Supporting Information for

A Plant Type III Polyketide Synthase that Produces Pentaketide Chromone

Ikuro Abe,* Yoriko Utsumi, Satoshi Oguro, Hiroyuki Morita, Yukie Sano, Hiroshi Noguchi

School of Pharmaceutical Sciences and the 21st Century COE Program,

University of Shizuoka, 52-1 Yada, Shizuoka 422-8526, Japan

Materials and Methods

Chemicals. [2-^{14}C]Malonyl-CoA (48 mCi/mmol) and [1-^{14}C]acetyl CoA (47 mCi/mmol) was purchased from Moravek Biochemicals (California). 5,7-Dihydroxy-2-methyl-chromone was obtained from Professor Takao Tanahashi (Kobe Pharmaceutical University), and SEK4 and SEK4b from Professor Chaitan Khosla (Stanford University).

cDNA Cloning. Roots of *Aloe arborescens* were harvested at the medicinal plant garden of University of Shizuoka in April 2001, and immediately frozen with liquid nitrogen. Total RNA was extracted by the acid guanidium thiocyanate-phenol-chloroform method, and reverse-transcribed using Reverscript (Wako) and oligo dT primer (RACE 32 = 5'-GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT T-3') according to manufacturer's protocol. The obtained cDNA mixture was diluted with TE (10 mM Tris, 1 mM EDTA, pH 8.0) and used as a template for the following PCR reactions. As described before (Abe, I., Takahashi, Y., Morita, H., and Noguchi, H. *Eur. J. Biochem.* **268**, 3354-3359, 2001), inosine-containing

degenerate oligonucleotide primers (112S, 174S, 368A and 380A, the number of primer indicates the amino acid number of corresponding *M. sativa* CHS) based on the highly conserved sequences of known CHSs were used for amplification of a core fragment of the cDNA. The sequences of the primers are as follows: 112S = 5'-(A/G)A(A/G) GCI ITI (A/C)A(A/G) GA(A/G) TGG GGI CA-3', 174S = 5'- GCI AA(A/G) GA(T/C) ITI GCI GA(A/G) AA(T/C) AA-3', 368A = 5'-CCC (C/A)(A/T)I TCI A(A/G)I CCI TCI CCI GTI GT-3', and 380A = 5'-TCI A(T/C)I GTI A(A/G)I CCI GGI CC(A/G) AA-3'. Following the conditions as described, nested PCR was carried out with the primer sets of 112S and 380A, and then with 174S and 368A, to amplify a 544-bp DNA fragment. For the PCR, 30 cycles of reactions (94 °C for 0.5 min, 42 °C for 0.5 min and 72 °C for 1 min) were performed each time with a 10 min final extension. The gel-purified PCR product was ligated into pT7Blue T-Vector (Novagen) and sequenced.

For the 3'-end amplification, two specific primers; 278S = 5'-GGA GCT CAC CAT CAT GAT GC-3' and 327S = 5'- CAT CTC GAC AAT GCC ATC GG-3', were designed based on the obtained core sequence. First RT-PCR was carried out with the primer set of 278S and RACE32, and the second PCR with 327S and RACE32, to amplify a 359-bp DNA fragment. For the PCR, 30 cycles of reactions (94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min) were performed each time. The 5'-end amplification was carried out using the Marathon™ cDNA Amplification kit (CLONTECH) and two specific primers; 331A = 5'-CCA ATG ATC AGT GCA GCA GC-3' and 232A = 5'- CCG ATG GCA TTG TCG AGA TG -3', based on the obtained core sequence. Thus, template cDNA was prepared according to manufacturer's instruction, and first PCR was carried out with the primer set of 331A and AP1, and the second

PCR with 232A and AP2, to amplify a 702-bp DNA fragment. The PCR conditions were the same as described above for the 3'-RACE. The nucleotide sequence has been deposited in the EMBL/DDBJ/GenBank™ data bases under accession no. AY823626.

