JOURNAL OF TROPICAL LIFE SCIENCE
20xx, Vol. 11, No. 3, 323 – 330
http://dx.doi.org/10.11594/jtls.11.03.09

Research Article

In silico Characterization of *UGT74G1* Protein in *Stevia rebaudiana* Bertoni Accession MS007

Afiqah Ramatullah Khan [1], Nor Iswani Mokthar [1, 2], Zarina Zainuddin [1], Nurul Hidayah Samsulrizal [1*]

[1] Department of Plant Science, Kulliyyah of Science, International Islamic University of Malaysia, Bandar Indera Mahkota, 25200, Kuantan, Pahang, Malaysia
[2] Department of Biological Sciences and Biotechnology, The National University of Malaysia 43600, Bangi, Selangor, Malaysia

Article history:
Submission March 2021
Revised April 2021
Accepted July 2021

**Corresponding author:*
E-mail:
hidayahsamsulrizal@iium.edu.my

ABSTRACT

Due to its low-calorie property, *Stevia rebaudiana* is being promoted as an alternative sweetener for diabetic and obese patients. The steady demand in the market for high-quality stevia extracts presents a challenge for the enhanced production of steviol glycosides that are safe for human consumption. This study characterized the structure and content of the gene involved in the production of *UGT74G1* protein for *Stevia rebaudiana* accession MS007 through *in silico* analysis using a transcriptome dataset of stevia MS007. Homologous search using BLASTp shows high similarity to Q6VAA6 RecName: Full=UDP-glycosyltransferase 74G1 (*S. rebaudiana*) as the top hit sequence. InterPro family and domain protein motif search revealed UDP-glucuronosyl/UDP-glucosyltransferase (IPR002213) and UDP-glycosyltransferase family, conserved site (IPR035595). The phylogenetic tree construction was done by selecting 14 out of 102 protein sequences from BLASTp search. The phylogenetic analysis revealed a high value of bootstrapping, which was 100, indicating the high similarity between *UGT74G1* (Q6VAA6.1 and Cluster-31069.45201) in *S. rebaudiana*. ProtParam ExPASy, PSIPRED, and Phyre2 computed the primary, secondary, and tertiary structures for *UGT74G1* protein. The *UGT74G1* predicted tertiary structure scored 100.0% confidence by the single highest scoring template and coverage of 96%. The model has dimensions (Å) of X: 57.609, Y: 70.386, and Z: 58.351. Outcomes of this research will help enhance understanding UDP-glycosyltransferase 74G1 (*S. rebaudiana* MS007) characteristics and enhance target identification processes to improve understanding of protein-protein interaction in *S. rebaudiana* MS007.

Keywords: Phylogenetic, Stevia, UGT74G1, 3D structure prediction

Introduction

Stevia rebaudiana Bertoni, a member of the Asteraceae family, has been used as a traditional sweetener for the longest time. This plant was first adopted in Japan in the 1970s as a popular commercial sweetener [1]. Stevia is cultivated mostly in Paraguay, Kenya, China, United States, Vietnam, Brazil, India, Argentina, and Colombia [2]. Steviol glycosides are the chemicals found in *Stevia* that give food a sweet flavour. Eleven major steviol glycosides are confirmed to be present in stevia, which rebaudioside A and stevioside are the highest. The sweetness of stevioside and rebaudioside A is 150 – 300 and 200 – 400 times sweeter than sucrose, respectively [3, 4]. This proves its potential to be an alternative sweetener for diabetic and obese patients. As a result, there is a steady pursuit for high quality, pure stevia leaf extracts with taste improvements [5].

In 2015, fourteen *S. rebaudiana* accessions from all across Malaysia and three accessions from Paraguay were collected and evaluated for advanced breeding developments. Accession MS007 was among the taller ones and showed good qualities in plant height, number of leaves,

How to cite:
Khan AR, Mokthar NI, Zainuddin Z, Samsulrizal NH (2021) *In silico* Characterization of *UGT74G1* Protein in *Stevia rebaudiana* Bertoni Accession MS007. Journal of Tropical Life Science 11 (3): 323 – 330. doi: 10.11594/jtls.11.03.09.

and leaf size [6]. The challenges for improving *S. rebaudiana* are the enhancement of the production of steviol glycosides while maintaining the safety of the consumption of this sweetener to humans [7]. To achieve this goal, a proper mechanistic understanding of the genes involved in the biosynthesis of steviol glycosides is needed. The latest data in UniProt 2019 shows that experimental evidence at protein level exists for UDP-glycosyltransferase 74G1 from the *UGT74G1* gene.

