Molecular And 3D-Structural Characterization Of Fructose-1,6-Bisphosphate Aldolase Derived From *Metroxylon Sagu*

Hairul Azman Roslan[1]*, Md. Anowar Hossain[2], Jerry Gerunsin[1].
[1]*Universiti Malaysia Sarawak - Department of Molecular Biology, Kota Samarahan, Sarawak, Malaysia;*
[2]*University of Rajshahi - Department of Biochemistry and Molecular Biology, Rajshahi, Bangladesh.*

ABSTRACT

Fructose-1,6-bisphosphate aldolase (FBAld) is an enzyme that catalyzes the cleavage of D-fructose-1,6-phosphate (FBP) to D-glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP), and plays vital role in glycolysis and gluconeogenesis. However, molecular characterization and functional roles of FBAld remain unknown in sago palm. Here we report a modified CTAB-RNA extraction method was developed for the isolation of good quality RNA (RIN>8) from sago leaves and the isolation of FBAld cDNA from sago palm. The isolated sago FBAld (msFBAld) cDNA has total length of 1288 bp with an open reading frame of 1020 bp and a predicted to encode for a protein of 340 amino acid resides. The predicted protein shared a high degree of homology with Class-I FBAld from other plants. Meanwhile, the msFBAld gene spanned 2322 bp and consisted of five exons. Conserved domain search identified fifteen catalytically important amino acids at the active site and phylogenetic tree revealed localization of msFBAld in the chloroplast. A molecular 3D-structure of msFBAld was generated by homology modeling and a Ramachandran plot with 86.7% of the residues in the core region, 13.4% in the allowed region with no residues in the disallowed region. The modeled structure is a homotetramer containing an α/β-TIM-barrel at the center. Superimposition of the model with Class-I aldolases identified a catalytic dyad, Lys209-Glu167, which could be involved in the Schiff's base formation and aldol condensation. Apart from that, overproduction of the recombinant msFBAld in Escherichia coli resulted in increased tolerance towards salinity.

Key words: *Metroxylon sagu*; Fructose-1,6-bisphospate aldolase; 3D-structure; stress; Schiff's base

*Author for correspondence: hairulroslan@hotmail.com

INTRODUCTION

Starch in sago palm (*Metroxylon sagu*) has been utilized for many years as an energy source for animal and human. Research in sago palm has been focused on alteration of the machinery that is involved in starch production. This plant became an interesting subject of research in this region because of its remarkable characteristics that can withstand among others physiological drought of salinity, prolonged flooding as well as soil acidity and yet still be able to produce a substantial amount of energy reserve. The cultivation of this tree favors the rural community because of its minimal care and can grow well in swampy areas, where most other plants could not. Moreover, it can withstand strong typhoon wind and forms a barrier between the sea and land [1]. Furthermore, the root is capable of trapping silt load as well as industrial pollutant [1]. Besides serving as a food source for the rural people, the physiochemical properties of sago starch also provided various value-added uses that are suitable for industrial applications [2]. For instance, in the food manufacturing industries that are currently utilizing starch from sago palm for the production of products such as high fructose syrup, vermicelli noodle, monosodium glutamate (MSG), glucose, maltodextrin as well as pharmaceutical products. In addition, modified sago starch in manufacturing industry could be an alternative substitute to corn and tapioca products such as in the manufacturing of MSG, textile, pulp and paper industry [1,3]. However, in the post-genomic era, sago molecular biology remains unknown which hampers further improvement of starch production and to understand the underlying mechanism of biotic and abiotic stress responses of sago palm. Nevertheless, recent publications in sago palm molecular biology includes characterization of stress-related genes [4,5], development of expressed sequence tags of young leaves [6] and the advancement of gene transformation method in sago palm [7].

Fructose 1,6-bisphosphate aldolase (EC 4.1.2.13; FBAld) is an enzyme that catalyzes a reversible reaction of glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) via Schiff base reaction that is crucial for both glycolysis/gluconeogenesis and the pentose phosphate pathways [8]. On the basis of the reaction mechanisms and distribution in the biosphere, FBAld are grouped into two classes, namely Class I and Class II FBAld. Class I FBAld occur as a homotetramer with a molecular mass of 160 kDa, is found in plants and animals, and utilizes Schiff base reaction mechanism. Meanwhile, Class II FBAld occur as a homodimer with approximately 80 kDa, can be found in bacteria and fungi, and uses divalent metal ions for catalysis [9,10]. FBAld in plants can exist in two isomeric forms: chloroplastic and cytosolic [11]. Chloroplastic FBAld is a crucial enzyme that generates metabolites for starch biosynthesis in the Calvin cycle. In contrast, cytosolic FBAld catalyzes the reaction of FBP production in sucrose biosynthetic pathway [12]. Moreover, cytosolic FBAld is also a key enzyme in the glycolytic and gluconeogenesis pathways in both source and sink plant tissues [13]. The gene encoding plastidic and cytosolic glycolytic isozymes is said to have evolved from a common ancestral nuclear gene, or have been transferred from genome to nucleus of prokaryotic symbiont that forms the plastids [14].

Researches involving aldolase in other plants have been shown to be of great interest especially in regulating environment adaptation to abiotic stresses, drought [15,16], responding to hormones such as gibberellin [17], involvement in fruit ripening process [18] and toxicity tolerance towards heavy metals [19,10]. Meanwhile, expression analysis of plastidic aldolase genes in green leaves of *Nicotiana* plants showed increased accumulation (250%) of mRNA, thus suggesting that chloroplast aldolase is involved in salt stress-tolerance [20]. Fan and co-workers [21] reported on the role of FBAld in responding to salt stress and related abiotic stimuli, and in

improving the survival ability in *Sesuvium portulacastrum* under high salinity and drought. Due to the importance of FBAld and sago palm, here we report on an optimized RNA isolation protocol from sago leaves and the isolation of a full-length cDNA of *Metroxylon sagu* FBAld (msFBAld) with high homology to other known chloroplastic FBAld genes. The full-length cDNA sequence was analyzed using various bioinformatics software for genetic characterization. The coding sequence of msFBAld was expressed in *E. coli* to characterize the recombinant enzyme. To further understand the structural features of the cDNA, we generated molecular 3D model structure by homology modeling methods. The knowledge from this work can be useful to study the mechanisms by which sago palm responds and adapts to its adverse environment.

MATERIALS AND METHODS

Plant Materials
Young spineless sago leaves were obtained from Bau area (Sarawak, Malaysia) and rinsed with distilled water. Blades of the leaves were removed and remaining sample was sterilized with DEPC-treated 70% ethanol. Leaf parts, other than the lamina, were discarded and cut into smaller pieces to ease homogenization process.

Extraction of Total RNA and genomic DNA
For total RNA isolation, all apparatus were soaked overnight with 0.1% (v/v) of diethylpyrocarbonate (DEPC)-treated water, drained and autoclaved for 15 minutes at 121°C. Mortars, pestles and metal apparatus were also treated with DEPC and further baked in the oven at 160°C . All buffers, solutions and reagents were prepared with DEPC-treated reverse osmosis water and autoclaved. Total RNA from sago leaves was isolated according to the method as described by Zeng and Yang [22] and a modified method of Azevado et al. [23]. Ten milliliters of cetyl trimethylammonium bromide (CTAB) extraction buffer was pre-warmed at 55°C for 30 min after the addition of 2% polyvinylpolypyrrolidone (PVPP) and 0.2% β-mercaptoethanol. After the addition of 1 g of homogenate from sago leaves, the mixture was incubated at 55°C for 30 min. Following that, one volume of CIA (CHCl$_3$: isoamyl alcohol at 24:1 ratio) was added and centrifuged at 5500×g for 30 min at 4°C to separate the phases. The top phase was collected in an aliquot of 1 ml into 2 ml microcentrifuge tubes, followed by another addition of 1 volume of CIA and centrifuged at 14400×g for 15 minutes at 4˚C. The top phase was collected into a new tube and 1/3 volume of 10 M ice-cold lithium chloride was added prior to overnight incubation at 4°C. The overnight sample was then centrifuged at 14400×g for 15 minutes at 4°C, followed by 3 times washing with 75% ethanol and centrifuged again at 14400×g for 15 minutes at 4°C. The pellet was air dried and re-suspended with RNAse-free water at a final volume of 60 μl. RNA with good integrity and quality was treated with RNAse-Free DNAse I (Fermentas) according to manufacturer instruction. Subsequently, RNA was either stored in -85°C for long term storage or immediately used for cDNA synthesis. Good quality genomic DNA was isolated and purified according to the method developed by Wee and Roslan [5].

Reverse Transcription-Polymerase Chain Reaction (RT-PCR)
A one-step RT-PCR was carried to synthesize cDNA from total RNA. Moloney Murine Leukemia Virus Reverse Transcriptase (MMuLV-RT, Fermentas) was used to synthesize the first strand cDNA from RNA of sago palm young leaves. An

oligo, (dT)$_{17}$ adaptor, was used to synthesize downstream of the gene of interest while a gene specific primer was used to synthesize the upstream region. The gene specific primers were designed on basis of our previously published EST sequence (Accession: JK731404.1 and GI:374530933 [5] (Table 1). The reaction mixture was prepared in a final volume of 20 μl, which was assembled on ice. The reaction included 5-10 μg of purified total RNA, 2 μl of oligo (dT)$_{17}$ adaptor or gene specific primer (10 μm) were added and topped up with RNAse-free water in final volume of 10 μl. The mixture was briefly spun followed by incubation at 65°C for 10 minutes prior to immediate cooling on ice for 2 minutes. Subsequently, 1 U of RNAse inhibitor (Fermentas), 4 μl of RT-buffer (Fermentas), 2 μl of 10 mM dNTPs was added and further incubated at 37°C for 5 minutes. Finally a 1 μl of RT was added and the mixture was incubated for an hour at 42°C. The reaction was heat-inactivated at 70°C for 10 minutes and transferred on ice until further use. Polymerase chain reaction (PCR) was carried out using BioEr GeneQ thermal cycler and/or Bio-Rad MyCycler™ Personal Thermal Cycler. PCR mix was prepared using GoTaq® Green Master Mix (Promega) for amplification of the sago palm cDNA. The PCR was carried out at an initial denaturation of 94°C for 3-5 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at temperatures optimized for each primer combination and elongation at 72°C for 1 minute, followed by final extension at 72°C for 5 minutes.

