Three-dimensional (3D) structure prediction of the American and African oil-palms β-ketoacyl-[ACP] synthase-II protein by comparative modelling

Edina Wang, Suresh Chinni & Subhash Janardhan Bhore*

Department of Biotechnology, Faculty of Applied Sciences, AIMST University, Bedong-Semeling Road, Bedong, 08100, Kedah, Malaysia; Subhash Janardhan Bhore – Email: subhash@aimst.edu.my, subhashbhore@gmail.com; Phone: +60-4-429 8176; Fax: +60-4-429 8109; *Corresponding author

Received March 03, 2014; Accepted March 06, 2014; Published March 19, 2014

Abstract

Background: The fatty-acid profile of the vegetable oils determines its properties and nutritional value. Palm-oil obtained from the African oil-palm [Elaeis guineensis Jacq. (Tenera)] contains 44% palmitic acid (C16:0), but, palm-oil obtained from the American oil-palm [Elaeis oleifera] contains only 25% C16:0. In part, the β-ketoacyl-[ACP] synthase II (KASII) [EC: 2.3.1.179] protein is responsible for the high level of C16:0 in palm-oil derived from the African oil-palm. To understand more about E. guineensis KASII (EgKASII) and E. oleifera KASII (EoKASII) proteins, it is essential to know its structures. Hence, this study was undertaken. Objective: The objective of this study was to predict three-dimensional (3D) structure of EgKASII and EoKASII proteins using molecular modelling tools. Materials and Methods: The amino-acid sequences for KASII proteins were retrieved from the protein database of National Center for Biotechnology Information (NCBI), USA. The 3D structures were predicted for both proteins using homology modelling and ab-initio technique approach of protein structure prediction. The molecular dynamics (MD) simulation was performed to refine the predicted structures. The predicted structure models were evaluated and root mean square deviation (RMSD) and root mean square fluctuation (RMSF) values were calculated. Results: The homology modelling showed that EgKASII and EoKASII proteins are 78% and 74% similar with Streptococcus pneumonia KASH and Brucella melitensis KASII, respectively. The EgKASII and EoKASII structures predicted by using ab-initio technique approach shows 6% and 9% deviation to its structures predicted by homology modelling, respectively. The structure refinement and validation confirmed that the predicted structures are accurate. Conclusion: The 3D structures for EgKASII and EoKASII proteins were predicted. However, further research is essential to understand the interaction of EgKASII and EoKASII proteins with its substrates.

Keywords: ab-initio, fatty acids, homology modelling, molecular dynamics, oil-palm, palm-oil.
biosynthesis pathway [8]. However, the level of C\textsubscript{16:0} in the palm-oil obtained from the American oil-palm fruit-mesocarp-tissue is only 25\%, which is about 43.2\% less in comparison to C\textsubscript{16:0} content in palm-oil obtained from the African oil-palm [8].

The understanding of KASII in the African and American oil-palms is important in order to understand the deviation in its activity in two oil palm species. As a part of it, the full length KASII cDNA clones were isolated from both \textit{E. guineensis} (EgKASII) and \textit{E. oleifera} (EoKASII) previously [2]. The EgKASII and EoKASII cDNAs are 2011 and 2138 base pair in length [2].

As of March 1, 2014, nobody has reported the structural features of oil-palms KASII protein; which is needed to understand the differences in KASII efficiency in the African and the American oil palm species. This warrants the study on oil palms KASII protein structures. Protein structures can be studied using X-ray crystallography and or NMR techniques [9, 10]. However, protein structures can be predicted using computational tools to quickly understand the protein structure [11]. Therefore, this study was undertaken to predict three-dimensional (3D) structures of the African and American oil-palms KASII protein by comparative modelling and to elucidate and understand their unique features. The predicted 3D structures of EgKASII and EoKASII proteins are being reported in this paper.

Figure 1: The predicted secondary structures of the African oil-palm [\textit{Elaeis guineensis} Jacq. (Tenera)] \(\beta\)-ketoacyl-[ACP] synthase II protein.

Figure 2: The predicted secondary structures of the American oil-palm [\textit{Elaeis oleifera}] \(\beta\)-ketoacyl-[ACP] synthase II protein.

Methodology:

\textbf{Protein sequence retrieval}

The nucleotide database of NCBI contains full-length KASII cDNA sequences for both EgKASII and EoKASII [Gene Bank Accession Numbers: AF220453 (EgKASII) and FJ940767 (EoKASII)] [2]. The deduced amino acid sequence for EgKASII and EoKASII proteins are available in protein database of NCBI. The amino acid sequence of EgKASII and EoKASII proteins were retrieved from the NCBI’s protein database. These retrieved protein sequences were used in the experiments to predict 3D structures.

