Molecular Biology Reports
https://doi.org/10.1007/s11033-023-08417-1

ORIGINAL ARTICLE

Molecular cloning and in silico analysis of chalcone isomerase from *Polygonum minus*

Fatin Lyana Azman Shah[1,2] · Syarul Nataqain Baharum[3] · Hoe-Han Goh[3] · Thean Chor Leow[1,4] · Ahmad Bazli Ramzi[3] · Siti Nurbaya Oslan[1,5] · Suriana Sabri[1,2]

Received: 24 February 2023 / Accepted: 29 March 2023

Abstract

Background Chalcone isomerase (CHI; EC 5.5.1.6) is one of the key enzymes in the flavonoid biosynthetic pathway that is responsible for the intramolecular cyclization of chalcones into specific 2*S*-flavanones.

Methods and results In this study, the open reading frame (ORF) of CHI was successfully isolated from the cDNA of *Polygonum minus* at 711-bp long, encoding for 236 amino acid residues, with a predicted molecular weight of 25.4 kDa. Multiple sequence alignment and phylogenetic analysis revealed that the conserved residues (Thr50, Tyr108, Asn115, and Ser192) in the cleft of CHI enzyme group active site are present in *Pm*CHI protein sequence and classified as type I. *Pm*CHI comprises more hydrophobic residues without a signal peptide and transmembrane helices. The three-dimensional (3D) structure of *Pm*CHI predicted through homology modeling was validated by Ramachandran plot and Verify3D, with values within the acceptable range of a good model. *PmCHI* was cloned into pET-28b(+) plasmid, expressed in *Escherichia coli* BL21(DE3) at 16 °C and partially purified.

Conclusion These findings contribute to a deeper understanding of the *Pm*CHI protein and its potential for further characterization of its functional properties in the flavonoid biosynthetic pathway.

Keywords Flavonoids · *Polygonum minus* · Chalcone isomerase · Bioinformatics

Introduction

Flavonoids are natural pigments mainly found in plants and these secondary metabolites are mostly derived via the phenylpropanoid pathway [1]. Flavonoids have various roles in plants, such as defense against microbial infection, ultraviolet radiation, and as activators of nodulation in plants [1–3]. Besides, flavonoids have the potentials as antiviral, anti-obesity, antioxidants, and anti-diabetics to be capitalized for the health and pharmaceutical industries [4, 5].

The biosynthetic pathway of flavonoids whereby flavanones such as naringenin (a universal precursor) is produced in plants involving several enzymes. Phenylalanine-ammonia lyase (PAL; EC 4.3.1.5), converts the starting precursor (phenylalanine) into cinnamic acid, which is oxidized by cinnamate-4-hydroxylase (C4H; EC 1.14.14.91) into 4-coumaric acid. This coumaric acid precursor is then converted into 4-coumaroyl-CoA by the action of 4-coumaroyl-CoA ligase (4CL; EC 6.2.1.12). Naringenin chalcone was then synthesized through the condensation reaction of malonyl-CoA and 4-coumaroyl-CoA, catalyzed by chalcone

✉ Suriana Sabri
suriana@upm.edu.my

1 Enzyme and Microbial Technology Research Centre, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Malaysia

2 Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Malaysia

3 Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

4 Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Malaysia

5 Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Malaysia

Published online: 06 May 2023

synthase (CHS; EC 2.3.1.74). The chalcone is then isomerized into 2*S*-naringenin via the enzymatic reaction of chalcone isomerase (CHI; EC 5.5.1.6) or nonenzymatically in an alkaline condition [6].

Polygonum minus (syn. *Persicaria minor*) is a member of the *Polygonaceae* family, commonly used in Southeast-Asian countries as one of the major ingredients in various delicacies. This plant has also been used in traditional remedies against digestive and dandruff problems [7]. The antioxidant properties of this plant are contributed from its high polyphenol, vitamin C, and β-carotene contents [7]. Its pungent aroma is mainly attributed to its unique blend of terpenes [8], which are induced by stress elicitors such as methyl jasmonate [9]. Flavonoids are another major compounds present in *P. minus*, in which the total flavonoid content was affected by different temperatures and populations [10, 11, 12]. The recent transcriptome analysis of *P. minus* reveals the genes involved in the phenylpropanoid and flavonoid pathways [13, 14]. Following the success of identified chalcone synthase (CHS) [14], a continuing study on the isolation of chalcone isomerase (CHI) is necessary to engineer microbial hosts for the biotechnological production of flavonoids [6, 15].

Chalcone isomerase (CHI; EC 5.5.1.6) is one of the actively studied enzymes crucial for the formation of flavonoids. To date, many researchers have successfully isolated and analyzed the chalcone isomerase genes from various plant species, such as *Ipomoea batatas*, *Carthamus tinctorius* L., and *Medicago sativa* [15–18]. There is no study on the chalcone isomerase enzyme from *P. minus*, hence this study was conducted for its identification and isolation. Protein structure prediction through homology modeling and heterologous expression of this enzyme in a prokaryotic host *Escherichia coli* (*E. coli*) were also performed. Physicochemical properties prediction on this enzyme and the generation of a three-dimensional (3D) model provide functional insights on this enzyme. Additionally, the recombinant expression of *CHI* demonstrated its potential to be incorporated into an engineered phenylpropanoid pathway.

Materials and methods

Bacterial strains and construction of recombinant plasmids

Escherichia coli TOP10 (Thermo Fisher Scientific, USA), competent *E. coli* BL21(DE3) (Transgen Biotech, China), and *E. coli Trans*1-T1 (Transgen Biotech, China) cells were purchased for use in gene cloning and protein expression. *E. coli* TOP10 competent cells were prepared according to [18]. Codon-optimization, gene synthesis, and cloning of *CHI* gene into pET-28b(+) plasmid were done by GenScript (China). The transformation of recombinant plasmids into *E. coli* was conducted via heat-shock method.

PCR amplification of *CHI* from *Polygonum minus*

The amplification of putative open reading frame (ORF) of *CHI* gene from *P. minus* was performed through the polymerase chain reaction (PCR) in a 50 µL reaction comprising EconoTaq® PLUS GREEN 2X Master Mix (Lucigen, USA), sterile distilled water, 0.5 µg cDNA from *P. minus*, and 0.5 µM of each primer (*CHI*-ORF Forward: 5′-ATGGCA TCACCGCATGCCGT-3′ and *CHI*-ORF Reverse: 5′-TCA TGCCTTGATCTCCACAC-3′). Thermocycler conditions were set accordingly: one cycle of initial denaturation at 94 °C for 2 min, followed by denaturation at 94 °C for 30 s, annealing at 53 °C for 30 s, and extension at 72 °C for 1 min. The process was repeated for 30 cycles and ended with one cycle of final extension at 72 °C for 7 min.

Confirmation of amplified full *CHI* ORF from *P. minus*

The amplified *CHI* fragment from *P. minus* was cloned into p*EASY*-T5-Zero plasmid (Transgen Biotech, China) and subsequently transformed into *E. coli Trans*1-T1 (Transgen Biotech, China). Bacterial clones harboring recombinant plasmid were identified through colony PCR. Then, extraction of plasmids was carried out from these colonies, and nucleotide sequencing was done to analyze the amplified *CHI* sequence.