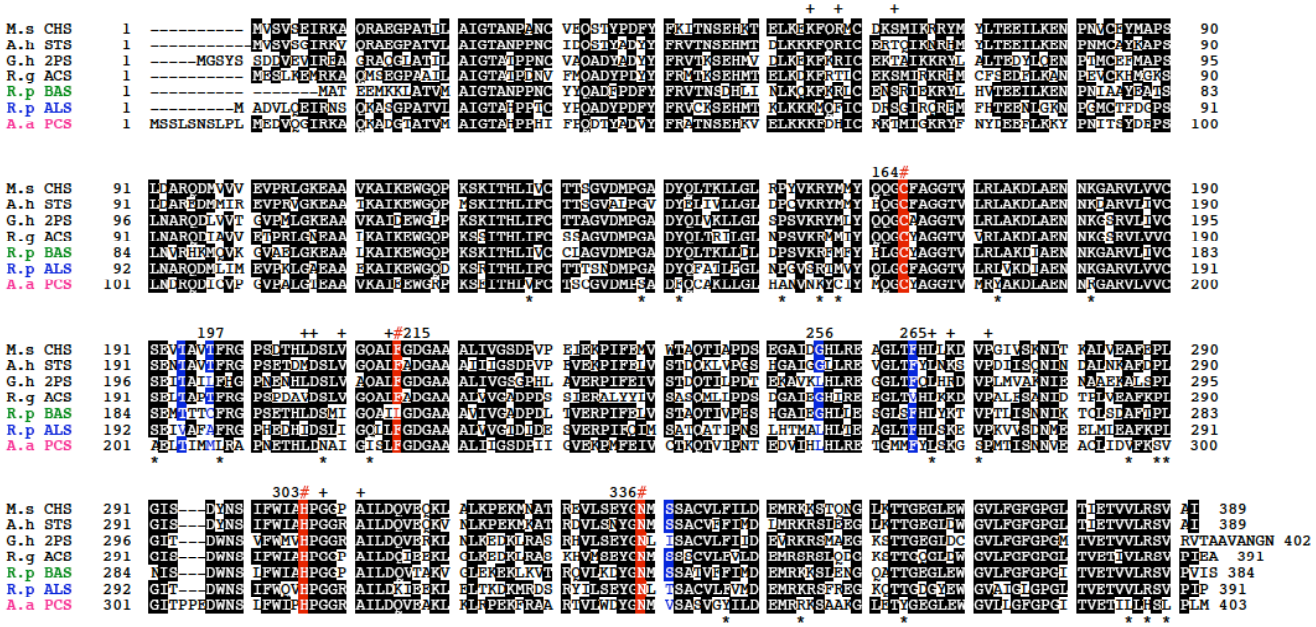

Fig. 1 Comparison of primary sequences of *Aloe arborescens* PCS and other CHS-superfamily enzymes. M.s CHS, *Medicago sativa* CHS; A.h STS, *Arachis hypogaea* stilbene synthase; G.h 2PS, *Gerbera hybrida* 2-pyrone synthase; R.g ACS, *Ruta gravenolens* acridone synthase; R.p BAS, *Rheum palmatum* benzalacetone synthase; R.p ALS, *R. palmatum* ALS. The active-site residues conserved in the CHS-superfamily enzymes (Cys164, Phe215, His303, and Asn336, numbering in *M. sativa* CHS) were marked with #, and residues for the CoA binding with +. Amino acid residues conserved in CHS-superfamily enzymes but absent in *A. arborescens* PCS are marked with *.

Expression of Full-length cDNA. A full length cDNA was obtained using N- and C-terminal PCR primers; 5'-CCT GCC GAA TTC ATG AGT TCA CTC TCC AAC-3' (sense, the *EcoR* I site is underlined) and 5'-ATA ATA CTC GAG TTA CAT GAG AGG CAG GCT-3' (antisense, the *Xho* I site is underlined). The amplified DNA was digested with *EcoR* I/*Xho* I, and cloned into the *EcoR* I/*Xho* I site of pET-41a(+) (Novagen). Thus, the recombinant enzyme was a fusion protein with GST at the N-terminal. After confirmation of the sequence,

the plasmid was transformed into *E. coli* BL21(DE3)pLysS. The cells harboring the plasmid were cultured to an A_{600} of 0.6 in Luria-Bertani medium containing 50 μg/mL of kanamycin at 30 °C. Then, 0.4 mM isopropyl thio-β-D-galactoside was added to induce protein expression, and the culture was incubated further at 30 °C for 16h.

