Protein modelling of triterpene synthase genes from mangrove plants using Phyre2 and Swiss-model

M Basyuni1*, R Wati1, N Sulistiyono1, R Hayati1, Sumardi2, H Oku3, S Baba4 and H Sagami5

1 Department of Forestry, Faculty of Forestry, Universitas Sumatera Utara, Jl. Tri Dharma Ujung No. 1 Medan, North Sumatera, 20155, Indonesia
2 Faculty of Pharmacy, Universitas Tjut Nyak Dhien, Jl. Rasmu No. 28 Medan 20123, Indonesia
3 Molecular Biotechnology Group, Tropical Biosphere Research Center, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan
4 International Society for Mangrove Ecosystems, Faculty of Agriculture, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan
5 Institute of Multidisciplinary Research for Advanced Material, Tohoku University, Sendai, 980-8577, Japan

*Email: m.basyuni@usu.ac.id

Abstract. Molecular cloning of five oxidosqualene cyclases (OSC) genes from Bruguiera gymnorrhiza, Kandelia candel, and Rhizophora stylosa had previously been cloned, characterized, and encoded mono and -multi triterpene synthases. The present study analyzed protein modelling of triterpene synthase genes from mangrove using Phyre2 and Swiss-model. The diversity was noted within protein modelling of triterpene synthases using Phyre2 from sequence identity (38-43%) and residue (696-703). RsM2 was distinguishable from others for template structure; it used lanosterol synthase as a template (PDB ID: w6j.1.A). By contrast, other genes used human lanosterol synthase (1w6k.1.A). The predicted bind sites were correlated with the product of triterpene synthase, the product of BgbAS was β-amyrin, while RsM1 contained a significant amount of β-amyrin. Similarly BgLUS and KcMS, both main products was lupeol, on the other hand, RsM2 with the outcome of taraxerol. Homology modelling revealed that 696 residues of BgbAS, BgLUS, RsM1, and RsM2 (91-92% of the amino acid sequence) had been modelled with 100% confidence by the single highest scoring template using Phyre2. This coverage was higher than Swiss-model (85-90%). The present study suggested that molecular cloning of triterpene genes provides useful tools for studying the protein modelling related regulation of isoprenoids biosynthesis in mangrove forests.

1. Introduction
Secondary metabolites including triterpenes and phytosterols are common chemical constituents in mangrove forest [1]. A variety of triterpenes and phytosterols extensively distributed in plants are biosynthesized from a precursor 2,3-oxidosqualene, with the contribution of oxidosqualene cyclases (OSCs). 2,3-Oxidosqualene, therefore, situates at the branching point of isoprenoid pathway toward triterpenes or phytosterols biosynthesis [2]. Molecular cloning of OSC genes from mangrove forests...
namely *Bruguiera gymnorrhiza*, *Kandelia candel*, and *Rhizophora stylosa* had previously been cloned, characterized, and encoded mono and multi triterpene synthases [3-4].

Recently it has been reported the bioinformatics analysis of the OSC genes and the amino acid sequence in mangrove plants [5]. Molecular mechanism of salinity tolerance in mangrove plants has been described: the level mRNA of triterpene synthase genes of mangrove leaves and roots increased with increasing salt concentration [6-8]. These studies suggested that triterpenes play an essential physiological role in the adaptation of mangroves trees as the self-protecting barrier against the external salt stress.

Furthermore, OSCs have been attracted our consideration because of their possibility to modify the chemical structures of triterpenoids, as well as, their significance as the primarily committed enzymes in the triterpene biosynthesis [3-4]. In higher plants, OSC family member triterpenes are involved in triterpenes synthesis, and other OSCs are responsible for sterol biosynthesis. Given the diversity of the triterpene genes found in mangrove plants, it became interesting to understand the function of a protein of each triterpene synthase. However, the predict proteins from mangrove forests and their modelling have not been studied yet. Thus, the present study aimed to describe the protein modelling of triterpene synthase genes from mangrove forests using online software of Phyre2 and Swiss-model.

