

Isolation, Sequencing *and in silico* Comparative Modeling of *Zea Mays* β -Glucosidase Isolated from Different Parts of Plant

Er. Santosh Kumar, Dr. O. P. Verma

Assistant Professor, Department of Life Sciences, ITM University, Gwalior, M.P, India

Associate Professor, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom University of Agriculture,
Technology and Sciences, Allahabad, UP, India

ABSTRACT: Maize (*Zea mays* L) is cultivated on nearly 150 m ha in about 160 countries having wider diversity of soil, climate, biodiversity and management practices that contributes 36 % (782 million tonnes) in the global grain production. β -Glucosidase (β -D-glucoside glucohydrolases; EC 3.2.1.21) occur ubiquitously in plants, fungi, mammals and microorganisms. It constitutes a major group in glucoside hydrolase families and catalyses the hydrolysis of aryl and alkyl β -D-glucosides such as cyanogenic glucosides, thioglucosides, glucose disaccharides or phenolic glycosides. In plants β -glucosidases have been implicated in regulating various aspects of development, e.g., phytohormone activation, pathogen defense reactions, cell wall degradation in the endosperm during germination and lignifications. β -Glucosidase was isolated from radicle, leaves and first leaf blade of *ze mays* using SDS-PAGE. Each protein sample was sequenced using Edman's Degradation method. Comparative modeling was performed to generate good quality models of each sequence using Geno3D. The assessment of the generated three dimensional structure against structure verification tools PROCHECK showed that model generated by Geno3D was acceptable. The predicted model can be used in future studies.

KEYWORDS: Maize, β -Glucosidase, SDS-PAGE, Comparative modeling, Geno3D

I. INTRODUCTION

Maize (*Zea mays* L) is one of the most versatile crops which have adaptability for various agro-chemical conditions. Maize is known as queen of cereals because it has the highest genetic yield potential among the cereals. It is cultivated on nearly 150 m ha in about 160 countries having wider diversity of soil, climate, biodiversity and management practices that contributes 36 % (782 million tonnes) in the global grain production.

β -Glucosidase (β -D-glucoside glucohydrolases; EC 3.2.1.21) occur ubiquitously in plants, fungi, mammals and microorganisms. It constitutes a major group in glucoside hydrolase families and catalyses the hydrolysis of aryl and alkyl β -D-glucosides such as cyanogenic glucosides, thioglucosides, glucose disaccharides or phenolic glycosides. In plants β -glucosidases have been implicated in regulating various aspects of development, e.g., phytohormone activation, pathogen defense reactions, cell wall degradation in the endosperm during germination and lignifications. The current work has been carried out with the aim to isolate, sequence and predict the tertiary structure of the sequenced enzyme β -Glucosidase. Radicle, leaves and first leaf blade has been selected for study.

II. MATERIALS AND METHODS

A. Isolation and sequencing of β -Glucosidase:

The Maize plant parts were grinded to fine powder in a chilled mortar and pestle using liquid nitrogen and extraction buffer (1 ml buffer/ 1 g tissue) was mixed to it. This was centrifuged at 12000 rpm / 4⁰C / 30 min. 50 μ g prepared sample was loaded into the wells for SDS-PAGE. First marker was loaded in the first lane to determine the size of desired protein. The gel was run at 60V till the dye reaches the end of the gel. Then the gel was stained with staining dye. After staining the gel was taken apart from the electrophoretic unit for destaining. De-staining was done until all the extra dye was removed from the gel except the dye bound to protein in the form of bands. Applied PVDF



sequencing membrane to gel size and wet with 10% methanol. Soaked the gel and membrane in CAPS (3[cyclohexylamino] 1-propane sulphonic acid) buffer. Place the gel on the cathode and the membrane on the anode. Run at 250mA contact current for 45 min (Matsudaira; P. 1987). Stain blotted with Coomassie Blue. A small piece of PVDF was cut and wet thoroughly with methanol followed by a water rinse. Wet blot paper with water and place membrane on top of blot paper. Pipette a small volume (~2µl) of concentrated pure sample onto membrane. With a dry blot paper, hold against sides of membrane (as absorbent material) until the protein disappears. Place the membrane in a container and cover with a small amount of Coomassie R-250 stain for blots for 60 seconds. Discard stain and destain blot for 2-5 min. with destain changes, until spot appears and background is light. Rinse with water and dry membrane. Excise spot with a razor blade and load onto sequencer for sequencing.

