Malaysian Journal of Chemistry, 2023, Vol. 25(3), 150-163
(Special Issue: *2nd Applied Science Research International Conference 2022*)

In Silico Structure Prediction and Molecular Docking Studies of the DGAT1_2 protein from *Elaeis guineensis* with Oleoyl-CoA

Jamunaa Ravindran[1] and Zaidah Rahmat[1,2*]
[1]Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[2]Institute of Bioproduct Development, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
*****Corresponding author (e-mail: zaidahrahmat@utm.my)

Diacylglycerol acyltransferase (DGAT) is a key enzyme that catalyses the last step of TAG biosynthesis, modulating lipid accumulation in plants. Parallel to growing demands for vegetable oils, *Elaeis guineensis* DGAT Type 1 (EgDGAT1) overexpression has been reported to improve fatty acid content, which could possibly serve as a target to enhance palm oil yield. As the EgDGAT1_2 isoform has been functionally characterised in a previous study, EgDGAT1_2 protein modelling and molecular docking were performed in this study to predict the interacting residues with oleoyl-CoA. In the absence of x-ray crystal structures, there is a need to unravel the three-dimensional (3D) structural conformations to study the binding mechanisms of EgDGAT1_2. Physical and chemical properties of EgDGAT1_2 were computed, followed by 3D structure prediction using the I-TASSER server. The model was validated using the ProSA web server, ERRAT program and PROCHECK server. The molecular docking analysis of EgDGAT1_2 with oleoyl-CoA during both blind docking and specific site docking showed favourable binding energies of -8.1 kcal/mol and -8.2 kcal/mol, respectively, and several residues were identified as potential interacting residues. Information from this study can be exploited for the molecular engineering of the native DGAT enzyme to enhance palm oil composition and yield further.

Keywords: Diacylglycerol acyltransferase; *Elaeis guineensis*; oil biosynthesis; protein model prediction; docking

Received: August 2022; Accepted: October 2022

Increased demand for vegetable oils, propelled by a growing human population, has prompted a variety of measures to boost oil crop yield. Numerous types of molecular research are being conducted to understand the process behind oil accumulation in various parts of plants, particularly in leaves [1], seeds [2], and fruits, as boosting oil accumulation in plants would be a profitable strategy to address this issue [3]. Due to its ability to produce more oil per unit of land than any other vegetable oil crop, African oil palm, also known as *Elaeis guineensis*, is being acknowledged as the most effective oil crop [4]. The two oil kinds extracted from oil palm, the mesocarp-derived palm oil and the seed-derived palm kernel oil, are exceptionally rich in triacylglycerols (TAG) of various chain lengths [5].

TAG biosynthesis is vastly studied and well-documented in many oil-bearing plants, contributing insights into the components and complex mechanisms involved in lipid metabolism [6-8]. Triacylglycerol production is initiated by *de novo* fatty acid bio-synthesis in the plastids and is followed by the Kennedy pathway's acylation of glycerol-3-phosphates (G3P) in the endoplasmic reticulum. Acyl-CoA molecules are utilised as a source of fatty acyl components for sequential acylation of G3P into phosphatidic acid (PA) which is then hydrolysed into diacylglycerol (DAG). In the final committed step, diacylglycerol acyltransferase (DGAT), which carries an acyl-CoA binding motif and DAG binding motif, uses acyl-CoA as an acyl donor substrate to convert DAG into TAG, driving TAG assembly and accumulation in plants [9,10]. Given that it controls the type and amount of acyl-CoA flux into TAG synthesis, DGAT is proposed as the rate-limiting enzyme for oil production in plants [11].

DGAT1, DGAT2, DGAT3, and wax ester synthase (WS)/DGAT are among the four types of DGAT identified until now. However, DGAT1 and DGAT2 are the ones most studied due to their significant role as membrane proteins in the endoplasmic reticulum. DGAT1 is mostly reported to be expressed in seeds leading to oil accumulation. Following DGAT1 overexpression in *Brassica juncea* [12], *Gossypium hirsutum* [13], and *Glycine max* [14], a substantial rise in oil content was seen, emphasising the enzyme expression to be seed-specific. DGAT2 is often linked to the production of uncommon fatty acids such as in *Vernicia fordii* producing large quantities of eleostearic

†Paper presented at the 2nd Applied Science Research International Conference 2022 (2nd ASRIC 2022)

acid [15], in *Ricinus communis* yielding ricinoleic acid (hydroxy fatty acids) [16] and in *Vernonia galamensis* producing vernolic acid (epoxy fatty acid) [17]. DGAT3 soluble enzyme represented a much more recent discovery than DGAT1 and DGAT2 and was reported to be expressed mainly in the cytosol, as revealed in Arabidopsis [18] and peanuts [19]. The bifunctional WS/DGAT has been characterised in *Acinetobacter* sp. [20] and *Arabidopsis* [21], suggesting its possible function mainly involved wax synthase activity, while also performing diacylglycerol acyltransferase activity in some organisms.

Elaeis guineensis DGAT (EgDGAT) has emerged as one of the essential genes governing oil accumulation in oil palm, and it could be a candidate gene for biotechnological techniques aimed at improving oil quality and quantity. The discovery of the genetic architecture of EgDGAT isoforms is a significant accomplishment in oil palm molecular research, allowing for a better knowledge of protein functional roles in improving oil yield [22]. EgDGAT1_1 transcript was discovered to be concentrated in the palm endosperm and embryo, whereas EgDGAT1_2 and EgDGAT2 transcripts were found to accumulate in the mesocarp, indicating that TAG biosynthesis occurs primarily in the fruits of oil palm [23,24]. EgDGAT1_1 (annotated EgDGAT1_2 in this work) expression has led to the accumulation of lauric acid in the seed endosperm, implying its role in storing fatty acids of the medium chain [25].

Many complex systems and gene expressions are involved in TAG production and oil build-up in plants. However, in the oil palm, higher transcription of DGAT1 has been shown to enhance TAG content in the seeds. Assays of substrate specificity revealed that DGAT1 prefers to bind to oleoyl-CoA and dioleoyl glycerol [26-28]. Because of its critical role in controlling oil accumulation, EgDGAT1 is a good candidate for molecular engineering to boost oil output. Given the diverse isoforms of EgDGAT1 proteins expressed at different sections of the fruits and at different developmental phases, understanding structural information and protein-ligand interactions are critical. A previous study reported on the substrate specificity of EgDGAT1_2, followed by a prediction of ten putative transmembrane domains and conserved motifs based on sequence alignment [25]. Unfortunately, the structural conformations of EgDGAT1 isoforms have not been determined, and the underlying interactions of EgDGAT1 protein isoforms' interactions with oleoyl CoA are still unknown. Therefore, the goal of our current research is to anticipate the three-dimensional molecular structure of EgDGAT1_2, and its interactions with oleoyl-CoA using computational approaches, as a previous study reported that oleic acid was one of the highly expressed fatty acids upon EgDGAT1_2 heterologous expression in *Yarrowia lipolytica* mutant strain [25]. Blind docking and specific docking were performed between EgDGAT1_2 and oleoyl-CoA to compare the potential interacting residues for enzyme catalysis.

MATERIAL AND METHODS

Primary Sequence and Secondary Structure Analysis

The amino acid sequence of EgDGAT1_2 (gene_id: p5.00_sc00018_p0155) was retrieved from the palm-xplore database (http://palmxplore.mpob.gov.my). The ProtParam tool via the ExPAsy server [29] was used to calculate the chemical and physical properties of the protein. The protein secondary structure was estimated by a self-optimised prediction method with an alignment (SOPMA) server [30]. Sequence alignment was executed using Clustal omega [31] and Espript 3 [32] using EgDGAT1_2 and *Homo sapiens* DGAT1 (HsDGAT1) (PDB ID: 6VP0_C) sequences.