According to Richman *et al.* in 2005, *UGT74G1* plays a part in the biosynthesis of steviol glycosides in leaves [8]. The glycosylation of C-19 carboxyl of steviolbioside catalysed by *UGT74G1* to form stevioside [8]. Until today there is no information on the genetic material of the local accession of *S. rebaudiana* MS007. Thus, there is a great need to characterize the genes and protein structures involved in the production of stevioside and rebaudioside A in local stevia MS007 plants. It would be a great advantage to manipulate the said structures in generating more artificial sweeteners. Thus, this study aims to characterize the protein structure of functional *UGT74G1* in *S. rebaudiana* accession MS007 using structural bioinformatics.

Material and Methods

Protein translation using ExPASy translate tool

Bioinformatics analyses were performed using the transcriptome data of stevia MS007 [9]. The dataset comprises raw sequence data derived from transcriptome sequencing of *S. rebaudiana* leaf samples accession MS007. Cluster-31069.45201 *UGT74G1 S. rebaudiana* MS007 was used in this study as the longest gene available in the dataset. The FASTA sequences of the transcriptome dataset for *UGT74G1* were uploaded to the ExPASy Translate tool available at web.expasy.org/translate/ to translate the nucleotide sequences into amino acid sequences [10]. The possible protein sequences were highlighted in red color. The longest open reading frame of *UGT74G1* started with M (Methionine) was selected and saved in FASTA file format.

Homology search

The protein sequence of *UGT74G1* of *Stevia rebaudiana* MS007 was analysed further via homology search using BLASTp at the National Center for Biotechnology Information (NCBI) blast.ncbi.nlm.nih.gov/Blast.cgi [11]. The parameters of BLASTp were maintained as default i.e., Expect threshold=0.05; Matrix= BLOSUM 62 and Gap extension=1.

Protein domains and families

The domain and protein families of *UGT74G1* (*S. rebaudiana* MS007) were analyzed by using InterPro database (www.ebi.ac.uk/interpro/search/sequence/) [12]. This database classifies protein sequences into its families and can predict the presence of important domains and sites. Other databases used to find protein families and domains were Pfam (pfam.xfam.org/search/sequence) and Simple Modular Architecture Research Tool (SMART) (smart.embl.de/) [13, 14].

Physico-chemical properties of UGT74G1 protein

The translated protein sequence of *UGT74G1* was inserted to ProtParam Server available at web.expasy.org/protparam/. This program was used to identify the molecular weight, theoretical pl and amino acid composition of protein sequence [15].

Constructing phylogenetic tree

Multiple sequence alignment was done involving 15 sequences selected from the homology search analysis as having a high percentage of identity (Table 1), including *UGT74G1* protein sequence available at www.ebi.ac.uk/Tools/msa/muscle/ [16]. Molecular Evolutionary Genetics Analysis (MEGA) software was downloaded at https://www.megasoftware.net/ [17]. This software was used to generate the phylogenetic tree of *UGT74G1*. The phylogenetic tree was constructed by using Maximum Likelihood Tree method with the bootstrapping value was set at 1,000 replications.

PSIPRED workbench

Amino acid sequence for the *UGT74G1* was uploaded to PSIPRED Workbench at http://bioinf.cs.ucl.ac.uk/psipred/. PSIPRED 4.0 (Predict Secondary Structure) was selected for "Popular Analyses" while DeepMetaPSICOV 1.0 (Structural Contact Prediction) was chosen for "Contact Analysis". Finally, DomPred (Protein Domain Prediction) was checked as the selection for "Domain Prediction"[18].

Phyre2 protein fold recognition server

Phyre2 Protein Fold Recognition Server is a web-based tool for predicting tertiary protein structure [19]. Therefore, this program was used to predict the tertiary protein structures of *UGT74G1* by uploading amino acid sequence at www.sbg.bio.ic.ac.uk/phyre2/ with normal modelling mode was being selected. The web tool showed the most suitable template to be used, which is 6L8Z from Protein Data Bank (PDB), as having 100.0% confidence in the model.