Bioinformatics analysis of amplified *Pm*CHI

The CHI protein sequence was subjected for analysis through protein BLAST (BLASTP) using the NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and by multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) against CHI sequences from various plants. The phylogenetic relationship of *Pm*CHI with other identified CHIs was also investigated by using the Neighbor-Joining method available in the MEGA7 software [19]. The predicted protein molecular weight (MW), theoretical isoelectric point (pI), amino acid, and atomic composition, instability index, and its Grand Average Hydropathicity (GRAVY) value were analyzed using ProtParam tool on the Expasy website (https://web.expasy.org/protparam/) [17].

The presence of signal peptide and transmembrane helices were predicted by SignalP V4.1 (http://www.cbs.dtu.dk/services/SignalP-4.1/) [20], and TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [21], respectively. Prediction of subcellular localization of the protein sequence was done using TargetP-2.0 (http://www.cbs.dtu.dk/services/TargetP/) [22], CELLO2GO (http://cello.life.nctu.edu.

tw/cello2go/) [23], and WoLFPSORT (https://wolfpsort.hgc.jp/) [24].

Three-dimensional (3D) structure prediction through homology modeling

YASARA program [25] was used to generate a three-dimensional (3D) structure of *Pm*CHI by using homology modeling, where a model with the best scoring (based on this program) was later selected. Validation of the generated 3D protein model was carried out using ERRAT [26], Verify3D, and PROCHECK [27] from PDBSum for the generation of a Ramachandran plot [28].

Expression of recombinant *CHI* from *P. minus* in *E. coli*

A single colony of recombinant *E. coli* BL21(DE3) harboring pET-28b(+)/co*CHI* was grown in 5 mL of Luria–Bertani (LB) medium supplemented with kanamycin (50 µg/mL) at 37 °C and 200 rpm. The overnight culture was used to inoculate 50 mL of LB medium supplemented with kanamycin (50 µg/mL) at 37 °C, with shaking until OD_{600} reaches 0.6. Then, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.55 mM for induction [29]. The cells were later incubated separately at three different temperatures (16 °C, 20 °C, and 25 °C) for 16 h, and subsequently harvested by centrifugation at 10,000×*g* for 10 min at 4 °C. The pellets were dissolved in 100 mM phosphate buffer (pH 7.0) and the suspended cells were lysed by sonication with an amplitude of 45% and a pulse time of 30 s for 5 min [30].

The disrupted cells were then centrifuged at 10,000×*g* for 10 min at 4 °C. The supernatants and pellets collected were analyzed using SDS–PAGE [31]. Western blot analysis for the confirmation of recombinant *Pm*CHI protein expression was performed according to the protocol by [32], with several modifications based on the blotting kit (Novagen, Germany). The expressed His-tagged recombinant CHI protein was later partially purified by using Ni-Sepharose FF column (Sigma Aldrich, Germany) through gravity-flow purification and analyzed using SDS–PAGE.

Results

Protein sequence analysis of *CHI* amplified from *Polygonum minus*

The open reading frame (ORF) of *CHI* from *P. minus* consists of 711 bp nucleotides encoding for 236 amino acids, with the predicted molecular weight (MW) of 25.4 kDa, and a theoretical isoelectric point (pI) of 5.56. From the BLASTP analysis performed on the amino acid sequence of *Pm*CHI, this sequence showed identity similarities to a few other *CHI* sequences from different species, such as *Fagopyrum esculentum* [ADT63063], *Fagopyrum dibotrys* [AHH84790], *Morus notabilis* [AHY35310], *Gossypium hirsutum* [NP_001314370], *Prunus avium* [AJO67964], and *Pyrus pyrifolia* [ADP09377] at 80.93%, 79.03%, 64.95%, 62.16%, 63.45%, and 63.68%, respectively. There are four conserved amino acid residues (Thr50, Tyr108, Asn 115, and Ser192) crucial for the catalytic function of CHI enzymes [16, 32], which are found in *Pm*CHI and other plants (Fig. 1). In addition, several other amino acid residues found in the active site cleft of this enzyme isolated from *Medicago sativa* [16] and CHI proteins from liverwort and *Selaginella moellendorffi* are also present in *Pm*CHI (Arg38, Gly39, Leu40, Phe49, Ile52, Lys111, Val112, Asn115, and Ile193). The identified ORF of *Pm*CHI was submitted to GenBank with the accession number MN193938.

Phylogenetic analysis of *Pm*CHI

*Pm*CHI was observed to be clustered in a larger group consisting of CHI sequences from other non-legumes (Type I CHI), such as *Arabidopsis thaliana*, *Hordeum vulgare*, *Ipomoea batatas*, and *Zea mays*, whilst leguminous (Type II) CHI sequences [33, 34] from *Medicago sativa*, *Glycine max*, *Pisum sativum*, and *Glycyrrhiza uralensis* were clustered together into a distinct group in the phylogenetic tree (Fig. 2). Hence, *Pm*CHI that clustered with non-legumes was identified to be a Type I CHI.

Amino acid, atomic composition, hydropathicity, and instability index of *Pm*CHI

Based on the analysis from ProtParam, the 236 residues constructing CHI are from 20 essential amino acids, from which 29 residues are negatively charged (Asp + Glu), and 25 of them are positively charged (Arg + Lys) at a neutral pH. The amino acid composition analysis of *Pm*CHI further revealed that there are 57 charged residues (24.15% of the total amino acid residues), 137 hydrophobic residues (58.05%), and 42 uncharged polar residues (17.80%). From this analysis, it was shown that *Pm*CHI is identified to have more hydrophobic than hydrophilic residues.

Besides, *Pm*CHI comprised of 3604 atoms, made up of carbon (1155), hydrogen (1810), nitrogen (294), oxygen (338), and sulfur (7), and this protein is represented in the chemical formula of $C_{1155}H_{1810}N_{294}O_{338}S_7$. The GRAVY value of CHI protein from *P. minus* was calculated at − 0.030, indicating that *Pm*CHI is a hydrophilic protein. Furthermore, *Pm*CHI was identified as an unstable protein as its instability index was observed to be at 40.61.