Protein Model Prediction and Model Validation

The tertiary structure of the protein was predicted using the I-TASSER [33] protein modelling web server and was saved in PDB format for further analysis. The generated model was quality-checked using the ProSA-web server [34] and ERRAT program [35]. The stereo-chemical qualities of the model were evaluated using a Ramachandran plot via the PROCHECK server [36].

Molecular Docking Studies

Water molecules were removed, Gasteiger charges were computed, AutoDock 4 atoms were assigned to the modelled protein structure, and the files were translated to PDBQT format using Autodock tools (ADT). The structure of oleoyl-CoA was obtained in SDF format from the RCSB PDB database (PDB ID: 3VV), converted into PDB format and assigned with Gasteiger charges. AutoDock Tools were used to convert the PDB format ligand structures to PDBQT format. Docking studies were executed using AutoDock Vina version 1.1.2 [37]. Blind docking was performed by adjusting the grid box to cover the whole protein structure, while specific site docking was performed by adjusting the grid box to cover corresponding interacting residues reported for HsDGAT1. The size and coordinates of the grid box corresponding to expected binding sites were generated by the Auto-Grid tool. A spacing of 0.375Å and grid box coordinates of 76.521, 76.967 and 76.668 in *x*, *y* and *z*, respectively, were set for blind docking. For specific site docking, a spacing of 0.375Å and grid box coordinates of 60.569, 56.274 and 61.384 in *x*, *y* and *z*, respectively, were set. Grid and configuration files were prepared with the information on grid box size and coordinates. The exhaustiveness value was set to 8. The AutoDock Vina program was run to dock EgDGAT1_2 protein with oleoyl-CoA using the command-line interface. AutoDock Vina 1.1.2, BIOVIA Discovery Studio 2021 Client, and UCSF Chimera

version 1.15 were used to view and analyse the docking postures with the lowest RMSD and binding energy (kcal/mol). The binding site for EgDGAT1_2 was also predicted using the COACH meta server for comparison of interacting binding residues [38]. The interacting residues from both docking types were compared with the HsDGAT1_2-oleoyl-CoA complex (PDB ID: 6VP0_C), as it is the only crystal structure resolved for DGAT1 protein.

RESULTS AND DISCUSSION

Primary Sequence and Secondary Structure Analysis

Totalling 512 amino acid residues in length, the DGAT1_2 enzyme from *E. guineensis* is a large protein. Using the Expasy Protparam tool, the physicochemical parameters of EgDGAT1_2 were calculated, and the findings are summarised in Table 1. Protparam analysis of the primary sequence revealed that the enzyme contains 8200 total atoms corresponding to 57726.11 Daltons of molecular weight with a molecular formula of $C_{2662}H_{4138}N_{668}O_{700}S_{32}$. The high theoretical pI (9.01) indicates that the enzyme exists in its neutral form at pH 9.01 only. The basic nature of the EgDGAT1_2 protein is in parallel with the increased number of basic residues (Arg and Lys = 49) than acidic residues (Asp and Glu = 36) in the primary sequence. The instability index of EgDGAT1_2 showed that the protein is unstable (48.99) *in-vitro* and has a high aliphatic index (104.65), indicating thermo-stability, which correlates to the capability of oil palm trees to withstand high temperatures [39,40].

Positively valued GRAVY (0.333) illustrates a protein's hydrophobic feature, which corresponds to the presence of a greater number of hydrophobic residues (285 residues), indicating that the protein has hydrophobic domains. EgDGAT1_2 is predicted to be a membrane-bound protein localised in the endoplasmic reticulum [22]. The presence of both hydrophobic and hydrophilic amino acids (130 residues) indicates that the proteins are amphipathic and could potentially work as transmembrane proteins [41]. A transmembrane protein spanning the lipid bilayer contains segments exposed to aqueous spaces and the hydrophobic core of the membrane. The hydrophobic core of integral membrane proteins consists of one or more α-helical regions mainly composed of hydrophobic residues exposed on the protein surface that interact and are submerged in the hydrophobic interior of the membranes [42,43]. The hydrophilic regions are mainly composed of hydrophilic, polar residues exposed to the cytosol or extracellular fluid for interaction with an aqueous phase.

Secondary structure prediction using the SOPMA server revealed that the EgDGAT1_2 protein is dominated by 45.12% (231 residues) alpha helices followed by 39.45% (202 residues) random coils, along with 12.11% (62 residues) extended strand and 3.32% (17 residues) beta turns. Helix packing influences membrane protein folding, stability, and interaction. Protein folding and deposition of the transmembrane portion into the biological membrane is influenced by the hydrogen bonding network of polar peptide bonds, which also stabilises α-helices. When a hydrophobic sequence is inserted into a membrane bilayer, helices spontaneously form because of the negative free energy associated with hydrogen bonding of the polar backbone carbonyls and amide groups [44]. Coils provide flexibility for proteins and are significant for protein function. By forming bridges between different secondary structure elements, amino acid residues from coil areas may help stabilise secondary structure elements that are located nearby [45].

Table 1. Chemical and physical properties of EgDGAT1_2 protein determined by the ExPASy's ProtParam tool.

Properties	EgDGAT1_2
Length (amino acids)	512
Molecular weight (Da)	57726.11
Molecular formula	$C_{2662}H_{4138}N_{668}O_{700}S_{32}$
Total number of atoms	8200
Theoretical pI	9.01
Positively charged residues (RK)	49
Negatively charged residues (DE)	36
Polar residues (NCQSTY)	130
Hydrophobic residues (AILFWVGPM)	285
Instability index	48.99
Aliphatic index	104.65
Grand average of hydropathicity (GRAVY)	0.333