Enzyme Purification. The *E. coli* cells were harvested by centrifugation and resuspended in 50 mM Tris-HCl buffer (pH 8.0), containintg 0.2 M NaCl, 5% Glycerol, and 2 mM DTT. Cell lysis was carried out by sonication, and centrifuged at 15,000 *g* for 40 min. The supernatant was passed through a column of Glutathione Sepharose 4B (Amersham). After washing with the Tris-HCl buffer, the resin was resuspended in 2 mL of 50 mM Tris-HCl buffer (pH 8.0), containintg 10% Glycerol, and subjected to protease digestion at 4 °C overnight. After the GST-tag cleavage, recombinant PCS was recovered in the buffer, and finally eluted with 50 mM Tris-HCl buffer (pH 8.0), containintg 10% Glycerol. Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as standard. Finally, to determine subunit composition, the purified enzyme was applied to HPLC gel filtration column (TSK-gel G3000SW, 7.5 x 600 mm, TOSOH), which was eluted with 0.1 M KPB, pH 6.8, containing 10% glycerol and 0.2 M KCl at a flow rate of 1.0 ml/min. The standard molecular weight markers used: β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome c (12.4 kDa).

Site-directed Mutagenesis. Met207 of PCS was replaced by Thr or Gly using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) and pair of primers as follows (mutated codons are underlined): M207T sense 5'-GAG CTC ACC ATA ATC <u>ACG</u> CTT CGA GGC

CCT-3', anti sense 5'-AGG GCC TCG AAG CGT GAT TAT GGT GAG CTC-3'; M207G sense 5'-GAG CTC ACC ATA ATC ATG CTT CGA GGC CCT-3', anti sense 5'-AGG GCC TCG AAG CAT GAT TAT GGT GAG CTC-3'.

Enzyme Reaction. The standard reaction mixture contained 54 nmol of malonyl-CoA, (and 27 nmol of other CoA ester), and 286 pmol of the purified recombinant enzyme in a final volume of 500 μL of 100 mM potassium phosphate buffer, pH 7.0. Incubations were carried out at 30 °C for 1 hour, and stopped by adding 50 μL of 20% HCl. The products were then extracted with 1,000 μL of ethyl acetate, and concentrated by N_2 flow. The residue was dissolved in MeOH containing 0.1% TFA, and analyzed by reverse-phase HPLC and LC-ESIMS. Column, TSKgel ODS-80Ts, 4.6 x 150 mm, TOSOH; gradient elution with H_2O and MeOH, both containing 0.1% TFA: 0-5 min, 30% MeOH; 5-17 min, 30 to 60% MeOH; 17-25 min, 60% MeOH; 25-27 min, 60 to 70% MeOH; flow rate, 0.8 mL/min; monitored by a multichannel UV detector, MULTI 340, JASCO; retention time, 5,7-Dihydroxy-2-methylchromone (24.4 min), SEK4 (21.2 min), and SEK4b (22.1 min).

On-line LC-ESIMS spectra were measured with a Hewlett-Packard HPLC 1100 series (Wilmington, DE) coupled to a Finnigan MAT LCQ ion trap mass spectrometer (San Jose, CA) fitted with an ESI source. HPLC separations were carried out under the same conditions as described above. The ESI capillary temperature and capillary voltage were 275 °C and 3.0 V, respectively. The tube lens offset was set at 20.0 V. All spectra were obtained in both negative and positive mode; over a mass range of m/z 120-350, at a range of one scan every 2 s. The collision gas was helium, and the relative collision energy scale was set at 30.0% (1.5 eV).

For large-scale enzyme reaction, 26 mg of purified enzyme was incubated with malonyl-CoA (20 mg, 30 μmol) in 100 mL of 100 mM phosphate buffer, pH 7.5, containing 1 mM EDTA, at 30 °C for 18 hours. The reaction was quenched by addition of 20% HCl (10 mL), and extracted with ethyl acetate (200 mL x 3). After HPLC separation, pure 5,7-Dihydroxy-2-methylchromone (ca 1.0 mg) was obtained.

