Molecular Modelling and Docking of Tomato *Beta Galactosidase enzyme-7* (A2JGX1)

Balakrishnan M.*, Soam S.K., Sumalatha K. and Ch. Srinivasa Rao
ICAR-National Academy of Agricultural Research Management, Hyderabad, Telangana, INDIA
*balakrishnan@naarm.org.in

Abstract
Beta galactosidase (TBG-7) is a fruit softening related enzyme and eliminates terminal non reducing beta residues in β-gals by hydrolysis. Beta-(1,4)-galactans play a vital role in cell wall growth and inhibition during ripening. Beta-galactosidases (βgals) (EC 3.2.1.23) have been found in different number of plant species. Tomato β-galactosidase 7 (A2JGX1) is responsible for fruit softening through the degradation of β-(1,4)-galactans. The present study represents prediction of secondary structure of the protein using PSIPRED, most of the structure occupied by coils. The 3D structure of protein prediction is through Homology modelling and Docking (Auto dock 4.2) studies. The predicted protein model validated and evaluated through Ramachandran plot PDBSUM and RAMPAGE analysis revealed that 91.9% and 93.3% were in favoured region providing support for quality and accuracy respectively.

The ligands CME and BMA are used for in silico docking reaction with the A2JGX1(TBG-7) enzyme and successfully docked in the binding sites of TBG-7 of tomato. Once the binding sites are predicted then Autodock4 performed to docking the TBG-7. The binding site pocket ID- 1 VAL[73], PRO[74], ASN[99], GLY[100], PRO[103], SER[104], TYR[109], GLY[112], PHE[114] residues are used to study the ligand interactions. Predominantly this process is significantly useful for increase in tomato fruit firmness.

Keywords: Docking, RMSD, Homology modelling, β-galactosidases.

Introduction

Beta galactosidases are glycosidic hydrolases (GH) existent in several types of micro-organisms, plants and animals. These hydrolases grouped into different families of 1,2,35 and 42[14]. TBG-7 belong to GH35 which includes several genes. Among them plant BGALS[7], especially in tomato, seven BGALs genes were shown to expressed during the fruit ripening stage. These enzymes have major role in cell wall growth, degradation and turnover of signalling molecules. These enzymes catalyse the hydrolysis of terminal non-reducing β-D-galacgtosyl groups related to carbohydrates, galacto lipids and glycol proteins[15].

Plant BGALs are focused more due to disassembly of pectin during the fruit ripening[8]. TBGs are responsible for fruit softening during plant developmental stage by break down of β -(1,4)- galactans (Figure 1) in the pericarp cell wall. The increased activity of beta –galactosidase and relatively the pectin solubility (loss of galactoses) enhance fruit ripening in many plant species[2].

The enzyme has been involved in plant physiological processes like developmental stages of plants including fruit ripening. These enzymes exist in many plant species and their organs and tissues but these protein functions are not yet characterised[5]. Fruit ripening related proteins were isolated from the tomato fruits[3], apple[13], mango[1]and also identified their biochemical properties. Present study represents the BGAL protein TBG7 involved in fruit ripening of tomato.

To know the structural characterisation, we performed the sequence alignment, Protein modelling and docking of Target-template alignment structure or model. We used different ligands to define the enzyme catalytic activity during fruit ripening. To study mechanism of enzyme (TBG-7) action, it is required to determine the 3D structure. To decrease the activity of TBG7, the substrate CME and BMA (Table 1) were used to increase the fruit firmness. Around 24 different sequences related to BGALS were used for construction and phylogenetic analysis (Figure 3).

Material and Methods

Sequence information: Target sequence beta galactosidase enzymes are used as source for sequence search. A2JGX1 protein sequence information is available in NCBI Reference ID NP_00123430.2.

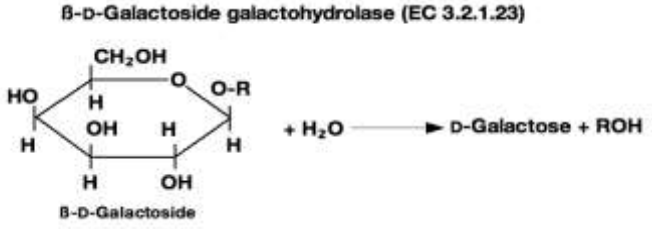


Fig. 1: Mechanism of beta galactosidase enzyme (http://www.toyobousa.com/images/gah201_chemical_formula.gif)

A2GJX1 related to GH35 family of Beta galactosidase gene group. The FASTA sequence of Template sequence for the query was retrieved from NCBI.

Secondary Structure prediction: PSIPRED server was used for secondary structure prediction of TBG-7(Figure 9). It is a highly accurate protein structure prediction method.