Results and Discussions

The whole genome sequence and annotation of *S. rebaudiana* were done in 2017. This project was led by scientists from PureCircle Stevia Institute and KeyGene. A huge breakthrough in research was revealed by presenting the annotated and high-quality genome sequences of three stevia cultivars. As a result, a total of 146 838 888 paired end reads consisting of 22.2 gigabases were obtained by sequencing leaves' DNA from commercially grown seedlings [2]. In 2018, the transcriptomic project of *S. rebaudiana* MS007 was started by researchers from International Islamic University Malaysia [9]. From a previous study, the UDP-glycosyltransferase 74G1 was identified as a protein involved in the biosynthesis of steviol glycosides in plant leaves. It involved a subsequent conversion from steviol to the mono-glycoside up to tetra-glycoside rebaudioside A [8]. The longest open reading frame of *UGT74G1* containing a cumulative length of 459 amino acids was chosen for potential study using this transcriptome dataset of stevia MS007.

Homology search

Table 1 shows all the sequences in the non-redundant database (*nr*) with significant sequence homology to our *UGT74G1* sequence. By default, the results are sorted according to the Expect value (E-value) in ascending order. The lower the e-value, the more significant the score. Based on the results, the most similar corresponding alignment for protein sequence of *UGT74G1* (*S. rebaudiana* MS007) is Q6VAA6 RecName: Full=UDP-glycosyltransferase 74G1 (*S. rebaudiana*) with percentage identity of 81.48%. The length of the subject sequence is 460 amino acids. The sequence identity for this alignment, which refers to the number of identical bases between the query and subject sequences is 374/459 which contributes to 81%. The value of positives shows 411/459 or 89% represents the number of residues that are identical or have the same chemical properties. In the BLASTp alignment analysis result, there is + sym-

Table 1. BLASTp analysis showing possible identity of *S. rebaudiana* MS007. The identity of *S. rebaudiana* MS007 matched with sequence of *S. rebaudiana* GeneBank ID Q6VAA6.1 from Canada.

Identical genus & species to *S.rebaudiana* MS007	Accession No.	Identity %	Reference
RecName: Full=UDP-glycosyltransferase 74G1 *S. rebaudiana*	Q6VAA6.1	81.48	[8]
UDP-glycosyltransferase 74G1 *Helianthus annuus*	XP_021980960.1	77.01	-
UDP-glycosyltransferase 74G1 *Helianthus annuus*	XP_021980962.1	77.53	-
UDP-glycosyltransferase 74G1 isoform X1 *Helianthus annuus*	XP_021980974.1	75.22	-
UDP-glycosyltransferase 74G1 isoform X3 *Helianthus annuus*	XP_021980975.1	75.00	-
hypothetical protein E3N88_09839 *Mikania micrantha*	KAD6455133.1	73.96	[20]
hypothetical protein E3N88_09838 *Mikania micrantha*	KAD6455132.1	74.56	[20]

Figure 1. The domain region found in multiple sequence alignment of *UGT74G1* of *S. rebaudiana* MS007.The most consensus region was highlighted in yellow color

Maximum Likelihood Estimate of Substitution Matrix				
	A	T/U	C	G
A	-	6.26	4.42	12.20
T/U	6.68	-	11.33	5.80
C	6.68	16.07	-	5.80
G	14.07	6.26	4.42	-

Figure 2. The maximum likelihood estimates substitution matrix for MSA of *UGT74G1* nucleotide sequences.

bol which indicates that some amino acids are different between the query and subject sequences, but the residues have similar chemical properties. This was the differences between protein sequence of *UGT74G1* of *S. rebaudiana* from MS007 Malaysia (query) and protein sequence of *UGT74G1* of *S. rebaudiana* (subject) from Canada [8]. Overall, it can be observed that both corresponding alignments were quite similar in terms of the number of nucleotides with only one letter difference.

Protein domains and families

The essential step in recognizing a protein's functional property is to assess its domain and relatives. Based on the data from InterProScan, InterPro family and InterPro domain revealed the presence of IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase [12] in *UGT74G1* amino acid sequence available at position 211 to 419 and IPR035595 UDP-glycosyltransferase family, conserved site [21] at position 335 to 378 respectively. This *UGT74G1* acts during catalysis of the transfer of a glycosyl group from a UDP-sugar to a small hydrophobic molecule. Database from PROSITE with ID PS00375 UDP-glycosyltransferases signature [22] was also included. It can be seen that the domain region was considered as a conserved site and the protein involved primarily in molecular function.

Multiple sequence alignment

Multiple sequence alignment (MSA) was performed using MUSCLE to align all the candidate sequences, which helps to describe the level of conserved regions within *S. rebaudiana* MS007. From the analysis, the asterisk (*) symbol indicates the conserved residues of the protein sequences, while symbol (:) colon shows conservation between amino acid groups of similar properties and (.) indicates conservation between amino acid groups of weakly similar properties [16]. Therefore, sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides can be assessed. The highlighted region shown in Figure 1 is the domain region. Based on the guide tree, when comparing with Cluster-31069.45201 (*S. rebaudiana* MS007), Q6VAA6.1 (*S. rebaudiana*) sequence showed high consensus characteristic while the sequence, XP_021632365.1 0.0503282 (*Manihot esculenta*), showed low consensus characteristic.