Rapid Amplification of cDNA Ends (RACE)
To amplify the msFBAld cDNA, RACE was carried out according to a modified method described by Frohman [24]. The forward primer (FBA5endF) was designed based on the conserved region from various FBAld available in NCBI database, while the reverse primer (5endFBAR) was designed from an EST-derived FBAld sequence (table 1). A second set of oligonucleotide primers [(dC)15_ATG _Adaptor & 5endFBA], were also designed for the PCR. The first strand of cDNA was synthesized using reverse primer (5endFBAR) according to the previously described method. The dGTP tailing of the first stand cDNA was performed using a terminal deoxynucleotidyl transferase (TdT) (Fermentas). The PCR mixture was prepared using GoTaq® Green Master Mix (Promega) and the PCR reaction was conducted as described in previous subsection.

Table 1 Primers used for cloning and sequencing of msFBAld.

Name	Sequence (5'>3')	Uses
FBA247-F	CTGCACAGGAAGCACTGTTG	PCR
FBA247R	ATTCCTTCCTTGGCCTCCTC	PCR
Oligo (dT)17	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT	RT-PCR
Oligo T)17 adapter	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT	RT-PCR
Oligo (dT)18	TTTTTTTTTTTTTTTTTT	RT-PCR
5'endFBA-R	CTGGTAAGGGAGATAGACCCTTG	RACE-PCR
(dC)15 ATG adapter	GATCTGTCGACCCCCCCCCCCCCCCCCATG	RACE-PCR
5'endFBA 2nd Nested	CTGGTAAGGGAGATAGACCCTTG	RACE-PCR
NdeI-msFBA-F	CGC **CAT ATG** AAA TCT ATT GCT TCA CC	Full-length
HindIII-msFBA-R	CGC **AAG CTT** GTA GGT GTA GCC TTT T	amplification

cDNA Cloning and Sequencing
For the analysis of PCR, the product was separated on 0.7% agarose gel and

fragment was purified with GF-1 Gel DNA Recovery Kit (Vivantis). The cDNA fragment was then ligated into pGEM-T vector (Promega) and selected on 100 µg/mL ampicillin, 80 µg/mL X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and 0.5 mM IPTG (Isopropyl-ß-D-1-thiogalactopyranoside). Plates were then incubated at 37°C for 16-22 hours. The plasmids were purified and sequenced by First BASE Laboratories (Malaysia).

Construction of an Expression Plasmid containing msFBAld
The full-length sago palm FBAld cDNA coding sequence was ligated into an expression vector, pET41a(+) via NdeI and HindIII sites. The recombinant pET41a(+) vector (pET41a(+)/msFBAld) was then transformed into *E. coli* BL21 (DE3) strain. The fused target protein contains histidine (His) affinity tag at the C-terminal to assist in purification using nickel column. The transformed cell of BL21 (DE3) was then selected on LB agar plate supplemented with 50 µl/ml kanamycin. Plates were then incubated at 37°C for 16-22 hours. The plasmid was purified with GF-1 Plasmid DNA Extraction Kit (Vivantis), visualized on 1% agarose gel pre-stained with ethidium bromide and viewed under UV. The plasmid was purified and sequenced by First BASE Laboratories (Malaysia).

Production and Purification of Recombinant msFBAld
A fresh BL21 (DE3) bacterial colony harboring pET41a(+)/msFBAld was inoculated into 10 ml of LB medium containing 50 mg/ml of kanamycin. The inoculum was incubated with agitation at 120 rpm overnight at 37°C and subsequently culture was transferred into 30 ml of fresh LB medium, containing kanamycin (50 mg/ml) and incubated at 37°C with vigorous shaking until OD600 reaches 0.6. Control culture was prepared by transferring 1 ml of the culture into a new 1.5 ml tube (not induced). Ni-NTA spin column (QIAGEN) was used to isolate the recombinant proteins. For this purpose, 1 mM IPTG was added to the culture and grown at 37°C with vigorous shaking for 4 hours. Cells were harvested via centrifugation at 4000×g for 15 minutes and stored at -20˚C either for longer storage or used immediately. Approximately 30 ml of cells were centrifuged and re-suspended in buffer B (7 M urea, 0.1 M sodium phosphate, 0.01 M Tris-HCl at pH 8.0). The mixture was agitated for 15 minutes at room temperature and the lysates was collected by centrifugation at 12000×g for 15 to 30 minutes. Lysates containing recombinant protein with histidine-tag (His-tag) were purified using Ni-NTA spin column (QIAGEN). The column was first equilibrated with 600 µl of buffer B and centrifuged at 890×g for 2 minutes. Samples were loaded into the column and centrifuge at 270×g for 5 minutes. Following that, Ni-NTA spin column were equilibrated 2 to 3 times with 600 µl of buffer C (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris-Cl at pH 6.3). Finally, the recombinant protein was eluted twice with 200 µl buffer E (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris-Cl at pH 4.5). Protein concentration was determined via Bradford method [25].

SDS-Polyacrylamide Gel Electrophoresis and MALDI-TOF Analysis
The Ni-NTA column-purified recombinant msFBAld was analyzed by SDS-PAGE (12% polyacrylamide) in 0.1 M Tris-glycine buffer using slab gel apparatus according to the method described by Laemmli [26]. A total of 15 µl of the sample was added with 1 volume of 2X SDS sample buffer and preceded with SDS-PAGE. Twenty microliters of the supernatant was added to each lane of the gel. After electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250 and de-stained by washing with a mixture of methanol: acetic acid: water (9:2:9, v/v/v).

PageRuler Prestained Protein Ladder (Thermo Scientific) was used to determine the molecular weight. Experimental result was documented using ImageQuant 400. MALDI-TOF analysis was conducted by First BASE Laboratories (Malaysia) to sequence the peptides and confirm the expression.

Enzyme Assay of Recombinant msFBAld

Aldolase activity of recombinant protein (msFBAld) was assayed according the method described by Fan et al. [21]. The 2,4-dinitrophenylhydrazine (DNPH) reacts with dihydroxyacetone phosphate (DHAP) derived from the decomposition product of D-fructose-1,6-bisphosphate (FBP) to form phenylhydrazone. Enzyme assay at various reaction times over the range 0~30 min were performed in 2 mL mixture (50 mM Tris-HCl buffer pH 7.4, 4 mM D-fructose-1,6-bisphosphate trisodium salt (FBPNa3), 5 mM DNPH). The reaction was allowed to proceed at 37°C after a suitable aliquot of the enzyme was added. Then, the optical density at 540 nm was monitored once every 3 minutes by spectrophotometric measurements.

Sequence Analysis

The nucleotide sequences were aligned using Basic Local Alignment Search Tool (BLAST) program [27] provided by National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute's (EBI) ClustalW multiple alignment software, to search for similarities. Molecular mass and isoelectric point (pI) were predicted using ExPASy software (http://web.expasy.org/compute_pi). The NCBI CD-search tool (CDD V3.0-44354 PSSMs) was used to identify the conserved domains (CD) within msFBAld protein sequence. Phylogenetic tree was constructed using the CLC Sequence Viewer 6 and SignalP 4.1 server was used to predict the signal peptide cleavage site(s).

Molecular Homology Modeling

Overall homology modeling was carried out according to Hossain et al. [28], with modification. The process of protein homology modeling involves several steps, which includes; (1) finding template for the unknown protein, (2) sequence-template alignment, (3) model building, and (4) model assessment [29,30]. Template search with BLAST and HHBlits was performed against the SWISS-MODEL template library. For each identified template, the template quality was predicted from features of the target-template alignment. Templates with the highest quality were selected for model building and were built based on the target-template alignment in the SWISS-MODEL workspace using Promod-II [31,32,33]. Coordinates, which are conserved between the target and the template, were copied from the template to the model. Insertions and deletions were remodeled using a fragment library and side chains were then rebuilt. Finally, the geometry of the resulting model was regularized using a force field. In loop modeling, ProMod-II was used but did not give satisfactory results; an alternative model was built with MODELLER. The global and per-residue model quality was assessed using the QMEAN scoring function [32]. For improved performance, weights of the individual QMEAN terms were trained specifically for SWISS-MODEL. Several programs, PROCHECK [34], WHAT_CHECK [35] and ModEval model evaluation server [36], were used to verify the quality of the modeled structure. Homo-oligomeric structure of the target protein was predicted based on the analysis of pairwise interfaces of the identified template structures. The prediction was performed with a random forest regressor using these features as input parameters to predict the probability of conservation for each interface. The QscoreOligomer of the whole complex was then calculated as the weight-averaged QscoreOligomer of the interfaces. The oligomeric state of the target was predicted to be the same as in the template when

QscoreOligomer was predicted to be higher or equal to 0.5. The COFACTOR, a structure-based method for biological function annotation of protein molecules, was used to identify highly similar protein structures and the functional insights including ligand-binding site, gene-ontology terms, and enzyme classification [37].