3D structure prediction: Modeller software was used for 3D structure prediction of A2JGX1(TBG7) protein**.** Target sequence submitted in BLAST-P was based on Multiple Sequence alignment template sequence based on percentage of identity and similarity of sequence.

Ramachandran Plot: The allowed regions in Ramachandran plot[11] (Figure 7) are showing residues of phi/psi torsion angles in tomato. Beta galactosidase-7 protein was detected by using PDB sum (PROCHECK).

Docking Analysis of TBG-7 Ligand and receptor interaction well studied in Auto dock software 4.2 version and Auto dock tools include MGL and Python software[9].

Results and Discussion

Multiple alignment: Protein sequence of β-galactosidase-7 of tomato was retrieved from NCBI[6]. PSI-BLAST was used for selection of homologous sequences of proteins. Highly conserved sequences were found in tomato β-galactosidase enzymes. Phylogenetic tree reveals that the hypothetical cladogram of tomato Beta galactosidases using Clustal Omega is shown in figure 2. The plant BGALS is evolved from bacteria.[12]

Secondary structure of TBG-7 protein: The secondary structure of TBG7 (Figure 4) contains highest percentage around 56.1% coils, 18.8% helix and 24% strands.

```
CLUSTAL 2.1 multiple sequence alignment

3W5F_A_23-702       ILISGSIHYPRSTPQMWPDLIQKAKDGGLDVIETYVFWNGHEPSPGKYNFEGR----YDL
6IK6_A_23-702       ILISGSIHYPRSTPQMWPDLIQKAKDGGLDVIETYVFWNGHEPSPGKYNFEGR----YDL
4MAD_A_33-323       -LLSGAMHYFRTVPEYWEDRLLKLKACGFNTVETYVAWNLHEPEEGQFVFEGI----ADI
6EON_A_46-333       ---AAEIHYPRIPKEYWEHRIKMCKALGMNTICLYVFWNFHEPEEGRYDFAGQ----KDI
5GSL_A_25-189       ----GTLQYFRVPRNYWEDRLRKMKSHGLNTVETYIAWNWHEPQEGVFDFTGETHPQRDL
5GSM_A_30-113       -----------VPADAWKDRLLKMREAGLNTVDTYVAWNWHEPEKGSFDFKGETHPQRNL
6EON_A_545-607      ------------------------LNMMDWSKGMVWVNGH--AIGRYWEIGP-------
5IAZ_A_4-119        ---------MRTVSGVCADVSEYHPNIKNWQIESYGEPEFHTAKVHLKCAPGQT---ISA
                                                      : *       *

3W5F_A_23-702       VRFIKMVQRAGLYVNLRIGPYVCAEWNFGGFPVWLKYVPGMEFRTNNQPFKVAMQGFVQK
6IK6_A_23-702       VRFIKMVQRAGLYVNLRIGPYVCAEWNFGGFPVWLKYVPGMEFRTNNQPFKVAMQGFVQK
4MAD_A_33-323       VRFIKTAEKVGLHVIVRPGPFICAEWEFGGFPYWLLTVPNIKLRCFNQPYLEKVDAYFDV
6EON_A_46-333       AAFCRLAQENGMYVIVRPGPYVCAEWEMGGLPWWLLKKKDIKLREQDPYYMERVKLFLNE
5GSL_A_25-189       IGFLELAQKLGLYVIIRPGPYICGEWKNGGIPDWLINSHPEILAKSPNGSFPRDVYYP--
5GSM_A_30-113       VGFLELADELGFYVIIRPGPYICGEWRNGGIPDWL------------------------
6EON_A_545-607      ----------------QQILYVPGCWLKKGENEIIILDMAGPSKAETEGLRQP-------
5IAZ_A_4-119        IKFASFGTPLG-----TCGTFQQGECHSINSNSVLERKCIGLERCVVAISPS--------