The estimation of substitution matrix for MSA of nucleotide sequences of *UGT74G1* is shown in Figure 2. Substitution patterns and rates were estimated under the Tamura-Nei (1993) model [17]. Rates of different transitional substitutions are shown in bold and those of transversion substitutions are shown in italics. Relative values of instantaneous r should be considered during evaluation [23]. For simplicity, sum of r values is made equal to 100. The nucleotide frequencies are A = 28.85%, T/U = 27.05%, C = 19.08%, and G = 25.02%. For estimating maximum Log likelihood values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -11267.048. Codon positions included were $1^{st}+2^{nd}+3^{rd}$+Noncoding. All positions containing gaps and missing data were eliminated (complete deletion option). There was a total of 1,333 positions in the final dataset.

Evolutionary analyses conducted by MEGA X also show the substitution matrix's estimated values for MSA of nucleotide sequences of *UGT74G1*. Based on the result obtained, the estimation rate for transitional substitution between nucleotide C and nucleotide T/U showed the highest value of 16.07. Meanwhile, the estimation rates for two transversion substitutions between nucleotide C and nucleotide A and nucleotide C and nucleotide G resulted in the lowest value which is 4.42. The lower the value, the easier for the nucleotide to change or substitute to the other form of nucleotides and the higher the value, the harder for the nucleotide to change or substitute due to long distance between nucleotides [17].

Phylogenetic tree analysis

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the taxa analysed [17]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed.

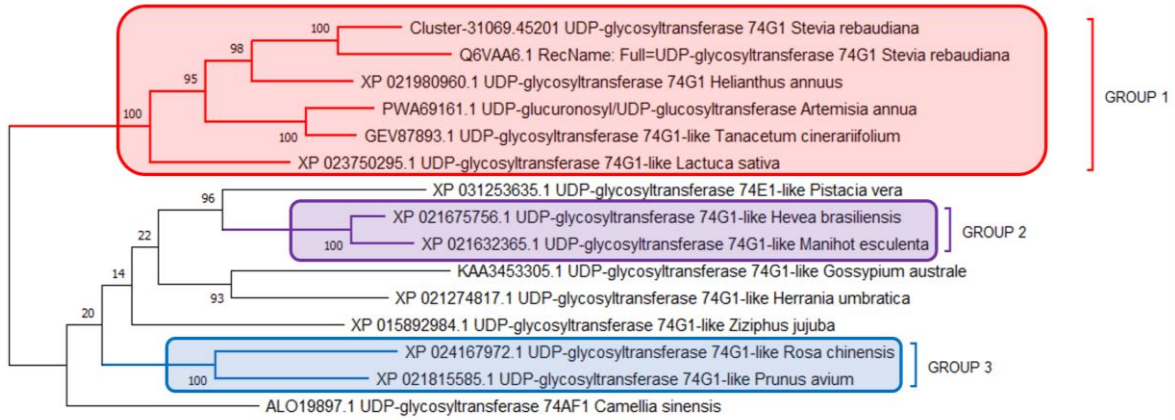


Figure 3. Maximum likelihood tree of *UGT74G1*. The branch lengths are in the same units as the developmental distances used to infer the evolution process of phylogenetic tree using the JTT model with amino acid substitutions per site

associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches [23]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 15 amino acid sequences which were selected from the homology search result. These sequences were selected as having a high percen-tage of identity among other species. All positions containing gaps and missing data were eliminated (complete deletion option). There was a total of 443 positions in the final dataset.

Figure 3 shows the phylogenetic tree by using Maximum Likelihood method. The number that is located at every branch indicates the bootstrap value, indicating the percentage occurrence of the same branch was observed when repeating the phylogenetic reconstruction on the sample data. The highest value of bootstrapping was 100 where it indicated the most similar and relatable relationship between these species based on all UGT74G1 sequences while the lowest boot-strapping value was 20. The lower the bootstrapping value, the more divergent of the species. The line segment with value of 0.01 (Figure 3) represents the length of branch which indicated an amount of genetic change of 0.01. The phylogenetic tree was divided into 3 groups based on the highest boot-strapping values.