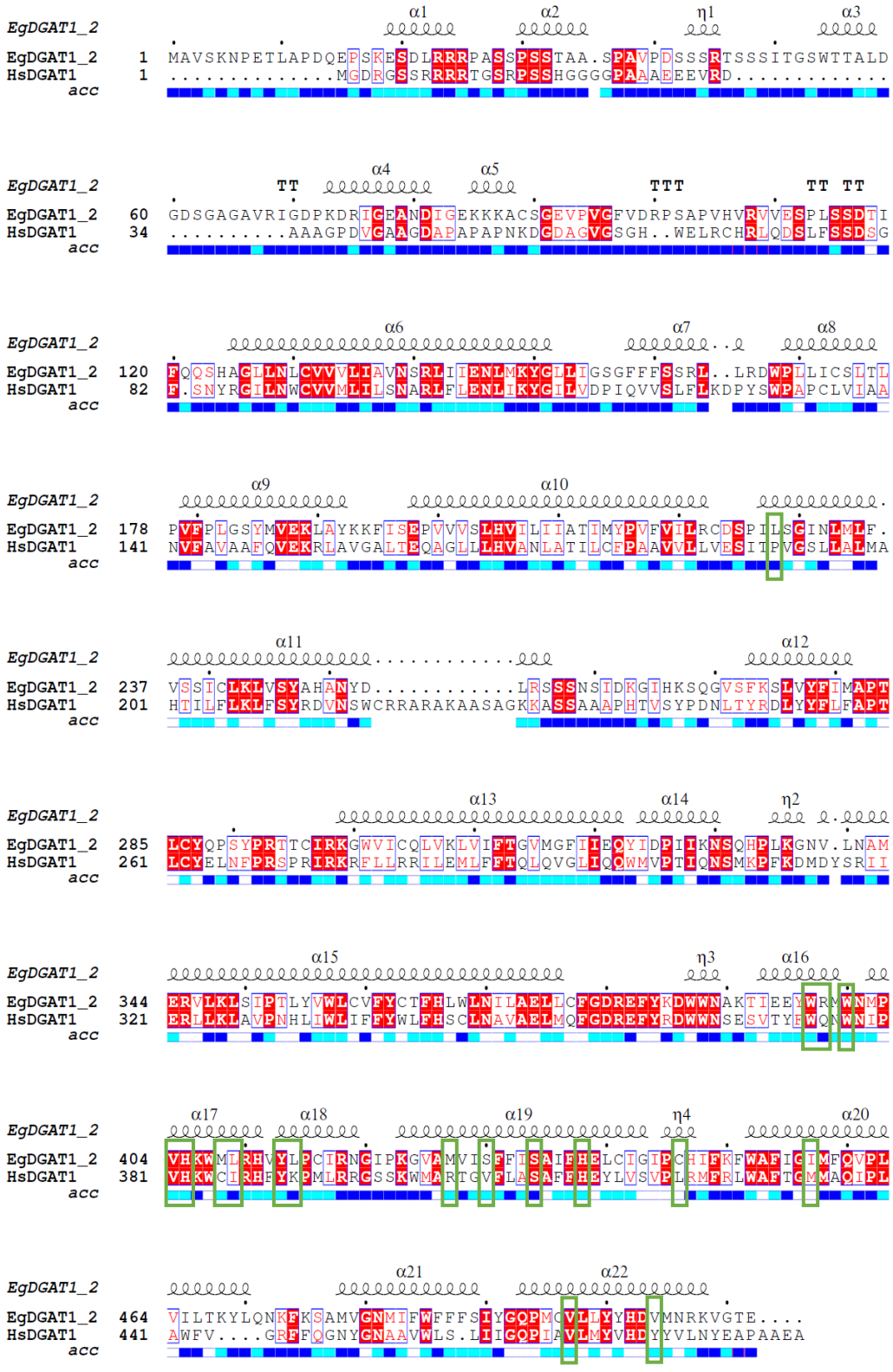


Figure 1. Sequence alignment of EgDGAT1_2 and HsDGAT1. Green boxes indicate HsDGAT1 interacting residues in the oleoyl-CoA binding and the corresponding residues in EgDGAT1_2.

A BLASTP program search against Protein Data Bank (PDB) revealed that EgDGAT1_2 has a 36.56% sequence identity with Human DGAT1 (PDB ID: 6VP0), which has been solved to a resolution of 3.1 Å. From the PDB structure of HsDGAT1, Pro191, Trp374, Gln375, Trp377, Val381, His382, Cys385, Ile386, Tyr390, Lys391, Arg404, Val407, Ser411, His415, Leu423, Met434, Val469 and Tyr476 were identified as interacting residues with oleoyl-CoA, and were compared with EgDGAT1_2 residues to establish

potential interacting residues for docking studies. An alignment of the two sequences revealed EgDGAT1_2 shares eight out of 18 common interacting residues (boxed in green) with HsDGAT1 to proceed with molecular docking (Figure 1). A conserved histidine residue common to most DGAT proteins in plants and mammals located near the C terminus of the enzyme was also detected in both sequences. In bovine DGAT1, the histidine (His416) residue is located on the extra-membranous loop and was proposed as the catalytic residue [46].

Protein Model Prediction and Model Validation

An attempt was made to model the EgDGAT1_2 protein as there is only one entry in the PDB database on the allosteric site of DGAT1 from *Brassica napus* (PDB ID: 5UZL) which had a sequence similarity of 65.62% with EgDGAT1_2 but with only 6% query cover [47]. However, no protein crystal structures are resolved on DGAT1 whole protein from plant sources, which makes this paper the first to carry out DGAT1 protein modelling. The three-dimensional (3D) structures of the EgDGAT1_2 protein was generated using the I-TASSER server. I-TASSER reassembles multiple threading fragments identified by LOMETS into full-length models to generate five final models with a higher confidence level, as measured by C-score [48]. The model with the highest C-score was selected for structural analysis and docking studies. C-score ranges from -5 to 2, whereby a higher C-score value corresponds to a better model quality [49]. Since oil palm DGAT1_2 is a membrane protein which is a member of the membrane-bound O-acyltransferase (MBOAT) super-family that are present in all kingdoms of life, the modelled protein is expected to acquire some structural similarity to the experimentally determined 3D structure of human DGAT1.

The EgDGAT1_2 protein appears to adapt a helix-loop-helix structure, as shown in Figure 2. This conformation can also be seen in the secondary structure analysis by YASARA software. From the model, EgDGAT1_2 protein lacks β-sheets, but is made up of a high percentage of α-helices (62.9%). Proteins having more α-helices are considered to be thermostable and often involved with lipid membrane interactions [50]. The protein is composed of 31.6% of coils followed by turns of 5.5% in the structure. In general, many integral membrane proteins have a bundle of hydrophobic α -helices in their transmembrane region, which corresponds to the structural conformation of EgDGAT1_2 protein [51].

The selected model was evaluated for quality and reliability using ERRAT, PROCHECK and ProSA-web server. ProSA-web server predicts the energy profiling of modelled structures in terms of *Z*-scores. The predicted *Z*-score was −3.94 (Figure 3), which falls within the rank of values perceived for native protein structures of similar sizes which are resolved experimentally (*z*-score ≤ 10), indicating the *Z*-score of this model is comparable to the scores of X-ray crystal structures. This data suggests that the model is of good quality. Statistical analysis of ERRAT shows that the overall quality of the modelled protein is 95.38% (Figure 4), which indicates the model is of high resolution (≥95%). The predicted model's stereochemical quality, which was examined by a Ramachandran plot from PROCHECK, identified that most of the residues (80.5%) are located in the most favoured region, whereas 17.8% residues are in the allowed region and 1.8% are in the disallowed region (Figure 5), further assuring that the predicted model is good and reliable. Altogether, this signifies that the constructed model is reliable and trustworthy for further docking experiments.

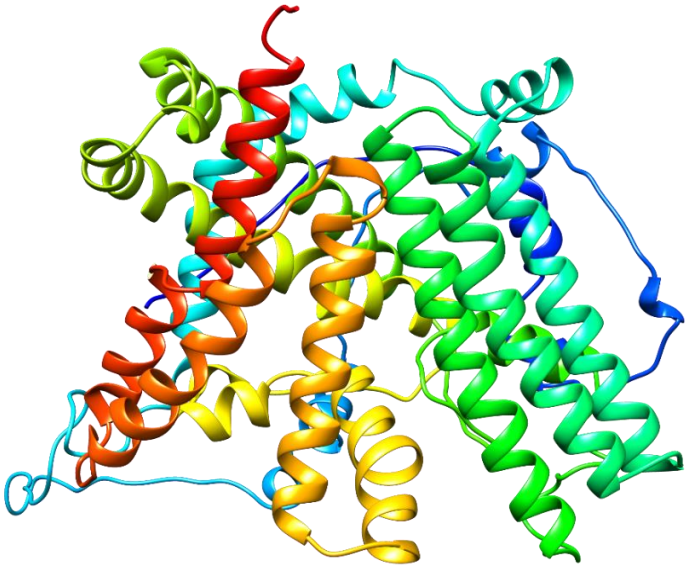


Figure 2. The EgDGAT1_2 modelled structure overview represented by rainbow colours.

