Malaysian Journal of Science **28** *(Special Edition)*: 115- 122 (2009)

Homology Modeling of B-Glucuronidases from *E. Coli* And *T. Maritima*

[1]**Ibrahim Ali Noorbatcha,** [2]**Avesha Masrur Khan,** [3]**Hamzah Mohd Salleh**

Biomolecular Engineering Research Unit (BioMERU), Department of Biotechnology Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia. [1]ibrahiman@iiu.edu.my, [2]ayeshamasrur@gmail.com, [3]hamzah@iiu.edu.my (Corresponding author)
Received in 15th September 2008, accepted in revised form 14th December 2008.

Abstract The enzyme β-glucuronidase (GUS), which belongs to the glycoside hydrolase family of enzymes, can hydrolyze any aglycone conjugated to D-glucuronic acid through a β-O-glycosidic linkage. It is present in almost all tissues of vertebrates and their resident intestinal flora, including *E. coli*. However, GUS enzymes obtained from different sources have different stability towards heat, resistance to detergents and varying catalytic activities. A good understanding of the reasons for this variation can lead to designing new enzymes with desired level of property, having great prospect in the industry. For this purpose, studies on the three-dimensional structure of GUS enzyme can offer insights on the structure-function correlations, and provide information on the distribution of certain residues both in *E. coli* and *T. maritima* enzymes. The structures of GUS enzymes from *E. coli* and *T. maritima* are not known experimentally. As such in the current work, homology modeling of the three-dimensional structure of both variants of the GUS enzyme was carried out based on the solved crystal structure of Human GUS enzyme. Multiple sequence alignment for both enzyme sequences was carried out in order to locate the most suitable template for homology modeling and the models thus prepared were found to contain 32-43% sequence identity with the template. Superposition of the model obtained with the template as well as structural alignment were carried out to classify the structural differences. This paper will also present an analysis and verification studies of the model based on various criteria. The current work offers a better understanding of the structural differences between GUS enzymes from different sources, as well as suggests regions for further modification using experimental and computational methods.

(**Keywords:** β-glucuronidase, homology modeling, verification)

INTRODUCTION

Structural studies of enzymes enable thorough understanding of key factors related to their activity. As the databases of solved structures are on the rise, techniques such as homology modeling has proven to be constructive in terms of reducing the time and cost involved in protein structure determination methods. The difficulty involved for proteins to be extracted, crystallized or solubilized can hamper experimental methods of structure determination. Therefore, the method of homology modeling strives to resolve the gap between number of available sequences and experimentally solved protein structures.

Since crystal structures of microbial GUS enzymes have yet to be solved, homology modeling can be used to construct three-dimensional structures of the enzymes from *E. coli* and *T. maritima* based on the known structure of human GUS enzyme. Previously, structural study of GUS enzymes from *E. coli* and *Staphylococcus* has revealed features explaining functional differences between them [1]. This and other structural and mutational studies on GUS enzyme from *E. coli* have relied on the solved X-ray structure of human GUS enzyme [2], whereby the sequence identity between both is 43.6%. Having sequence identity above 40% enables simple alignment, and 90% of main-chain atoms can be modeled with a RMSD of about 1 Å [3].

The enzyme β-glucuronidases (GUS; EC 3.2.1.31) belongs to glycoside hydrolase family of enzymes and can hydrolyze a wide variety of glucuronides. Due to the absence of significant GUS activity in plants, it is used as a transformation marker in transgenic plants. Other reported uses of GUS are in medical diagnostics and therapeutics [4]. There has been achievement in obtaining a mutant enzyme with thermostable GUS activity from wild type β-galactosidase via directed evolution [5], as well as mutant of GUS with high thermostability [6]. *T. maritima* has significant potential as a source of glucuronidases which are thermostable and suitable for use in the industry, as well as a starting point for mutation studies for synthesis of oligosaccharides [7] and attainment of activity comparable to *E. coli*. In this study, we report structural data and analysis of both *E. coli* and *T. maritima* GUS enzymes, with the aim of identifying important regions and differences between them. It is intended that the outcome of this study will provide a basis for further computational studies related to probing the activity and stability of these enzymes.