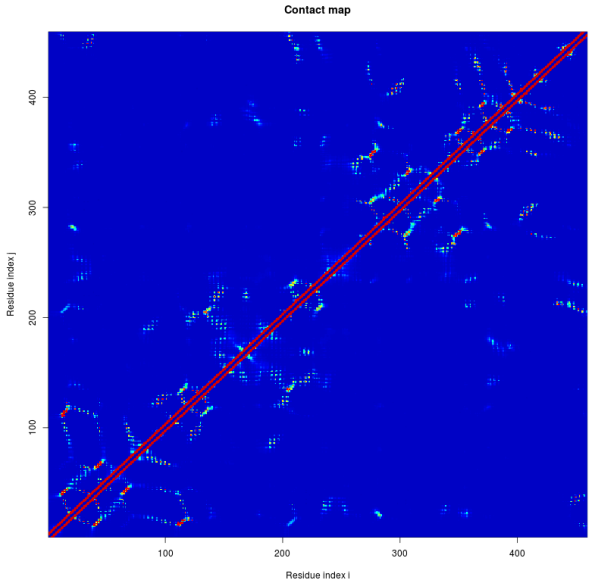
Based on Figure 3, it can be inferred that group 1 consists of UGT74G1 (Q6VAA6.1 and Cluster-31069.45201 MS007 accession) in *S. rebaudiana*, UGT74G1 in *Helianthus annuus*, UDP-glucuronosyl/UDP-glucosyltransferase in *Artemisia annua*, UGT74G1-like in *Tanacetum cinerariifoium* and UGT74G1-like in *Lactuca sativa*. This group are classified as having the same plant family, Asteraceae. The bootstrapping value for this group is 100. In *S. rebaudiana*, UGTs enzymes are involved in steviol glycosides production which is a compound that produces sweet taste [8]. UGT74G1 in *S. rebaudiana* is involve in stress tolerance, for example in water deficit condition [24]. Dehydration treatment which indicates with environmental response might decreased the UGT74G1 transcription [25].

In Asteraceae family, rapid diversification is associated with a polyploidization event [26]. *A. annua* appears as the only natural source of artemisinin, which is an effective anti-malarial compound [27]. In addition, *T. cinerariifolium* was found having the heterozygosity characteristic which is likely to be contributed to be required outcrosser [28]. Sporophytic self-incompatibility (SSI) is promoted in outcrossing in the Asteraceae family [28].

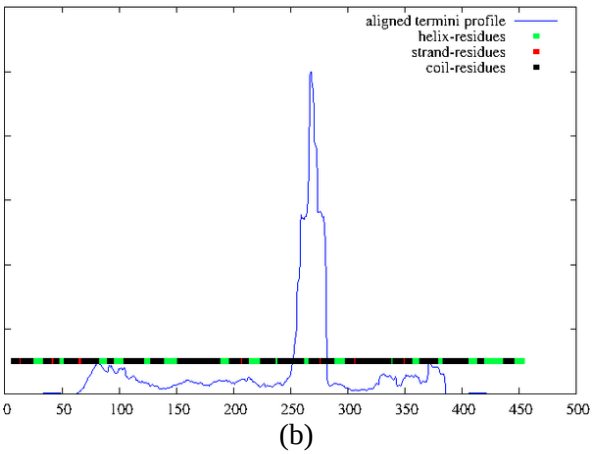
Besides, group 2 consists of UGT74G1-like in Hevea brasiliensis and *Manihot esculenta*. This group has a high bootstrapping value, 100. Both *H. brasiliensis* and *M. esculenta* are from Euphorbiaceae [29] which is a vast and diverse plant family with over 300 genera that is recognized for having a significant source of medicines and toxins [30]. H. brasiliensis also known as ru bber tree is important in the production of natural rubber. Cassava (*M. esculenta*) is an important staple food crop. Millions of people in tropical

Table 2. Physicochemical characteristics by ExPASy ProtParam for *UGT74G1 S. rebaudiana* MS007

Type	Value
Number of amino acids	459
Molecular weight	51796.23
Theoretical pI	5.67
Formula	$C_{2334}H_{3615}N_{607}O_{686}S_{20}$
Total number of atoms	7262
Extinction coefficient (EC)	88690
Extinction coefficient	87890
Instability index (II)	41.11 (Unstable)
Aliphatic index (AI)	86.58
Grand average of hydrophaticy (GRAVY)	-0.217



(a)



(b)

Figure 4. Secondary structure prediction using PSIPRED (a) DeepMetaPSICOV Contact Map (b) DOMPred results of *UGT74G1 S. rebaudiana* MS007

areas rely on this crop as a primary source of nutrition [31].