B. Insilco Comparative Modeling:

Comparative modeling usually starts by searching the database of known protein structures (Protein Data Bank PDB) (Abola, *et al*; 1987) using target sequence as the query. This is generally done by comparing the target sequence with the sequence of each of the structures in the database. Once potential templates has been obtained using PSI-BLAST the templates were selected that were appropriate for the particular modeling problem. The higher the overall sequence similarity between the target and the template sequences the better the template is likely to be. Other factors should also be taken into account when selecting a template:

1. The family of proteins that includes the target and the templates frequently can be organized in subfamilies. The construction of a multiple alignment and a phylogenetic tree (Felsenstein J., 1985) can help in the template from the subfamily that is closest to the target.
2. The similarity between the environment of the template and the environment in which the target needs to be modelled should also be considered.
3. The quality of the experimental template structure is another important factor in template selection. The resolution and R factor of a crystallographic structure and the number of restraints per residue for a nuclear magnetic resonance (NMR) structure are indicative of the accuracy of the structure.

A specialized method should be used to align template with the target sequence. The alignment is relatively simple to obtain when the target-template sequence identity is above 40%. An accurate alignment can be obtained automatically using standard sequence-sequence alignment methods. A variety of methods can be used to construct a 3D model for the target protein. The original and still most widely used method is modeling by rigid body assembly (Blundell, Sibanda *et al*; 1987, Browne, North *et al*; 1969, and Greer J., 1981). This method constructs the model from a few core regions and from loops and side chains that are obtained from dissecting related structures. Geno3D software was used to generate the 3D model of the protein

After generation of the model two types of evaluation should be carried out: (1) Internal evaluation of self consistency that checks whether or not the model satisfies the restraints used to calculate it and (2) 'External' evaluation that relies on information that was not used in calculating the model (Luthy and Bowie, *et al*; 1992, Sippl. M. J., 1993). PROCHECK (Laskowski and McArthur *et al*; 1993) was used for finding unreliable regions of a model to evaluate the stereochemistry of the model.

III. RESULTS & DISCUSSION

The three sequences obtained after sequencing using Edmann's degradation method for the samples of radicle, leaves and first leaf blade are as mentioned below in FASTA format.

1. >beta-D-glucosidase [*Zea mays*] from Radicle (538AA)

```
NHATHSVLRSHLGPESFSRHHLPQSSKRRFNLSFTPRSARVGNENGVQLLSPSEIPRRDWFPDFIFGTSAYQIE
GAWNEDGKGESNWDHFCHNFPERIMDGSNADIGANSYHMYKTDVRLKEMGMDAYRFSISWPRLPKGTVE
INQDGDIDYKRLINENGIPEYVTFIFWDPVQALKYFLDKTHKRIVNDYKNFAKVCDFDNFGDKVKNWLTNFP
QTFTSFSYGTGVFAPGRCSPLDCAIPTGNSLVEPYIAGHNILLAHAEAVDLYNKYYKGENGRIGLAFDVMGR
VPYGTSLFDEQAKERSMDINLGFLEPVVRGDYPFSMRSLARERLPFFSDKQKELVGSYNMLGINYYTSIFS
KHIDISPKYSPVLNTDDAYASQGTYPGDPGKPIGPPMGNPWYLYPEGLKDILMIMKNKYGNPPIYITENGIGDV
DTKEKPLPMEAAALNDYKRLDYIQRHISTLKESIDLGANVHGYFAWSLLDNFEWYAGYTERYGIAYVDRKNSY
TRYMKESAKWLKEFNATAKKPSKKIITPA
```

2. >beta-D-glucosidase [*Zea mays*] from Leaves (543AA)

```
MATHSVLRSHLGPNNESFSRHHLPQSSKRRFNLSFTPRSARVGNENGVQLLSPSEIPRRDWFPDFIFGTSAYQI
EGAWNEDGKGESNWDHFCHNFPERIMDGSNADIGANSYHMYKTDVRLKEMGMDAYRFSISWPRLPKGTVE
```