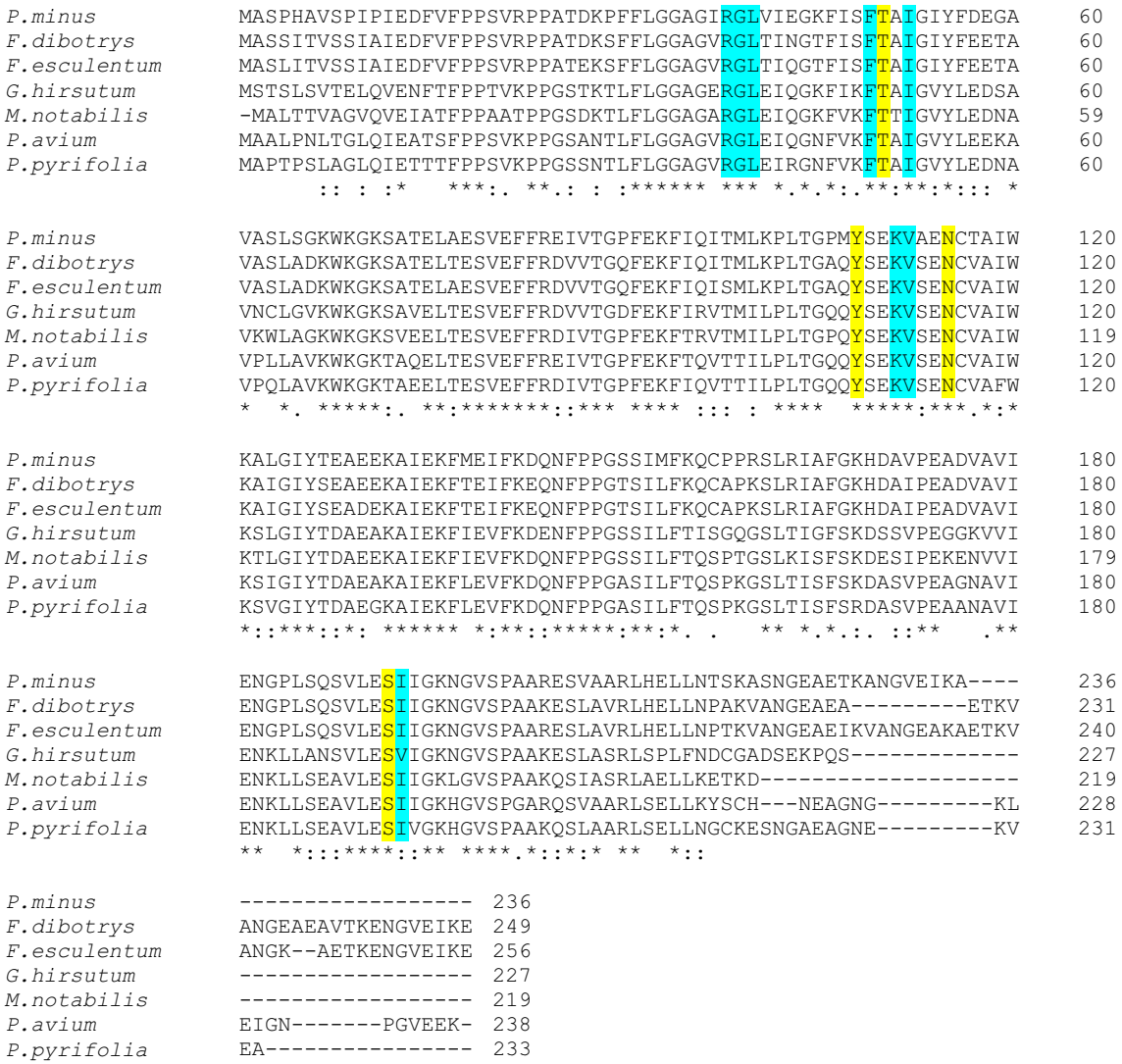

Fig. 1 Multiple sequence alignment of *Pm*CHI with CHI sequences from other plants. Cyan and yellow highlighted residues are present in the active site cleft of functional CHI, with the yellow residues being the active site residues conserved in most of the plant CHI sequences. Symbols (*) denote the highly conserved residues, (:) the conserved residues, and (.) the less conserved residues. (Color figure online)

Prediction of signal peptide, subcellular localization, and transmembrane helices

There is no signal peptide cleavage site predicted in *Pm*CHI as the *C*-, *S*- and *Y*-scores were calculated at 0.109, 0.155, and 0.120, respectively. Meanwhile, TargetP-2.0 server suggested that *Pm*CHI is not located in the chloroplast (or other plastids), mitochondria, or the secretory pathway. The highest score obtained at 0.992 for other locations when compared to that of chloroplast transfer peptide, mitochondrial transfer peptide, signal peptide, and thylakoid luminal transfer peptide (0.0072, 0.0001, 0.0005, and 0.0002, respectively). On the other hand, CELLO2GO results calculated scores in which *Pm*CHI was predicted to be in the cytoplasm or chloroplast (with scores of 1.944 and 1.418, respectively), with expected isomerase activity. Lastly, WolfPSORT analysis showed the identity percentages of 14 nearest neighbors of *Pm*CHI, where the highest percentages were observed at 15.1% [cytoplasmic protein cysteine synthase; UniProtKB ID: Q00834] and 16.1% [thioredoxin protein from *A. thaliana*; At5g39950.1]. In addition, *Pm*CHI is observed to lack any transmembrane helices in its structure based on the output information from TMHMM server [21].

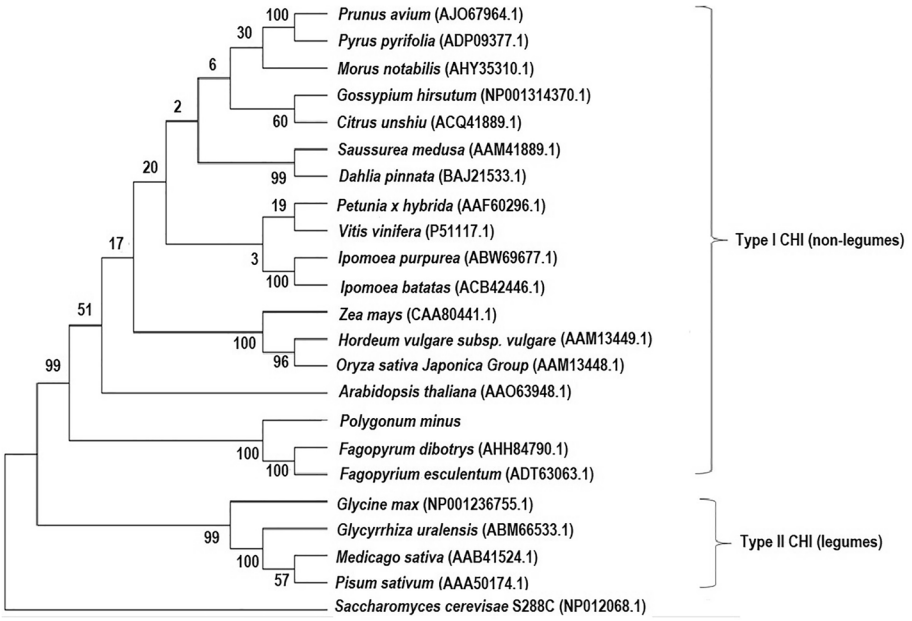

Fig. 2 Evolutionary relationships of various CHI genes based on deduced amino acid sequences. The evolutionary history was inferred using the Neighbor-Joining method. The analysis involved 23 sequences and the phylogenetic analysis was conducted in MEGA7 [19]. *Pm*CHI is identified to be a Type I CHI