The group 3 in Figure 3 consists of UGT74G1-like in *Prunus avium* and *Rosa chinensis* which both come from Rosaceae family with bootstrapping value equal to 100. Polyploidy in the Rosaceae is associated with the reproductive biology [32]. Phenolics and anthocyanins can be provided to the total of antioxidant activity in *P. avium* [33].

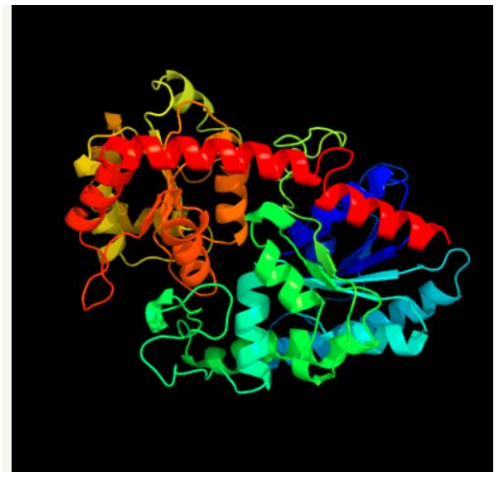
Physicochemical properties of UGT74G1

The study of physicochemical characteristics showed that the significance of the isoelectric point (pI) for *UGT74G1* MS007 proteins that varies from 5.67. These requirements are needed, especially for means of experimental handling, mainly for protein isolation and purification, to understand the condition of the protein sequence [15]. The highest extinction coefficient (EC) belonging to *UGT74G1* MS007 was 88690 M–1 cm–1 and in terms of the instability index (II) *UGT74G1* MS007 was predicted to be unstable inside a test tube. Unstable proteins require substantial steps before isolation and purification, such as denaturation [15]. Further description of the *UGT74G1* MS007 protein parameters, like molecular weight, aliphatic index, and GRAVY, are provided in Table 2.

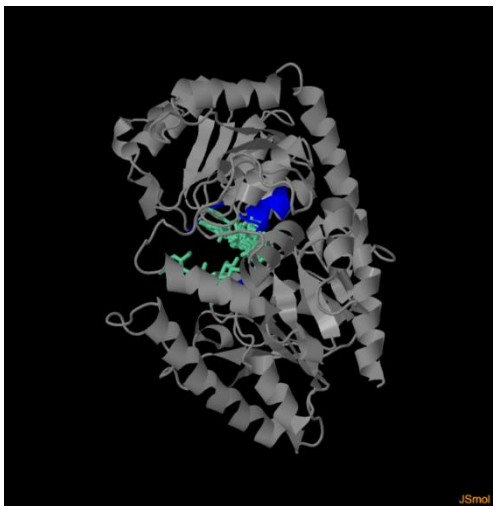
Secondary and tertiary structure prediction

Secondary structure was predicted using PSIPRED, where sequence plot, PSIPRED cartoon, DeepMetaPSICOV Contact Map and DomPred Results were obtained. Figure 4 (a) and (b), show the results obtained using PSIPRED. DeepMetaPSICOV contact maps of *UGT74G1* shows a directly proportional relationship between the x-axis and the y-axis of the graph. DOMPred results of *UGT74G1* show aligned termini profile at its peak at the scale of approximately 270.

The two-and three-dimensional structure seems to be another significant feature to be considered when reviewing protein functional properties [19]. The whole estimation uncovered a possible form or folding of a query protein from its own amino acid sequence (loops, helices, and strands). Protein structure data helps even more recognition of important protein features, such as active sites and binding ligands [19]. Besides, structural refinement is also essential to improve the predicted structure in order to minimize energy



(a)



(b)

Figure 5. (a) Tertiary structure prediction of UGT74G1 *S. rebaudiana* MS007. (b)Structural model prediction with the predicted binding site (blue) and other residues (grey).

thus providing more native protein folding [34].

The results of *UGT74G1* tertiary structure prediction were shown in Figure 5 (a) and (b). The model was based on template 6L8Z from Protein Data Bank (PDB) with the PDB title as crystal structure of UGT transferase mutant in complex with UPG [10]. The template is a glycosyltransferase molecule containing chain A. The ligand prediction in Figure 5 (b) showed the site where uridine diphosphate (UDP) - dependent glucosyltransferases will bind to the *UGT74G1* enzyme. In this study, *UGT74G1 S. rebaudiana* MS007 structure was successfully predicted. However, for the next step of study, refining, active site estimation, and structural validation might also be considered.