GGINQDGIDYKRLINLENGIEPYVTIFHWDVPQALEEKYGGFLDKTHKRIVNDYKNFAKVCFDNFGDKVKNW
LTFNEPQTFTSFSYGTGVFAPGRCSPLDCAIPTGNSLVEPYIAGHNILLAHAEAVDLYNKYKGENGRIGLAF
DVMGRVPYGTSTFLDEQAKERSMDINLWGFLEPVVRGDYPFMSRSLARERLPFFSDKQEQELVGSYNMLGINY
YTSIFSCHKIDISPKYSPVLNTDDAYASQGTYPDGKPIGPPMGNPWYLYPEGLKDILMIMKNKYGNPPIYITEN
GIGDVDTKKPLPMEALNDYKRLDYIQRHISTLKESIDLGANVHGYPFAWSLLDNFEWYAGYTERYGIAYVD
RKNSYTRYMKESAKWLKEFNATKPSKKIITPA

3. >beta-D-glucosidase [*Zea mays*] from First Leaf Blade (533AA)

NHTHSVLRSHLGPESFSRLPQSSKRRFNLSFTPR SARVGNENGVQLLSPSEIPRRDWFPSDFIFGTSAYQIGAWN
EDGKGESNWDHFCNHFPERIMDGSNADIGANSYHMYKTDVRLKEMGMDAYRFSISWPRILPKGTVEINQDG
IDYYKRLINLENGIEPYVTIFHWDVPQALKYFLDKHKRIVNDYKNFAKVCFDNFGDKVKNWLTFTNEPQTFTSFS
YGTGVFAPGRCSPLDCAIPTGNSLVEPYIAGHNILLAHAEAVDLYNKYKGENGRIGLAFDVMGRVPYGTSTF
LDEQAKERSMDINLWGFLEPVVRGDYPFMSRSLARERLPFFSDKQEQELVGSYNMLGINY YTSIFSCHKIDISPK
YSPVLNTDDAYASQGTYPDGKPIGPPMGNPWYLYPEGLKDILMIMKNKYGNPPIYITEN GIGDVDTKKPLP
MEALNDYKRLDYIQRHISTLKESIDLGANVHGYPFAWSLLDNFEWYAGYTERYGIAYVDRKNSYTRYMKES
AKWLKEFNATKPSKKIITPA

Comparative modeling of beta-D-glucosidase

The *in silico* comparative modeling of the three protein sequences isolated from coleoptile, radicle and roots using Geno3D software were the most stable structures (Fig. 1- Fig. 3) with lower energy levels and highest identity according to PSI-BLAST software.