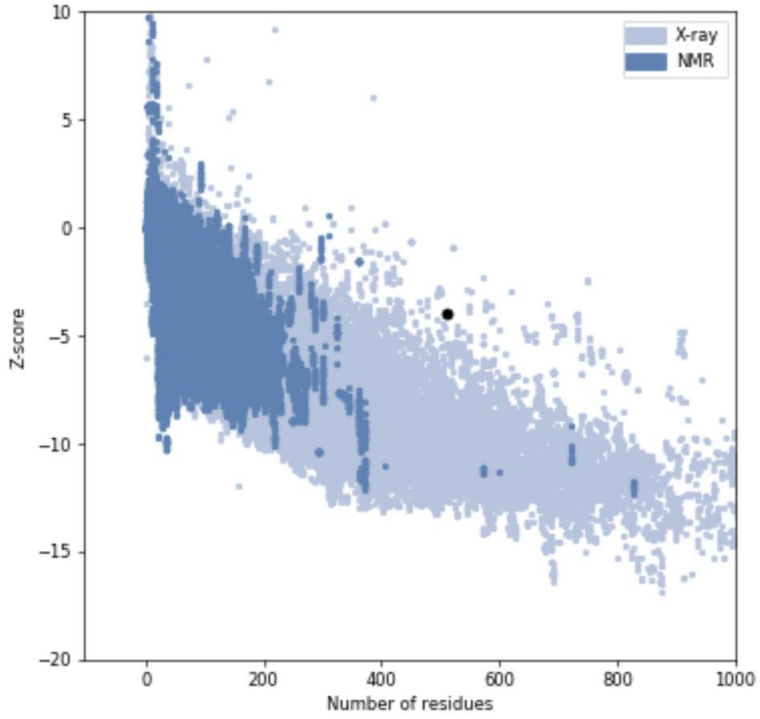


Figure 3. The ProSA-web *Z*-score of modelled EgDGAT1_2 is determined with respect to length.

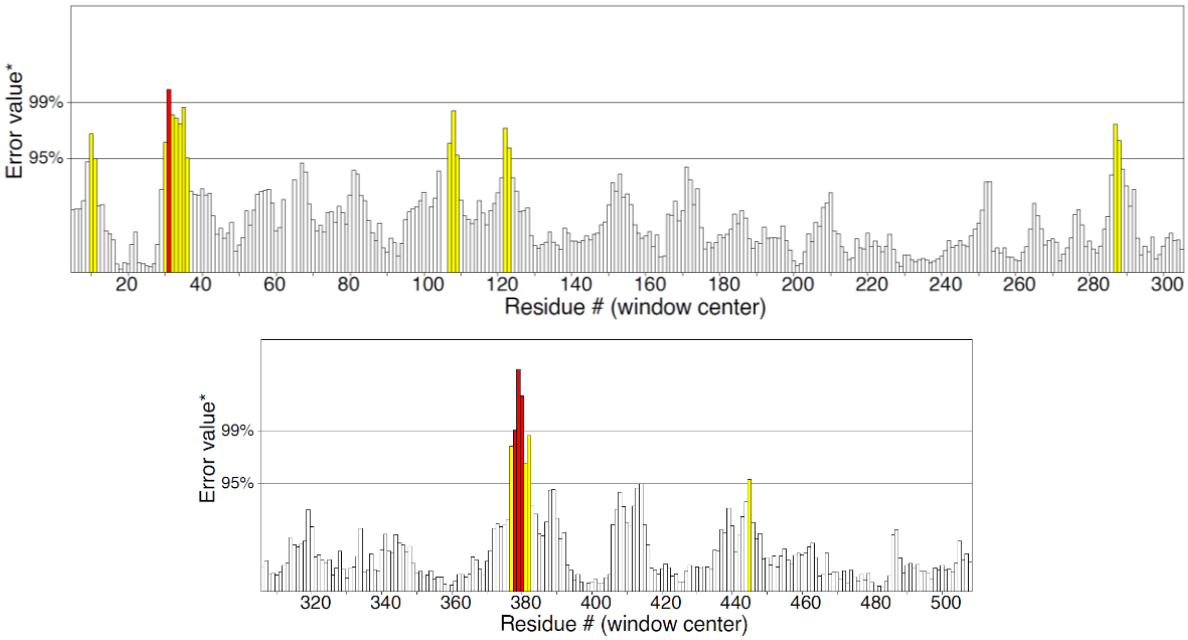


Figure 4. The ERRAT quality verification plot of modelled EgDGAT1_2.

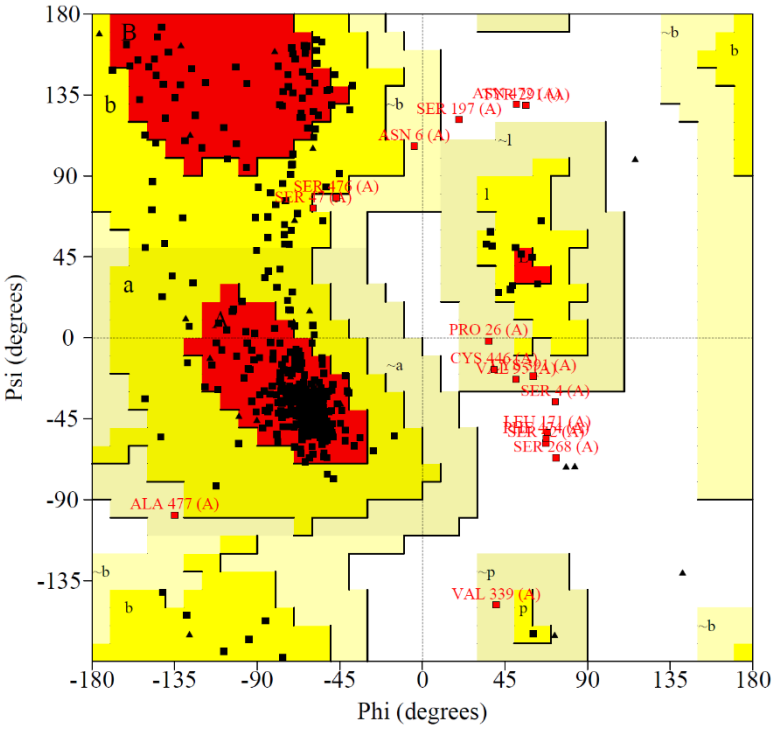


Figure 5. The Ramachandran plot of a polypeptide backbone torsion angles psi against phi of amino acid present in the modelled EgDGAT1_2. Red, yellow, pale yellow and white areas represent the most favoured, additionally allowed, generously allowed and disallowed regions, respectively. Black squares represent non-glycine and proline residues, while black triangles represent glycine (non-end) residues.

Molecular Docking Studies

Computational docking studies contribute a vital role in gaining insights into receptor-ligand binding interactions at the atomic level [52]. Blind docking involves docking without prior information regarding the binding region. When the binding site is actually known before docking ligands into it, specific site docking is used [53]. Blind docking and specific site docking were performed on EgDGAT1_2 and oleoyl-CoA to compare the interacting binding residues. Complexes with the lowest binding energy were selected for further analysis as it represents a basic estimation of a favourable substrate affinity towards the enzyme for enzyme-substrate interactions.