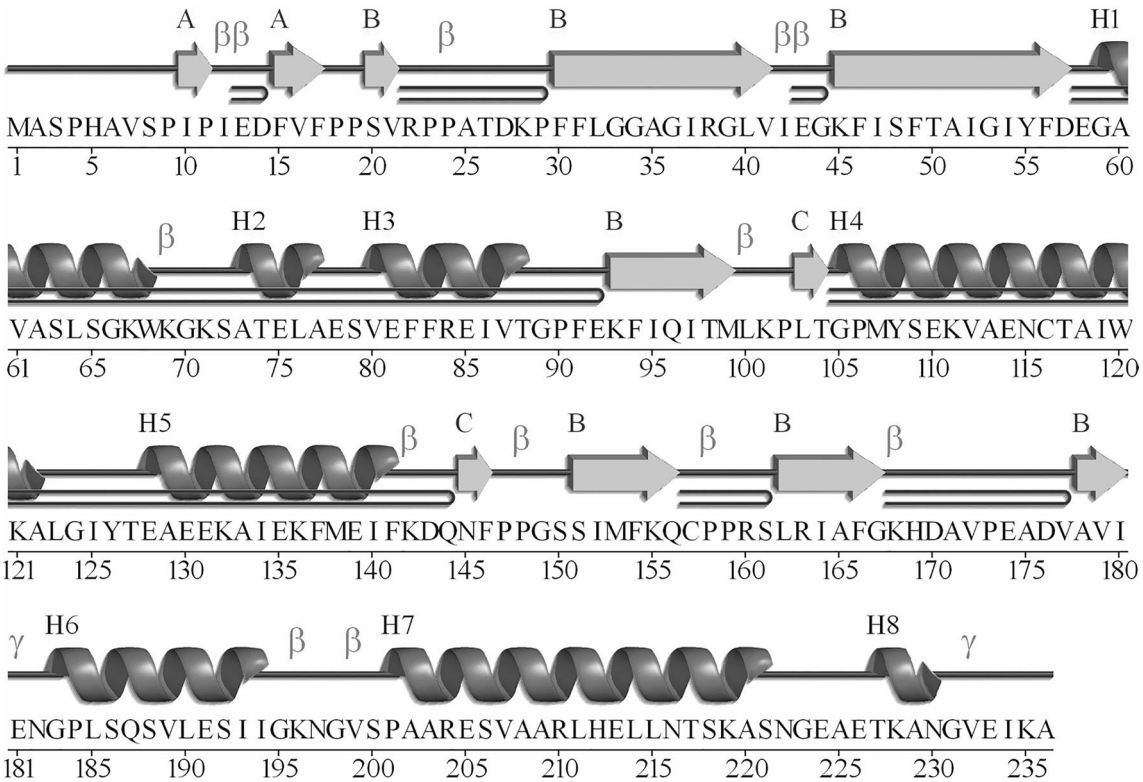

Fig. 3 Secondary structure of *Pm*CHI. This protein consists of eight α-helices, three β-sheets, seven β-hairpins, five β-bulges, 11 strands, 11 helix–helix interactions, 13-β turns, and 2-γ turns

Secondary structure analysis of *Pm*CHI

From the analysis output, it was observed that the composition of the secondary structure without ligand of *Pm*CHI protein comprises 236 amino acid residues with 8 helices, 3 sheets, 7 β-hairpins, 5 β-bulges, 11 strands, 11 helix–helix interactions, 13-β turns, and 2-γ turns as seen in Fig. 3.

The three beta-sheets were labeled as β-A, β-B, and β-C, with all the sheets stabilized by hydrogen bonds, and are shown to be in the antiparallel direction. The beta-sheets are made up of 11-β strands in this predicted protein model. β-A sheet consisted of strands with residues starting from Ile10–Ile12 and Phe15–Phe17. Majority of the strands are located in the β-B sheet, with seven strands in total including residues Ser20-Val21, Phe30-Ile42, Lys45-Asp57, Lys93-Met99, Ser151-Gln156, Leu162-Gly167 and Ala178-Ile180, and the final two β-strands (starting from residues Leu103-Thr104 and Asn145-Phe146) were responsible for the formation of the final beta-sheet (β-C) in the generated secondary structure of CHI protein. On the contrary, in the *Pm*CHI secondary structure, there are eight helices observed, all belonging to the α-helix type. The longest helix comprises 21 residues starting from residues Pro201-Ser221, and the shortest helix was made up of 4 residues (Thr227-Asn230).

Homology modeling of *Pm*CHI

Homology modeling was used in predicting the three-dimensional (3D) structure of chalcone isomerase protein from *P. minus*. Sequence identity of the query protein (*Pm*CHI) and the template protein sequence (PDB ID: 5YX4) was calculated to be at 52.4%, and the protein template used was a successfully crystallized CHI protein from *Deschampsia antartica* Desv. (mutated at Ser189) complexed with isoliquiritigenin substrate [34]. The reasoning behind in using isoliquiritigenin rather than naringenin chalcone is that based on the study conducted by [34], which shows the DaCHI protein can utilize both naringenin chalcone and isoliquiritigenin, albeit at a low catalytic activity. The predicted *Pm*CHI protein model based on this template (Fig. 4) had the best *Z*-score of 0.072 compared to other templates.

Active site residues including Arg38 (Arg36), Thr50 (Thr48), Tyr108 (Tyr106), Asn115 (Asn113), and Ser192 (Thr190) were found to be conserved in this *Pm*CHI protein structure (residues in parentheses referring to residues present in *Ms*CHI) [16]. These conserved residues are located between strands β4 to β5 and helices α4 and α6 (Fig. 4).

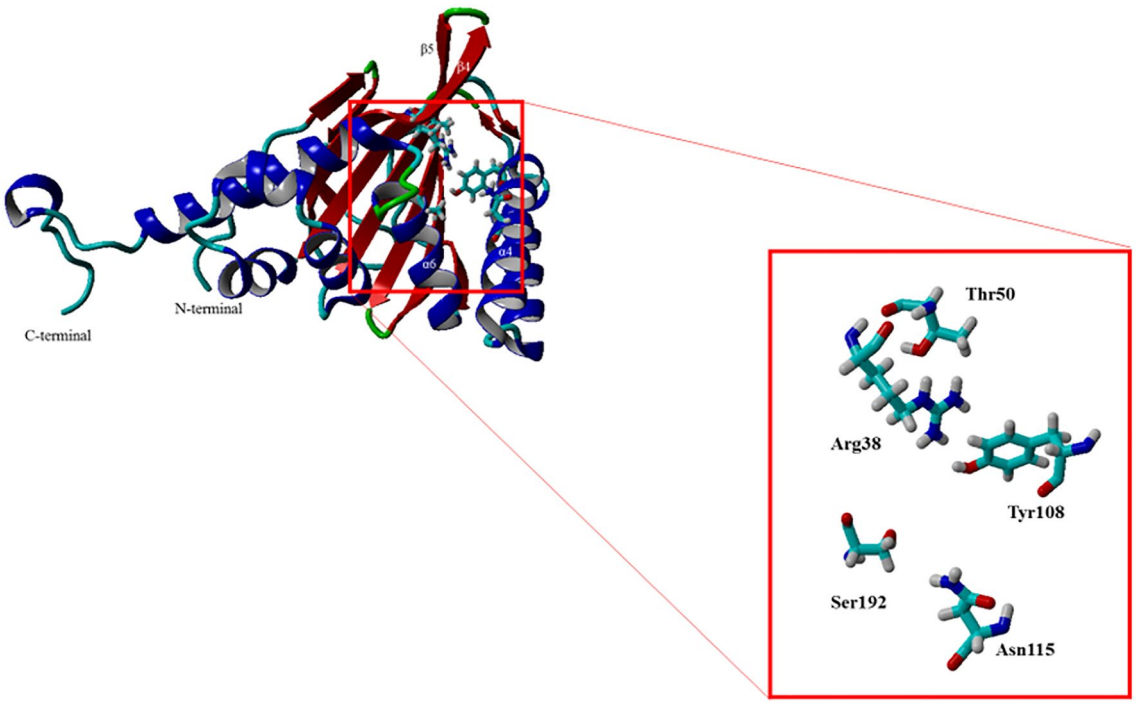

Fig. 4 The 3D structure of *Pm*CHI predicted through YASARA. There are 11 β-strands and eight α-helices in the structure. N- and C-terminals of this protein are in proximity in this model. The larger box indicates the suggested active site residues in the catalytic mechanism of *Pm*CHI, which are Arg38, Thr50, Tyr108, Asn115, and Ser192