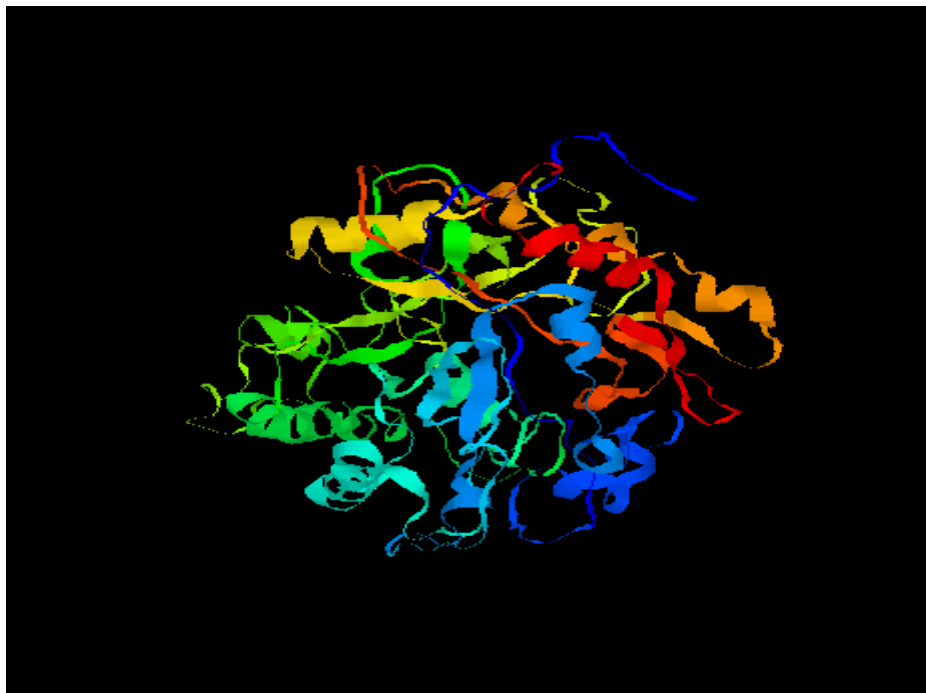


Fig. 1: Structure of beta-D-glucosidase isolated from radicle generated by Geno3D