5,7-Dihydroxy-2-methylchromone: LC-ESIMS: *m/z* 193 $[M+H]^+$. UV: λ_{max} 276 nm. ^{1}H NMR (400 MHz, acetone-d_6): δ 6.34 (1H, d, *J* = 2.1 Hz), 6.21 (1H, d, *J* = 2.1 Hz), 6.06 (1H, s), 2.36 (3H, s). In the similar manner, SEK4 and SEK4b were obtained. SEK4: LC-ESIMS: *m/z* 319 $[M+H]^+$. UV: λ_{max} 280 nm. ^{1}H NMR (400 MHz, DMSO-d_6): δ 6.28 (1H, d, *J* = 2.0 Hz), 6.21 (1H, d, *J* = 2.0 Hz), 5.33 (1H, s), 4.06 (1H, d, *J* = 15.7 Hz), 3.95 (1H, d, *J* = 15.7 Hz), 2.89 (1H, d, *J* = 15.9 Hz), 2.50 (1H, d, *J* = 15.9 Hz), 1.54 (3H, s). SEK4b: LC-ESIMS: *m/z* 319 $[M+H]^+$. UV: λ_{max} 280 nm. ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.18 (1H, s), 6.25 (1H, d, *J* = 2.0 Hz), 6.15 (1H, d, *J* = 2.0 Hz), 6.01 (1H, s), 5.13 (1H, bs), 3.02 (1H, d, *J* = 14.4 Hz), 2.98 (1H, d, *J* = 16.2 Hz), 2.94 (1H, d, *J* = 14.4 Hz), 2.54 (1H, d, *J* = 16.2 Hz), 2.44 (3H, s).

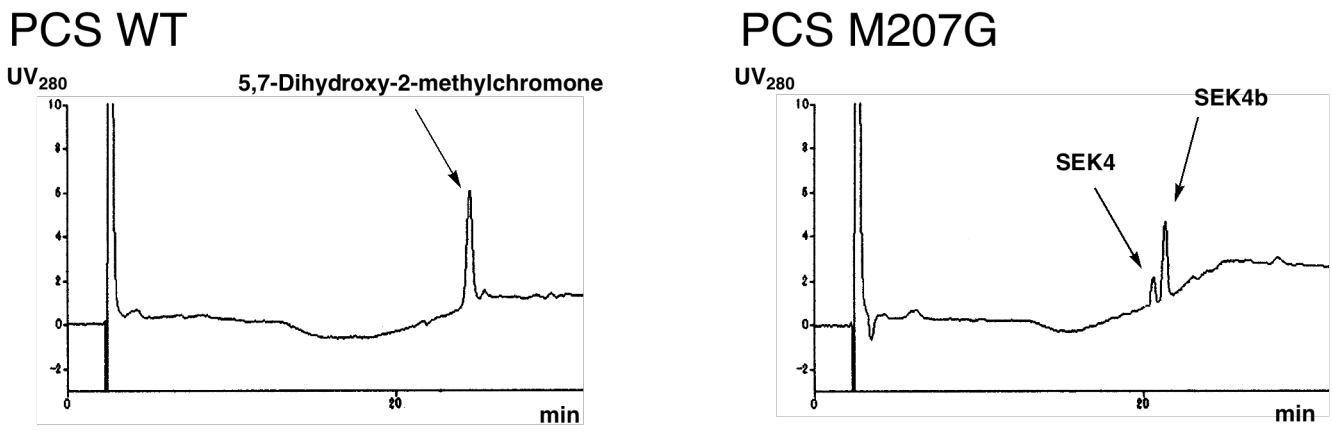

Fig. 2 HPLC profile of the enzyme reaction products of wild-type PCS (left) and PCS M207G mutant (right).

Determination of Starter Substrate. Acetyl-CoA, resulting from decarboxylation of malonyl-CoA, was also accepted as a starter substrate, but not so efficiently as in the case of *R. palmatum* ALS (Abe, I., Utsumi, Y,, Oguro, S., and Noguchi, H. *FEBS Lett.* **562**, 171-176, 2004). This was confirmed by the ^{14}C incorporation rate from [1-^{14}C]acetyl CoA in the presence of cold malonyl-CoA, while the yield of the pentaketide from [2-^{14}C]malonyl-CoA was almost at the same level in the presence or absence of cold acetyl-CoA in the reaction mixture. Theoretical ^{14}C specific incorporation from [1-^{14}C]acetyl-CoA should be 20% of those from [2-^{14}C]malonyl-CoA if acetyl CoA serves as a starter unit of the pentaketide forming reaction.