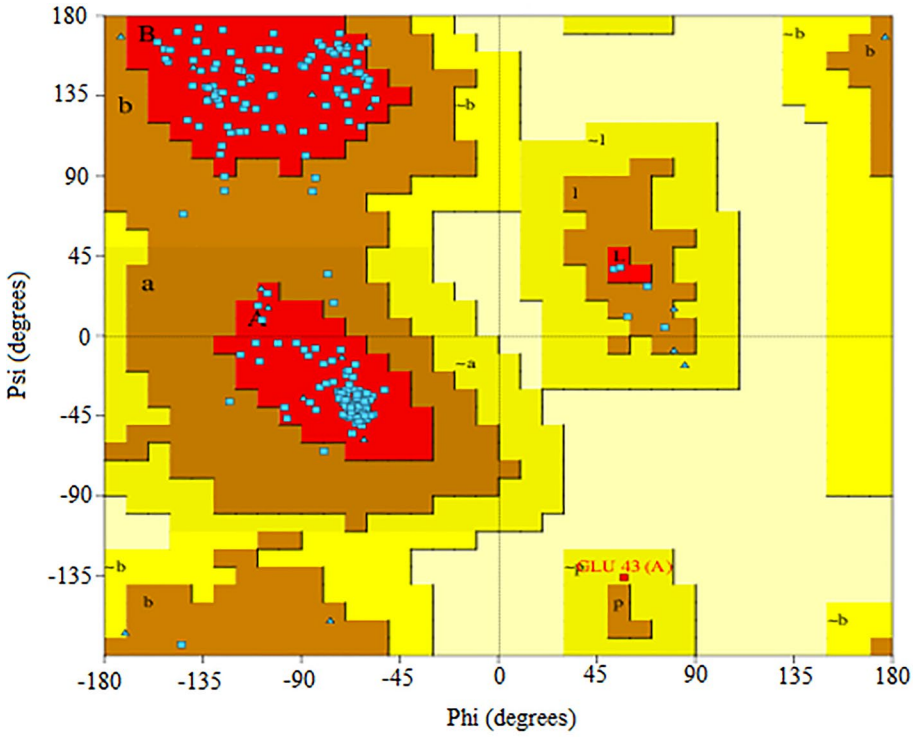

Fig. 5 Ramachandran plot of predicted *Pm*CHI protein model. The quality of this model is based on the most favored region represented in the red-colored area, additional allowed region in brown-colored, generously allowed region in yellow-colored, and disallowed region in the light, yellow-colored area. There are more than 90% of residues reside in the most favored region based on this model. (Color figure online)

Verification of predicted *Pm*CHI protein structure

From the Ramachandran plot (Fig. 5), it was observed that 93.9% of the residues resided were in the most favored regions, followed by 5.6% and 0.5% (a single Glu43 residue) in the additional and generously allowed regions, respectively (excluding the glycine and proline residues).

From Verify3D analysis, 93.22% of the residues have an averaged 3D–1D score ≥ 0.2 (Supplementary File: Fig. S1). There are 16 amino acid residues observed from the plot that scored below 0.2 (of the averaged 3D–1D score), and the lowest and highest values recorded at 0.10 and 0.69, respectively. Meanwhile, the ERRAT analysis showed that the predicted 3D *Pm*CHI protein model comprised of several residues (residues 6, 7, 8, 13, 15, 20, 22, 23, 24, 25, 26, 27, and 149) with error confidence percentages of above 95% and the overall quality factor of this model was calculated at 94.12% (Supplementary File: Fig. S2).

Expression of chalcone isomerase in *E. coli* BL21(DE3), Western blot analysis, and partial purification of *Pm*CHI protein

The expression of recombinant chalcone isomerase from *P. minus* in *E. coli* after IPTG induction at 0.55 mM for 16 h at three different temperatures setting (16 °C, 20 °C, and 25 °C) were subsequently analyzed by using SDS–PAGE. The enzyme was expressed at the expected size of ~ 27 kDa with the inclusion of His-tag and thrombin cleavage sequence (Fig. 6). Western blot analysis confirmed the expression of *Pm*CHI proteins with the protein band at size ~ 27 kDa observed on the nitrocellulose membrane with no band in the negative control. These results corroborated SDS–PAGE

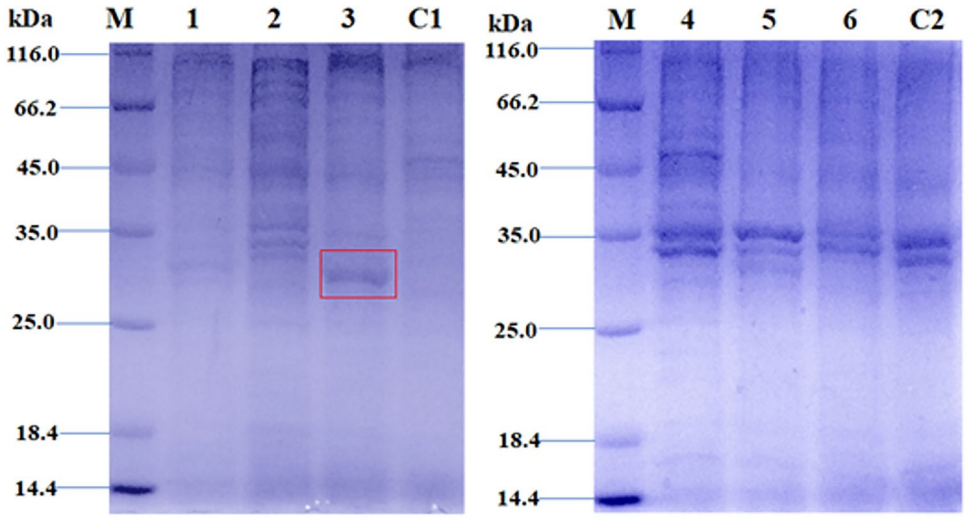

Fig. 6 SDS–PAGE analysis of strain *E. coli* BL21(DE3) harboring pET-28b(+)/co*CHI* at 25 °C, 20 °C and 16 °C and 0.55 mM IPTG. Protein expression was analyzed on 12% SDS–PAGE gel. Lane M: Unstained Protein Molecular Weight Marker (Thermo Scientific, USA), Lane 1, 2, and 3: supernatants from strain *E. coli* BL21(DE3) harboring pET-28b(+)/co*CHI* from expression at 25 °C, 20 °C and 16 °C respectively, Lane C1: supernatant from strain *E. coli* BL21(DE3) harboring pET-28b(+), Lane 4, 5, and 6: pellets from strain *E. coli* BL21(DE3) harboring pET-28b(+)/co*CHI* from expression at 25 °C, 20 °C and 16 °C, Lane C2: pellet from strain *E. coli* BL21(DE3) harboring pET-28b(+). Label with box is the expected band of CHI protein

analysis (Fig. 7) supporting that the CHI protein was successfully expressed in the *E. coli* system. Meanwhile, the analysis of the fractions (from the partial purification) using SDS–PAGE, showed distinct bands for chalcone isomerase at ~ 27 kDa, with the presence of other faint protein bands (Fig. 8) and further optimizations on the purification protocol will be needed to obtain a fully purified protein in future studies.