\textbf{Secondary structure prediction}

The secondary structures in EgKASII and EoKASII proteins were predicted using PHYRE server [12] and visualized using Pymol [13]. The MEMSAT-SVM server [14] was used to predict the protein topology and to identify the presence of signal peptide and transmembrane helices within the EgKASII and EoKASII proteins.

\textbf{Template selection and 3D structure prediction}

Two comparative molecular modelling approaches namely, homology modelling by MODELLER [15] and \textit{ab-initio} by 1-
TASSER (server) were used in this study to predict the 3D structure of EgKASII and EoKASII proteins [16, 17].

In homology modelling, the templates were identified based on position-specific profile search method which improves the accuracy of sequence alignments and also extends the boundaries of detectable sequence similarity. Position-specific iterative basic local alignment search tool (PSI-BLAST) [18] was used to derive a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment (MSA) of sequences using protein–protein BLAST. After templates identification, global alignment was carried out between the query sequence and the identified templates. The best template was selected based on the E-value (lowest), highest score, highest matching secondary structures and the most aligned region between the query and the template.

In ab-initio (by I-TASSER) approach of 3D structure prediction, the 3D structure models were built based on multiple-threading alignments and iterative template fragment assembly simulations by Local Meta-Threading Server (LOMETS) [19]. Five top decoys were predicted for EgKASII and EoKASII proteins. The structures with the lowest c-score were selected as the best model for EgKASII and EoKASII proteins.

Refinement and validation of predicted 3D structures

The predicted 3D structures of EgKASII and EoKASII were processed for the refinement and the validation. Using GROMACS v4.5.4 [20] and GROMOS96 53a6 force field on a Linux system, energy minimization and molecular dynamics (MD) simulation was performed for the predicted 3D structures. Energy minimization was performed using steepest descent algorithm and was allowed to run until it converged to machine precision or to a maximum force on each atom less than 100 kJ/mol/nm. The 3D structures were centered in a rhombic dodecahedral cell filled with simple point charge (SPC) water with a box edge set at 1.0 nanometer (nm). Sodium or chloride ions were added accordingly to neutralize the overall charge of the system. The position restraints were applied to (all) protein and heavy atoms and simulations were performed for NVT equilibration ensemble where number of particles, volume of the system and temperature were kept constant at 300K for 100 picoseconds (ps) using velocity rescaling method [21] followed by 100 ps of NPT equilibration ensemble where number of particles, pressure and temperature were kept constant at 1 bar. The temperature and pressure were controlled by Nose-Hoover thermostat and Parrinello-Rahman barostat, respectively [22, 23]. After the system has been well-equilibrated, we run a 30 nanosecond (ns) of MD simulation for our protein structures predicted by ab-initio approach. A time-step of 2 femtoseconds (fs) was used where all bonds were constrained using the linear constraint solver (LINCS) algorithm [24]. Coulombs potentials were calculated using Particle Mesh Ewald (PME) electrostatics [25] using a cubic-spline interpolated grid with 0.16 nm grid spacing. The stereochemical quality of the models was determined using PROCHECK [26].

Results:
The EgKASII and EoKASII protein sequences were retrieved, and its secondary structures were predicted. The topology of both EgKASII and EoKASII proteins to show secondary structures is shown in Figure 1 & Figure 2, respectively. The predicted 3D structure produced for EgKASII and EoKASII proteins by homology modelling using MODELLER software covers their 120 to 567 and 145 to 562 regions (amino acids), respectively. The predicted 3D structure produced for EgKASII and EoKASII proteins by homology modelling using MODELLER as well as by using ab-initio method (using I-TASSER) are shown in Figure 3.

Superimposition showing active sites of EgKASII (Cys316, His456, His492) and EoKASII (Cys316, His453, His489) with their respective templates used in homology modelling is depicted in Figures 4A & B. Superimposition of EgKASII on EoKASII active sites from the structures generated by MODELLER and I-TASSER is shown in Figure 4C & D, respectively.

The total energy for the protein models after energy minimization and equilibration is shown in supplementary Figure 1. All the four models built were evaluated with PROCHECK for stereochemistry quality of protein structures. The comparative study of these structures by using Ramachandran plot is shown in supplementary Figure 2. Other details such as the main chain, side chain, bond length, bond angle and planar groups within limits obtained are shown in Table 1 (see supplementary material). The RMSF for ab-initio generated models of EgKASII and EoKASII is shown in supplementary Figure 3. Most of the residues and the active residues fluctuate within the range of 0.1 to 0.2 nm which is in the acceptable range. The radius of gyration plot (supplementary Figure 4) shows that EgKASII and EoKASII protein remains compact and stably folded after 30 ns (or 30,000 ps) of simulation.