Salt-Tolerance Assay of pET41a(+)/msFBAld *E. coli*

Two 15 ml fresh culture of *E. coli* transformants harboring pET41a(+) and pET41a(+)/msFBAld were prepared by overnight incubation at 37°C with agitation at 180 rpm on an orbital shaker. Six 50 ml falcon tubes (each contained 5 ml LB medium, 1 mM IPTG and OD_{600} ~ 0.5) were prepared for each transformant with fresh culture and incubated for 30 min. NaCl at final concentrations of 0 M, 0.2 M, 0.4 M, 0.6 M, 0.8 M and 1.0 M, was directly added to the 5 ml medium of all cultures, and continued to be incubated at 37°C for 2 h. Cell salt-tolerance assay was performed by measuring the OD_{600} value and inoculating onto LB-agar plates with 50 mg/L kanamycin after 10,000-folds dilution. The survival probability of *E. coli* = the number of colonies in medium with different NaCl concentrations/the number of colonies in control medium × 100%. Each treatment was conducted with three repetitions.

RESULTS AND DISCUSSION
Optimization of Good Quality RNA Extraction Protocol for Sago Palm
A good starting RNA material is crucial for downstream application and is one of the prerequisite in subsequent mRNA manipulation. For that reasons, we attempted two RNA extraction methods: the original method developed by Zeng and Yang [22], and our modified method from Zeng and Yang [22] and Azevado et al. [23]. The modified method for good quality RNA from sago leaves was found to be much better than the original method, which was described in the materials and methods section. Using the modified method, good quality of total RNA was obtained with A260/A230 ratio greater than 1.8 and A260/A280 in the range of 1.8-2.0. The relative yield of total RNA per gram tissue can be up to 200 μg/g of leaves sample and 140 μg/g of pith sample of fresh weight. Furthermore, the integrity of ribosomal bands observed with 1% ethidium bromide stained agarose gel showed two distinct bands of 28S rRNA, which was more distinct than 18s rRNA (Fig. 1). RNA isolation protocol as described by Zeng and Yang [22] produced partially degraded total RNA from sago leaves, which were not suitable for downstream application. Significant modification in the protocol includes addition of insoluble polyvinylpolypyrrolidone (PVPP) prior to sample incubation and longer centrifugation time. In addition to polyvinylpyrrolidone (PVP) used in the original protocol, the use of insoluble PVPP also prevents oxidation of polyphenols and binding to nucleic acids as well as protein latter [38]. Salzman et al. [39] reported that the addition of PVPP works moderately in the isolation of RNA, and PVP polymers require lengthy ultracentrifugation [38]. Addition of both polymers into CTAB buffer in this study was found to work well in isolating RNA from young leaves of sago palm. Furthermore, using the modified method, no significant RNA degradation was found with unfavorable protein and carbohydrate contamination in the extracted RNA.
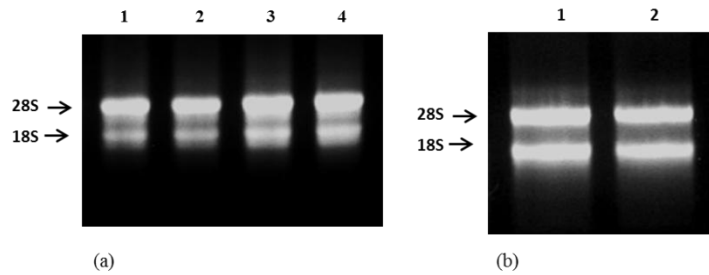
Figure 1. Visual analysis of RNA extracted using modified method. The RNAs were resolved with 1% agarose gel stained with ethidium bromide in 1X TAE. RNA from young leaves (a) and pith (b) of sago palm extracted with 5 µl of RNA sample loaded in each well each containing 2 µg of extracted sago leaves RNA. Clear and distinct 28S and 18S rRNA could be seen indicating good quality RNA.

RT-PCR Amplification, Cloning and Identification of msFBAld

A set of specific primer, 247F and 247R, were designed based on the conserved region of a partial msFBAld derived from our previously published EST data (Accession no. JK731404.1). The purpose of this step was to verify the sequence prior to isolating the 3' and 5' ends of the msFBAld. After the first round gradient PCR amplification using this primer pair, PCR product with approximate size of 255 bp was obtained with annealing temperature at 53°C. Sequence analysis of this fragment showed 100% identity with sequence of predicted EST sequence. Three oligo (dT) reverse primers, namely Oligo(dT)17, oligo(dT)18 and Oligo(dT)17 adapters, with gene specific forward primer (247F) were then used to amplify the 3'end of msFBAld cDNA. However, no satisfactory results were generated using the first two primers. Non-specific binding of the primers and templates may have been the cause where rather multiple bands or no PCR product were detected when analyzed on agarose gel electrophoresis. On the contrary, oligo(dT)17 adapter with forward primer (247F) was able to generate a PCR product with the size of approximately 390 bp. Sequence alignment of the 390 bp PCR product revealed sequence continuation of the 3'-end of EST nucleotide sequence. Then we used RACE-PCR to amplify the 5'end of msFBAld. Using the primers FBA5endF and 5endFBAR, a total of 283 bp of nucleotides was obtained from the sequencing whereas 149 bp of nucleotides obtained from sequencing result of primer (dC)15 ATG Adaptor and 5endFBA 2nd nested. When both 5'RACE sequences were assembled, a total of 368 nucleotides were obtained from the upstream region of the 5' cDNA of msFBAld. Finally, assembly of the partial cDNAs resulted in the generation of a full-length msFBAld cDNA that contained 1288 nucleotides with an open reading frame of 1020 nucleotides (Fig. 2). It also contained start-stop codons and 238 bp 3'-UTR (untranslated region) including a poly-A tail at the 3'end. The nucleotide sequence of msFBAld mRNA was registered under the GenBank accession no. KP271949. Nucleotide BLAST search results showed the maximum identities of msFBAld mRNA with mRNA sequences of *Phoenix dactylifera* (93%, XM_008813494.1), *Musa acuminata* subsp. *malaccensis* (83%, XM_009388988.1), *Ricinus communis* (93%, XM_002526262.1), *Nelumbo nucifera* (82%, XM_010281106.1) and *Vitis vinifera* (82%, XM_010663987.1). To determine the genetic organization of *msFBAld*, the genomic DNA was amplified from sago leaves using NdeI-msFBAld-F and HindIII-msFBAld-R primers and sequenced. The genomic DNA amplification yielded 2322 bp of nucleotides. The sequence generated showed that the *msFBAld* gene consisted of five exons: exon-1 (1~274 bp), exon-2 (388-498 bp), exon-3 (619~709 bp), exon-4 (1558~1826 bp) and exon-5 (2047~2322 bp) (Fig. 3).

```
   1 - ATGAAATCTATTGCTTCCCCTGGACGTGGTATCCTTGCAATTGATGAATCAAATGCTACA  -  60
   1 - M   K   S   I   A   S   P   G   R   G   I   L   A   I   D   E   S   N   A   T  -  20
  61 - TGTGGGAAGAGGCTTGCTTCAATTGGTTTGGAAAATACAGAAGCAAATCGCCAAGCTTAC  - 120
  21 - C   G   K   R   L   A   S   I   G   L   E   N   T   E   A   N   R   Q   A   Y  -  40
 121 - AGACAGCTTCTGCTGACAACTCGTGGTTTAGGTGAATACATCTCTGGTGCCATTCTTTTT  - 180
  41 - R   Q   L   L   L   T   T   R   G   L   G   E   Y   I   S   G   A   I   L   F  -  60
 181 - GAGGAAACACTTTACCAGTCGACAACAGATGGAAAGAAGTTTGTAGATTGTCTGTGTGAA  - 240
  61 - E   E   T   L   Y   Q   S   T   T   D   G   K   K   F   V   D   C   L   C   E  -  80
 241 - GAGAAAATGATCCCTGGTATTAAAGTTGACAAGGGTCTATCTCCCTTACCAGGATCAAAT  - 300
  81 - E   K   M   I   P   G   I   K   V   D   K   G   L   S   P   L   P   G   S   N  - 100
 301 - AATGAATCCTGGTGTCAAGGTCTGGATGGATTGGCATCACGATGTGCTGAATACTACAAG  - 360
 101 - N   E   S   W   C   Q   G   L   D   G   L   A   S   R   C   A   E   Y   Y   K  - 120
 361 - CAAGGGGCACGTTTTGCTAAGTGGCGAACTGTAGTTAGCATTCCTTGTGGTCCTTCTGCC  - 420
 121 - Q   G   A   R   F   A   K   W   R   T   V   V   S   I   P   C   G   P   S   A  - 140
 421 - CTGGCTGTTAAGGAGGCCGCATGGGGGCTTGCTCGTTATGCTGCTATTTCTCAGGACAAT  - 480
 141 - L   A   V   K   E   A   A   W   G   L   A   R   Y   A   A   I   S   Q   D   N  - 160
 481 - GGTCTTGTGCCAATCGTCGAACCTGAGATTCTCCTTGATGGAGACCACTCTATTGAAAGG  - 540
 161 - G   L   V   P   I   V   E   P   E   I   L   L   D   G   D   H   S   I   E   R  - 180
 541 - ACACTTGAAGTGGCAGAGAATGTCTGGGCAGAAGTATTCTACTACCTTGCGCAAAATAAT  - 600
 181 - T   L   E   V   A   E   N   V   W   A   E   V   F   Y   Y   L   A   Q   N   N  - 200
 601 - GTGGTGTTTGAGGGCATCCTTCTGAAACCCAGCATGGTGACTCCAGGTGCTGAACACAAG  - 660
 201 - V   V   F   E   G   I   L   L   K   P   S   M   V   T   P   G   A   E   H   K  - 220
 661 - CAGAAAGCATCTCCAGAGACCATTGCCAAGTTCACTCTCAAAATGCTGCACAGGAGGGTA  - 720
 221 - Q   K   A   S   P   E   T   I   A   K   F   T   L   K   M   L   H   R   R   V  - 240
 721 - CCACCTGCAGTTCCTGGAATCATGTTTCTGTCGGGAGGGCAGTCCGAATTGGAAGCAACA  - 780
 241 - P   P   A   V   P   G   I   M   F   L   S   G   G   Q   S   E   L   E   A   T  - 260
 781 - TTGAACTTGAATGCAATGAACCAAGGCCCCAACCCTTGGCATGTCTCTTTCTCGTATGCT  - 840
 261 - L   N   L   N   A   M   N   Q   G   P   N   P   W   H   V   S   F   S   Y   A  - 280
 841 - CGAGCTCTTCAAAATACGGTCTTGAAGACATGGCAGGGGCGTCCAGAGAATGTGGAAGCT  - 900
 281 - R   A   L   Q   N   T   V   L   K   T   W   Q   G   R   P   E   N   V   E   A  - 300
 901 - GCACAGGAAGCACTGTTGGTGCGTGCCAAGGCTAACTCTCAGGCTCAGCTTGGCCGCTAC  - 960
 301 - A   Q   E   A   L   L   V   R   A   K   A   N   S   Q   A   Q   L   G   R   Y  - 320
 961 - TCTGCCGAGGGTGAGAGCGAGGAGGCCAAGGAAGGAATGTTCCAAAAAGGCTACACCTAC  -1020
 321 - S   A   E   G   E   S   E   E   A   K   E   G   M   F   Q   K   G   Y   T   Y  - 340
1021 - TAAAACGGGTGGACGGCACTCTTAAGTCTAAGTAGTTTTCATGAAGCTCAGAGAGTTGTT  -1080
 341 - *
1081 - TTGAGTAAATTATTTTATCAGAGATGGTCTATTAAATGACAATAATAAGTATGCTGTTCT  -1140
1141 - CTTCGTCGTAACAGTGAACTGCAAGACAAGATATGACTTTTGGATTGATTGAACGAAATG  -1200
1201 - CATTCTGTAACCATGGATATTACTATCATATTCGAGAATATTTGAGGATTTTGAGAAAAA  -1260
1261 - AAAAAAAAAAAAAATCGATGTCGACCGTC                                -1288
```