                                    :  .    .   :

3W5F_A_23-702       IVNMMKSENLFESQGGPIIMAQIENEYGPVEWEIGAPGKAYTKWAAQMAVGLKTGVPWIM
6IK6_A_23-702       IVNMMKSENLFESQGGPIIMAQIENAYGPVEWEIGAPGKAYTKWAAQMAVGLKTGVPWIM
4MAD_A_33-323       LFERLR--PLLSSNGGPIIALQIENEYG----SFGNDQKYLQYLRDGIKKRVGNELLFTS
6EON_A_46-333       VGKQLA--DLQISKGGNIIMVQVENAYG----AFGIDKPYISEIRDMVKQAGFTGVPLFQ
5GSL_A_25-189       ----------------PITYLHPTYLEY----AMKWYEEVLPIIRDYLYSNGGSIISVTI
5GSM_A_30-113       ------------------------------------------------------------
6EON_A_545-607      ------------------------------------------------------------
5IAZ_A_4-119        --------------------------------NFGGDPCPEVMKRVAVEAVCS-------

3W5F_A_23-702       CKQEDAPDFVIDTCNGFYCEGFRP------------NKPYKPKMWTEVWTGWYTKFGGP
6IK6_A_23-702       CKQEDAPDFVIDTCNGFYCEGFRP------------NKPYKPKMWTEVWTGWYTKFGGP
4MAD_A_33-323       DGPEPSMLSGGMIEGIFETVNFGS--RAESAFAQLKQYQPNAPLMCMEFWHGWFDHNGEE
6EON_A_46-333       CDWN-SNFENNALDDLLWTINFGTGANIDEQFKRLKELRPDIPLMCSEFWSGWFDHWGAK
5GSL_A_25-189       DDEP----------------------------------------------SYWEIIF------
5GSM_A_30-113       ------------------------------------------------------------
6EON_A_545-607      ------------------------------------------------------------
5IAZ_A_4-119        ------------------------------------------------------------

3W5F_A_23-702       IPQRPAEDIAFSVARFVQNNGSFFNYYMYHGGTNFGRTSSG-------LFIATSYDYDAP
6IK6_A_23-702       IPQRPAEDIAFSVARFVQNNGSFFNYYMYHGGTNFGRTSSG-------LFIATSYDYDAP
4MAD_A_33-323       HHTRSAESVVETLEEILKQNGSVN-FYMAHGGTNFGFYNGANHNETDYQPTITSYDYDGL
6EON_A_46-333       HETRSAEELVKGMKEMLDRNISFS-LYMTHGGTSFGHWGGANFPN--FSPTCTSYDYDAP
5GSL_A_25-189       ------------------------------------------------------------
5GSM_A_30-113       ------------------------------------------------------------
6EON_A_545-607      ------------------------------------------------------------
5IAZ_A_4-119        ------------------------------------------------------------

3W5F_A_23-702       LDEYGLLNEPKYGHLRDLHKAIKLSEPALVSSYAAVTSLGSNQEAHVYRSKSGACAAFLS
6IK6_A_23-702       LDEYGLLNEPKYGHLRDLHKAIKLSEPALVSSYAAVTSLGSNQEAHVYRSKSGACAAFLS
4MAD_A_33-323       LTESG---------------------------------------------------------
6EON_A_46-333       INESG---------------------------------------------------------
5GSL_A_25-189       ------------------------------------------------------------
5GSM_A_30-113       ------------------------------------------------------------
6EON_A_545-607      ------------------------------------------------------------
5IAZ_A_4-119        ------------------------------------------------------------
```