2. Materials and Method

2.1. Materials

A total of five mangrove oxidosqualene cyclase genes deposited in NCBI (https://www.ncbi.nlm.nih.gov/) were analyzed. The GenBank accession numbers of the DNA sequence and amino acid sequence of used these analyses are as follows: AB289585, BAF80443 (*B. gymnorrhiza* BgbAS), AB289586, BAF80444 (*B. gymnorrhiza* BgLus), AB257507, BAF35580 (*K. obovata* KcMS), AB263203, BAF80441 (*R. stylosa* RsM1), AB263204, BAF80442 (*R. stylosa* RsM2).

2.2. Template search, model building, and quality assessment

Template search was performed using Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [9] and against the Swiss-model template library (https://swissmodel.expasy.org/) [10]. The BgbAS, BgLUS, KcMS, RsM1 and RsM2 were searched with BLAST [12] against the amino acid sequence contained in the Protein Data Bank (PDB). The templates with the highest scoring crystal structure have then been selected for model building. Models are made according to the target-template alignment using ProMod3 [10]. The global and per-residue model quality was assessed using the QMEAN scoring function [12].

2.3. Prediction of ligand binding site

The prediction of ligand binding site in the modelled protein structure was made using 3DLigandSite server (http://www.sbg.bio.ic.ac.uk/3dligandsite/) [13].

2.4. Transmembrane helix prediction

Transmembrane helices were predicted using the sequence to adopt the topology using Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [9]. The extracellular and cytoplasmic sides of the membrane are labelled and the beginning and stopping of each transmembrane helix illustrated with a number representing the residue index [9].

3. Results and Discussion

3.1. Protein modelling using Phyre2 and Swiss-model

Table 1 depicts protein modelling using Phyre2 of five triterpene synthase genes: *B. gymnorrhiza* β-amyrin synthase (BgbAS), *B. gymnorrhiza* lupeol synthase (BgLUS), *R. stylosa* multifunctional triterpene synthase (RsM1 produced germanicol (63%), β-amyrin (33%), and lupeol (4%), *R. stylosa*
multifunctional triterpene synthase (RsM2 produced taraxerol, β-amyrin, and lupeol in the ratio of 70:17:13), and K. obovata multifunctional triterpene synthase with product of lupeol (50%), β-amyrin (50%), and α-amyrin (50%).

The diversity was noted within protein modelling of triterpene synthases using Phyre2 from sequence identity (38-43%) and residue (696-703). The human lanosterol synthase was used for protein model (PDB ID: c1w6kA as template structure with confidence 100% and coverage 91-92%).

Similar results were obtained using Swiss-model as displayed in Table 2. RsM1 had the relatively highest value in sequence identity, sequence similarity, and range of protein sequence. BgbAS, RsM2, and RsM2 were covered by 0.90, while BgLUS and RsM1 was 0.85. It is noteworthy that Qmean was varied among the genes (0.67-0.70). RsM2 was distinguishable from others for template structure; it used lanosterol synthase as a template (w6j.1.A). By contrast, other genes used human lanosterol synthase (1w6k.1.A). This result supported the previous result on the high relative molecular weight of RsM2 (84.8 kDa) [5]. The remaining triterpene synthases had a molecular weight lower (83.4 to 83.7 kDa) than that RsM2.

3.2. Predicted binding site
The predicted ligands bind particular residue, amino acid, contact, and average distance. BgbAS and RsM1 had the same parameter of the binding site, BgLUS and KcMS had four parameters, and RsM2 had six parameters. BgbAShad seven residues from 679, 680, 681, 684, and 725 which corresponded to amino acids of Leu, Val, His, Trp, and Thr. The longest distance belongs to Leu679. The contact was varied from 12 to 22 (Table 3). RsM1 possed five residues Leu679, Val680, His681, Trp684, and Thr725 with contact value was 13-22. The longest distance belongs to Leu679.

In contrast to this observation, BgLUS and KcMS had Leu679, Val679, Gln680, Ser724 and Leu678, Ile679, His680, and Ser724; respectively. Both had the longest distance of Leu678 and contact values 14-22. RsM2 was found to have six binding sites: Asn680, Leu681, Val682, Gln683, Trp686, and Thr727. RsM2 had the contact with 12-24 and longest average distance 0.67 (Table 3).