A previous study predicted the putative acyl-CoA binding site to be located closer to the N-terminal region [25]. In comparison with *Brassica napus* DGAT1 (BnaDGAT1), the reported binding site of acyl-CoA is located towards the end of the N-terminal region, demonstrating the function of the N-terminal is in enzyme regulation rather than catalysis [47]. The same was reported in a recent study whereby the N-terminal region in *Chromochloris zofingiensis* DGAT1 (CzDGAT1) was reported to not be required for enzyme catalysis but is vital for sustaining favourable enzyme activity and may serve as a regulatory domain rather than as an acyl-CoA binding site [54]. As previous studies reported that full-length enzyme exhibits more normal enzyme activity in CzDGAT1 and BnaDGAT1 than proteins with truncated regions [47,54], therefore, it is rational to look for the potential interacting residues involving whole EgDGAT1_2 protein with oleoyl-CoA as substrates to gain further functional and structural insights. Therefore, blind docking was performed using an EgDGAT1_2 protein and oleoyl-CoA, a preferred substrate, to predict the possible interacting residues with the ligand. The binding affinities of oleoyl-CoA towards EgDGAT1_2 for both docking types are presented in Table 2.

Docking studies reveal that oleoyl-CoA interacted with 15 various residues when binding sites were not specified. The identified interacting residues were comparable to the previously resolved DGAT1 from *Homo sapiens* which reported 18 residues (Pro191, Trp374, Gln375, Trp377, Val381, His382, Cys385, Ile386, Tyr390, Lys391, Arg404, Val407, Ser411, His415, Leu423, Met434, Val469 and Tyr476) as interacting residues with oleoyl-CoA, which affirms that the substrate in EgDGAT1_2 is close to the enzyme catalytic site. In EgDGAT1_2 blind docked complex, up to six interacting residues (Trp397, Trp400, Leu409, Tyr413, Met427, His438) were observed to correspond to reported HsDGAT1 interacting residues in the sequence alignment. In addition to these six residues, there are 12 more possible residues predicted to be involved in oleoyl-COA binding.

In the blind-docked structure of EgDGAT1_2, His438 (corresponding to His415 of HsDGAT1 in sequence alignment) interacts with the acyl part of oleoyl-CoA by forming a hydrogen bond, which appears to occupy the tunnel-like region of the membrane protein, as shown in Figure 6. In HsDGAT1, the conserved catalytic histidine residue (His415), which is buried inside the central cavity of the membrane protein, is being stabilised by the formation of a hydrogen bond with the acyl region of the ligand. It was reported that hydrogen bond formation between oleoyl-CoA and HsDGAT1 catalytic histidine residues induces a conformational change to DGAT1 active centre upon oleoyl-CoA binding to accommodate the acyl-CoA substrate in the membrane cavity [55]. Histidine residues were also reported as important residues in enzyme function and catalysis in human acyl-CoA: cholesterol acyltransferase (ACAT) [56,57]. Due to the reactive imidazolium nitrogen atom's role as a hydrogen bond donor/acceptor, histidine residues are commonly engaged in enzyme catalysis [58].

It is important to note that His438 of EgDGAT1_2 was found to be interacting closer to the thioester bond of oleoyl-CoA where Trp400 of EgDGAT_2 interacts electrostatically (4.89Å) (Figure 7a). Trp377 in HsDGAT1 (corresponding to Trp400 in EgDGAT1_2) was observed to be closer to the thioester bond and His415 of HsDGAT1 followed by a scissile bond in the thioester group during the acyl-transfer reaction [55], implying the possibility of these two residues in EgDGAT1_2 exhibiting catalytic activity. Several hydrophobic residues were also observed to interact, such as methionine, phenylalanine and tryptophan, which would strongly bind low water-soluble molecules such as acyl-CoAs.

Specific site docking was performed using the same protein and ligand to ascertain the interacting residues. Arg398, Trp400, Met408 and Ser434 His438 were identified as interacting residues corresponding to reported HsDGAT1 interacting residues in sequence alignment, which suggests that these residues are vital catalytic residues of EgDGAT1_2. Apart from that, Ser239, His438, Cys441, His405 and Trp400 are detected as interacting residues in both blind and specific docking, which indicates that these residues have potential roles in oleoyl-CoA binding and catalysis by EgDGAT1_2. However, unlike blind docking, His438 and Trp400 were spotted far away from the thioester bond of oleoyl-CoA, but Ser239 was observed closer by forming a hydrogen bond with the docked oleoyl-CoA near the thioester bond (Figure7b). This could probably be due to different orientations of the oleoyl-CoA binding pose in blind and specific site docking, as shown in Figure 8. For both docking types, the docked conformation carrying the lowest docking energy depicted interacting residues similar to potential residues predicted by the COACH meta server, affirming the tunnel of EgDGAT1_2 as the binding cavity for oleoyl-CoA. With common residues being detected between both docking types and favourable interactions due to the lowest binding energies, the results suggest that the residues are potential catalytic residues of EgDGAT1_2 towards oleoyl-CoA.

Table 2. Molecular docking result of EgDGAT1_2 with oleoyl-CoA. The residues in bold represent similar interacting residues between both blind and specific site docking.

Type of docking	Score (kcal/mol)	Interacting residues			
		Hydrogen bonding	Hydrophobic interaction	Electrostatic interaction	Unfavourable
Blind docking	-8.1	**SER239**	TRP397	LYS243	ASN401
		THR364	**HIS405**	MET281	
		HIS438	TYR413	**TRP400**	
		CYS441	MET427		
		PHE236	LEU409		
Specific site docking	-8.2	**SER239**	**HIS405**	MET408	
		HIS438	LEU369		
		CYS441	LEU235		
		GLN493	TRP357		
		SER434	PHE361		
			PHE365		
			TRP400		
			ARG398		