Fig. 2: Multiple sequence alignment using Clustal W

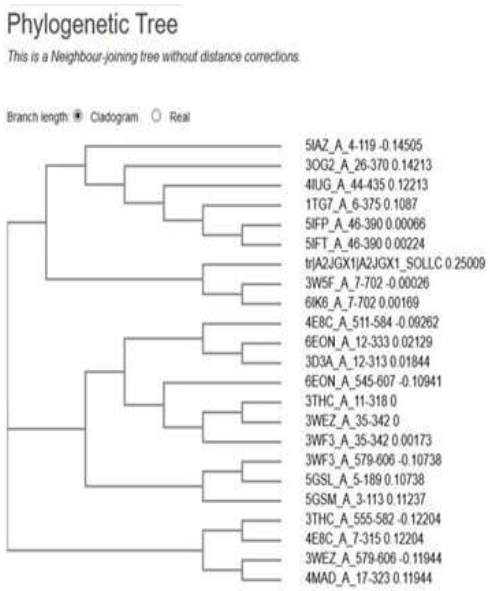


Fig. 3: Phylogenetic analysis

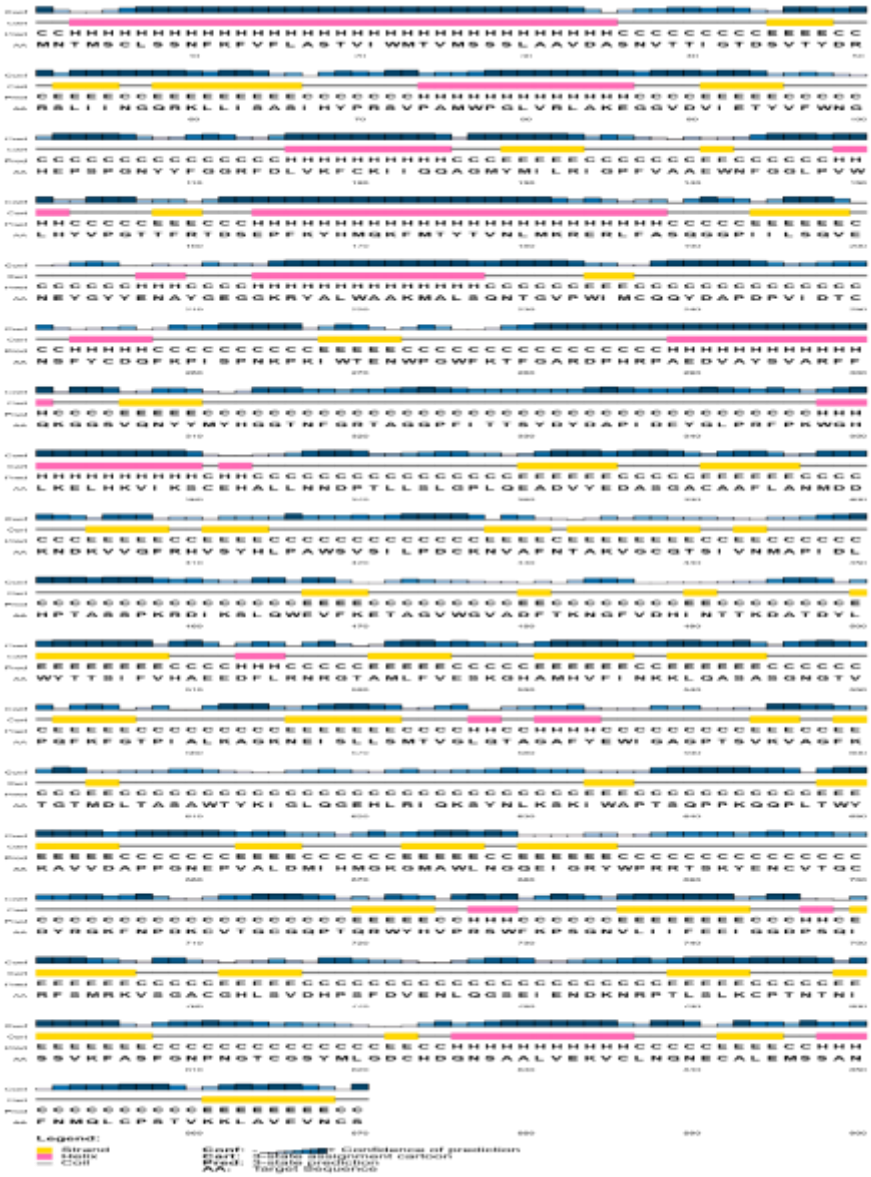


Fig. 4: Secondary structure prediction using PSIPRED server

3D structure Prediction using Modeller: The protein structure (3D) was modelled and generated using modeller 9.22. The sequence alignment (Figure 5) is between target and template visualised by Clustal Omega. The best model was carefully chosen for evaluation studies which has the lowest discrete optimized protein energy (DOPE) score from model number 1 to 10 score. These ten different DOPE values were executed through modeller. The finest model-3 was selected on the basis of lowest DOPE value -88953.

```
6ik6        ----------------------------------------EFSVSYDDRAIIINGKRK
A2JGX1      MNTMSCLSSNFKFVFLASTVIWMTVMSSSLAAVDASNVTTIGTDSVTYDRRSLIINGQRK
                                                    **:** *::****:**

6ik6        ILISGSIHYPRSTPQMWPDLIQKAKDGGLDVIETYVFWNGHEPSPGKYNFEGRYDLVRFI
A2JGX1      LLISASIHYPRSVPAMWPGLVRLAKEGGVDVIETYVFWNGHEPSPGNYYFGGRFDLVKFC
            :***.*******.* ***.*:: **:**:******************:* * **:***:*

6ik6        KMVQRAGLYVNLRIGPYVCAEWNFGGFPVWLKYVPGMEFRTNNQPFKVAMQGFVQKIVNM
A2JGX1      KIIQQAGMYMILRIGPFVAAEWNFGGLPVWLHYVPGTTFRTDSEPFKYHMQKFMTYTVNL
            *::*:**:*: *****:*.*******:****:****  ***:.:***  ** *:   **:

6ik6        MKSENLFESQGGPIIMAQIENAYGPVEWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQE
A2JGX1      MKRERLFASQGGPIILSQVENEYGYYENAYGEGGKRYALWAAKMALSQNTGVPWIMCQQY
            ** *.** *******::*:** **  *   *  ** *: ***:**:. :********:*

6ik6        DAPDPVIDTCNGFYCEGFRPNKPYKPKMWTEVWTGWYTKFGGPIPQRPAEDIAFSVARFV
A2JGX1      DAPDPVIDTCNSFYCDQFKPISPNKPKIWTENWPGWFKTFGARDPHRPAEDVAYSVARFF
            ***********.***: *:* .* ***:*** * **:..**.  *:*****:*:*****.

6ik6        QNNGSFFNYYMYHGGTNFGRTSSGLFIATSYDYDAPLDEYGLLNEPKYGHLRDLHKAIKL
A2JGX1      QKGGSVQNYYMYHGGTNFGRTAGGPFITTSYDYDAPIDEYGLPRFPKWGHLKELHKVIKS
            *:.**. ***************:.* **:********:***** . **:***::***.**

6ik6        SEPALVSSYAAVTSLGSNQEAHVYRSKSGACAAFLSNYDSRYSVKVTFQNRPYNLPPWSI
A2JGX1      CEHALLNNDPTLLSLGPLQEADVYEDASGACAAFLANMDDKNDKVVQFRHVSYHLPAWSV
            .* **:..  :: ***  ***.**.. ********:* *.: .  * *::  *:** **:|

6ik6        SILPDCKTAVYNTAQVNSQSSSIKMTPAG------------GGLSWQSYNEETPTADDSD
A2JGX1      SILPDCKNVAFNTAKVGCQTSIVNMAPIDLHPTASSPKRDIKSLQWEVFKETAGVWGVAD