MATERIALS AND METHODS

Sequence Alignment and Physicochemical Properties Estimation

Amino acid sequences of β-glucuronidase for *E. coli* (AAC74869) and *T. maritima* (AAD36143) in FASTA format were downloaded from NCBI [8]. The BLAST search protocol of Discovery Studio (Accelrys Software Inc, version 2.1) was carried out for each amino acid sequence using the BLOSUM62 scoring matrix, based on which multiple sequence alignment was also carried out, using the Human GUS enzyme (1BHG_A) as the input sequence for each amino acid sequence. The protein analysis tool ProtParam [9] was used to calculate various physicochemical properties of both protein sequences.

Secondary Structure Analysis

Analysis of the predicted secondary structures of the *E. coli* and *T. maritima* GUS enzymes sequences were carried out using GORIV secondary structure prediction method [10].

Homology Modeling

The 3-D structures for *E. coli* and *T. maritima* GUS enzymes were obtained via homology modeling based on the available crystal structure of Human GUS enzyme (1BHG_A) as a template using Discovery Studio.

Model Verification

The obtained homology models were validated using Ramachandran plot. The sequence-structure compatibility of the models was evaluated by calculating the Verify score using the Profiles-3D analysis tool of Discovery Studio. Assessment of stereo-chemical qualities of the models were carried out by the 'protein model check' option of WHAT_IF [11].

RESULTS AND DISCUSSIONS

From the BLAST search, the Human GUS structure (1BHG_A) having a resolution of 2.53Å was chosen as a template, based on the high bit score and low E-value obtained. Multiple sequence alignment for *E. coli* and *T. maritima* GUS sequences showed sequence identities of 45% and 33.5% and sequence similarities of 62.2% and 54.5% (Figure 1 and Figure 2). Superimposition of the models obtained with the original template yielded a RMSD of 1.56 Å and 2.37 Å. RMSD measures the difference between Cα atom positions between two proteins (Equation 1). The smaller the deviation, the more spatially equivalent the two proteins is. The models obtained also contain the common structural features from some of the family 2 enzymes of GH such as the TIM barrel catalytic domain, a jelly roll barrel and an immunoglobulin constant domain (Figure 3).

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left| r_i^{model} - r_i^{real} \right|^2}$$

(Equation 1)