Figure 2. Nucleotide sequence of msFBAld cDNA showing the coding sequence and a 207 nucleotide 3'-untranslated region.
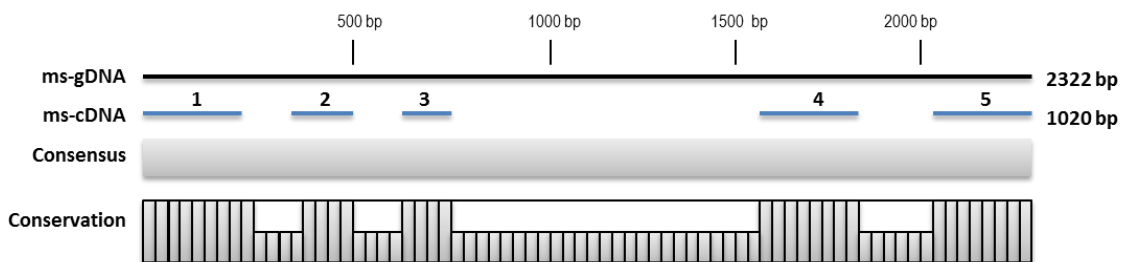

Figure 3. Organization of msFBAld gene. Schematic diagram between cDNA and genomic sequences indicating the positions of exons and introns. The full length msFBAld gene spans 2322 bp containing five exons and four introns.

Sequence Analysis and Localization of msFBAld

The NCBI conserved domain (CD) database analysis showed that msFBAld contained fructose-1,6-bisphosphate aldolase domain which belonged to aldolase Class 1 group and a TIM-barrel phosphate binding superfamily domain (Fig. 4). Based on CD database available and three-dimensional structure-activity relationship, the amino acid residues Asp(15), Glu(16), Ser(17), Thr(20), Lys(88), Lys(127), Arg(129), Glu(168), Lys(209), Leu(250), Ser(251), Gly(252), Ser(278) , Ala(280) and Arg(281) were predicted to be present at the active site of msFBAld, which were conserved in other FBAld sequences. Deduced amino acid sequence from expressed msFBAld cDNA contains 355 residues with a predicted molecular

mass of 39.14 kDa and isoelectric point of 6.49, determined using ExPASy online software (http://web.expasy.org/compute_pi). This predicted molecular mass of msFBAld, which was in good agreement with molecular masses of other aldolases that ranges from 35-40 kDa (Moorhead and Plaxton, 1990). Further analysis by PredictProtein revealed that msFBAld sequence contained a stretch of ten amino acids (VVFEGILLKP) conserved within the active site of fructose-1, 6-bisphosphate aldolase class 1, for example in FBA of *Sesuvium portulacastrum* [21]. The lysine residue (K) located at the active site is essential for aldol condensation reaction by catalyzing Schiff's base intermediate to form dihydroxyacetone-phosphate [21,40]. PredictProtein also identified an acidic region as the *N*-myristoylation sites and three potential amidation sites in the amino acid sequence of msFBAld (data not shown). Meanwhile, SignalP4.1 server could not find any signal peptide in the amino acid sequence of msFBAld, suggesting that msFBAld is not a secreted protein. Since aldolase class I exist in cytosol and plastid [21,18], NCBI database search of msFBAld protein resulted in a mixed of both cytoplasm and plastids. Hence, a phylogenetic tree was reconstructed to analyze the compartmentalization of msFBAld with close resemblance to twelve higher plants FBA available in the GenBank. The phylogenetic tree clearly showed that msFBAld is classified under plastidic aldolases and belongs to the same cluster as *Nicotiana paniculata* (BAA77604/BAA77403), *Trifolium pretense* (AAR10885) and *Oryza sativa* (BAA02730) (Fig. 5).

```
Feature 1                                  ###  #
4ALD          12 .[8].HRI.[3].GKGILAADESTGSIAKRL.[2].IGTENTEENRRFYRQLLLTADDR.[2].PCIGGVILFHETLYQK 86
query          1 .[1].KSI.[3].GRGILAIDESNATCGKRL.[2].IGLENTEANRQAYRQLLLTTRGL.[1].EYISGAILFEETLYQS 67
gi 113617      9 .[8].KYI.[3].GRGILAADESTETIGKRF.[2].INVENTESNRQAYRELLFTSPGS.[1].PCLSGVILFEETLYQK 82
gi 113622      9 .[8].AYI.[3].GKGILAADESTGTIGKRF.[2].INVENVEENRRSLRELLFCTPGA.[1].QYLSGVILFEETLYQK 82
gi 1703244    13 .[8].QRI.[3].GKGILAADESTGSMAKRL.[2].IGVENTEENRRLYRQLLFTADER.[2].KCIGGVIFFHETLYQK 87
gi 1703248    13 .[8].QAI.[3].GKGILAADESTATMGKRL.[2].IGVENNEENRRLYRQLLFSADHK.[2].ENISGVILFEETLHQK 87
gi 3913018    42 .[8].KTI.[3].GRGILAMDESNATCGKRL.[2].IGLENTEANRQAYRTLLVTAPGL.[1].QYISGAILFEETLYQS 115
gi 3928511    13 .[8].QRI.[3].GKGILAADESVGSMDKRL.[2].IGVDNTEENRRLYRQLLFKGDER.[2].ECIGGVIFFHETMYQQ 87
gi 19075034    7 .[8].KKI.[3].GKGILAADETPKTLGRRF.[2].LGITNTEENRRKFREILFSTKGI.[1].RYIGGVILNQETFEQT 80
gi 25989716   13 .[8].RKI.[3].GKGILAADESTGTIKKRF.[2].IGVENTEANRAFYRDLLFSTKGL.[1].QYISGAILFEETLYQK 86

Feature 1                        #                             # #
4ALD          87 ADDGRPFPQVIKSKGGVVGIKVDKGVVPL.[5].ETTTQGLDGLSERCAQYKK.[1].GADFAKWRCVLKIG.[6].LAI 163
query         68 TTDGKKFVDCLCEEKMIPGIKVDKGLSPL.[5].ESWCQGLDGLASRCAEYYK.[1].GARFAKWRTVVSIP.[5].LAV 143
gi 113617     83 TSDGKPFVDLLMENGVIPGIKVDKGLVDL.[5].ETTTQGLDSLGARCQQYYE.[1].GGRFAKWRAFFKIG.[6].LSI 159
gi 113622     83 TKDGKPFVDVLKEGGVLPGIKVDKGTIEV.[5].ETTTQGHDDLGKRCAKYYE.[1].GARFAKWRAVLKIG.[6].LAI 159
gi 1703244    88 ADDGTPFAKMIKDRGIVVGIKVDKGVVPL.[5].ETTTQGLDGLSERCAQYKK.[1].GADFAKWRSVLKIS.[6].LAI 164
gi 1703248    88 SDDGKTLPTLLAERNIIPGIKVDKGVVPL.[5].ETTTQGLDDLASRCAEYWR.[1].GCRFAKWRCVLKIS.[6].LAM 164
gi 3913018   116 TVDGKKIVDILTEQKIVPGIKVDKGLVPL.[5].ESWCQGRDGLASREAAYYQ.[1].GARFAKWRTVVSIP.[5].LAV 191
gi 3928511    88 GDCGTDFVKMIKDKGIVVGIKVDKGVVPL.[5].ETTTQGLDGLSERCAQYKK.[1].GADFAKWRSVLKIS.[6].LAI 164
gi 19075034   81 SGSGVPLTELLKKKGIEIGIKLDKGLIDY.[3].EKISVGLEDLDLRCKSSAF.[1].DATFAKWRSLFYFY.[6].DCI 155
gi 25989716   87 SPSGVPMVDLLKAEGIIPGIKVDKGLETL.[5].EKATMGLDGLSERCKKYYE.[1].GARFAKWRAVLSID.[8].LSI 165

Feature 1                                #                                   #
4ALD         164 MENANVLARYASICQQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLLKPNMVTPG.[7].SH 245
query        144 KEAAWGLARYAAISQDNGLVPIVEPEILLDGDHSIERTLEVAENVWAEVFYYLAQNNVVFEGILLKPSMVTPG.[7].SP 225
gi 113617    160 QEDARVLARYAIICQENGLVPIVEPEVLTGGSHDIKKCAAVTETVLAAVFKALNYHHVLLEGTLLKPNMVTPG.[6].AP 240
gi 113622    160 HLNAQGLARYAIICQENGLVPIVEPEILVDGPHDIDPCAYVSEVVLAACYKALNEHHVLLEGTLLKPNMVTPG.[6].SP 240
gi 1703244   165 MENANVLARYASICQQNGLVPIVEPEILPDGDHDLKRCQYVTEKVLAACYKALSDHHVYLEGTLLKPNMVTAG.[7].SN 246
gi 1703248   165 LENANVLARYASICQQNGLVPIVEPEVLPDGDHDLLTAQRVTEQVLAFVYKALADHHVYLEGTLLKPNMVTAG.[7].TP 246
gi 3913018   192 KEAAWGLARYAAISQDNGLVPIVEPEILDGDEHGIDRTFEVAQKVWAETFFYMAENNVMFEGILLKPSMVTPG.[7].TP 273
gi 3928511   165 LENANVLARYASICQQNGIVPIVEPEILPDGEHDLKRCQYITEKVLAAVYKALSDHHVVYLEGTLLKPNMVTAG.[7].SA 246
gi 19075034  156 NENCSILAKYAIICQKNGLVPIVEPEVFLEGDYSMKRSYEVTRQILSTLMKYLNYELVYIPGVLIKASYVTSG.[7].TP 237
gi 25989716  166 TEVAHGLARYAAICQANRLVPIVEPEILTDGSHDITVCAEVTERVLAAVFKALNDHHVLLEGALLKPNMVTHG.[7].SH 247

Feature 1                                ###                        # ##
4ALD         246 EEIAMATVTALRRTVP.[2].VTGITFLSGGQSEEEASINLNAIN.[6].PWALTFSYGRALQASALKAWGG.[2].ENL 320
query        226 ETIAKFTLKMLHRRVP.[2].VPGIMFLSGGQSELEATLNLNAMN.[4].PWHVSFSYARALQNTVLKTWQG.[2].ENV 298
gi 113617    241 ELIAEYTVTALRRTVP.[2].IPGIVFLSGIQREEQATLNLNAMN.[6].PWTLTFSFGGALQQSAIKAWAG.[2].ENV 315
gi 113622    241 EVIAEYTVRTLQRTVP.[2].VPAIVFLSGGKSEEATRNLNAMN.[6].PWSLSFSFGRALQQSTLKAWSG.[2].ANI 315
gi 1703244   247 QEIAMATVTALRRTVP.[2].VTGVTFLSGGQSEEEASINLNAIN.[6].PWALTFSYGRALQASALSAWRG.[2].ENE 321
gi 1703248   247 QENALATVRALQRTVP.[2].VPGITFLSGGQSELDATKNLNEIN.[6].PWALTFSYGRALQASVLATWKG.[2].ENV 321
gi 3913018   274 EQVSDYTLKLLHRRIP.[2].VPAIMFLSGGQSEVEATQNLNAMN.[4].PWHVSFSYARALQNTCLKTWGG.[2].ENV 346
gi 3928511   247 EEVAMATVTALRRSVP.[2].VPGVTFLSGGQSEEEATINLNAIN.[6].PWALTFSYGRALQASALNAWRG.[2].DNE 321
gi 19075034  238 KKVATFTLRALLSTIP.[2].IPGIVFLSGGHGSEDAIGFLNAIN.[6].TWSLSFSFARALTDGVLETWRG.[2].SNI 312
gi 25989716  248 EEIAFYTVRSLKRTVP.[2].LPGVMFLSGGQSEEDASLNLNEMN.[5].PFQLSFSYGRALQASCLKAWKG.[2].ENK 321

Feature 1
4ALD         321 KAAQEEYVKRALANSLACQGKY 342
query        299 EAAQEALLVRAKANSQAQLGRY 320
gi 113617    316 AKAQAKFLTRCKANKDATLGKY 337
gi 113622    316 EKARAAFLTRCKANSEATLGTY 337
gi 1703244   322 KAATEEFLKRAEANGLAAQGKY 343
gi 1703248   322 HAAQEELLKLAKANGAAAVGKF 343
gi 3913018   347 KAAQDRLLLRAKANSLAQLGKY 368
gi 3928511   322 DAATKVFLTRAEANGLAALGKY 343
gi 19075034  313 EEAQKILLETSFKACRGAEGKL 334
gi 25989716  322 AKAQQVLMERARANGEAQLGKY 343
```