            *******...:***:*..*:* ::*:* .             .*.*: ::* : . . :*

6ik6        TLTANGLWEQKNVTRDSSDYLWYMTNVNIASNEGFLKNGKDPYLTVMSAGHVLHVFVNGK
A2JGX1      -FTKNGFVDHINTTKDATDYLWYTTSIFVHAEEDFLRNRGTAMLFVESKGHAMHVFINKK
             :* **: :: *.*:*::***** *.: : ::*.**:*     * * * **.:***:* *

6ik6        LSGTVYGTLDNPKLTYSGNVKLRAGINKISLLSVSVGLPNVGVHYDTWNAGVLGPVTLSG
A2JGX1      LQASASGNGTVPQFKFGTPIALKAGKNEISLLSMTVGLQTAGAFYE-WIGAGPTSVKVAG
            *..:. *.   *::.:.  : *:** *:*****::*** ..*..*: * ..    *.::*

6ik6        LNEGSRNLAKQKWSYKVGLKGESLSLHSLSGSSSVEWVRGSLMAQKQPLTWYKATFNAPG
A2JGX1      FKTGTMDLTASAWTYKIGLQGEHLRIQKSYNLKSKIWAPTSQPPKQQPLTWYKAVVDAPP
            :: *: :*: . *:**:**:** * ::.  . .*  *.  *   ::********..:**

6ik6        GNDPLALDMASMGKGQIWINGEGVGRHWPGYIAQG--DCSKCSYAGTFNEKKCQTNCGQP
A2JGX1      GNEPVALDMIHMGKGMAWLNGQEIGRYWPRRTSKYENCVTQCDYRGKFNPDKCVTGCGQP
            **:*:****  ****  *:**: :**:**   ::     ::*.* *.** .** *.****

6ik6        SQRWYHVPRSWLKPSGNLLVVFEEWGGNPTGISLVRRSRSA------------------
A2JGX1      TQRWYHVPRSWFKPSGNVLIIFEEIGGDPSQIRFSMRKVSGACGHLSVDHPSFDVENLQG
            :**********:*****:*::*** **:*: * :  *. *.

6ik6        ------------------------------------------------------------
A2JGX1      SEIENDKNRPTLSLKCPTNTNISSVKFASFGNPNGTCGSYMLGDCHDQNSAALVEKVCLN


6ik6        --------------------------------
A2JGX1      QNECALEMSSANFNMQLCPSTVKKLAVEVNCS
```

Fig. 5: Sequence alignment of TBG-7 with Template (PDB code:6ik6) the conserved regions are indicated by '*'

The model is used for further docking studies. The developed protein model (target and template proteins) by homology modelling is schematically represented here in figure 6 using Chimera software by match maker.

Validation and evaluation of Model

Ramachandran Plot: The validation step using PUBSUM ensures that 91.9 % in TBG7 lies in allowed region, 6.5% residues in more allowed region, 1% in generously allowed region and 0.6% present in disallowed regions. RAMPAGE reveals the 93.3% of residues in favoured region and 4.5% residues are in the allowed region and 2.2% in outlier. The high number of residues present in allowed region suggests the best quality of structure having stereo chemical quality to build model. Errat statistical analysis (Figure 8) indicates that non –bonded interactions were observed in different atom types. The present 3D structure of TBG7 shows overall quality factor 84.68.