Figure 3: The predicted three-dimensional (3D) structures of oil-palm β-ketoacyl-[ACP] synthase II (KASII) protein; A) EgKASII 3D structure predicted by MODELLER; B) EoKASII 3D structure predicted by MODELLER; C) EgKASII 3D structure predicted by I-TASSER; D) EoKASII 3D structure predicted by I-TASSER. The structures are colored to show their secondary structures. Red color represents helices, yellow color represents strands, and green color represents loops.

ISSN 0973-2063 (online) 0973-8894 (print)
Biominformation 10(3):130-137 (2014) 132 © 2014 Biomedical Informatics
structures of these respective proteins. However, we suggest the further wet lab experimental work to validate these predicted structures using X-ray crystallography or NMR technique [9, 10]. Similarly, the active-site residues of KASII proteins has been determined successfully but have not been tested experimentally. In order to confirm the predicted structures and their active sites, molecular docking and simulations for the formation of complex between the predicted protein structure and their respective substrates needs to be carried out for the structures predicted by MODELLER; so that comparative analysis can be done for the 3D structures predicted by using MODELLER and I-TASSER.

Oil palm derived from the African oil palm contains high amount (~54%) of saturated fatty acids [3] in comparison to palm oil obtained from the American oil palm [4]. A systematic study of oil palms key genes involved in fatty acid biosynthesis pathway and a comparative modelling of key proteins important in fatty acid biosynthesis will help to elucidate and understand their unique features. Molecular modelling can be used for the understanding and prediction of the microscopic and macroscopic properties of the proteins [29]. It is also useful in the study of enzyme’s binding affinity [30], in virtual screening of natural products [31], in predicting molecular interaction [32], and saves the time and money. Our aim was to utilize molecular modelling tools to predict the 3D structures of the African and American oil palm KASII protein.

In the recent past, Malaysian Palm Oil Board (MPOB) and collaborators published oil palms (Elaeis guineensis Jacq. and E. oleifera) genome [33]. They have also reported a unique gene that controls the oil yield in oil palm fruits [34]. These advances in oil palm research will have a significant impact in oil palm industry. It is estimated that there are at least 34,802 genes in oil palm genome. If we want to understand structures of all proteins in oil palm by doing wet lab work then this is too much experimental work. However, to understand the structural features of important proteins in short time molecular modelling will be useful. For the real understanding of the protein structures, structures should be determined by using NMR or X-ray crystallography. However, the 3D structures predicted in this study could serve as foundation for the further research on oil palm KASII protein and could be useful in clear understanding of the fatty acid biosynthesis pathway in oil palm.

Conclusion:
We determined the three-dimensional structure for E. guineensis and E. oleifera KASII protein. The RMSF value for the three active residues of EgKASII and EoKASII were around 0.1nm. Both the structures remain compact and stably folded after 30 ns of simulation at an average of 2.41 nm and 2.38 nm for EgKASII and EoKASII, respectively. Molecular docking and simulation study is required to understand the interactions between the predicted KASII proteins and their substrates. In addition, further research is required in wet lab to validate the predicted structures.

Acknowledgement:
SJB acknowledges the financial support from the Malaysia’s Ministry of Education to EW under the MyMaster programme. Authors are grateful to the University Putra Malaysia, University Technology Malaysia, and University Technical Malaysia (Melaka) for providing access to their facilities.
Conflict of interests:
Authors attest that there are no conflicts of interest to declare.