Figure 4. Conserved domain of *Metroxylon sagu* fructose bisphosphate aldolase (msFBAld) aligned with nine other FBAlds from other organisms. Input sequence, msFBAld, is represented as query. The alignment showed that conserved active site residues in all FBAlds, (indicated in yellow highlight and designated # above the sequences). The alignment was generated with Conserved Domain Database [41]; 4ALD, Human muscle fructose 1,6-bisphosphate aldolase complexed with fructose 1,6-bisphosphate; gi113617, fructose-bisphosphate aldolase, cytoplasmic isozyme from *Arabidopsis thaliana*; gi113622, fructose-bisphosphate aldolase, cytoplasmic isozyme from *Oryza sativa*; gi1703244, brain-type aldolase from *Carassius auratus*; gi1703248, fructose-bisphosphate aldolase from *Schistosoma mansoni*; gi3913018, fructose-bisphosphate aldolase, chloroplast precursor from *Oryza sativa*; gi3928511, aldolase from *Xenopus laevis*; gi19075034, fructose-bisphosphate aldolase b from *Encephalitozoon cuniculi* and gi25989716, fructose-1,6-bisphosphate aldolase from *Toxoplasma gondii*. Color bits are shown as identity.
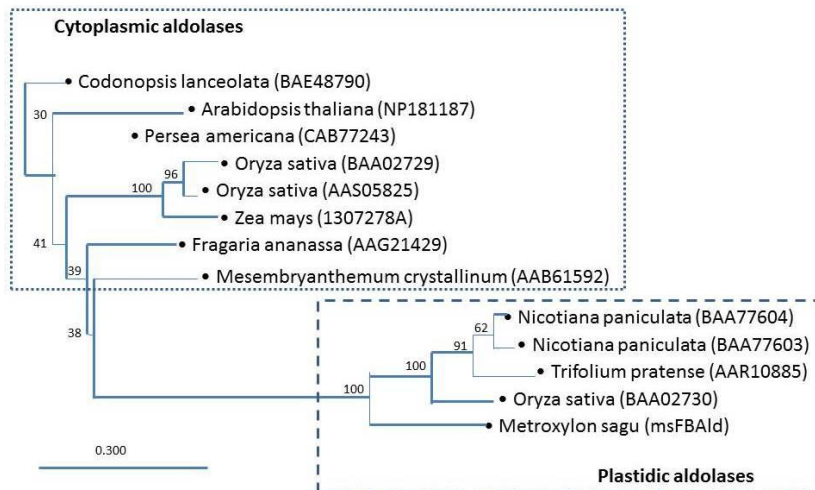
Figure 5. Phylogenetic tree based on amino acids sequence of aldolases, showing the relationships between cytoplasmic and plastidic aldolase of other plants. Neighbor joining method was used and a bar representing 0.30 substitutions per amino acids position. The phylogenetic tree was reconstructed using CLC Sequence Viewer 6.

### Homology Modeling, Quality Control and Resolved Three-Dimensional Structure

To better understand the structure-function of proteins, the 3D modeled structure of msFBAld was generated using SWISS-MODEL web server. A template search was performed using amino acid sequence of msFBAld against the SWISS-MODEL template library resulting in 170 matching templates. The quality of templates was checked manually on the basis of their E-value, sequence identity, target-template alignment and classification. The best quality template was found to be fructose-1,6-bisphosphate aldolase (pdb: 1j4e.1.A, resolution: 2.65Å) from rabbit muscle tissue, which had 55.8% sequence identity with msFBAld and belonged to FBA class I. Accuracy of model structure depended much on the template-target sequence homology or identity. Homology modeling can accurately predict protein structure if sequence identity of target-template is more than 30%. Model was built based on the target-template alignment using ProMod-II in SWISS-MODEL workspace. In case of loop modeling, an alternative model was built with MODELLER [42].

The quality of model was checked through a series of tests such as DFire [43], QMEAN [32], PROCHECK [34], WHAT_CHECK [35] and ModEval model evaluation server [36]. DFire and QMEAN6 analyzes of the msFBAld-modeled structure were determined to be -1861.64 and 0.72, respectively, which were good indicator for a reliable modeled structure. The Ramachandran produced by PROCHEK showed that 86.7% of the residues were in the "core" region with 12.9% in the "additional allowed region". Further 0.5% residues were in the "generally allowed" region and no residues were in the "disallowed" region (Fig. 6). The Overall G-factor was 0.17 (dihedrals: 0.02 and covalent: 0.41). The main chain and side chain parameters were also better. The M/C bond length and bond angle were 99.4% within the normal limits. The overall structural assessment by PROCHECK was very good and reliable. WHAT_CHECK showed that Ramachandran Z-score (-0.68) was acceptable for modeled structure. The score indicates how well the backbone conformations of all residues correspond to the known allowed areas in Ramachandran plot were within expected ranges for well-refined structures. All B-factors fell in the range of 0.0-100.0. Structural average packing environment was

satisfactory and none of the residues was within the bad packaging environments. The overall second generation packing quality was acceptable. The RMSD (all C-alpha atoms) and z-Dope scores as determined by model evaluation server were found to be 1.47 and -1.37, respectively, which were within the normal limits. The GA341 (model score) and predicted native overlap (3.5 Å) were 1.0 and 0.95, respectively. A model is predicted to be reliable when the model score (GA341) is higher than a pre-specified cutoff (0.7). Therefore, modeled structure was reliable according to results obtained from model evaluation server.