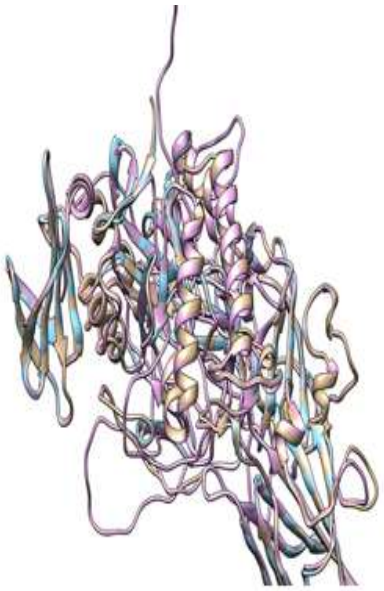


Fig. 6: Modelling using Moeller software A2JGX1(TBG7):6iK6

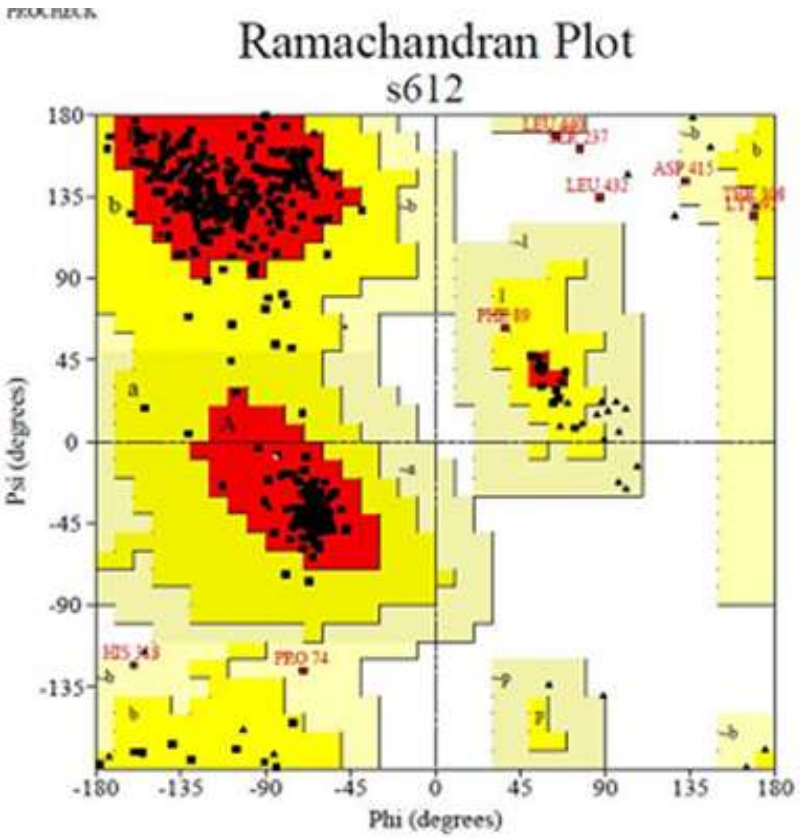


Fig. 7: Ramachandran plot 3D model of TBG-7(A2JGX1) derived from PROCHECK

Docking Study of TBG-7: The chemical reaction of enzyme and substrate complex is useful to study the interactions between substrate and active sites of enzymes. The Receptor -ligand complex gives information to understand the interactions and structural exact fit between the substrate and active site of an enzyme. The docking process provides possible confirmation of a complementarity between receptor and ligand, where generally the receptor is considered as an enzyme or a protein and ligand is a small organic molecule[4].

In this study, we represented the Auto dock method of studies for protein model of TBG7 (PDB id: A2JGX1) to determine the enzyme – inhibitor interactions. The binding sites on a protein(enzymes) are very specific to substrates or inhibitors. The binding sites include characteristics of chemical specificity bond strength and type of ligand and its affinity. The binding pockets for TBG-7 were predicted by server CASTp (Computer Atlas of Surface Topography of proteins) analysis[16]. Protein ligand interaction profiler (PLIP) is used for the analysis, visualisation and characterisation of non-covalent bonds in protein and ligand complexes[10]. The pocket -1 with 1469.71 area and 1708 .96 volume was selected for docking process.

The RMSD values of CME and BMA are 90.22 and 71.17 respectively. The molecular docking of TBG7 enzyme to generate the various feasible poses of ligand is observed in receptor – binding pocket. Binding energies of TBG7 used CME and BMA ligands, the best docking scores (–AG) selected -4.04 and -3.65 respectively based on the enzyme substrate interactions.