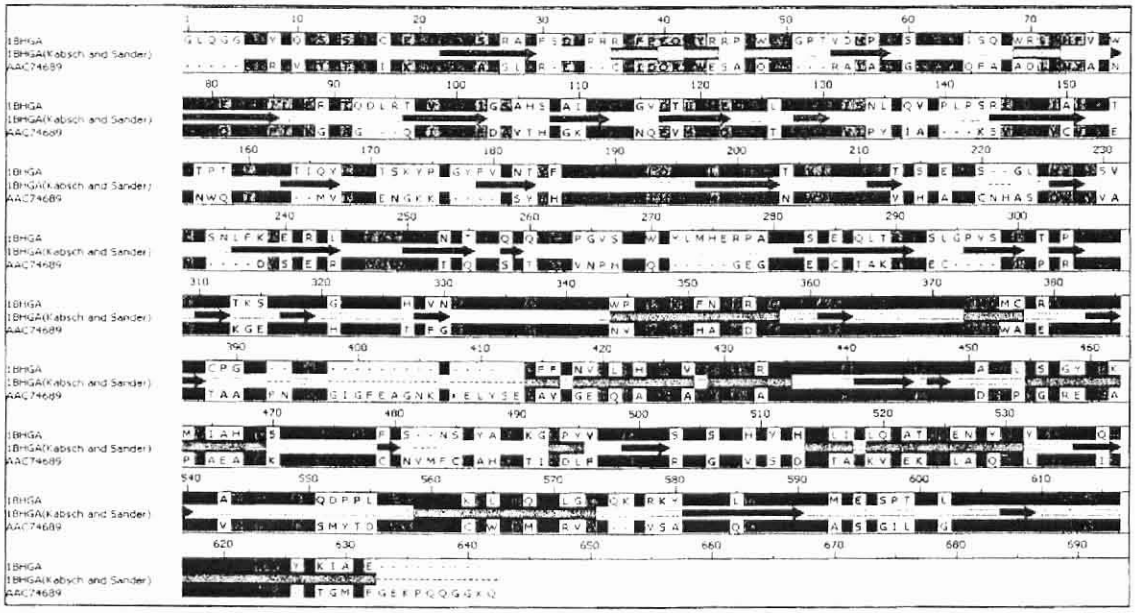

Figure 1. Multiple sequence alignment of *E.coli* GUS (AAC74689) with Human GUS (1BHGA), with red bars indicating helices and blue arrows indicating β-strands

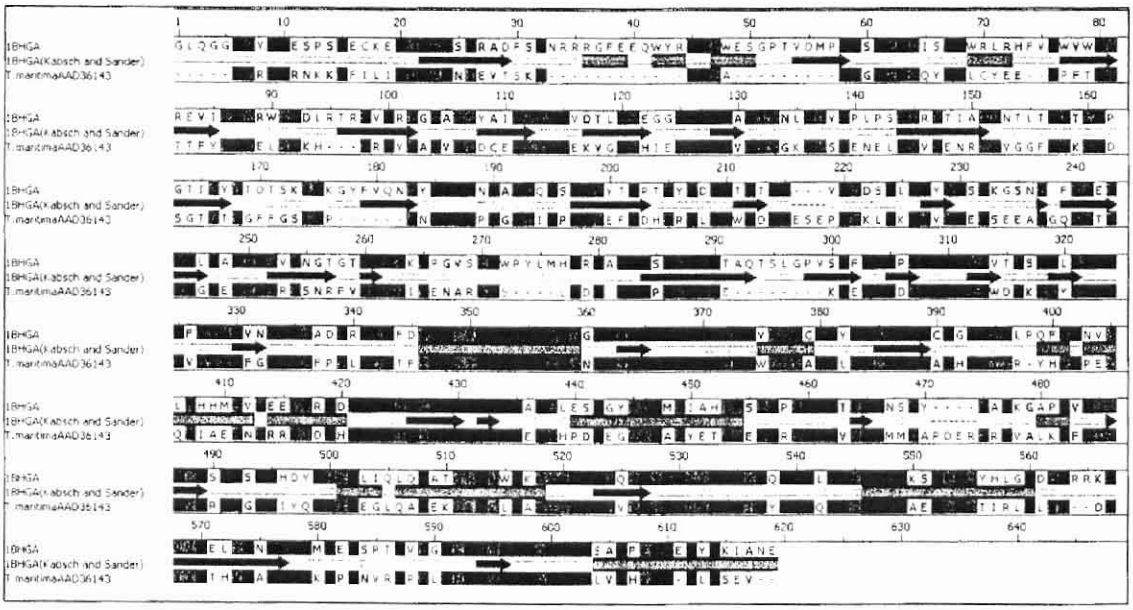

Figure 2. Multiple sequence alignment of *T.maritima* GUS (AAD36143) with Human GUS (1BHGA), with red bars indicating helices and blue arrows indicating β-strands.

(a) (b) (c)

Figure 3. (a) Solved structure of Human GUS enzyme, and homology models of (b) *E.coli* GUS enyzme and (c) *T.maritima* GUS enzyme, with three domains; jelly-roll barrel in dark blue, immunoglobulin constant domain in light blue and TIM barrel domain in red.

The Ramachandran plot, which shows the φ-ψ torsion angles for all residues in the structure (Figures 4, 5 and 6), showed only a small percentage of the residues to be in the disallowed regions. The stereo-chemical checks on the models using WHAT_IF gave Z scores of constraint-independent quality indicators, and RMS Z-scores of model conformity to common refinement constraint values (Table 1). The standard bond lengths and bond angles for amino acid residues are used to evaluate the same in the protein models [12].