The predicted bind sites were correlated with a product of triterpene synthase, the product of BgbAS was β-amyrin, while RsM1 contained the significant amount of β-amyrin [4], both genes sit together in the phylogenetic tree [4]. Similarly BgLUS and KcMS, both main products was lupeol and were close to the branch of lupeol synthase [3, 4]. On the other hand, RsM2 with the outcome of taraxerol was scattered in the phylogenetic tree [4].

There are three motifs of critical residues of plant OSCs namely MW(L)CYCR, MQSFGSQ, and FIKKSQ motifs. These motifs added the established one: DCTAE and QW motifs in the plant OSCs. The BgbAS, RsM1 and RsM2 had the essential residues of Trp257, BgLUS and KcMS had Leu257 [4].

### Table 1. Results of protein modelling using Phyre2

| Gene   | Template   | Sequence Identity | Residue | Confidence | Coverage (%) |
|--------|------------|------------------|---------|------------|--------------|
| BgbAS  | c1w6kA     | 43               | 696     | 100        | 92           |
| BgLUS  | c1w6kA     | 39               | 696     | 100        | 91           |
| RsM1   | c1w6kA     | 43               | 696     | 100        | 92           |
| RsM2   | c1w6kA     | 38               | 703     | 100        | 91           |
| KcMS   | c1w6kA     | 39               | 696     | 100        | 91           |

### Table 2. Results of protein modelling using Swiss-model

| Gene   | Template   | Sequence Identity | Sequence similarity | Range | Coverage | QMEAN |
|--------|------------|------------------|---------------------|-------|----------|-------|
| BgbAS  | 1w6k.1.A   | 43.50            | 0.42                | 23-756| 0.90     | 0.70  |
| BgLUS  | 1w6k.1.A   | 39.75            | 0.40                | 90-765| 0.85     | 0.67  |
| RsM1   | 1w6K.1.A   | 44.50            | 0.43                | 93-756| 0.85     | 0.68  |
| RsM2   | w6j.1.A    | 39.16            | 0.40                | 32-753| 0.90     | 0.68  |
| KcMS   | 1w6k.1.A   | 41.20            | 0.41                | 54-753| 0.90     | 0.68  |
β-amyrin and lupeol have three critical residues SerPhe in the MQSFGSQ motifs as previously described [4]. The last amino acid residue was Lys449 in the FIKKSQ motifs has been reported to control triterpene synthase product [4].

| Clones  | Residue | Amino Acid | Contact | Average distance |
|---------|---------|------------|---------|-----------------|
| BgbAS   | 679     | LEU        | 13      | 0.69            |
|         | 680     | VAL        | 22      | 0.00            |
|         | 681     | HIS        | 19      | 0.30            |
|         | 684     | TRP        | 12      | 0.45            |
|         | 725     | THR        | 22      | 0.00            |
| BgLUS   | 678     | LEU        | 14      | 0.69            |
|         | 679     | VAL        | 22      | 0.00            |
|         | 680     | GLN        | 21      | 0.34            |
|         | 724     | SER        | 22      | 0.21            |
| RsM1    | 679     | LEU        | 14      | 0.69            |
|         | 680     | VAL        | 22      | 0.00            |
|         | 681     | HIS        | 21      | 0.34            |
|         | 684     | TRP        | 13      | 0.50            |
|         | 725     | THR        | 22      | 0.00            |
| RsM2    | 680     | ASN        | 24      | 0.00            |
|         | 681     | LEU        | 12      | 0.67            |
|         | 682     | VAL        | 24      | 0.00            |
|         | 683     | GLN        | 23      | 0.29            |
|         | 686     | TRP        | 14      | 0.47            |
|         | 727     | THR        | 24      | 0.00            |
| KcMS    | 678     | LEU        | 14      | 0.69            |
|         | 679     | ILE        | 22      | 0.00            |
|         | 680     | HIS        | 21      | 0.34            |
|         | 724     | SER        | 22      | 0.21            |

3.3. Phyre2 and Swiss-protein model

Homology modelling revealed that 696 residues of BgbAS, BgLUS, RsM1, and RsM2 (91-92% of the amino acid sequence) had been modelled with 100% confidence by the single highest scoring template using Phyre2 (Figure 1A-E). This coverage was higher than Swiss-model (85-90%) as depicted in Figure 2A-E. The slight different protein model may be due to distinct zonation of mangrove species [5].