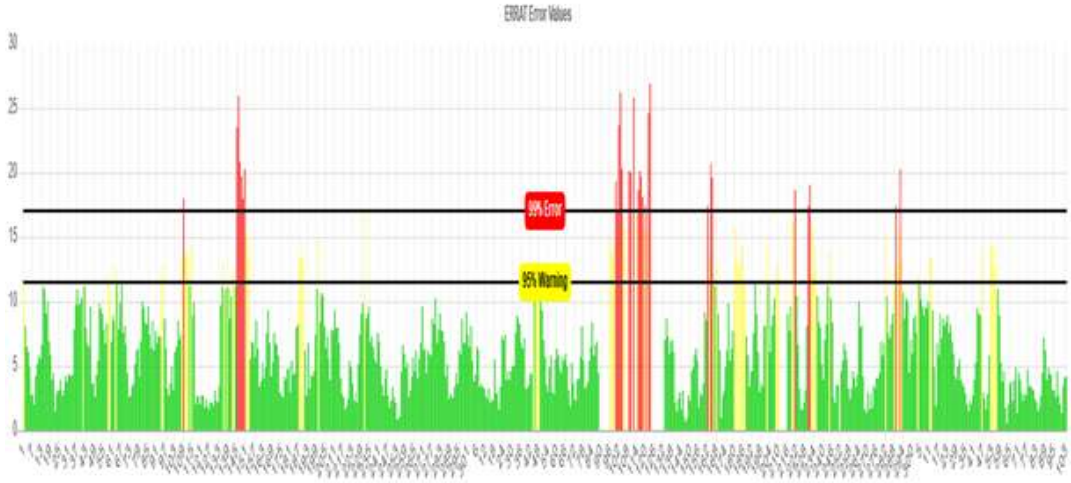
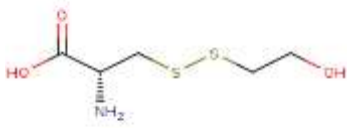
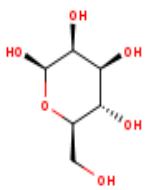


Fig. 8: ERRAT server for validation 3D structure of TBG7

Table 1
Ligand chemical structure.

S.N.	LIGAND NAME	FORMULA	CHEMICAL STRUCTURE
1	S, S-(2-HYDROXYETHYL) THYOCYSTEINE(CME)	C6H11NO3S2	
2	BETA-D-MANNOSE(BMA)	C6H12O6	

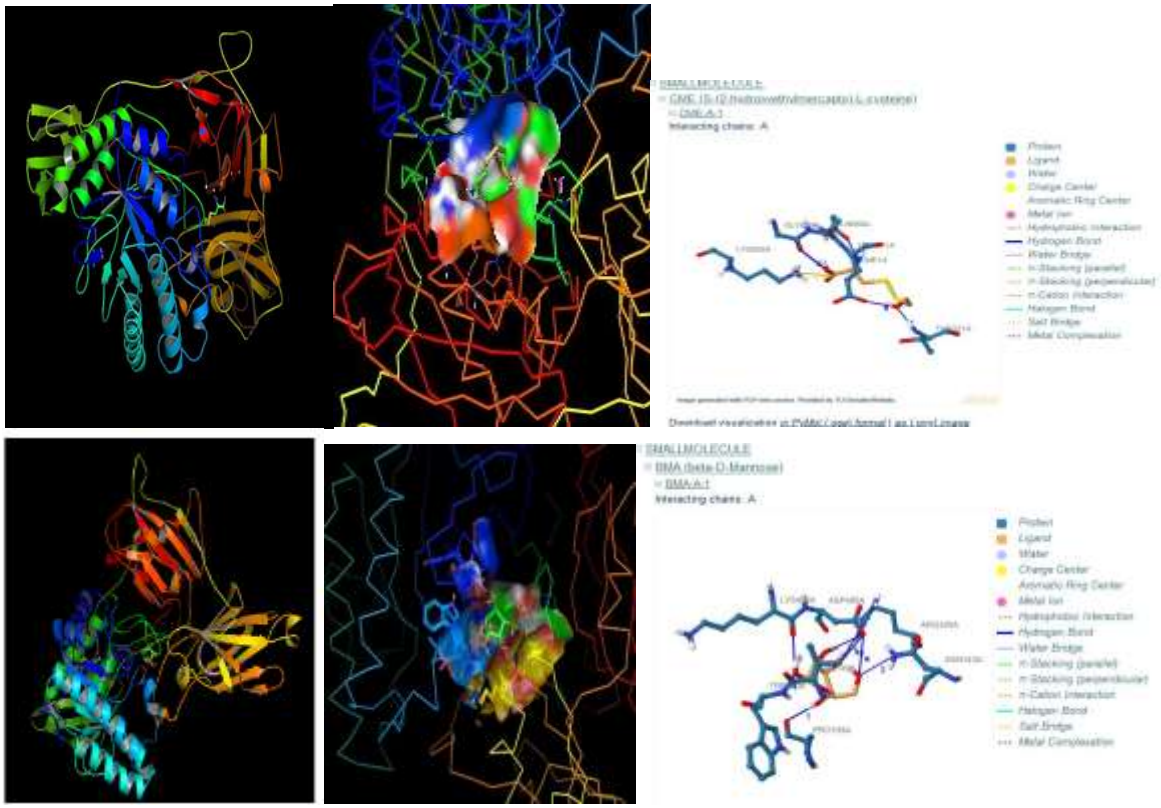


Fig. 9: TBG7-CME and BMA ligand complexes docking images from PLIP[10]