PROCHECK

Ramachandran Plot
of msFBAld
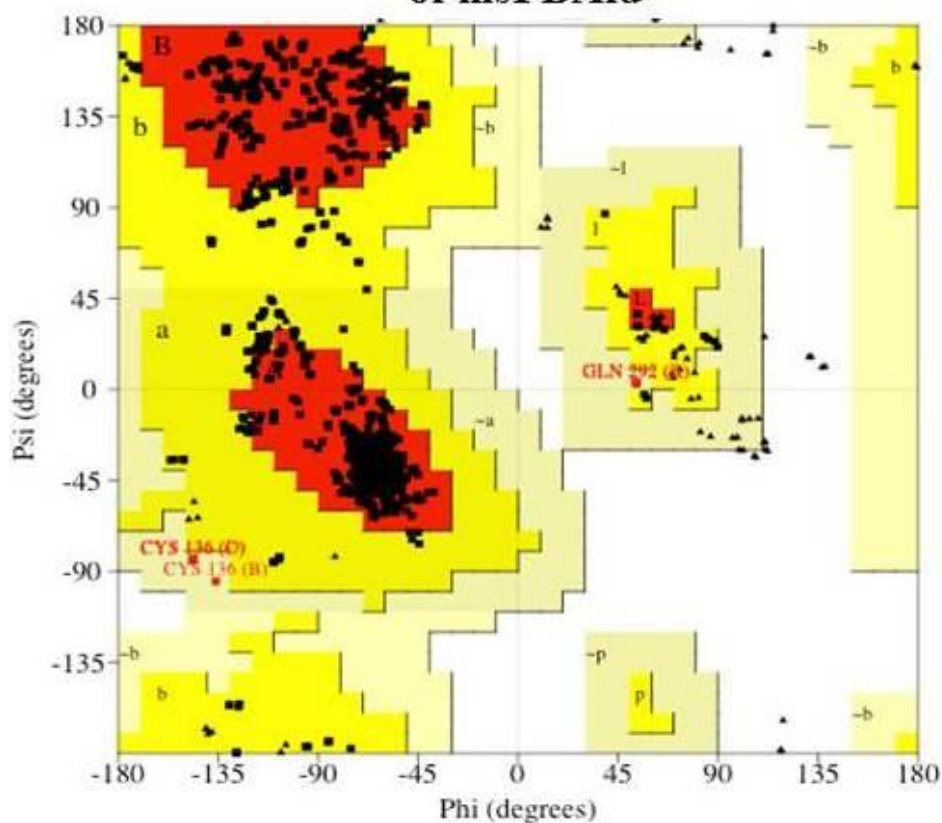
Plot statistics

Residues in most favoured regions [A.B.I]	957	86.7%
Residues in additional allowed regions [a,b,1,p]	142	12.9%
Residues in generously allowed regions [~a,~b,~1,~p]	5	0.5%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	1104	100%
Number of end-residues (excl. Gly and Pro)	8	
Number of glycine residues (shown as triangles)	104	
Number of proline residues	68	
Total number of residues	1284	

Figure 6. Ramachandran plot of the modeled structure of msFBAld generated via PROCHECK.

Rabbit muscle fructose-1,6-bisphosphate aldolase (pdb:1j4e) was used as a template for homology modeling and from this, the modeled structure of msFBAld was determined to contain four monomers in an asymmetric unit. The overall modeled tetrameric structure of msFBAld is shown in figure 7a. It contains 321 residues (position 2-322) although full-length protein had 340 residues. The 18C-terminal residues were not included in the modeled structure (msFBAld) because of the absence of similar residues in the target structure (pdb:1j4e). The results also showed that the 19C-terminal residues were disordered and not included in the model (pdb:1j4e) [44]. Furthermore, it was suggested that this C-terminal region of fructose-1,6-bisphosphate aldolase was involved in substrate specificity, based on the results of chemical modification, truncation, and site-directed mutagenesis experiments [44]. Therefore, it is speculated that the 18C-terminal residues of msFBAld is involved in substrate specificity in sago palm. The modeled structure consists of eleven α-helices (16.2%) and ten β-pleated sheets (42.4%). The central part of the modeled structure contains the β/α TIM-barrel housing active residues. The modeled structure was found as tetrameric structure as generated in template structure (Fig. 7b). Superimposition of modeled structure with template structure (pdb:1j4e) showed perfect fit between the structures (Fig. 7c), suggesting reliability of the generated modeled structure. Superimposition of structures such as msFBAld, rabbit aldolase (1j4e) and human muscle aldolase allowed for the identification of catalytically important conserved residues at the active site of msFBAld-modeled structure that includes; Ala(13), Asp(15), Lys(127), Glu(167), Lys(188), Lys(209), Ser(251), Gly(252), Ser(278), Ala(280) and Arg(281) (Fig. 7d). Three conserved residues (Lys209, Glu167 and Asp15) from msFBAld overlapped with the conserved residues Lys(229), Glu(187), Asp(33) of rabbit and human aldolases [44,45] (Fig. 7). The residue Lys(209) was assumed to form Schiff's base with substrate, DHAP, in msFBAld as compared with in rabbit aldolase [44]. Meanwhile, Asp(15) might play a role in deprotonation of the C4-hydroxyl leading to C-C bond cleavage. Meanwhile, the position of Glu(167), in relation to DHAP Schiff's base with Lys(209), is consistent with a role of protonation of the hydroxyl group of the carbinolamine in dehydration step, catalyzing the formation of Schiff's base. The positions of catalytic dyad, Lys(209)-Glu(167) were very much conserved between the three overlaid structures. In the ligand DHAP-docked structure, the distances from Asp15 and Lys127 to C3 of DHAP were 3.87 and 2.22 Å, respectively. On the other hand, C4-hydroxy group of docked fructose-1,6-bisphosphate (FBP) was located at 4.78 Å from the carboxylate oxygen of Asp15 and 2.29 Å from the epsilon-amino group of Lys127, each of which could be predicted to be catalytic bases that initiate C-C bond cleavage as compared with rabbit aldolase [44]. The 1-phosphate group of FBP is associated with the main-chain nitrogen of Ser(251) and Gly(252). The interaction between the two adjacent main-chain nitrogens indicates a strained conformation of the protein backbone at this position. Both of these amino acids are conserved throughout the type I aldolase sequences [45].
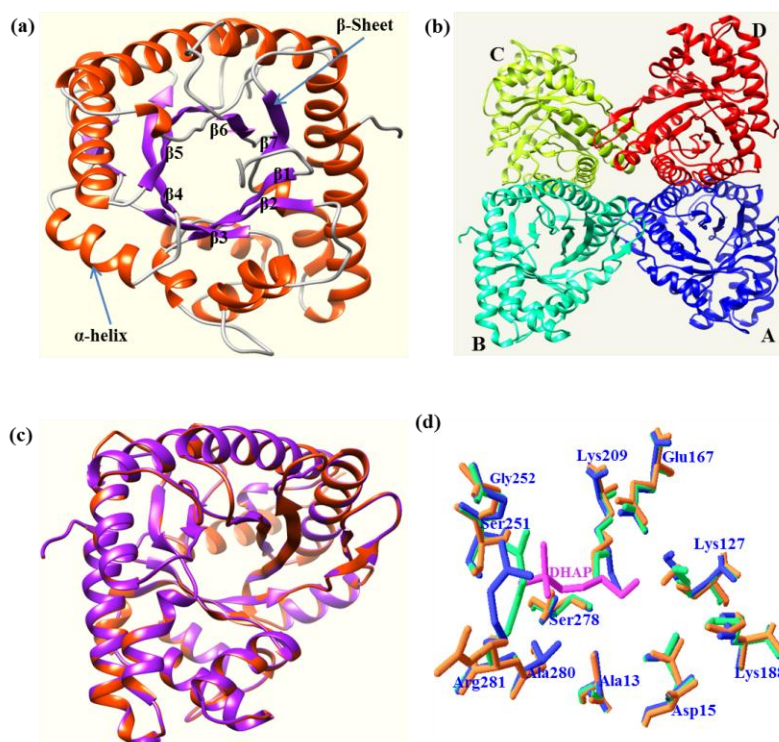
Figure 7. Molecular modeled structure of msFBAld generated via SWISS-MODEL web server. (a) Monomeric structure of msFBAld containing β/α TIM-barrel. (b) Modeled structure containing four monomeric units of msFBAld such as A (blue), B (cyan), C (green) and D (red) as found in rabbit muscle aldolase. (c) Magic fit alignment of the modeled structure (purple) of msFBAld with template structure (red) of rabbit aldolase (pdb:1j4e). (d) Superimposition of active site residues of modeled msFBAld (blue), rabbit muscle aldolase (green) with bound DHAP (pink) complex, and human muscle aldolase (orange). Bound dihydroxyacetone phosphate (ligand) is indicated as DHAP (pink). Images were prepared using Chimera and SWISS-PdbViewer.