It has been shown that the composition of triterpene may be a reflection of the distribution of triterpene synthase in the cells of mangrove plants [4]. Furthermore, a significant proportion of triterpene is found in the external parts of the root and may provide an advance indication for the shielding roles of triterpenoids in mangroves tree species [1]. The invention of several triterpene synthase genes from mangroves [3-4], whose molar ratio is defined by multifunctional OSCs, may be useful to the plant by depiction it more tolerant to the environmental stress, such as salinity stress [3-4,8]. Furthermore, previous works have demonstrated that the expression PgTPS terpene synthase increased by salt stress in Panax ginseng [14]. Our previous studies also showed salinity-dependent increases in the content of triterpenes and triterpene synthase gene expression in both secreting and
non-secreting mangrove roots and leaves [6-8, 15-16].

![A](image1.png)  ![B](image2.png)  ![C](image3.png)  ![D](image4.png)  ![E](image5.png)

**Figure 1.** Phyre protein model for *BgbAS* (A), *BgLUS* (B), *RsM1* (C), *RsM2* (D), and *KcMS* (E)

It has been disputed for a long time whether triterpene synthases ensuing different products are discrete proteins or if they are produced post-translational from one gene product to another [3]. This condition because the activities of OSCs are subject to changes in pH, detergents and electrolyte concentrations [3]. Some successful cloning of triterpene synthases has proved the presence of multiple OSCs as distinct proteins even in one plant species [3-4]. These changes may be because of differences in catalytic efficiency and specificity of multifunctional triterpene synthases such as *KcMS*, *RsM1*, and *RsM2* or other plant multifunctional triterpene synthases by the alteration in tertiary protein structure with different functions by transcriptional regulation [3-4].

3.4. Transmembrane helices
Figure 3 shows transmembrane (TM) helices in five triterpene synthases. *BgbAS*, *BgLUS*, *RsM1*, and *KcMS* had four TM helices (S1-S4), only *RsM2* had five TM helices (S1-S5) both in the extracellular
and cytoplasmic. These results were parallel with protein model and binding sites of triterpene synthase genes from mangroves.

Figure 2. Swiss protein model for BgBAS (A), BgLUS, (B) RsM1 (C), RsM2 (D), and KcMS (E).

To gain more imminent into structural and functional roles of triterpenes, the well-defined and simple in vitro membrane models (liposomes) and in vitro study showed that triterpene, as well as phytosterol, was incorporated into the lipid bilayer of liposome [17], suggesting that triterpene was also a modulator of membrane permeability. Triterpene alcohols have been structurally distinguished from phytosterols, even though they share common biosynthetic pathway and hence the similar chemical structure. However, there has been no scientific basis to functionally identify these compounds as the structural membrane lipid [16-17]. Changes in the membrane lipid composition and properties represent an essential factor in the adaptation to high salt concentration [18]. These observations strongly suggested that the triterpene and triterpene synthase contribute to the salt tolerance of mangrove plants probably by changing the property of the cell membrane [6-8].

Recently, it has been revealed that the expression of two triterpene synthases, BgbAS and RsM1 under a GAL1 promoter in GIL77 increased the triterpene content of both whole cell body and plasma membrane fractions [19]. In previous results [5], both gene BgbAS and RsM1 were placed at the plasma membrane, supported earlier results on their subcellular localization situated in the plasma membrane [19].
Figure 3. Transmembrane helices of BgBAS (A), BgLUS (B), RsM1(C), RsM2 (D), and KcMS (E)

4. Conclusion
Molecular cloning of triterpene genes provides useful tools for studying the protein modelling related regulation of isoprenoids biosynthesis in mangrove plants. Some successful cloning of triterpene synthases has proved the presence of multiple OSCs as distinct proteins even in one plant species. The present study suggested that the triterpene and triterpene synthase contribute to the salt tolerance of mangrove forests probably by changing the property of the cell membrane.

Acknowledgment
This study was supported by an International Research Collaboration and Scientific Publication Grant 2017 (No. 003/SP2H/LT/DRPM/IV/2017) from the Directorate for Research and Community Service, Ministry of Research, Technology and Higher Education, Republic of Indonesia.