Discussion

Flavonoids, comprising a large family of plant secondary metabolites, are well-known for their antioxidant capacity and various health benefits [35]. Chalcone isomerase (CHI; EC 5.5.1.6) is an enzyme found in the flavonoid biosynthesis pathway, which is responsible for catalyzing the intramolecular cyclization of chalcones into functional 2*S*-flavanones [17, 36]. The ORF of CHI was successfully amplified from the cDNA of *P. minus* and the identified amino acid sequence is highly homologous with other CHI genes from various plant species.

The structural analysis of CHI from *M. sativa* has revealed several residues (Thr48, Tyr106, Asn113, and Thr190) that are important for the catalytic mechanism of this enzyme [17, 37] and based on the multiple sequence alignment of *Pm*CHI with various CHI sequences, these residues (Thr50, Tyr106, Asn113, and Ser192 in *Pm*CHI) appeared to be conserved among different plant species (Fig. 1). Residues Thr48 and Tyr106 are conserved in all

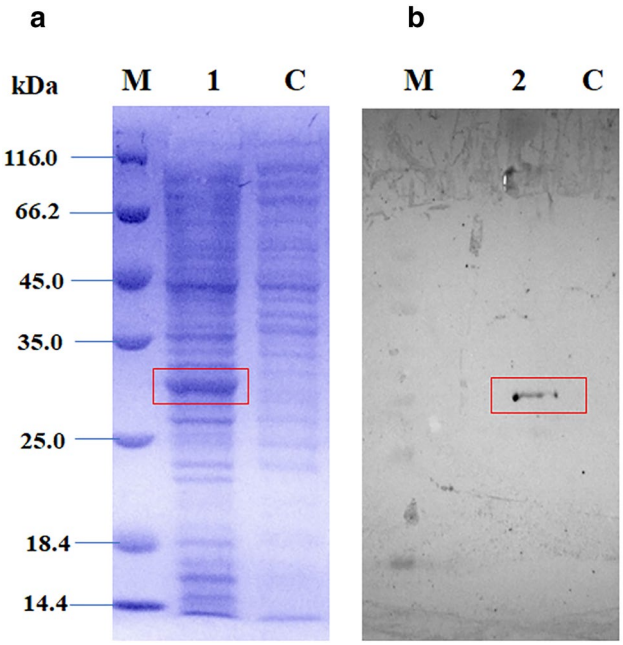

Fig. 7 SDS–PAGE and Western blot analyses of recombinant *E. coli* BL21(DE3) harboring pET-28b(+)/co*CHI*. Protein expression was analyzed on 12% SDS–PAGE gel. The *E. coli* cells harboring recombinant and empty plasmids were induced at 0.55 mM IPTG for 16 h at 16 °C for protein expression. **a** SDS–PAGE analysis: Lane M: Unstained Protein Molecular Weight Marker (Thermo Scientific, USA), Lane 1: supernatant of *E. coli* BL21(DE3) expressing CHI protein. **b** Western blot analysis: Lane 2: CHI protein, Lane C: supernatant of *E. coli* BL21(DE3) harboring empty pET-28b(+) plasmid (negative control). Label with box is the expected band of CHI protein

CHI sequences from different species, but in non-legumes, Thr190 and Met191 residues are mostly replaced by serine and isoleucine, respectively [17, 33]. The catalytic activities of CHIs differ in leguminous and non-leguminous plants. Type I CHI in non-legumes converts only naringenin chalcone to naringenin [34], whereas type II CHI involves an additional isomerizing reaction of isoliquiritigenin into liquiritigenin [16, 38, 39]. The phylogenetic analysis with functional CHIs from different plant species showed that *Pm*CHI is a Type I CHI (Fig. 2).

The predicted molecular weight of *Pm*CHI was similar to CHI proteins from other plants such as *Carthamus tinctorius* L., *Ophiorrhiza japonica*, and *Chamaemelum nobile*, which are in the range of 25 kDa [36, 40, 41]. This enzyme consisted of more hydrophobic residues, similarly to chalcone isomerase isolated from *G. max* [42, 43]. The hydrophilicity of *Pm*CHI was observed as the GRAVY value calculated resulted in a negative value and it was identified previously that CHI from *C. tinctorius* L. is a hydrophilic protein with its GRAVY value at − 0.203 [17].

On the other hand, *Pm*CHI is categorized as an unstable protein based on the instability index [44] similar to several other CHI proteins identified from *Ananas comosus*, *Paeonia suffruticosa*, *Lonicera japonica*, *Morus alba*, *Lilium speciosum*, and *Scutellaria baicalensis* [37]. *Pm*CHI does not have any transmembrane helical structure. There is also no signal peptide sequence detected, hence, *Pm*CHI was deduced to be a non-secretory protein as observed with identified CHI protein from other organisms [18, 36]. Subcellular localization

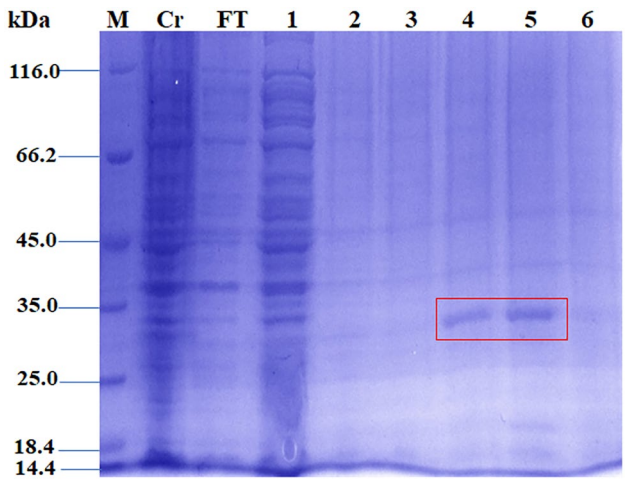

Fig. 8 SDS–PAGE analysis of the partial protein purification of chalcone isomerase. Protein expression was analyzed on 12% SDS–PAGE gel. Lane M: Unstained Protein Molecular Weight Marker (Thermo Scientific, USA), Lane Cr: Crude CHI, Lane FT: Flow-through, Lane 1: Fraction 2, Lane 2: Fraction 5, Lane 3: Fraction 6, Lane 4: Fraction 8, Lane 5: Fraction 9, and Lane 6: Fraction 10. Lane 5 corresponded to Fraction 9 that showed the highest absorbance at 280 nm during the elution phase

of *Pm*CHI showed the possibility of this protein located in the cytoplasm as were highlighted from identified CHI protein from previous studies [18, 36]. Chalcone isomerase is the enzyme responsible for the cyclization of chalcones into various flavanones, and this process occurs in the cytoplasm of plant cells [45]. These properties predicted from CHI of *P. minus* through computational analyses provided new information on *Pm*CHI that has never been studied previously.

The predicted 3D model of *Pm*CHI selected was generated from YASARA software program [24] based on the template from crystallized CHI protein of *D. antartica* Desv. with the highest *Z*-score. A *Z*-score of > 0 suggested an optimal or a perfectly folded protein [45], and a higher *Z*-score indicates a better protein model. The *Pm*CHI model was seen to resemble an upside-down bouquet that adopted an open-faced β-sandwich fold (Fig. 4) and comprises 8 α-helices and 11 β-strands, as reported previously in [17, 34]. This 3D model showed similarities with other crystal structures of CHI proteins deposited in PDB [17, 34] indicating that it is representative of *Pm*CHI.