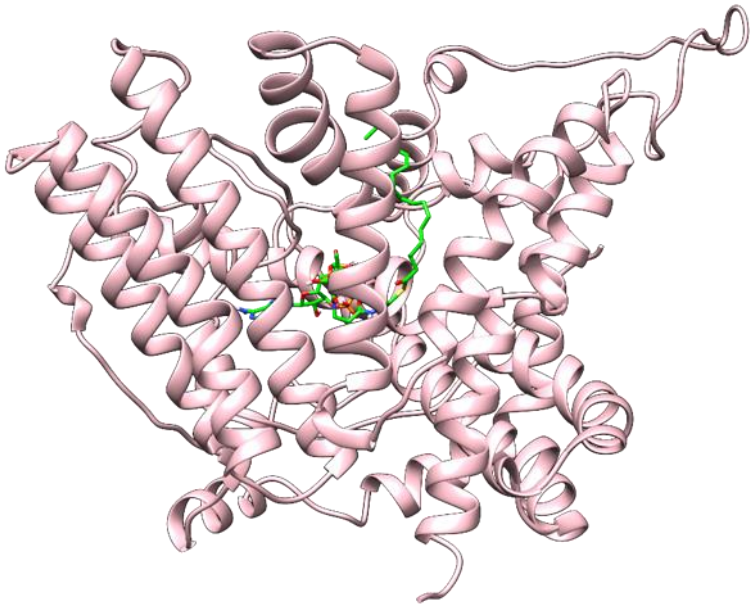
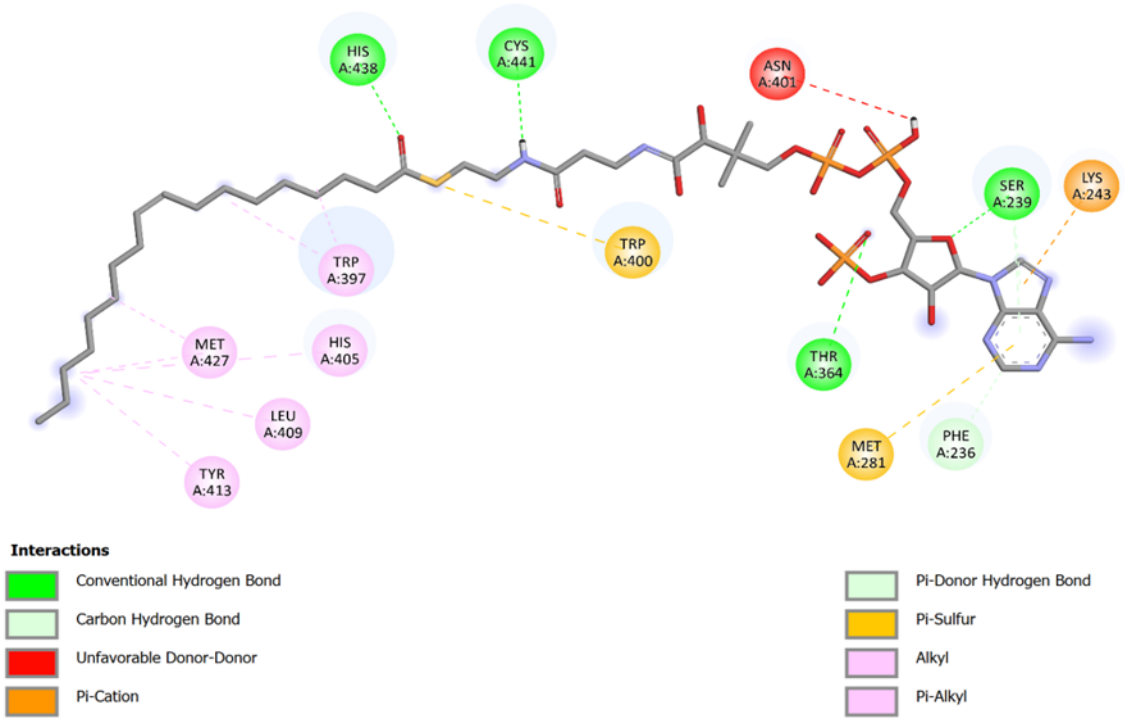


Figure 6. The EgDGAT1_2 (pink) and oleoyl-CoA (green) blind docked complex. Oxygen in red, nitrogen in blue and hydrogen in white.

(a)



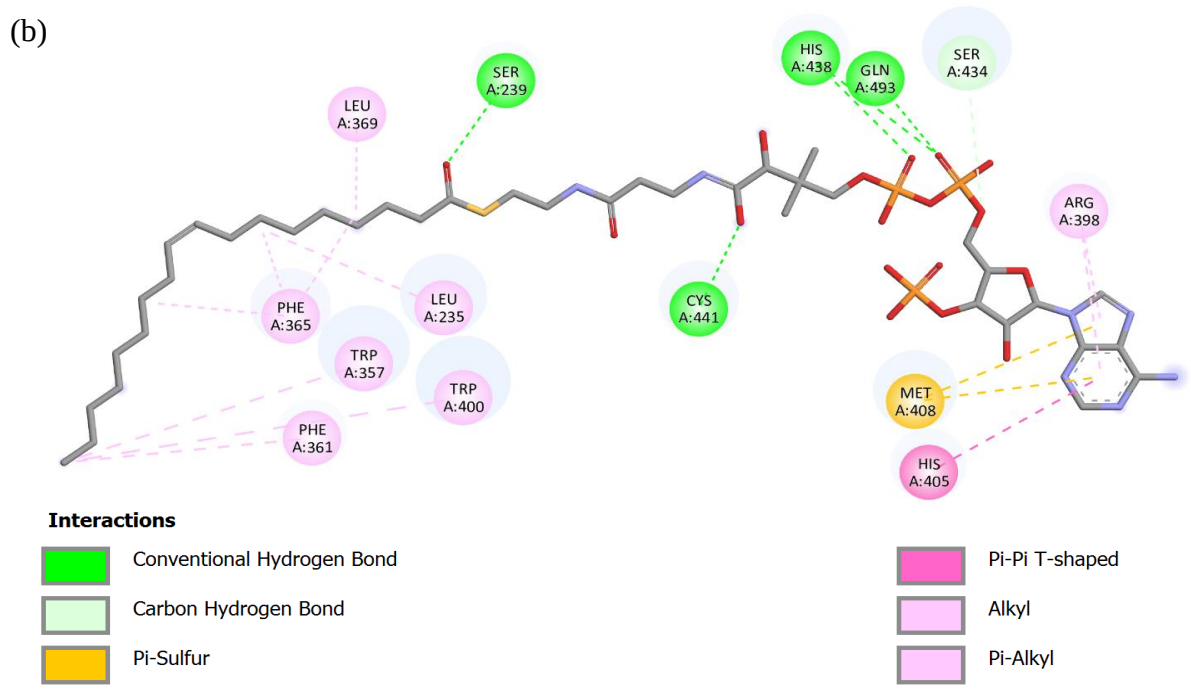


Figure 7. The molecular docking two-dimensional (2D) interaction and orientation of oleoyl-CoA generated using BIOVIA Discovery Studio 2021 Client. (a) blind docking (b) specific site docking. Amino acids are coloured according to the type of interaction.

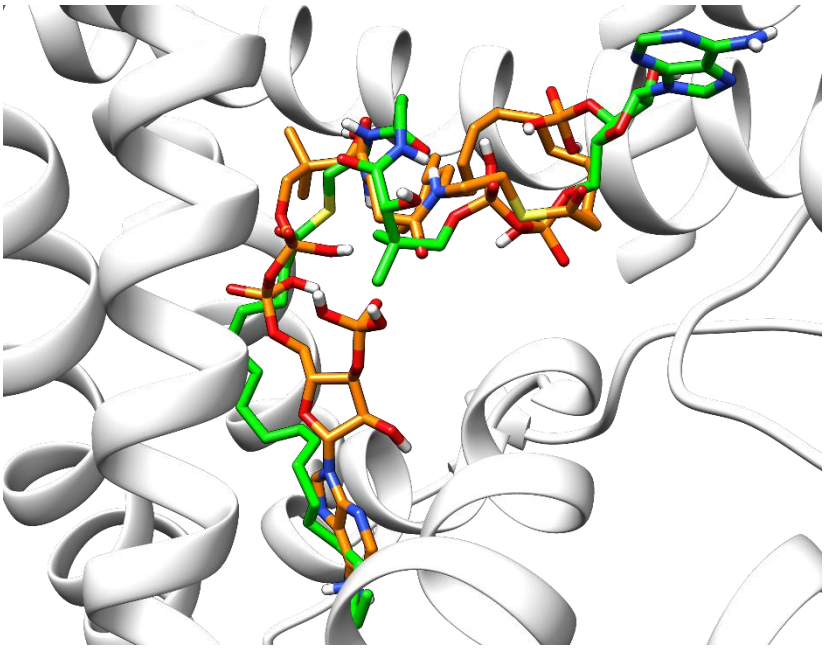


Figure 8. The superimposed binding poses of oleoyl-CoA from blind (green) and specific site docking (orange). Oxygen in red, nitrogen in blue and hydrogen in white.