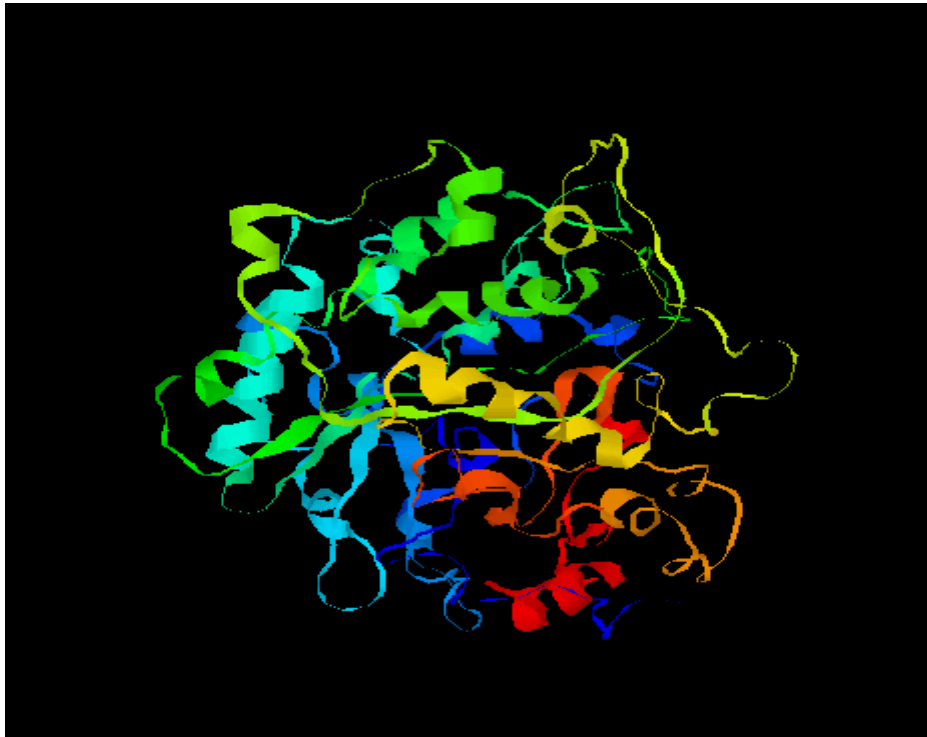


Fig. 2: Structure of beta-D-glucosidase isolated from leaves generated by Geno3D

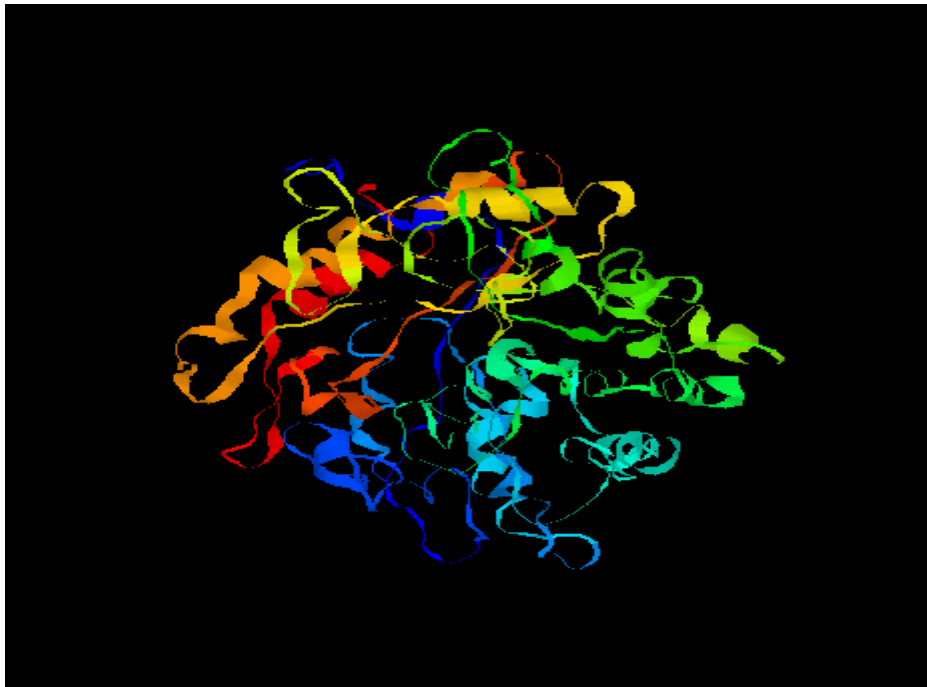


Fig. 3: Structure of beta-D-glucosidase isolated from first leaf blade generated by Geno3D

Ramchandran Plot

S. No.	Sample	Core (%)	Allowed (%)	Generously Allowed (%)	Disallowed (%)
1.	Radicle	59.7	28.8	7.0	4.6
2.	Leaves	60.8	29.8	4.7	4.7
3.	First leaf blade	59.9	26.6	8.1	5.4