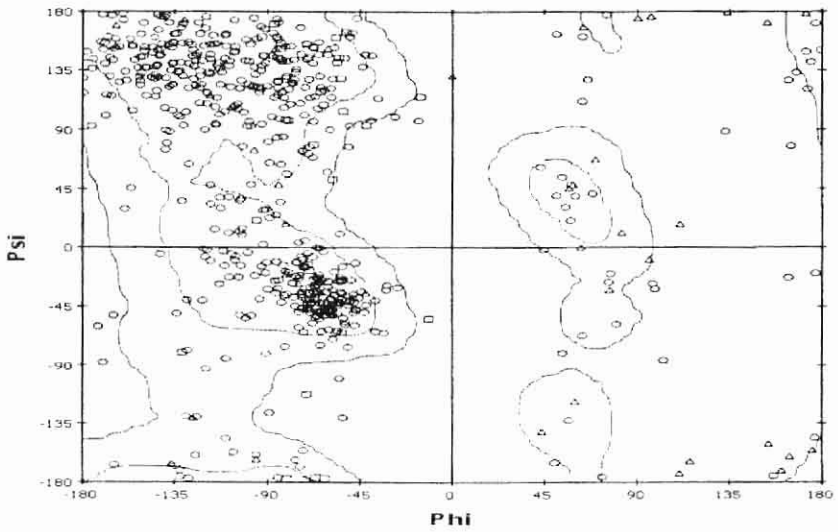

Figure 4. Ramachandran plot of the available crystal structure of Human GUS enzyme.

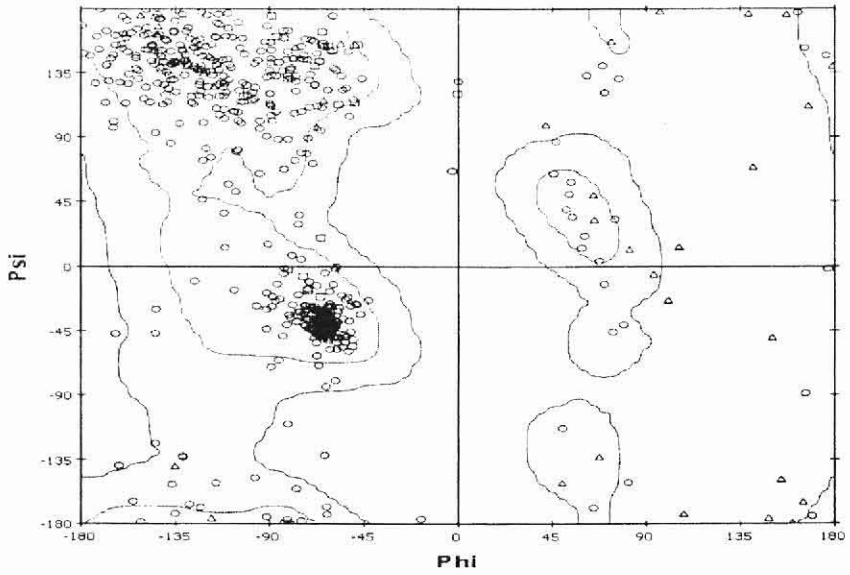

Figure 5. Ramachandran plot of *E.coli* GUS enzyme structure obtained from homology modeling.

