but only 1Y9G, 1Y9M class of structures shown highest.

Table 1 Active Site Cavity Sequences Found in Modelled Query 1 Structure

Cavities	
cavity_1_NTQKPILDEVRAIWGFSF	cavity_2_GLVTDMSIKYREPANWQHFR
cavity_3_DTLAGSKEYRFPNVIQH	cavity_4_IVTPDELSMQGKAWHFR
cavity_5_TKNQRIPEDYSGLWVAM	cavity_6_LTDMSYREKAPANVIWGGQ
cavity_7_KNGSLRIVEFYTHAMQPD	cavity_8_SAGMKRPEYDVILNWQ
cavity_9_EGRKSATPFYVLMIDH	cavity_10_MILAQETSRPDVGHKW
cavity_11_TDEPRYAFQVGIHKS	cavity_12_TFVNLHMISAGQREKPD
cavity_13_GNQTIPREDAWYSLVM	cavity_14_ADWTYQVGMKPRSNE
cavity_15_GNKQPTRILEVDAYWSM	cavity_16_DHPETFRVNSGIQAKW
cavity_17_RPEDAYWLVTGQSMIKN	cavity_18_EFYVDQLGTHKPRSNIA
cavity_19_KDTGSALRFVFNHMIQ	cavity_20_WTDGMYKNAEVPISQR
cavity_21_LVKDHEGTRIASPFW	cavity_22_KDLTRAGSVYEFPNMI
cavity_23_TFLHMVSAIQKREPDG	cavity_24_NKDLQTPGAISVEYRF
cavity_25_KGDLNTSAVYFHMQR	cavity_26_LKVTFNHMSAIQREP
cavity_27_IDKRVTYNFWESMQAP	cavity_28_ATYQVGDSEIKRE
cavity_29_TNSFDPWVQRI	

Fig.1 Template Structure and Sequence of 1Y9M used for the Reciprocal pBLAST Analysis

For Educational Use Only



Template Sequence

```
FNYDQPYRGGYHFSQKRWMDPNGLLYHNGTYHLFFQYNFSGGIEWGNISWGHAISEDLTHWEEKPVALLARGFGSDVTEMYFSGS
AVADVNTSGFGRDGTFLVAMYSYYPVQTLFSGQTVQEDPQQSQSIAYSLLDGLTWTTYDAANFVFNPFSPVYAEYQNRDFF
VFWHDESQRWVVTSLAELKLAITYSDMLDKLVSEFQFNAGQGVWECPLVLELSDNSSTKWLITSGLNPGFPSTVGSST
QIFVGFEDTFFPDADTVYFGNSRANWMDNGDFFYAAAGYNSLNDHVIHGWNNWQYGANIPYFRSAMAIPRHMALTKTIGS
KATLVQQPQEAWSISNRRPIYSRTFKTLSEGSTNTTTTGETFKVDLSFSAKSKASTFAIALRASANFTQTLVGYDFARQQIFLD
RTHSGDVSFDETFASVYHGRLFPDSTGVVKLISIFVDRSSVEVFGGQETTLLTAQIFPSSDAVHARLASTGGTTEDVDRADIYKIAST
WN
```