The COFACTOR [37] analysis was used to identify highly similar protein structures and analysis of functional insights including ligand-binding site, gene-ontology terms, and enzyme classification. Ten (10) proteins with highly similar structure in protein data bank (PDB) were identified by the COFACTOR that includes 1j4e, 1a5c, 1n30, 3kx6, 2qdh, 3mmt, 1f2j, 1fba, 3qrh and 3mhg. From these, rabbit muscle aldolase (1j4e), which was then used as template for modeling, had the top score for root means square (rms) deviation, identity and coverage (data not shown). The rms deviation between the modeled structure, msFBAld and rabbit enzyme complex for the 341 common Cα atoms was 0.33 Å. This indicated that the modeled structure is very similar in conformation to that of rabbit aldolase A [46]. COFACTOR also identified 13 residues as ligand-binding or active sites in the modeled structure of msFBAld and predicted fructose-1,6-bisphosphate (Fru-1,6-P2) as a ligand or substrate for this enzyme. The residues identified are Ala(13), Asp(15), Ser(17), Thr(20), Arg(24), Lys(127), Arg(129), Lys(209), Leu(250), Ser(251), Gly(252), Ala(280) and Arg(281) (Fig. 8). Of these residues, Lys(127) can form Schiff base with natural substrate Fru-1,6-P2 thereby catalyzes the cleavage of the C3-C4 bonds to form DHAP and G3P. The C1-phosphate binding site involves the three backbone nitrogens of Ser (251), Gly(252), and Ala(280), and the epsilon-amino group of Lys(209). Meanwhile the C6-phosphate binding site involves three basic side chains, Arg(281), Arg(24), and Lys(23), as compared with the rabbit muscle aldolase [46]. To predict the functions of modeled structure of msFBAld, we used COFACTOR and identified 27 gene ontology (GO) terms. The

consensus prediction of GO terms includes molecular function, biological process and cellular location. For molecular function, the GO:0004332 was found top ranked among the predicted seven GO terms, clustered in the fructose-bisphosphate aldolase activity. Meanwhile, the top scoring GO term for biological process was GO:0006096, indicating that the enzyme is involved in glycolytic process. Similarly, cellular location of this enzyme by top scoring GO term (GO:0005829) was found to predict the enzyme to be in the cytosol, although the possibility of localization in plastid was not excluded (data not shown). These predicted results were consistent with the expected molecular, cellular and biological functions of msFBAld. The COFACTOR also identified msFBAld structure with the classification EC 4.1.2.13 and predicted that amino acid residues Asp(15), Glu(167) and Lys(209) plays important role in enzymatic reaction, which were supported by the findings in structural superimposition. The modeling studies suggested that two lysine residues, Lys(127) and Lys(209), in msFBAld can form Schiff's base with the help of Asp(15), Glu(167) for the catalyzation of aldol condensation reaction.
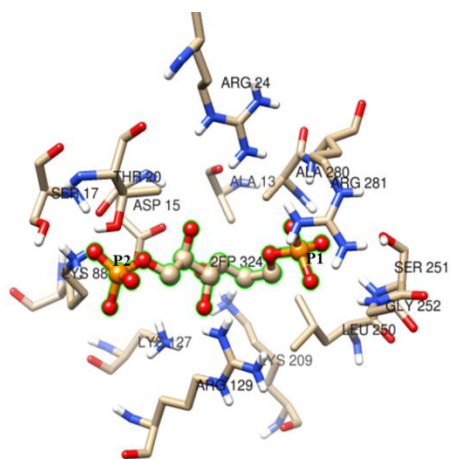

Figure 8. Predicted binding residues at the active site of msFBAld identified by COFACTOR in the template protein-complexes for Schiff's base intermediates, analyzed using rabbit muscle aldolase structures (pdb id: 2quv, 2ot0, 6ald and 3tu9) with similar binding. Fructose-1,6-bisphosphate (natural substrate) was predicted to form Schiff's base with surrounding residues. The C1- and C6-phosphate groups are labeled P1 and P6, respectively. Image was prepared using Chimera.

Expression and Characterization of Recombinant msFBAld cDNA in *E. coli*

Sequence analysis showed that the full-length msFBAld cDNA sequence had an open reading frame (ORF) of 1020 bp and encoded for 341 amino acids. The msFBAld cDNA (1020 bp) amplified by PCR using the primers NdeI-F and HindIII-R, was cloned into the pET41a(+) expression vector. The histidine (His) affinity tag was fused to the C-terminal to assist in the purification using nickel column. The recombinant vector, pET41a(+)/msFBAld, was transformed into competent cells of *E. coli* BL21 (DE3). The transformants were analyzed and inserted DNA sequence was confirmed by PCR amplification and sequence analysis. For the analysis of expression, an inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture to a final concentration of 1 mM and grown at 37°C with vigorous shaking for 4 hours. Harvested cells were centrifuged at 4000×g for 15 minutes and cell lysate was prepared for expression analysis. Cell lysate was analyzed on SDS-PAGE and Ni-NTA spin column was used to purify the samples. Protein-staining revealed a band indicating protein expression of msFBAld (Fig. 9). Recombinant msFBAld was largely induced under the control of T7 promoter in *E. coli* BL21 (DE3) after the addition of 1 mM IPTG

as shown in figure 9. Non-denaturing SDS-PAGE analysis showed that the recombinant msFBAld (lanes 3 and 4) derived from pET41a(+)/msFBAld, was 39 kDa which was in good agreement with the predicted molecular weight. On the other hand, the control sample which is the pET41a(+) without msFBAld, when induced with IPTG showed no sharp band at ~39 kDa (lanes 1 and 2). The result demonstrated that msFBAld protein was actively expressed in *E. coli* BL21 (DE3). However, analysis of His-tag purified sample showed the presence of two protein bands (lanes 5 and 6), which had molecular weights of 39 and 48 kDa. The 48 kDa sized protein might be eluted from non-specific binding of His-tag column to other endogenous protein of *E. coli* BL21 (DE3). Subsequent MALDI-TOF analysis of the 39 kDa band identified the peptides that had significant sequence similarities with Fructose 1 6-bisphosphate aldolases of other plants such *Vitis vinifera*, *Musa acuminata* subsp. *malaccensis*, *Glycine max*, *Lotus japonicus* etc. This result suggested that the lower band of 39 kDa on SDS-PAGE (lane 5 and 6) was determined to be the recombinant msFBAld.
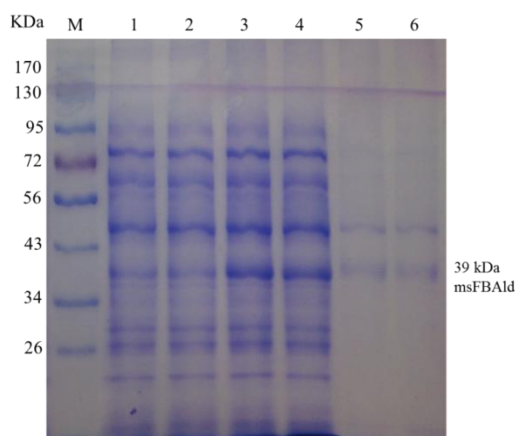

Figure 9. SDS-PAGE (12%) analysis of expressed msFBAld. M, marker proteins, lanes 1 and 2, pET41a(+) without msFBAld; lanes 3 and 4, 1 mM IPTG induced pET41a(+)/msFBAld; lanes 5 and 6, His-tag purified proteins which contained two bands at 48 and 39 kDa proteins. MALDI-TOF analysis showed the peptide sequences at the 39 kDa matched with other FBAld protein.

Salt-Tolerance Ability of the pET41a(+)/msFBAld *E. coli*

To analyze the potential functions of msFBAld in response to salt tolerance, the pET41a(+)/msFBAld *E. coli* (DE3) was cultured in LB media in various concentrations of NaCl at 37°C for 2 hr. The cell salt-tolerance was measured by OD_{600} value and showed that the medium with the pET41a(+)/msFBAld transformants were approximately 1.83-fold in OD600 value compared to the control (pET41a(+) minus msFBAld) under 1.0 M NaCl stress for 2 h; and 1.4-fold under 0.6 M NaCl stress (Fig. 10a). To analyse cell survival on salt stress, the medium with containing pET41a(+) and pET41a(+)/msFBAld transformants after addition of different concentrations of NaCl for 2 hrs were diluted and spread onto the LB-agar plates. Cell survival probability was calculated as percentage of colonies relative to the starting number of colonies. The results showed that the colony count of pET41a(+) were reduced to approximately 85% after the addition of 1.0 M NaCl whereas the pET41a(+)/msFBAld transformants were retained at 35% compared to 0.0 M concentration of NaCl (Fig. 10b). These results indicated that over-production of recombinant msFBAld increases the salt-tolerance and survival ability of *E. coli* cells at salt-stress thus suggesting that msFBAld could provide salt-tolerance ability in sago palm.
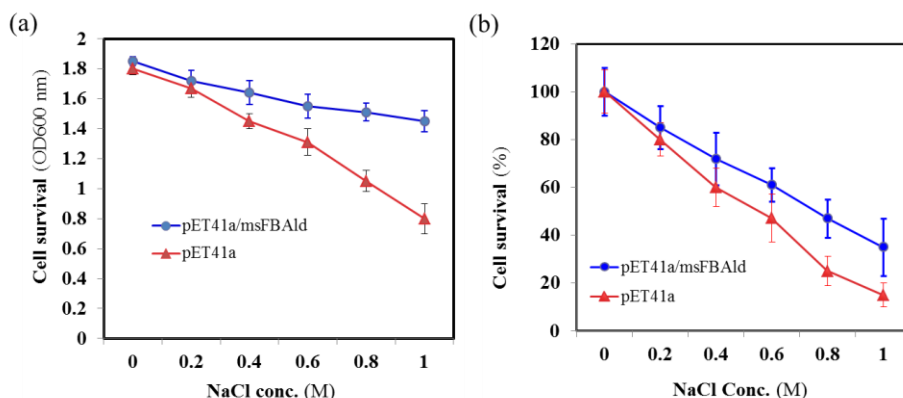
Figure 10. The salt-tolerance ability of *E. coli* BL21 (DE3) containing pET41a/msFBAld. (a) The pET41a(+) and pET41a(+)/msFBAld transformants were transferred to the fresh medium with a final concentration of 0, 0.2, 0.4, 0.6, 0.8, and 1.0 M NaCl for 2 h incubation at 37°C, and cell survival probability was evaluated by measuring the OD600 value. (b) The experiment was repeated and the cells were then plated (10^{-5} x dilution) on Luria–Bertani medium plates supplemented with kanamycin. Cell survival probability was plotted as the percentage of colonies relative to the starting number of colonies.