References:
[1] Dos Santos Corrêa IN et al. Enzyme Res. 2011 2011: 814507 [PMID: 21687622]
[2] Bhore SJ et al. J Nat Sc Biol Med. 2014 5: 73 [PMID: Awaiting]
[3] Bhore SJ & Shah FH, Bioinformation 2011 6: 212 [PMID: 21738318]
[4] Bhore SJ & Shah FH, World Appl Sci J. 2012 16: 362
[5] Wei Q et al. J Plant Physiol. 2012 169: 816 [PMID: 22424763]
[6] Othman A et al. Biochem Soc Trans. 2000 28: 619 [PMID: 11171146]
[7] Saed Taha R et al. J Plant Physiol. 2012 169: 1290 [PMID: 22658816]
[8] Parveez GK et al. Biochem Soc Trans. 2000 28: 969 [PMID: 11171275]
[9] Wanke R et al. Int J Antimicrob Agents. 2013 42: 443 [PMID: 24051055]
[10] Caliandro R et al. Future Med Chem. 2013 5: 1121 [PMID: 23795969]
[11] Pavlopoulou A & Michalopoulos I, Int J Mol Med. 2011 28: 295 [PMID: 21617841]
[12] Kelley LA & Sternberg MJ, Nat Protoc. 2009 4: 363 [PMID: 19247286]
[13] http://www.pymol.org/
[14] Nugent T & Jones DT, BMC Bioinformatics 2009 10: 159 [PMID: 19470175]
[15] Eswar N et al. Methods Mol Biol. 2008 426: 145 [PMID: 18542861]
[16] Zhang Y, BMC Bioinformatics 2008 9: 40 [PMID: 18215316]
[17] Roy A et al. Nat Protoc. 2010 5: 725 [PMID: 20360767]
[18] Altschul SF et al. Nucleic Acids Res. 1997 25: 3389 [PMID: 9254694]
[19] Wu S & Zhang Y, Nucleic Acids Res. 2007 35: 3375 [PMID: 17478507]
[20] Van Der Spoel D et al. J Comput Chem. 2005 26: 1701 [PMID: 16211538]
[21] Bussi G et al. J Chem Phys. 2007 126: 014101 [PMID: 17212484]
[22] Hoover WG, Phys Rev A. 1985 31: 1695 [PMID: 9895674]
[23] Braga C & Travis KP, J Chem Phys. 2005 123: 134101 [PMID: 16223269]
[24] Hess B et al. J Comput Chem. 1997 18: 1463
[25] Essmann U et al. J Chem Phys. 1995 103: 8577
[26] Laskowski RA et al. J Biomol NMR. 1996 8: 477 [PMID: 9008363]
[27] Blouin C et al. Protein Sci. 2004 13: 608 [PMID: 14978301]
[28] Potapov V et al. BMC Bioinformatics 2010 11: 374 [PMID: 20624289]
[29] Morrison AM et al. Biochim Biophys Acta. 2013 1830: 24361620
[30] Wong CF & Bairy S, Curr Pharm Des. 2013 19: 4739 [PMID: 23260021]
[31] Wang L et al. Bioorg Med Chem Lett. 2014 24: 1261 [PMID: 24530004]
[32] Lytle AK et al. J Mol Biol. 2014 [PMID: 24576606]
[33] Singh R et al. Nature 2013 500: 335 [PMID: 23883927]
[34] Singh R & et al. Nature 2013 500: 340 [PMID: 23883930]

Edited by P Kanguane

Citation: Wang et al. Bioinformation 10(3): 130-137 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited
Supplementary material:

**Supplementary Figure 1:** Total energy of predicted structures after its energy minimization and equilibration; A) Total energy of predicted EgKASII structure generated by MODELLER; B) Total energy of predicted EoKASII structure generated by MODELLER; C) Total energy of predicted EgKASII structure generated by I-TASSER; D) Total energy of predicted EoKASII structure generated by I-TASSER.

**Supplementary Figure 2:** Ramachandran plot for; A) EgKASII model generated by MODELLER; B) EoKASII model generated by MODELLER; C) EgKASII model generated by I-TASSER; and D) EoKASII model generated by I-TASSER.
**Supplementary Figure 3:** The root mean square fluctuation (RMSF) values for the predicted structure of EgKASII A) and EoKASII B).

**Supplementary Figure 4:** Radius of gyration plot for predicted structure of EgKASII A) and EoKASII B) proteins. Rg stands for radius of gyration.
Table 1: The main and side chain, bond length, bond angle and planar group values from Ramachandran plot for A) EgKASII model generated by MODELLER; B) EoKASII model generated by MODELLER; C) EgKASII model generated by I-TASSER; and D) EoKASII model generated by I-TASSER.

| Model | Ramachandran Plot | Main Chain | Side Chain | Bond Length (in limits) | Bond Angle (in limits) | Planar Groups (in limits) |
|-------|-------------------|------------|------------|-------------------------|------------------------|---------------------------|
| A     | 91.0 core, 7.4 allow, generous, 0.4 disallowed | 1.2 | 6 | 5 | 98.8 | 91.3 | 100 |
| B     | 92.9 core, 9.9 allow, generous, 1.2 disallowed | 3.9 | 6 | 5 | 97.6 | 84.6 | 100 |
| C     | 85.0 core, 9.9 allow, generous, 1.2 disallowed | 3.9 | 6 | 5 | 98.3 | 85.0 | 100 |
| D     | 83.9 core, 12.3 allow, generous, 1.7 disallowed | 2.1 | 6 | 5 | 97.6 | 84.6 | 100 |