Fig.4 Ramachandran Plots for the Evaluation of Quality of the Modelled Proteins

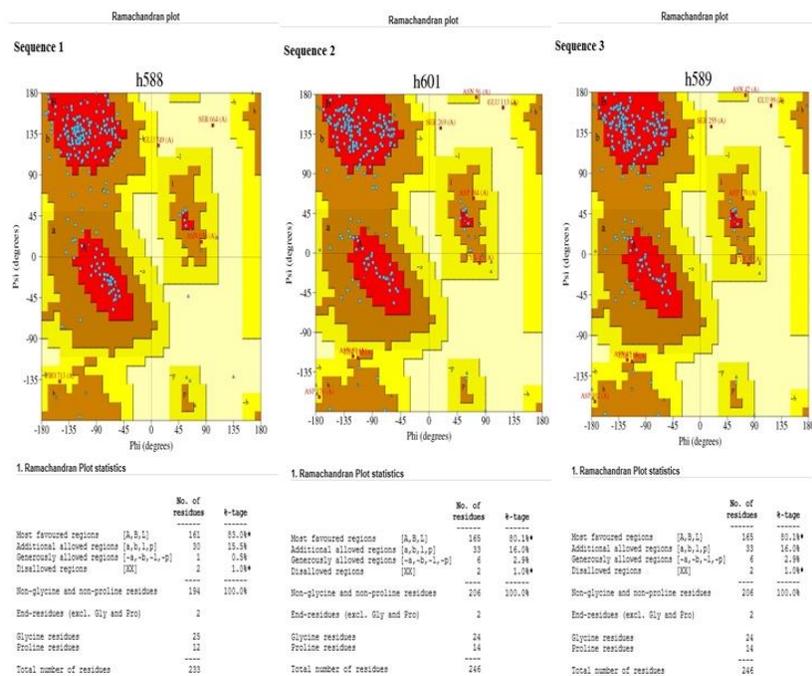


Fig.6 Active Site Cavities Position in Modelled Query 1 Structure

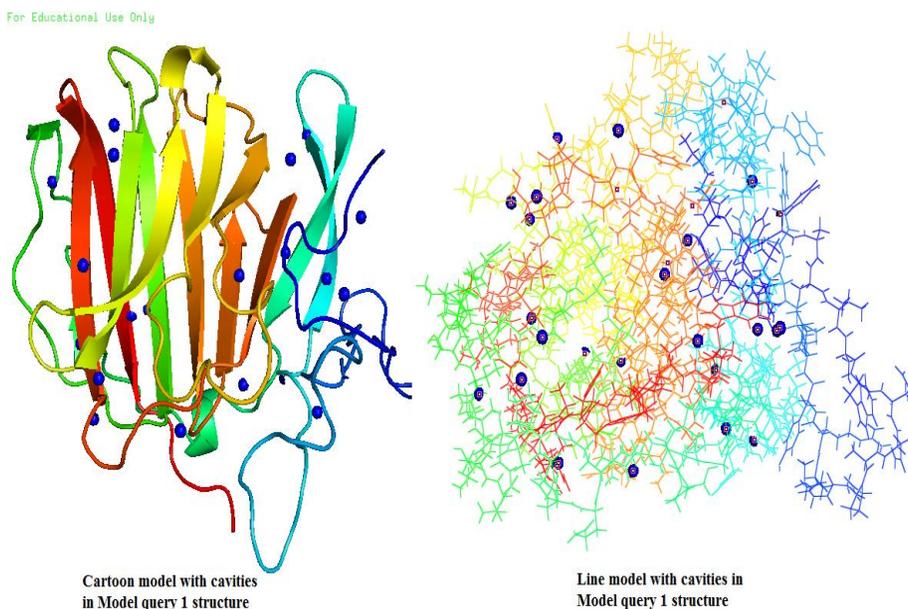


Fig.7 Transmembrane Domain Locations in the Query Sequences by TmHMM Server

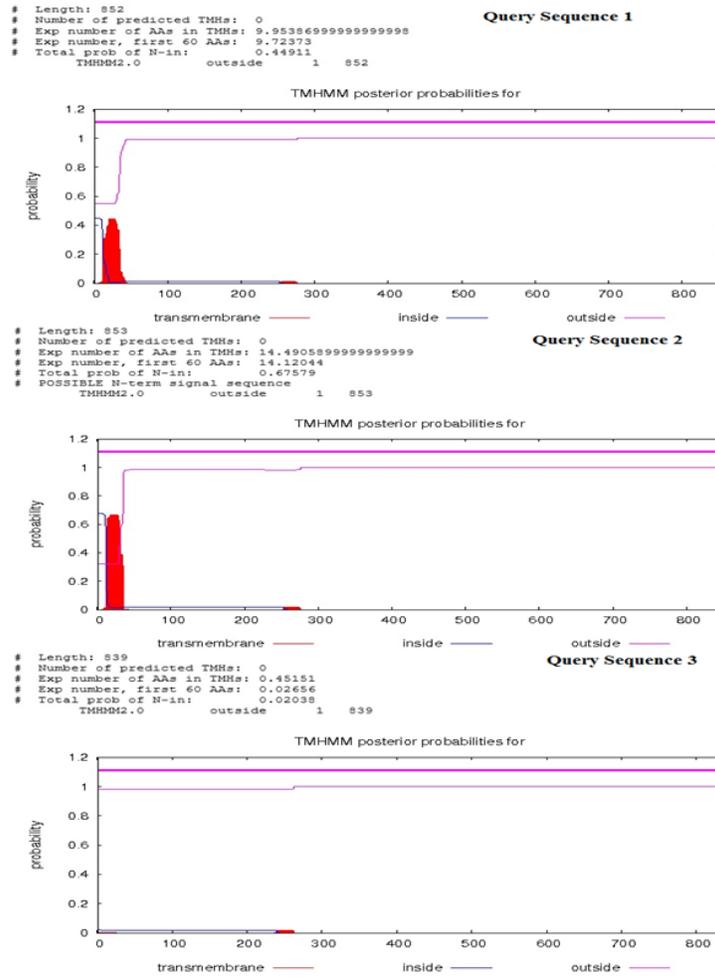


Fig.8 Functional Domain Analysis of Three Retrieved Query Sequences by InterPro





The structure developed in the present study, has low percentage in the most favoured regions of Ramachandran plot (83%), instead of the standard >90%. Thus there is need of good template structure which matches the sequences of *Streptomyces* sp. Because of much thermo-stable, non-pathogenic, most resistive microorganism only belongs to the class actinobacteria which are of high demand in the industrial sector.

The identification of the active sites and functional domain studies provides the insight knowledge of actions of

biomolecules in and outside of the microbial environments.

In conclusion, Industrial exploitation of exo-inulinases from *Streptomyces* sp. was not developed in the recent years, also there are very few reports from this genus. *Aspergillus* sp. has much exploitation in production of these enzymes and three dimeric structures are also reported from this genus, thus we took an opportunity in building 3D models of Glycosyl hydrolase family 32 and sequence analysis of *Streptomyces* sp. There is no availability of compatible template structures in PDB for *in*

silico protein modelling, in turn our experimental enzyme was subjected for NMR and XRC (results are under progress) to get deposited in the PDB to reach this compatibility.

Reference

1. Negro MJ, Ballesteros I, Manzanares P, Oliva JM, Saez F, Ballesteros M (2006). Inulin-containing biomass for ethanol production carbohydrate extraction and ethanol fermentation. *Appl Biochem Biotechnol.* 132:922–932.