Conclusion

The analyses done on *UGT74G1* gene revealed that the protein is classified under UDP-glycosyltransferase 74G1-like as the closest hit sequences with support from domain or family revealed the presence and entry of IPR002213 and IPR035595. Furthermore, the sequence- and structure- based assessments also showed that *UGT74G1* predicted structure scored 100.0% confidence with the template 6L8Z in PDB database and coverage of 96%. The model has dimensions (Å) of X: 57.609, Y: 70.386, and Z: 58.351 and was the best candidates to be further studied. This research has successfully filled the knowledge void of previously un-annotated essential of *UGT74G1* proteins in *S. rebaudiana* MS007 that use *in-silico* sequence-and structure-based strategies. It would be a great advantage to use this information in the ability to manipulate, the said structure, in steps to generate artificial sweeteners.

Acknowledgement

We sincerely thank the staffs and administration of Kulliyyah of Science, International Islamic University Malaysia for their assistance and support in this study.

References

1. Ashwell M (2015) Stevia, Nature's Zero-Calorie Sustainable Sweetener: A New Player in the Fight Against Obesity. Nutrition Today 50 (3): 129–134. doi:10.1097/nt.0000000000000094
2. PureCircle Stevia Institute (2019) Where Does Stevia Come from? https://www.purecirclesteviainstitute.com/resources/infographics/stevia-facts/where-does-stevia-come-from/. Accessed date: October 2019.
3. Razali A, Samsulrizal NH, Zainuddin Z (2020) Identification of genes involved in flowering in *Stevia rebaudiana* using expressed sequence tags (ESTs). Asia-Pacific Journal of Molecular Biology and Biotechnology 28 (2): 105–112. doi: 10.35118/apjmbb.2020.028.2.09
4. Aranda-González I, Betancur-Ancona D, Chel-Guerrero L, Moguel-Ordóñez Y (2017) Effect of Different Drying Methods on the Composition of Steviol Glycosides in *Stevia rebaudiana* Bertoni leaves. International Agrophysics 31 (1): 139 – 144. doi: 10.1515/intag-2016-0036
5. Samsulrizal NH, Zainuddin Z, Noh AL, Sundram TC (2019) A Review of Approaches in Steviol Glycosides Synthesis. International Journal of Life Sciences and Biotechnology 2 (3): 145–157. doi: 10.38001/ijlsb.577338
6. Osman HS, Osman M, Zainuddin Z (2018) Genetic variabilities of *Stevia rebaudiana* Bertoni Cultivated in Malaysia as Revealed by Morphological, Chemical and Molecular Characterisations. Agrivita 40 (2): 267-283. doi: 10.17503/agrivita. v40i2.1365
7. Pang MD, Goossens GH, Blaak EE (2020) The Impact of Artificial Sweeteners on Body Weight Control and Glucose Homeostasis. Frontiers in Nutrition 7.