Table 1 Ramchandran plot calculations with PROCHEK program

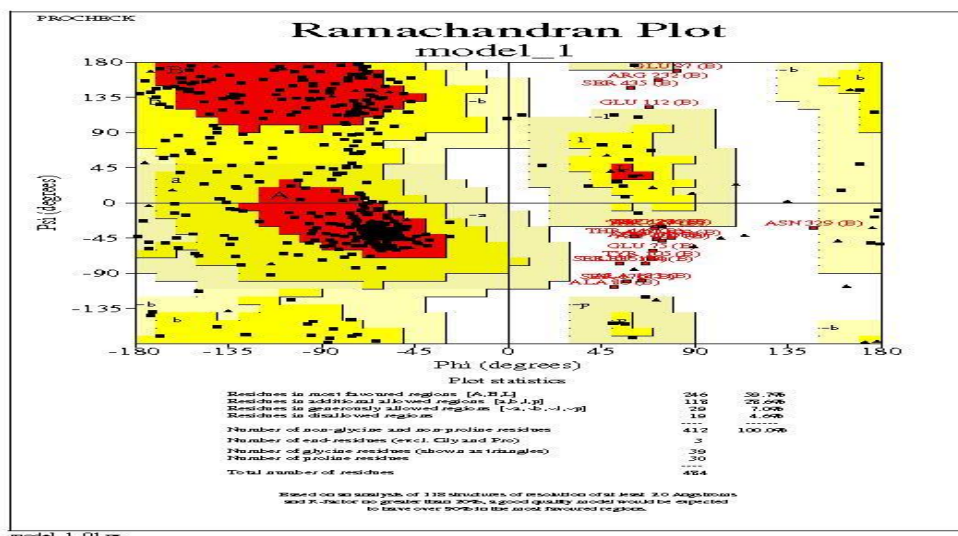


Fig. 4. Ramchandran plot for beta glucosidase isolated from radicle

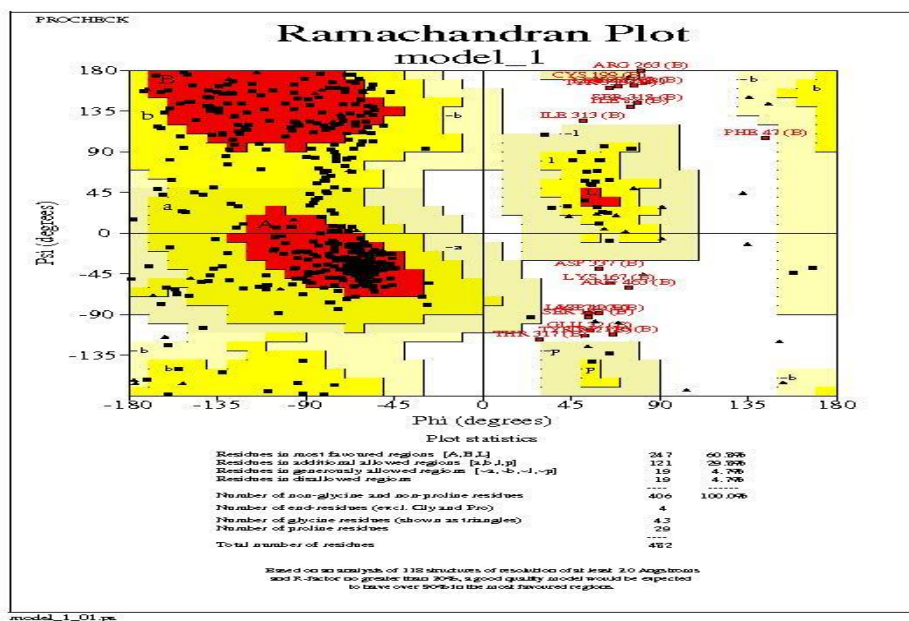


Fig. 5: Ramchandran plot for beta glucosidase isolated from leaves

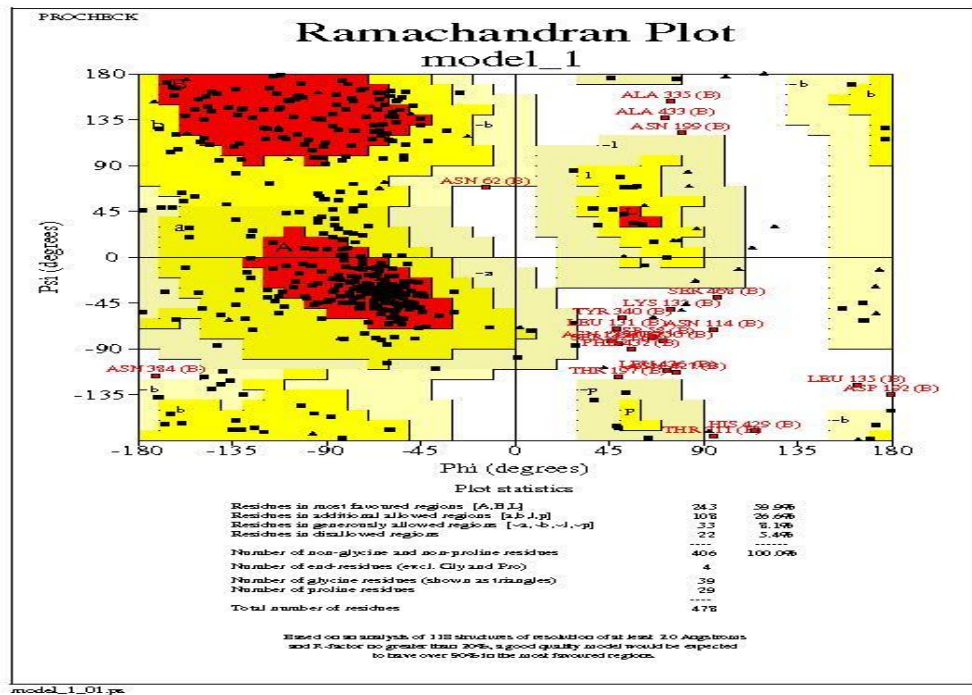


Fig. 6: Ramchandran plot for beta glucosidase isolated from first leaf blade

S. No.	Sample	Identity (%)	Energy (Kcal/mol)	RMSD
1.	Radicle	88.6	696.53	1.60
2.	Leaves	86.3	10706.50	1.55
3.	First leaf blade	88.4	707.18	1.75

Table 2. Percentage identity, energy and root mean square deviation (RMSD) of the models