CONCLUSION

Fructose-1,6-bisphosphate aldolase is an enzyme that catalyzes the cleavage of fructose-1,6-bisphosphate to glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) via Schiff base reaction, which is crucial for both glycolysis/gluconeogenesis and pentose phosphate pathways. Researches on aldolase in other plants have shown great interest because of its involvements in the regulation of environment adaptation to abiotic stress, drought. In this study we have developed a modified RNA extraction method from sago leaves and isolated the msFBAld cDNA. Homology modeling studies generated 3D structure of msFBAld which provided important structural information. Structural superimposition revealed catalytically important residues including the critical lysine residues, Lys127 and Lys209, which can form Schiff's base with substrates and catalyze aldol condensation reaction. Overproduction of recombinant msFBAld in *Escherichia coli* BL21 (DE3) showed in an increased salinity tolerance. The structure-functional characterization of msFBAld and analysis of salt-stress tolerance will be useful and formed a basis to explore into the physiological function of sago aldolase as well as towards the development of salt-tolerant crops.

REFERENCES

1.   Flores MD. Greening Asia: The sago invasion. Paper presented at 2nd ASEAN Sago Symposium, UNIMAS, Kota Samarahan, 29th-30th October 2012 (p6)

2.   Dos Mohammed AM. Importance of sago starch quality: Development of sago   starch standard and diversification of value added sago starch. Paper presented at 2nd ASEAN Sago Symposium, UNIMAS, Kota Samarahan, 29th-30th October 2012 (p9)

3.   Slattery CJ, Kavakli HI, Okita TW. Engineering starch for increased quantity and quality. *Trends Plant Sci*. 2000; 5(7): 291-298.

4.   Roslan HA, Anji SB. Characterization of inflorescence-predominant chitinase gene in Metroxylon sagu via differential display. *3Biotech*. 2011; 1(1): 27-33 DOI: 10.1007/s13205-011-0004-x

5.   Wee CC, Roslan HA. Expressed Sequence Tags (ESTs) from young leaves of Metroxylon sagu. *3Biotech*. 2012a; 2(3): 211-218. DOI: 10.1007/s13205-012-0048-6

6.   Wee CC, Roslan HA. Isolation of alcohol dehydrogenase cDNA and basal regulatory region from Metroxylon sagu. *ISRN Mol Biol*. 2012b. doi:10.5402/2012/839427

7.   Ibrahim ER, Hossain MA, Roslan HA. Genetic transformation of Metroxylon sagu (Rottb.) cultures via Agrobacterium-mediated and particle bombardment, *Biomed Res Intl*. 2014. doi:10.1155/2014/348140

8.   Flechner A, Gross W, Martin WF, Schnarrenberger WF. Chloroplast class I and class II aldolases are bifunctional for fructose-1,6-biphosphate and sedoheptulose-1,7-biphosphate cleavage in the Calvin cycle. *FEBS Lett*. 1999; 447(2-3): 200-202.

9.   Lal A, Plaxton WC, Kayastha AM. Purification and characterization of an allosteric fructose-1, 6-bisphosphate aldolase from germinating mung beans (Vigna radiata). *Phytochemistry* 2005; 66(9): 968-974.

10.   Wang W, Baker P, Seah SY. Comparison of two metal-dependent pyruvate aldolases related by convergent evolution: substrate specificity, kinetic mechanism, and substrate channeling. *Biochemistry*. 2010; 49(17):3774-3782.

11.   Lebherz HG, Leadbetter MM, Bradshaw RA. Isolation and characterization of cytosolic and chloroplastic forms of spinach leaf fructose diphosphate aldolase. *J Biol Chem*. 1984; 259: 1011-1017.

12.   Sonnewald U, Lerchl J, Zrenner R, Frommer W. Manipulation of source-sink relations in transgenic plants. *Plant Cell Environ*. 1994; 17(5): 649–658.

13.   Gross W, Lenze D, Nowitzki U, Weiske J, Schnarrenberger C. Characterization, cloning and evolutionary history of the chloroplast and cytosolic class I aldolases of the red alga, Galdieria sulphuraria. *Gene*. 1999; 230: 7–14.

14.   Emes MJ, Tobin AK. Control of metabolism and development in higher plant plastids. *Int Rev Cytol*. 1993; 145: 149-216.

15.   Ge P. Ma C, Wang S, Gao L, Li X, Guo G, et al. Comparative proteomic analysis of grain development in two spring wheat varieties under drought stress. *Anal Bioanal Chem*. 2012; 402(3): 1297-1313.

16.   Thapa G, Dey M, Sahoo L, Panda SK. An insight into the drought stress induced alterations in plants. *Biol Plantarum*. 2011; 55(4): 603-613.

17.   Konishi H, Maeshima M, Komatsu S. Characterization of vacuolar membrane proteins changed in rice root treated with gibberellin. *J Proteome Res*. 2005; 4(5): 1775-1780.

18.   Schwab W, Aharoni A, Raab T, Pérez AG, Sanz C. Cytosolic aldolase is a ripening related enzyme in strawberry fruits (Fragaria×ananassa). *Phytochemistry*. 2001; 56(5): 407-415.

19.   Mano JI, Miyatake F, Hiraoka E, Tamoi M. Evaluation of the toxicity of stress-related aldehydes to photosynthesis in chloroplasts. *Planta*. 2009; 230(4): 639-648.

20.   Yamada S, Komori T, Hashimoto A, Kuwata S, Imaseki H, Kubo T. Differential expression of plastidic aldolase genes in Nicotiana plants under salt stress. *Plant Sci*. 2000; 154(1): 61-69.

21.   Fan W, Zhang Z, Zhang Y. Cloning and molecular characterization of fructose-1,6-bisphosphate aldolase gene regulated by high-salinity and drought in Sesuvium portulacastrum. *Plant Cell Rep*. 2009; 28(6): 975-984.

22.   Zeng Y, Yang T. RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. *Plant Mol Biol Rep*. 2002; 20: 417a-417e.

23.   Azevado H, Lino-Neto T, Tavares RM. An improved method for high quality RNA isolation from needles of adult maritime pine trees. *Plant Mol Biol Rep*. 2003; 21(4): 333-338.

24.   Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad USA*. 1998; 85(23), 8998-9002.

25.   Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principles of protein-dye binding. *Anal Biochem*. 1976; 72(1): 248-254.

26.   Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970; 227: 680-685.

27.   Altschul S, Madden TL, Schaffer AA, Zheng-Shang J, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: generation of protein database search programs. *Nucleic Acids Res*. 1997; 25: 3389-3402.

28.   Hossain MA, Rana MM, Kimura Y, Roslan HA. Changes in biochemical characteristics and activities of ripening associated enzymes in mango fruits during the storage at different temperatures. *BioMed Res Intl (Biochemistry)*. 2014; Article ID 232969, 11 pages. DOI: 10.1155/2014/232969

29.   Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. 2001; 5; 294(5540): 93-6.

30.   Cymerman I, Feder M, Pawlowski M, Kurowski MA, Bujnicki JM. Computational Methods for Protein Structure Prediction and Fold Recognition. In Practical Bioinformatics. Janusz M. Bujnicki (Ed.). Springer-Verlag Berlin Heidelberg 2004.

31.   Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014; 42(W1): W252-258.

32.   Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011; 27: 343-350.

33.   Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006; 22: 195-201.

34.   Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993; 26: 283-291.

35.   Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature*. 1996; 381-272.

36.   Melo F, Sanchez R, Sali A. Statistical potentials for fold assessment. *Protein Sci*. 2002; 11: 430-448.

37.   Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res*. 2012; 40(1): W471–W477.

38.   Woodhead M, Taylor MA, Davies HV, Brennan RM, McNicol RJ. Isolation of RNA from blackcurrant (Ribes nigrum L.) fruit. *Mol Biotech*. 1997; 7: 1-4.

39.   Salzman RA, Fujita T, Hasegawa PM. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Mol Biol*. 1999; 17(765): 11-17.

40.   Sauve V, Sygusch J. Molecular cloning, expression, purification, and characterization of fructose-1,6-bisphosphate aldolase from Thermus aquaticus. *Protein Expr Purif*. 2001; 21: 293–302.

41.   Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res*. 2013; 41(D1): D348-352.

42.   Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993; 234: 779-815.

43.   Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci*. 2002; 11(11): 2714-2726.

44.   Choi KH, Shi J, Hopkins CE, Tolan DR, Allen KN. Snapshots of catalysis: the structure of fructose-1,6-(bis)phosphate aldolase covalently bound to the substrate dihydroxyacetone phosphate. *Biochemistry*. 2001; 40: 13868-13875.

45.   Dalby A, Dauter Z, Littlechild JA. Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: Mechanistic Implications. *Protein Sci*. 1999; 8: 291-297.

46.   Choi KH, Mazurkie AS, Morris AJ, Utheza D, Tolan DR, Allen KN. Structure of a fructose-1,6-bis(phosphate) aldolase liganded to its natural substrate in a cleavage defective mutant at 2.3 Å. *Biochemistry*. 1999; 38: 12655-12664.