CONCLUSION

In the present study, the protein model for EgDGAT1_2 and the target sites for oleoyl-CoA binding were successfully predicted. The predicted model of EgDGAT1_2 provides valuable insights into structural conformation and secondary structural arrangement for the proteins. The amino acid sequences provided information about the structural arrangement and its resemblance to the integral membrane protein of the DGAT1 family. The docking studies suggest that oleoyl-CoA have a promising association with predicted

binding residues in EgDGAT1_2, further emphasising its function in TAG biosynthesis. This work was done using a molecular docking technique; however, to validate the results of the current *in silico* investigation, molecular dynamic simulations lasting at least 50 ns would be required. It is strongly recommended to conduct additional research (*in vitro* and *in vivo*) to confirm the interacting residues and ligand binding poses, as this *in silico* study is only a prediction of protein-ligand interaction.

ACKNOWLEDGEMENTS

We would like to thank MRUN-LRGS (vote no. 4L925) and Universiti Teknologi Malaysia for the support and platform to conduct this study.

REFERENCES

1. Slocombe, S. P., Cornah, J., Pinfield-Wells, H., Soady, K., Zhang, Q., Gilday, A., Dyer, J. M., and Graham, I. A. (2009) Oil accumulation in leaves directed by modification of fatty acid breakdown and lipid synthesis pathways. *Plant Biotechnol J.*, **7(7)**, 694–703.

2. Lee, E.-J., Oh, M., Hwang, J.-U., Li-Beisson, Y., Nishida, I., and Lee, Y. (2017) Seed-Specific Overexpression of the Pyruvate Transporter BASS2 Increases Oil Content in *Arabidopsis* Seeds. *Frontiers in plant science*, **8**

3. Salmani, A., Seifi, E., Alizadeh, M., Ziaiifar, A., and Fereydooni, H. (2016) The growth pattern and oil accumulation in fruit tissues of olive cultivar Kroneiki. *Asian Jr. of Microbiol. Biotech. Env. Sc.*, **18**, 153–156.

4. Barcelos, E., Rios, S. d. A., Cunha, R. N. V., Lopes, R., Motoike, S. Y., Babiychuk, E., Skirycz, A., and Kushnir, S. (2015) Oil palm natural diversity and the potential for yield improvement. *Frontiers in plant science*, **6**, 190–190.

5. Montoya, C., Cochard, B., Flori, A., Cros, D., Lopes, R., Cuellar, T., Espeout, S., Syaputra, I., Villeneuve, P., Pina, M., Ritter, E., Leroy, T., and Billotte, N. (2014) Genetic architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis Jacq.*) compared to its wild relative *E. oleifera (H.B.K) Cortés*. *PLoS One*, **9(5)**, e95412–e95412.

6. Cagliari, A., Margis, R., dos Santos Maraschin, F., Turchetto-Zolet, A. C., Loss, G., and Margis-Pinheiro, M. (2011) Biosynthesis of triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology*, **2(1)**, e10.

7. Voelker, T. (2011) Secrets of palm oil biosynthesis revealed. *Proceedings of the National Academy of Sciences*, **108(30)**, 12193–12194.

8. Salas, J., Harwood, J., and Martínez Force, E., (2013) *Lipid Metabolism in Olive: Biosynthesis of Triacylglycerols and Aroma Components*, 97–127.

9. Chapman, K. D., and Ohlrogge, J. B. (2012) Compartmentation of triacylglycerol accumulation in plants. *The Journal of biological chemistry*, **287(4)**, 2288–2294.

10. Kong, Q., Yang, Y., Guo, L., Yuan, L., and Ma, W. (2020) Molecular Basis of Plant Oil Biosynthesis: Insights Gained From Studying the WRINKLED1 Transcription Factor. *Frontiers in plant science*, **11**

11. Turchetto-Zolet, A. C., Christoff, A. P., Kulcheski, F. R., Loss-Morais, G., Margis, R., and Margis-Pinheiro, M. (2016) Diversity and evolution of plant diacylglycerol acyltransferase (DGATs) unveiled by phylogenetic, gene structure and expression analyses. *Genetics and molecular biology*, **39(4)**, 524–538.

12. Savadi, S., Naresh, V., Kumar, V., and Bhat, S. R. (2016) Seed-specific overexpression of *Arabidopsis* DGAT1 in Indian mustard (*Brassica juncea*) increases seed oil content and seed weight. *Botany*, **94(3)**, 177–184.

13. Wu, P., Xu, X., Li, J., Zhang, J., Chang, S., Yang, X., and Guo, X. (2021) Seed-specific overexpression of cotton GhDGAT1 gene leads to increased oil accumulation in cottonseed. *The Crop Journal*, **9(2)**, 487–490.

14. Zhang, F., Gao, X., Zhang, J., Liu, B., Zhang, H., Xue, J., and Li, R. (2018) Seed-specific expression of heterologous gene DGAT1 increase soybean seed oil content and nutritional quality. *Sheng Wu Gong Cheng Xue Bao*, **34(9)**, 1478–1490.

15. Shockey, J. M., Gidda, S. K., Chapital, D. C., Kuan, J. C., Dhanoa, P. K., Bland, J. M., Rothstein, S. J., Mullen, R. T., and Dyer, J. M. (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell*, **18(9)**, 2294–2313.

16. Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I., and Browse, J. (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol J*, **6(8)**, 819–831.

17. Li, R., Yu, K., Hatanaka, T., and Hildebrand, D. F. (2010) Vernonia DGATs increase accumulation of epoxy fatty acids in oil. *Plant Biotechnol J.*, **8(2)**, 184–195.

18. Hernández, M. L., Whitehead, L., He, Z., Gazda, V., Gilday, A., Kozhevnikova, E., Vaistij, F. E., Larson, T. R., and Graham, I. A. (2012) A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued *Arabidopsis* seed oil catabolism mutants. *Plant physiology*, **160(1)**, 215–225.

19. Saha, S., Enugutti, B., Rajakumari, S., and Rajasekharan, R. (2006) Cytosolic Triacylglycerol Biosynthetic Pathway in Oilseeds. Molecular Cloning and Expression of Peanut Cytosolic Diacylglycerol Acyltransferase. *Plant physiology*, **141(4)**, 1533–1543.

20. Stöveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., and Steinbüchel, A. (2005) The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from *Acinetobacter sp. strain ADP1*: characterization of a novel type of acyltransferase. *Journal of bacteriology*, **187(4)**, 1369–1376.

21. Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L. (2008) Identification of the Wax Ester Synthase/Acyl-Coenzyme A:Diacylglycerol Acyltransferase WSD1 Required for Stem Wax Ester Biosynthesis in *Arabidopsis Plant physiology*, **148(1)**, 97–107.

22. Rosli, R., Chan, P. -L., Chan, K. -L., Amiruddin, N., Low, E. -T. L., Singh, R., Harwood, J. L. and Murphy, D. J. (2018) *In silico* characterization and expression profiling of the diacylglycerol acyltransferase gene family (DGAT1, DGAT2, DGAT3 and WS/DGAT) from oil palm, *Elaeis guineensis*. *Plant Science*, **275**, 84–96.

23. Dussert, S., Guerin, C., Andersson, M., Joët, T., Tranbarger, T. J., Pizot, M., Sarah, G., Omore, A., Durand-Gasselin, T., and Morcillo, F. (2013) Comparative Transcriptome Analysis of Three Oil Palm Fruit and Seed Tissues That Differ in Oil Content and Fatty Acid Composition. *Plant physiology*, **162(3)**, 1337–1358.