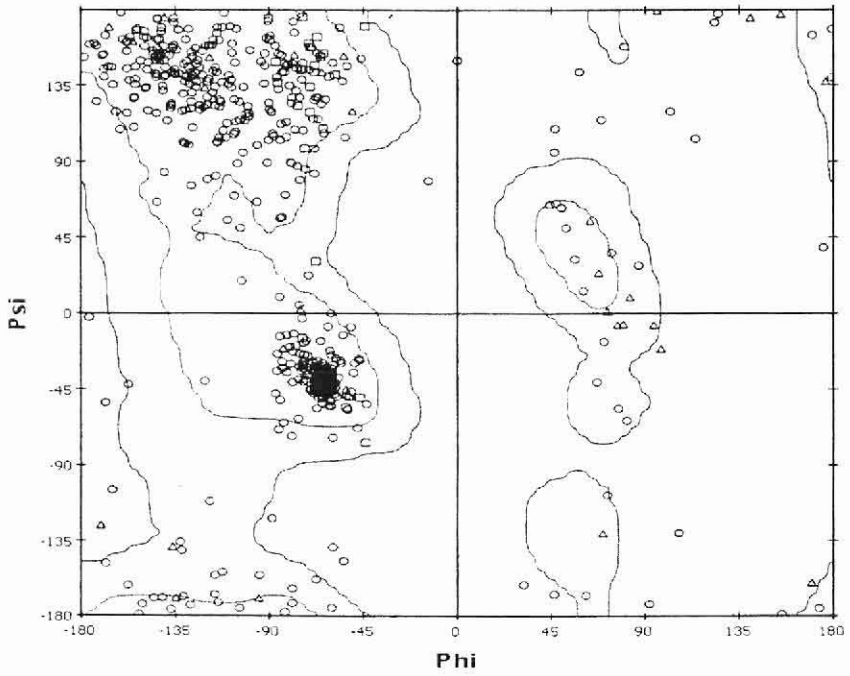

Figure 6. Ramachandran plot of *T.maritima* GUS enzyme structure obtained from homology modeling.

Table 1. Evaluation scores as obtained by WHAT_IF and Profiles-3D.

Evaluation scores	**1BHGA (Human)**	**AAC74869 (*E.coli*)**	**AAD36143 (*T.maritima*)**
Z-scores			
Ramachandran plot appearance	-5.628	-3.069	-3.442
χ_1/χ_2 Rotamer normality	-4.961	-1.615	-1.816
Backbone conformation	-4.046	-5.104	-4.873
1st generation packing quality	-1.939	-2.099	-2.157
2nd generation packing quality	-2.097	-2.243	-2.403
RMS Z-scores			
Bond lengths	0.657	0.964	0.959
Bond angles	0.955	1.358	1.365
Omega angle restraints	0.060	0.163	0.161
Side-chain planarity	0.733	0.323	0.349
Improper dihedral distribution	1.165	0.995	1.012
Inside/Outside distribution	1.095	1.105	1.111
Verify scores			
Verify expected high score	279.616	271.314	257.481
Verify score	262.23	233.04	218.24
Verify expected low score	125.827	122.091	115.866

The Verify scores were used to assess the structure-to-sequence fitness of the models obtained. These were originally developed for measuring accuracy of protein structure, regardless of the procedure used to obtain the model [13, 14]. Verify scores have been proven to be reliable indicators for further improvement of models as presented in other studies [15]. As given in Table 1, the Verify scores for both *E. coli* and *T. maritima* GUS enzyme models are well above the expected low score. However, further improvement can be carried out until the Verify scores cannot increase any further towards the expected high score.

Table 2. Predicted physicochemical properties of GUS from *Human*, *E.coli* and *T.maritima*.

Properties	1BHGA (Human)	AAC74869 (*E.coli*)	AAD36143 (*T.maritima*)
Number of Amino Acids	613	603	563
Molecular Weight (Dalton)	70686.2	68447.0	65682.9
Theoretical pI	6.31	5.24	5.77
Total number of negatively charged residues (Asp+Glu)	66	80	87
Total number of positively charged residues (Arg+Lys)	60	56	74
Total number of atoms	9880	9478	9233
Computed Instability Index	43.05	26.68	41.96
Aliphatic Index	82.95	77.74	83.91

In Table 2, the predicted physicochemical properties of GUS from all three sources, such as total number of positively and negatively charged residues can be seen. From sequence comparison and secondary structure evaluation, the results, as shown in Table 3 revealed that both sequences predominantly contain random coils, with GUS enzyme from *T. maritima* having a higher content of random coils.

Table 3. Composition of secondary structure elements for GUS from Human, *E.coli* and *T.maritima*

Composition	1BHGA (Human)	AAC74869 (*E.coli*)	AAD36143 (*T.maritima*)
Alpha helix (%)	25.00	21.72	28.42
Beta strand (%)	24.67	31.34	22.38
Random coil (%)	50.33	46.93	49.20

As previously discussed [16], arginine content is favoured in α-helices in thermophilic proteins, compared to mesophilic ones. From our analysis,

we found that *T. maritima* has a higher content of arginine residues (14) compared to *E. coli* (10) through examination of α-helices.