The three dimensional structures were generated using Geno3D software and validated by PROCHECK software with the Ramchandran Plots for each protein. In this study the model with the best identity (Radicle 88.6%, Leaves 86.3% and First leaf blade 88.4%), lowest energy in Kcal/mol (Radicle 696.53, Leaves 10706.5 and First leaf blade 707.18 Kcal/mol) and root mean square deviation (RMSD) (Radicle 1.60, Leaves 1.55 and First leaf blade 1.75) shown in table 2 were considered as the best optimized and superimposed model (Sateesh, et al, 2010). The Ramchandran plot data in the table 1 shows that nearly 60% of the residues for each protein, Radicle 59.7%, Leaves 60.8% and First leaf blade 59.9% lies in the core region. The maximum percentage of the residues lying in the core region represents the best stereo chemical quality of the protein structure (Morris et al, 1992) which is further supported by its nearly 29% residues lying in the allowed region, Radicle 28.8%, Leaves 29.8% and First leaf blade 26.6% and generously allowed regions, Radicle 7.0%, Leaves 4.7% and First leaf blade 8.1% Only 4.6 to 5.4 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best stereochemical quality and stability. *In silico* Comparative modeling (Kumar, et al., 2013) of the β-glucosidase sequences isolated from different parts of the zea mays justifies the stability of the structures generated through Geno3D software.



IV. CONCLUSION

The SDS PAGE was developed and analyzed for the protein bands visualized and their molecular weight was determined in comparison to standard molecular markers. Molecular weights of different bands were found in range 33 KDa to 66 KDa.

The β -glucosidase isolated from different parts of *zea mays* was sequenced using Edmann's degradation method. The isolated protein had different amino acid lengths and each was observed to be radicle -538 amino acids, leaves -543 amino acids and first leaf blade 533 amino acids.

The three dimensional structures generated using Geno3D software and validated by PROCHECK software showed the best identity above 88%, lowest energy in Kcal/mol and root mean square deviation (RMSD) were considered as the best optimized model. The Ramchandran plot data showed that nearly 60% of the residues for each protein lies in the core region. The maximum percentage of the residues lying in the core region represents the best stereo chemical quality of the protein structure which is further supported by its nearly 33% residues lying in the allowed region, and generously allowed regions. Only 4.6 to 5.4 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best stereochemical quality and stability.

V. ACKNOWLEDGEMENTS

We would like to acknowledge Sam Higginbottom University of Agriculture, Technology and Science, Allahabad which supported us to carry out our research work. We would also like to acknowledge ITM University, Gwalior for its timely support and also the groups who supported us directly or indirectly in making this research a success. We would also thank the softwares without whom this work would have not been completed.

REFERENCES

1. Matsudaira; P. Journal of Biol. Chem. (1987) 262, 10035-10038
2. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T and Wang, J. (1987). Protein Databank, in Crystallographic Databases-Information, contents, software systems, Scientific Applications. Data Commission of the International Union of Crystallography, Cambridge. 107-132
3. Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39: 783-791.
4. Blundell, T. L., Sibanda, B. L., Sternberg, M.J. E. and Thornton, J. M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. Nature. 326: 347-352.
5. Browne, W. J., North, A. C. T., Philips, D. C., Brew, K., Vanaman, T. C. and Hill, R. C. (1969). A possible three dimensional structure of bovine - lactalbumin based on that of hen's egg white lysozyme. Journal of Molecular Biology. 42: 65-86.
6. Greer, J. (1981). Comparative model-building of the mammalian serine proteases. Journal of Molecular Biology. 153: 1027-1042.
7. Luthy, R., Bowie, J. U. and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature. 356: 83-85.
8. Sippl, M. J., (1993). Recognition of errors in three-dimensional structures of proteins. Proteins. 17: 355-362.
9. Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 26: 283-291.
10. Patchikolla Sateesh, Allam Appa Rao, Suresh Kumar Sangeeta, . M. Naresh Babu, R. S.Datta, Teja Gandhi (2010). Homology modelling and sequence analysis of anxC3.1. International Journal of Engineering, Science and Technology. 2(5): 1125-1130.
11. Morris A L, MacArthur M.W., Hutchinson E. G. and Thornton J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins 12: 345-364.
12. Santosh Kumar*, Sachin Gupta, Sarika Saxena, Rajkumar Yadav, Ankur Thakur, O. P. Verma, Chandrabhan Seniya (2013). Insilico analysis and 3D structure prediction of matrix protein of nipah virus. Journal of Advanced Bioinformatics Applications and Research, 04 (01): 391-395