24. Jin, Y., Yuan, Y., Gao, L., Sun, R., Chen, L., Li, D., and Zheng, Y. (2017) Characterization and Functional Analysis of a Type 2 Diacylglycerol Acyltransferase (DGAT2) Gene from Oil Palm (*Elaeis guineensis Jacq.*) Mesocarp in *Saccharomyces cerevisiae* and Transgenic *Arabidopsis thaliana*. *Frontiers in plant science*, **8**.

25. Aymé, L., Jolivet, P., Nicaud, J. -M. and Chardot, T. (2015) Molecular Characterization of the *Elaeis guineensis* Medium-Chain Fatty Acid Diacylglycerol Acyltransferase DGAT1-1 by Heterologous Expression in *Yarrowia lipolytica*. *PLoS One*, **10(11)**, e0143113.

26. Yen, C. -L. E., Stone, S. J., Koliwad, S., Harris, C. and Farese, R. V. Jr. (2008) Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *Journal of lipid research*, **49(11)**, 2283–2301.

27. Lager, I., Jeppson, S., Gippert, A. -L., Feussner, I., Stymne, S., and Marmon, S. (2020) Acyltransferases Regulate Oil Quality in *Camelina sativa* Through Both Acyl Donor and Acyl Acceptor Specificities. *Frontiers in plant science*, **11**.

28. Roesler, K., Shen, B., Bermudez, E., Li, C., Hunt, J., Damude, H. G., Ripp, K. G., Everard, J. D., Booth, J. R., Castaneda, L., Feng, L., and Meyer, K. (2016) An Improved Variant of Soybean Type 1 Diacylglycerol Acyltransferase Increases the Oil Content and Decreases the Soluble Carbohydrate Content of Soybeans. *Plant physiology*, **171(2)**, 878–893.

29. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D. and Bairoch, A., 2005. The Proteomics Protocols Handbook. *Protein Identification and Analysis Tools on the ExPASy Server*, 571–607.

30. Geourjon, C., and Deléage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, **11(6)**, 681–684.

31. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*, **27(1)**, 135–145.

32. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, **42(W1)**, W320–W324.

33. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, **5(4)**, 725–738.

34. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, **35(Web Server issue)**, W407–W410.

35. Colovos, C., and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, **2(9)**, 1511–1519.

36. Laskowski, R. A., MacArthur, 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(2)**, 283–291.

37. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, **31(2)**, 455–461.

38. Yang, J., Roy, A., and Zhang, Y. (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, **29(20)**, 2588–2595.

39. Paterson, R. R. M., and Lima, N. (2018) Climate change affecting oil palm agronomy, and oil palm cultivation increasing climate change, require amelioration. *Ecol Evol*, **8(1)**, 452–461.

40. Ahmed, A., Ishak, M. Y., Uddin, M. K., Abd Samad, M. Y., Mukhtar, S., and Danhassan, S. S. (2021) Effects of Some Weather Parameters on Oil Palm Production in the Peninsular Malaysia.

41. Shyam Mohan, A. H., Rao, S. N., D, S., and Rajeswari, N. (2022) *In silico* structural, phylogenetic and drug target analysis of putrescine monooxygenase from *Shewanella putrefaciens-95*. *Journal of Genetic Engineering and Biotechnology*, **20(1)**, 57.

42. Bañó-Polo, M., Baeza-Delgado, C., Tamborero, S., Hazel, A., Grau, B., Nilsson, I., Whitley, P., Gumbart, J. C., von Heijne, G., and Mingarro, I. (2018) Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nature communications*, **9(1)**, 5246.

43. Sadaf, A., Cho, K. H., Byrne, B., and Chae, P. S., 2015. Methods in Enzymology. *Chapter Four - Amphipathic Agents for Membrane Protein Study*, **557**, 57–94.

44. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O., and Fleming, P. J. (2000) Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences*, **97(11)**, 5796–5801.

45. Khrustalev, V. V., Khrustaleva, T. A. and Barkovsky, E. V. (2013) Random coil structures in bacterial proteins. Relationships of their amino acid compositions to flanking structures and corresponding genic base compositions. *Biochimie*, **95(9)**, 1745–1754.

46. Lopes, J. L. S., Nobre, T. M., Cilli, E. M., Beltramini, L. M., Araújo, A. P. U., and Wallace, B. A. (2014) Deconstructing the DGAT1 enzyme: Binding sites and substrate interactions. *Biochimica et Biophysica Acta (BBA) – Bio-membranes*, **1838(12)**, 3145–3152.

47. Caldo, K. M. P., Acedo, J. Z., Panigrahi, R., Vederas, J. C., Weselake, R. J. and Lemieux, M. J. (2017) Diacylglycerol Acyltransferase 1 Is Regulated by Its N-Terminal Domain in Response to Allosteric Effectors. *Plant physiology*, **175(2)**, 667–680.

48. Yang, J. and Zhang, Y. (2015) Protein Structure and Function Prediction Using I-TASSER. *Current protocols in bioinformatics*, **52**, 5.8.1–5.8.15.

49. Zhang, Y. (2008) I-TASSER server for protein 3D structure predictions. *BMC bioinformatics*, **9**, 40.

50. Yakimov, A. P., Afanaseva, A. S., Khodorkovskiy, M. A. and Petukhov, M. G. (2016) Design of Stable α-Helical Peptides and Thermostable Proteins in Biotechnology and Biomedicine. *Acta naturae*, **8(4)**, 70–81.

51. Popot, J. -L. (1993) Integral membrane protein structure: transmembrane α-helices as autonomous folding domains: Current opinion in structural biology 1993, 3: 532–540. *Current Opinion in Structural Biology*, **3(4)**, 532–540.

52. Mishra, R., Mazumder, A., Mazumder, R., Mishra, P. and Chaudhary, P. (2019) Docking study and result conclusion of heterocyclic derivatives having urea and acyl moiety. *Asian Journal of Biomedical and Pharmaceutical Sciences*, **9**.

53. Meng, X. Y., Zhang, H. X., Mezei, M. and Cui, M. (2011) Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, **7(2)**, 146–157.

54. Xu, Y., Caldo, K. M. P., Falarz, L., Jayawardhane, K., and Chen, G. (2020) Kinetic improvement of an algal diacylglycerol acyltransferase 1 via fusion with an acyl-CoA binding protein. *The Plant Journal*, **102(4)**, 856–871.

55. Sui, X., Wang, K., Gluchowski, N. L., Elliott, S. D., Liao, M., Walther, T. C. and Farese, R. V. (2020) Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme. *Nature*, **581(7808)**, 323–328.

56. Guo, Z. -Y., Lin, S., Heinen, J. A., Chang, C. C. Y. and Chang, T. -Y. (2005) The Active Site His-460 of Human Acyl-coenzyme A:Cholesterol Acyltransferase 1 Resides in a Hitherto Undisclosed Transmembrane Domain*. *Journal of Biological Chemistry*, **280(45)**, 37814–37826.

57. Lin, S., Lu, X., Chang, C. C. Y. and Chang, T. -Y. (2003) Human Acyl-Coenzyme A:Cholesterol

Acyltransferase Expressed in Chinese Hamster Ovary Cells: Membrane Topology and Active Site Location. *Molecular Biology of the Cell*, **14(6)**, 2447–2460.

58. An, S., Cho, K. -H., Lee, W. S., Lee, J. -O., Paik, Y. -K. and Jeong, T. -S. (2006) A critical role for the histidine residues in the catalytic function of acyl-CoA:cholesterol acyltransferase catalysis: Evidence for catalytic difference between ACAT1 and ACAT2. *FEBS Letters*, **580(11)**, 2741–2749.