Inspection of salt bridges is also a necessary evaluation, since it has been found to be an important stabilizing factor in many studies [16]. Yet it was found that the number of salt bridges in *E. coli* GUS enzyme structure was 15, which was more than that of *T. maritima*, which was 13.

The occurrence of proline residues in thermostable proteins has been affirmed as an entropic stabilization factor, especially at second positions of β–turns and N1 position of α-helices [17]. Frequency of proline residues in the α-helices for both sequences was found to be the same.

CONCLUSION

We have obtained homology models of GUS enzymes from *E. coli* and *T. maritima*, and have identified the evaluation criteria which can support further improvement of the structures. There is a need for a reliable model of *T. maritima* GUS enzyme which can be used for in-depth study of the thermostability features of this enzyme. With additional measures, such as energy minimization using molecular dynamics, a more stable model of GUS enzyme from *T.maritima* may be attained.

REFERENCES

1. Arul L., Benita G. and Balasubramanian P. (2008). *Functional insight for β-glucuronidase in Escherichia coli* and *Staphylococcus sp.* RLH1. Bioinformation 2(8): 339-343.
2. Jain S., Drendel W.B., Chen Z-W., Matthews F.S., Sly W.S. and Grubb J.H. (1996). *Structure of human β-glucuronidase reveals candidate lysosomal targeting and active-site motifs.* Nature Structural Biology 3: 375-381.
3. Xiang Z. (2006). *Advances in homology protein structure modeling.* Curr Protein Pept Sci. 7(3): 217-227.
4. Jefferson R. A. and Mayer J. E. (2000). *Microbial β-glucuronidase genes, gene products and uses thereof.* World Intellectual Property Organization. International Publication Number WO 00/55333 A1.
5. Xiong A., Peng R., Zhuang J., Li X., Xue Y., Liu J., Gao F., Cai B., Chen J. and Yao Q. (2007*). Directed evolution of a beta-galactosidase from Pyrococcus woesei resulting in increased thermostable beta-glucuronidase activity.* Appl Microbiol Biotechnol.77: 569-578.
6. Flores H. and Ellington A.D. (2002). *Increasing the thermal stability of an oligomeric protein, beta-glucuronidase.* J. Mol. Biol. 315: 325-337.
7. Salleh H.M., Mullegger J., Reid S.P., Chan W.Y., Hwang J., Warren R.A.J. and Withers S.G. (2006). *Cloning and characterization of Thermotoga maritima β-glucuronidase.* Carbohydrate Research 341: 49-59.
8. http://ncbi.nlm.nih.gov
9. (http://www.expasy.org/tools/protparam.html)
10. Combet C., Blanchet C., Geourjon C. and Deléage G. (2000) NPS@: Network Protein Sequence Analysis. TIBS 25(3) [291]:147-150
11. http://swift.cmbi.ru.nl/servers/html/index.html
12. Engh R. and Huber R. (1991).*Accurate bond and angle parameters for X-ray protein structure refinement.* Acta Crystallogr. A47, 392—400.
13. Lüthy R., Bowie J.U., Eisenberg D. (1992). *Assessment of protein models with three dimensional profiles.* Nature 356: 83.
14. Bowie J.U., Lüthy R., Eisenberg D. (1991). *A Method to Identify Protein Sequences that fold into a known three-dimensional structure.* Science 253: 164.
15. Dunty J.M., Gabarra-Niecko V., King M.L., Ceccarelli D.F.J., Eck M.J., Schaller M.D. (2004). *FERM Domain Interaction Promotes FAK Signaling. Molecular and Cellular Biology.* American Society for Microbiology 24(12) 5353–5368.
16. Kumar S., Tsai C-J. and Nussinov R. (2000). *Factors enhancing thermostability.* Protein Engineering 13(3): 179-191.
17. Watanabe K. and Suzuki Y. (1998). *Protein thermostabilization by proline substitutions.* Journal of Molecular Catalysis B: Enzymatic 4: 167-180.