8. Richman A, Swanson A, Humphrey T et al. (2005) Functional Genomics Uncovers Three Glucosyltransferases Involved in The Synthesis of The Major Sweet Glucosides of *Stevia rebaudiana*. Plant Journal 41(1): 56–67. doi:10.1111/j.1365-313X. 2004.02275.x.
9. Samsulrizal NH, Khadzran KS, Shaarani SH et al. (2020) De novo Transcriptome Dataset of *Stevia rebaudiana* Accession MS007. Data in Brief 28 (104811). doi.org/10.1016/j.dib.2019.104811.
10. Artimo P, Jonnalagedda M, Arnold K et al. (2012) ExPASy: SIB Bioinformatics Resource Portal. Nucleic Acids Research 40 (W1): W597-W603. doi: 10.1093/nar/gks400
11. NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 46 (D1): D8–D13. doi: 10.1093/nar/gkx1095
12. Blum M, Chang H, Chuguransky S et al. (2020) The InterPro Protein Families and Domains Database: 20 years on. Nucleic Acids Research 49 (D1): D344 – D354. doi: 10.1093/nar/gkaa977
13. Mistry J, Chuguransky S, Williams L et al. (2020) Pfam: The Protein Families Database in 2021. Nucleic Acids Research 4 (D1): D412 – D419.doi: 10.1093/nar/gkaa913
14. Letunic I, Khedkar S, Bork P (2020) SMART: Recent Updates, New Developments and Status in 2020. Nucleic Acids Research 49 (D1): D458-D460. doi: 10.1093/nar/gkaa937
15. Gasteiger E, Hoogland C, Gattiker A et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook. Humana Press: 571-607. Accessed date: July 2020.
16. Madeira F, Park YM, Lee J et al. (2019) The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. Nucleic Acids Research 47 (W1): W636-W641. doi: 10.1093/nar/gkz268.
17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35: 1547-1549. doi:10.1093/molbev/msy096
18. Buchan DWA, Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research 47 (W1): W402 – W407. doi: 10.1093/nar/gkz297
19. Kelley LA, Mezulis S, Yates CM et al. (2015) The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. Nature Protocols (10): 845 – 858. doi: 10.1038/nprot.2015.053
20. Liu B, Yan J, Li W et al. (2020) *Mikania micrantha* Genome Provides Insights Into the Molecular Mechanism of Rapid Growth. Nature Communications 11 (1). doi: 10.1038/s41467-019-13926-4
21. Bock KW (2016). The UDP-glycosyltransferase (UGT) Superfamily Expressed in Humans, Insects and Plants: Animal Plant Arms-Race and Co-evolution. Biochemical pharmacology 99: 11-17.
22. Sigrist CJA, de Castro E, Cerutti L et al. (2020) New and Continuing Developments at PROSITE. Nucleic Acids Research doi: 10.1093/nar/gks1067
23. Mello B (2018) Estimating Timetrees with MEGA and the TimeTree resource. Molecular Biology and Evolution 35 (9): 2334-2342.
24. Behroozi P, Baghizadeh A, Saei A, Kharazmi S (2017) Quantitative Analysis of Uridine Diphosphate Glycosyltransferase UGT85C2, UGT74G1 and UGT76G1 Genes Expression in *Stevia rebaudiana* Under Different Irrigations. Russian Journal of Plant Physiology 64 (1): 67–72. doi: 10.1134/S1021443717010034
25. Xiang Z, Tang X, Liu W, Song C (2019) A Comparative Morphological and Transcriptomic Study on Autotetraploid *Stevia rebaudiana* (Bertoni) and Its Diploid. Plant Physiology and Biochemistry 143 (June): 154–164. doi: 10.1016/j.plaphy.2019.09.003.
26. Scaglione D, Reyes-chin-wo S, Acquadro A et al. (2016) The Genome Sequence of The Outbreeding Globe Artichoke Constructed De Novo Incorporating a Phase-Aware Low-Pass Sequencing Strategy of F1 Progeny. Nature Publishing Group (August 2015): 1–17. doi: 10.1038/srep19427
27. Shen Q, Zhang L, Liao Z et al. (2018) The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis: 1–13. doi: 10.1016/j.molp.2018.03.015
28. Malenica N, Mccann J, Mlinarec J et al. (2019) The Repetitive DNA Composition in the Natural Pesticide Producer *Tanacetum cinerariifolium*: Interindividual Variation of Subtelomeric Tandem Repeats 10 (May): 1–14. doi: 10.3389/fpls.2019.00613
29. Maghuly F, Vollmann J, Laimer M (2015). Biotechnology of Euphorbiaceae (*Jatropha curcas*, *Manihot esculenta*, *Ricinus communis*). In Applied Plant Genomics and Biotechnology. doi: 10.1016/B978-0-08-100068-7.00006-9
30. Ramalho SD, Pinto MEF, Ferreira D, Bolzani VS (2018) Biologically Active Orbitides from the Euphorbiaceae Family. Planta Medica 84 (9–10): 558–567. doi: 10.1055/s-0043-122604
31. McCallum EJ, Anjanappa RB, Gruissem W (2017) Tackling agriculturally relevant diseases in the staple crop cassava (*Manihot esculenta*). Current Opinion in Plant Biology 38: 50–58. doi: 10.1016/j.pbi.2017.04.008
32. Dickinson TA, Lo E, Talent N (2007) Polyploidy, Reproductive Biology, and Rosaceae: Understanding evolution and making classifications. Plant Systematics and Evolution 266 (1–2): 59–78. doi: 10.1007/s00606-007-0541-2
33. Usenik V, Fabčič J, Štampar F (2008) Sugars, Organic Acids, Phenolic Composition and Antioxidant Activity of Sweet Cherry (*Prunus avium* L.). Food Chemistry 107 (1): 185–192. doi: 10.1016/j.foodchem.2007.08.004.
34. Rydzewski J, Jakubowski R, Nowak W (2015) Communication: Entropic Measure to Prevent Energy Over-Minimization in Molecular Dynamics Simulations. Journal of Chemical Physics 143 (17). doi: 10.1063/1.4935370