References
[1] Basyuni M, Oku H, Baba S, Takara K and Iwasaki H 2007 Isoprenoids of Okinawan mangroves as lipid input into estuarine ecosystem J. Oceanography 63 601–8.
[2] Abe I, Rohmer M and Prestwich GD 1993 Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes Chem. Rev. 93 2189–2208.
[3] Basyuni M, Oku H, Inafuku M, Baba S, Iwasaki H, Oshiro K, Okabe T, Shibuya M and Ebizuka Y 2006 Molecular cloning and functional expression of a multifunctional triterpene synthase cDNA from a mangrove species Kandelia candel (L.) Druce Phytochemistry 67 2517–2524.
[4] Basyuni M, Oku H, Tsujimoto E, Kinjo K, Baba S and Takara K 2007 Triterpene synthases from the Okinawan mangrove tribe, Rhizophoraceae FEBS J. 274 5028–5042.
[5] Basyuni M, R Wati 2017 Bioinformatics analysis of the oxidosqualene gene and the amino acid sequence in mangrove plants J. Phys. : Conf. Ser. 801 012011.
[6] Basyuni M, Baba S, Inafuku M, Iwasaki H, Kinjo K, Oku H 2009 Expression of terpenoid synthase mRNA and terpenoid content in salt-stressed mangrove J. Plant Physiol. 166 1786–1800.
[7] Basyuni M, Baba S, Kinjo Y, Putri LAP, Hakim L, Oku H 2012 Salt-dependent increase in triterpenoids is reversible upon transfer to fresh water in mangrove plants Kandelia candel and Bruguiera gymnorrhiza J. Plant Physiol. 169 1903–1908.
[8] Basyuni M, Baba S, Kinjo Y, Oku H 2012 Salinity increases the triterpenoid content of a salt secretor and a non-salt secreting mangrove Aquat. Bot. 97 17–23.
[9] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg, MJE 2015 The Phyre2 web portal for protein modeling, prediction and analysis Nat. Protocols 10 845–858.
[10] Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. 2006 The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling Bioinformatics 22 195–201.
[11] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Res. 25 3389–3402.
[12] Benkert P, Biasini M, Schwede T 2011 Toward the estimation of the absolute quality of individual protein structure models Bioinformatics 27 343–350.
[13] Wass M.N., Kelley LA, Sternberg MJ 2010 3DLigandSite: predicting ligand-binding sites using similar structures Nucleic Acid Res. 38 W469–473.
[14] Kim YJ, Ham AR, Shim JS, Lee JH, Jung DY, In JG, Lee BS, Yang DC 2008 Isolation and characterization of terpene synthase gene from Panax ginseng J. Ginseng Res. 32 114–119.
[15] Basyuni M, Kinjo Y, Baba S, Shinzato N, Iwasaki H, Siregar EBM, Oku H 2011 Isolation of salt stress tolerance genes from roots of mangrove plant, Rhizophora stylosa Griff., using PCR-based suppression subtractive hybridization Plant Mol. Biol. Rep. 29 533–543.
[16] Oku H, Baba S, Koga H, Takara K, Iwasaki H 2003 Lipid composition of mangrove and its relevance to salt tolerance J. Plant Res. 116 37–45.
[17] Basyuni M, Oku H, Tsujimoto E, Baba S 2007 Cloning and functional expression of cycloartenol synthases from mangrove species Rhizophora stylosa Griff. and Kandelia candel (L.) Druce Biosci. Biotechnol. Biochem. 71 1788–1792.
[18] Turk M, Mejanelle L, Šentjurc M, Grimalt JO, Gunde-Cimerman N, Plemenitaš A 2004 Salt-induced changes in lipid composition and membrane fluidity of halophilic yeast-like melanized fungi Extremophiles 8 53–61.
[19] Inafuku M, Basyuni M, Oku H 2016 Triterpenoid modulates the salt tolerance of lanosterol synthase deficient Saccharomyces cerevisiae, GIL77 Saudi J. Biol. Sci. http://dx.doi.org/10.1016/j.sjbsj.2016.10.009.