Conclusion

The role of enzymes is implicated in the plant developmental stages specifically fruit ripening. These β-galactosidases are belonging to same family GH-35 enzymes. The modelled TBG-7 revealed that the structural domains are very similar to the 6ik6(TBG-4) protein. The results obtained from above evaluation and validation methods revealed the homology modelling of TBG7 satisfactory and the build model used further docking studies.

The docking studies of β-galactosidases of tomato provided the interaction of enzyme–substrate and their substrate specificities. The substrates CME and BMA might be potential substrates for TBG-7. The present study is useful for the physiological function of β-galactosidase enzyme in tomato fruit ripening to enhance the fruit shelf life.

Acknowledgement

We are grateful to Director and Joint Director NAARM, ICM Information Communication Management Department in ICAR National Academy of Agricultural research management allowing us to access their facilities.

References

1. Ali Z.M., Armugam S. and Lazan H., β- Galactosidase and its significance in ripening mango fruit, *Photochemistry*, **38(5)**, 1109-1114 **(1995)**

2. Brummell D.A., Cell wall disassembly in ripening fruit, *Functional Plant Biology*, **33(2)**, 103-119 **(2006)**

3. Carey A.T., Holt K., Picard S., Wilde R., Tucker G.A., Bird C.R., Schuch W. and Seymour G.B., Tomato exo-(1->4)-[beta]-D-galactanase (isolation, changes during ripening in normal and mutant tomato fruit and characterization of a related c DNA clone), *Plant Physiology*, **108(3)**, 1099-1107 **(1995)**

4. Hernandez-Santino A., Tentoria-Barajas A.Y., Altuzar V., Vivanco-Cid H. and Mendoza-Barrera C., Protein-protein and protein- ligand docking, Protien Engineering – Technology and Application, Edited by Tomohisa Ogawa, Publisher, Intec, Chapter 3, 66 **(2013)**

5. Hossain M.A., Roslan H.A., Karim M.R. and Kimura Y., Molecular phylogeny, 3D-Structual insights, docking and mechanisms of action of plant beta- galactosidases, *International Journal of Bioinformatics Research and Applications*, **12(2)**, 149-179 **(2016)**

6. https:www.Ncbi.nim.nih.gov **(2014)**

7. https://en.wikipedia.org/wiki/Beta-galactosidase

8. Moctezuma E., Smith D.L. and Gross K.C., Antisense suppression of a β- galactosidase gene(TBG6) in tomato increases fruit cracking, *Journal of Experimental Botany*, **54(390)**, 2025-2033 **(2003)**

9. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Good Sell D.S. and Olson A.J., Auto Dock4 and Auto Dock Tools4; Automated docking with selective receptor flexibility, *Journal of Computational Chemistry*, **30(16)**, 2785 **(2009)**

10. Salentin S. et al, PLIP: fully automated protein- ligand profiler, *Nucl. Acids Res.*, **43(W1)**, W443-W447 **(2015)**

11. Ramachandran G.N., Stereochemistry of polypeptide chain configurations, *J. Mol. Biol*., **7**, 95-99 **(1963)**

12. Rani V. and Dev K., molecular evolution of βgalactosidase in Thermophiles, Psychrophilies, Mesophiles, Plants and Mammals by insilico approach, *Research Journal of Recent Sciences*, **5(2),** 1-11 **(2016)**

13. Ross G.S., Wegrzyn T., MacRae E.A. and Redgwell R.J., Apple [beta]-Galactosidase (Activity against Cell wall Polysaccharides and Characterization of a Related C DNA Clone), *Plant Physiology*, **106(2)**, 521-528 **(1994)**

14. Shipkowski S. and Benchley J.E., Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 β-galactosidases are arabinogalactan type I oligomer hydrolases, *Appl. Environ. Microbial*., **72(12)**, 7730-7738 **(2006)**

15. Smith D.L., Abbott J.A. and Gross K.C., Down-regulation of tomato β-galactosidase 4 results in decreased fruit softening, *Plant Physiology*, **129(4)**, 1755-1762 **(2002)**

16. Tian W., Chen C., Lei X., Zhao J. and Liang J., CASTP,3.0: computed atlas of surface topography of proteins, *Nucleic Acids Research*, **46(W1)**, W363-W367 **(2018)**.

(Received 29th October 2019, accepted 01st February 2020)