Based on previous studies, the catalytic reaction of naringenin chalcone occurs through a combination of electrostatic catalysis and water-mediated charge stabilization, during the stereospecific Michael-type addition reaction [17, 44–47]. Several residues (Arg38, Thr50, Tyr108, Asn115, and Ser192) involved in the catalytic activity were also identified to be conserved in *Pm*CHI (Fig. 4). The proximity of these active site residues is observed in this predicted model and further studies could be conducted to observe the catalysis reaction of this enzyme.

Subsequent validation analysis of the generated model was performed to verify its quality. Three different analyses were conducted, starting with the use of PROCHECK program [27] for generating a Ramachandran plot [28]. From the plot, the predicted *Pm*CHI model was deduced to be of good quality as 93.9% of its residues are in the most favored regions (red-colored area) (Fig. 5). The *Pm*CHI structure was also subjected to validation analysis using Verify3D and this model was calculated to have 93.2% of its residues to have an averaged 3D–1D score ≥ 0.2. A Verify3D score of more than 80% was expected from a satisfactory predicted model [48]. There were low scoring residues located at the end of the protein structure between residue numbered 225 until 236, as observed from the plot (Supplementary File: Fig. S1) and this may suggest that this specific area in the predicted model may be incorrectly folded. However, as mentioned above, most of the residues (93.22%) have scored above the 0.2 marks for their averaged 3D–1D score, indicating that this predicted model has passed this evaluation. Meanwhile, ERRAT analysis [26] conducted on the *Pm*CHI 3D model gave a slightly unfavorable result as the overall quality factor of this model was calculated at 94.1% (Supplementary File: Fig. S2), where the resolution of this protein

structure is slightly low based on this analysis, as opposed to the benchmark value of the quality factor at ~ 95% for high resolution [48, 49, 50]. However, the previous validation analyses performed showed that the generated *Pm*CHI model by YASARA was the best structure predicted.

This is the first study to investigate the heterologous expression of native chalcone isomerase protein identified from *P. minus* in a prokaryotic system (*E. coli*). Several temperatures were tested during the expression study, and it was concluded that the recombinant chalcone isomerase protein was expressed in the soluble fraction at 16 °C due to the band (~ 27 kDa) that was observed in the supernatant (Fig. 6) compared to the insoluble fractions for other tested temperatures. Furthermore, a lower temperature setting after induction will often lead to soluble heterologous protein production [51]. The finding in this study was also confirmed by [37] as IPTG induction at a low temperature led to the heterologous expression of CHI from *O. japonica* as a major soluble protein.

The faint protein band warranted a subsequent Western blot analysis to confirm the recombinant expression of *Pm*CHI by *E. coli*. The expected protein band at ~ 27 kDa was viewed on the nitrocellulose membrane (Fig. 7) and this confirmed that *Pm*CHI was successfully expressed. Partial purification of *Pm*CHI using gravity-flow was conducted to observe the ability of this protein to be purified. *Pm*CHI that was expressed with His-tag at its N-terminal was then partially purified with Ni-Sepharose and the presence of this tag facilitated in the purification procedure. The protein band at ~ 27 kDa was detected when the fraction (detected with the highest absorbance reading at 280 nm during the elution phase) was analyzed with SDS–PAGE (Fig. 8). Hence, this suggests the ability of *Pm*CHI to be fully purified for further biochemical characterization on the functionality and characteristics of this important enzyme.

Conclusion

In conclusion, the present study sheds light on the molecular characteristics of *Pm*CHI, a key enzyme involved in the flavonoid biosynthetic pathway. The successful isolation of its ORF and the prediction of its three-dimensional structure through homology modelling provides a solid foundation for further biochemical studies of this protein. The partial purification of *Pm*CHI expressed in *E. coli* highlights its potential for future applications in the production of flavonoids, which could potentially reduce the dependence on plant extractions or chemical synthesis. However, further optimizations are necessary to fully realize the full potential of *Pm*CHI in this regard.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-023-08417-1.

Acknowledgements Fatin Lyana Azman Shah was supported by a scholarship from Jabatan Perkhidmatan Awam (JPA) Malaysia.

Author contributions All authors listed have made a considerable, direct, and logical contribution to the work and approved it for publication. The experiments were designed by FLAS, SNB, HHG, TCL, ABR, SNO and SS. FLAS and SS drafted the manuscript. All authors discussed the results, edited the manuscript, and approved the final version.

Funding The work was supported through the Special Allocation for Agencies Under Academy of Science Malaysia-Ministry of Science, Technology and Innovation Grant (PKA0514F004, 6300824) and Universiti Putra Malaysia Putra Graduate Initiative (GP/IPS/2016/9482200).

Data availability Data will be made available upon reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest regarding the publication of this article.

Ethical approval This article does not contain any studies involving human participants or animals performed by any of the authors.

Informed consent Not applicable.

Consent for publication Not applicable.

References

1. Pandey RP, Parajuli P, Koffas MAG, Sohng JK (2016) Microbial production of natural and non-natural flavonoids: pathway engineering, directed evolution and systems/synthetic biology. Biotechnol Adv 34:634–662. https://doi.org/10.1016/j.biotechadv.2016.02.012
2. Kumar S, Pandey AK (2013) Chemistry and biological activities of flavonoids: an overview. Sci World J. https://doi.org/10.1155/2013/162750
3. Simkhada D, Kurumbang NP, Lee HC, Sohng JK (2010) Exploration of glycosylated flavonoids from metabolically engineered *E. coli*. Biotechnol Bioprocess Eng 15:754–760. https://doi.org/10.1007/s12257-010-0012-4
4. Hossain MK, Dayem AA, Han J et al (2016) Molecular mechanisms of the anti-obesity and anti-diabetic properties of flavonoids. Int J Mol Sci. https://doi.org/10.3390/ijms17040569
5. Shah FLA, Ramzi AB, Baharum SN et al (2019) Recent advancement of engineering microbial hosts for the biotechnological production of flavonoids. Mol Biol Rep 46:6647–6659. https://doi.org/10.1007/s11033-019-05066-1
6. Kaneko M, Hwang EI, Ohnishi Y, Horinouchi S (2003) Heterologous production of flavanones in *Escherichia coli*: potential for combinatorial biosynthesis of flavonoids in bacteria. J Ind Microbiol Biotechnol 30:456–461. https://doi.org/10.1007/s10295-003-0061-1
7. Christapher P, Parasuraman S, Christina JA et al (2015) Review on *Polygonum minus*. Huds, a commonly used food additive in Southeast Asia. Pharmacogn Res 7:1. https://doi.org/10.4103/0974-8490.147125