2. Singh RS, Dhaliwal R, Puri M (2007). Production of high fructose syrup from Asparagus inulin using immobilized exoinulinase from *Kluyveromyces marxianus* YS-1. *J Ind Microbiol Biotechnol.* 34:649–655.
3. Kango N, Jain SC (2011) Production and properties of microbial inulinases: recent advances. *Food Biotechnol.* 25:165–212.
4. Hyun-Ju Kwon, Sung-Jong Jeon, Dong-Ju You, *et al.*, (2003). Cloning and characterization of an exoinulinase from *Bacillus polymyxa*. *Biotechnology Letters.* 25: 155–159.
5. Vandamme EJ, Derycke DG (1983). Microbial inulinases: fermentation process, properties and applications. *Adv. Appl. Microbiol.* 29: 139–176.
6. A. Sali & T.L. Blundell (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.
7. B. Webb, A. Sali (2014). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., 5.6.1-5.6.32.
8. Nagem RA, Rojas AL, Golubev AM, Korneeva OS, Eneyskaya EV, Kulminkaya AA, Neustroev KN, Polikarpov I. Crystal structure of exoinulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition (2004) *J.Mol.Biol.* 344: 471-480
9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
10. Bhusan K Kuntal, Polamarasetty Aparoy, Pallu Reddanna, (2010). EasyModeller: A graphical interface to MODELLER. *BMC Res Notes.* 3: 226.
11. de Beer T A P, Berka K, Thornton J M, Laskowski R A, (2014). PDBsum additions. *Nucleic Acids Res.*, 42, D292-D296.
12. Tanya Singh, D. Biswas, and B. Jayaram, (2011). AADS - An automated active site identification, docking and scoring protocol for protein targets based on physico-chemical descriptors, *J. Chem. Inf. Modeling*, 51, 2515-2527.
13. W.L. DeLano, DeLano Scientific, San Carlos, California CCP4 Newsletter on Protein Crystallography (2002).
14. Krogh A1, Larsson B, von Heijne G, Sonnhammer EL, (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305(3): 567-80.
15. Alex Mitchell1, Hsin-Yu Chang1, Louise Daugherty *et. al.*, (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, 43, D213–D221.
16. Amrishi Roy, Jianyi Yang and Yang Zhang, (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research*, 1–7.