8. Rusdi NA, Goh HH, Baharum SN (2016) GC–MS/olfactometric characterisation and aroma extraction dilution analysis of aroma active compounds in *Polygonum minus* essential oil. Plant Omics 9:289–294. https://doi.org/10.21475/poj.16.09.04.p7901
9. Rahnamaie-Tajadod R, Goh H-H, Mohd Noor N (2019) Methyl jasmonate-induced compositional changes of volatile organic compounds in *Polygonum minus* leaves. J Plant Physiol 240:152994. https://doi.org/10.1016/j.jplph.2019.152994
10. Khairudin K, Sukiran NA, Goh H-H et al (2013) Direct discrimination of different plant populations and study on temperature effects by Fourier transform infrared spectroscopy. Metabolomics 10:11306. https://doi.org/10.1007/s11306-013-0570-5
11. Loke K-K, Rahnamaie-Tajadod R, Yeoh C-C et al (2017) Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. PeerJ 5:e2938. https://doi.org/10.7717/peerj.2938
12. Goh H-H, Khairudin K, Sukiran NA et al (2016) Metabolite profiling reveals temperature effects on the VOCs and flavonoids of different plant populations. Plant Biol 18:130–139. https://doi.org/10.1111/plb.12403
13. Loke KK, Rahnamaie-Tajadod R, Yeoh CC et al (2016) RNA-seq analysis for secondary metabolite pathway gene discovery in *Polygonum minus*. Genomics Data 7:12–13. https://doi.org/10.1016/j.gdata.2015.11.003
14. Roslan ND, Tan C, Ismail I, Zainal Z (2013) cDNA cloning and expression analysis of the chalcone synthase gene (CHS) from *Polygonum minus*. Aust J Crop Sci 7:777–783
15. Guo J, Zhou W, Lu Z et al (2015) Isolation and functional analysis of chalcone isomerase gene from purple-fleshed sweet potato. Plant Mol Biol Rep 33:1451–1463. https://doi.org/10.1007/s11105-014-0842-x
16. Jez JM, Bowman ME, Dixon RA, Noel JP (2000) Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. Nat Struct Biol 7:786–791. https://doi.org/10.1038/79025
17. Ren C, Tang X, Chen C et al (2019) Cloning and expression analysis of a new chalcone isomerase gene during flowering in safflower. Turk J Bot 43:143–150. https://doi.org/10.3906/bot-1809-25
18. Chang AY, Chau VWY, Landas JA, Pang Y (2017) Preparation of calcium competent *Escherichia coli* and heat-shock transformation. JEMI Methods 1:22–25
19. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054
20. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786
21. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315
22. Armenteros JJA, Salvatore M, Emanuelsson O et al (2019) Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2:1–14. https://doi.org/10.26508/lsa.201900429
23. Yu CS, Cheng CW, Su WC et al (2014) CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS ONE 9:e99368. https://doi.org/10.1371/journal.pone.0099368
24. Horton P, Park KJ, Obayashi T et al (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35:585–587. https://doi.org/10.1093/nar/gkm259
25. Krieger E, Darden T, Nabuurs SB et al (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins 57:678–683. https://doi.org/10.1002/prot.20251
26. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519. https://doi.org/10.1002/pro.5560020916
27. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291. https://doi.org/10.1107/s0021889892009944
28. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99. https://doi.org/10.1016/S0022-2836(63)80023-6
29. Sun W, Meng X, Liang L et al (2015) Molecular and biochemical analysis of chalcone synthase from *Freesia hybrida* in flavonoid biosynthetic pathway. PLoS ONE 10:1–18. https://doi.org/10.1371/journal.pone.0119054
30. Das P, Rawal SK (2016) Cloning, expression and purification of chalcone synthase from *Solanum tuberosum*. IOSR J Biotechnol Biochem 2(3):7–11
31. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685. https://doi.org/10.1038/227680a0
32. Mahmood T, Yang P-C (2012) Western blot: technique, theory, and trouble shooting. N Am J Med Sci 4:429–434. https://doi.org/10.4103/1947-2714.100998
33. Cheng H, Li L, Cheng S et al (2011) Molecular cloning and function assay of a chalcone isomerase gene (GbCHI) from *Ginkgo biloba*. Plant Cell Rep 30:49–62. https://doi.org/10.1007/s00299-010-0943-4
34. Park SH, Lee CW, Cho SM et al (2018) Crystal structure and enzymatic properties of chalcone isomerase from the Antarctic vascular plant *Deschampsia antarctica* Desv. PLoS ONE 13:1–17. https://doi.org/10.1371/journal.pone.0192415
35. Ku Bahaudin KNA, Ramzi AB, Baharum SN et al (2018) Current progress in production of flavonoids using systems and synthetic biology platforms. Sains Malays 47:3077–3084. https://doi.org/10.17576/jsm-2018-4712-18
36. Yin YC, Zhang XD, Gao ZQ et al (2019) The research progress of chalcone isomerase (CHI) in plants. Mol Biotechnol 61:32–52. https://doi.org/10.1007/s12033-018-0130-3
37. Sun W, Shen H, Xu H et al (2019) Chalcone isomerase a key enzyme for anthocyanin biosynthesis in *Ophiorrhiza japonica*. Front Plant Sci 10:1–12. https://doi.org/10.3389/fpls.2019.00865
38. Jez JM, Noel JP (2002) Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. J Biol Chem 277:1361–1369. https://doi.org/10.1074/jbc.M109224200
39. Wan Q, Bai T, Liu M, Liu Y, Xie Y, Zhang T, Huang M, Zhang J (2022) Comparative analysis of the chalcone-flavanone isomerase genes in six citrus species and their expression analysis in sweet orange (*Citrus sinensis*). Front Genet 13:1–13. https://doi.org/10.3389/fgene.2022.848141
40. Dixon RA, Richard Blyden E, Robbins MP et al (1988) Comparative biochemistry of chalcone isomerases. Phytochemistry 27:2801–2808. https://doi.org/10.1016/0031-9422(88)80666-6
41. Guo D, Gao Y, Liu F et al (2019) Integrating molecular characterization and metabolites profile revealed CtCHI1's significant role in *Carthamus tinctorius* L. BMC Plant Biol 19:1–13. https://doi.org/10.1186/s12870-019-1962-0
42. Wang L, Liu X, Meng X et al (2018) Cloning and expression analysis of a chalcone isomerase (CnCHI) gene from *Chamaemelum nobile*. Biotechnology 17:19–25. https://doi.org/10.3923/biotech.2018.19.25

43. Bednar RA, Hadcock JR (1998) Purification and characterization of chalcone isomerase from soybeans. J Biol Chem 273:30003–30011. https://doi.org/10.1074/jbc.273.45.30003
44. Zhou X, Li J, Fan Z (2012) Cloning and expression analysis of chalcone isomerase gene cDNA from Camellia nitidissima. Forest Res Beijing 25(1):93–99
45. Shahat AA, Marzouk MS (2013) 13—tannins and related compounds from medicinal plants of Africa. In: Kuete V (ed) Medicinal plant research in Africa. Elsevier, Oxford, pp 479–555
46. Jez JM, Bowman ME, Noel JP (2002) Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry 41:5168–5176. https://doi.org/10.1021/bi0255266
47. Furumura S, Ozaki T, Sugawara A, Morishita Y, Tsukada K, Ikuta T, Inoue A, Asai T (2022) Identification and functional characterization of fungal chalcone synthase and chalcone isomerase. J Nat Prod 86:398–405. https://doi.org/10.1021/acs.jnatprod.2c01027
48. Khor BY, Tye GJ, Lim TS et al (2014) The structure and dynamics of BmR1 protein from *Brugia malayi*: in silico approaches. Int J Mol Sci 15:11082–11099. https://doi.org/10.3390/ijms150611082
49. Ngaki MN, Louie GV, Philippe RN et al (2012) Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature 485:530–533. https://doi.org/10.1038/nature11009
50. Barh D, Barve N, Gupta K et al (2013) Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds. PLoS ONE 8:e52773. https://doi.org/10.1371/journal.pone.0052773
51. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol 10:411–421. https://doi.org/10.1016/S0958-1669(99)00003-8