A. arborescens PCS also accepted aromatic (4-coumaroyl, cinnamoyl, and benzoyl) and aliphatic (*n*-hexanoyl, *n*-octanoyl, and *n*-decanoyl) CoA esters as a starter substrate, however, yielded only triketide and teraketide α-pyrones. Both M207T and M207G mutant also accepted 4-coumaroyl-CoA to produce CTAL and BNY, but did not yield naringenin chalcone.

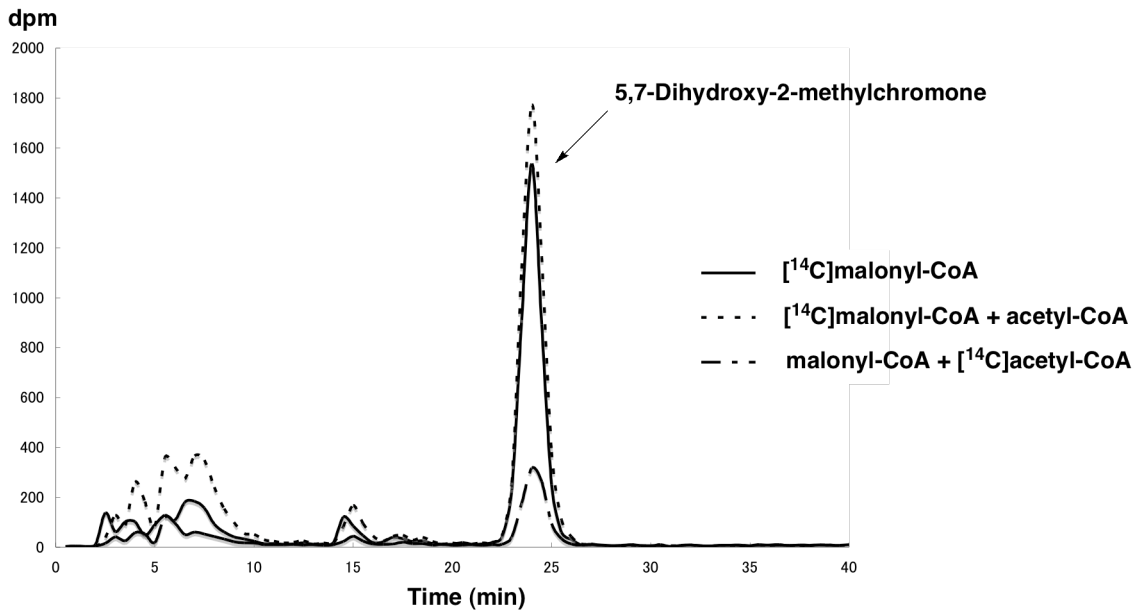

Fig. 3 HPLC analysis of PCS products by the ^{14}C incorporation rate from (1) 108 μM [2-^{14}C]malonyl-CoA (33,000 dpm) only, (2) 108 μM [2-^{14}C]malonyl-CoA (33,000 dpm) and 108 μM cold acetyl-CoA, and (3) 108 μM [1-^{14}C]acetyl CoA (33,000 dpm) and 108 μM cold malonyl-CoA, .

Enzyme Kinetics. Steady state kinetic parameters were determined by using [2-^{14}C]malonyl-CoA (1.8 mCi/mmol) as a substrate. The experiments were carried out in triplicate using five concentrations (from 6.5 to 117.8 μM) in the assay mixture, containing 2 μg of purified enzyme, 1 mM EDTA, in a final volume of 500 μL of 100 mM Tis-HCl buffer, pH 7.0. Incubations were carried out at 30 °C for 15 min. The reaction products were extracted and separated by TLC (Merck Art. 1.11798 Silica gel 60 F_{254}; ethyl acetate/hexane/AcOH = 63:27:5, v/v/v). Radioactivities were quantified by autoradiography using a bioimaging analyzer BAS-2000II (FUJIFILM). Lineweaver-Burk plots of data were employed to derive the apparent K_M and k_{cat} values (average of triplicates) using EnzFitter software (BIOSOFT).

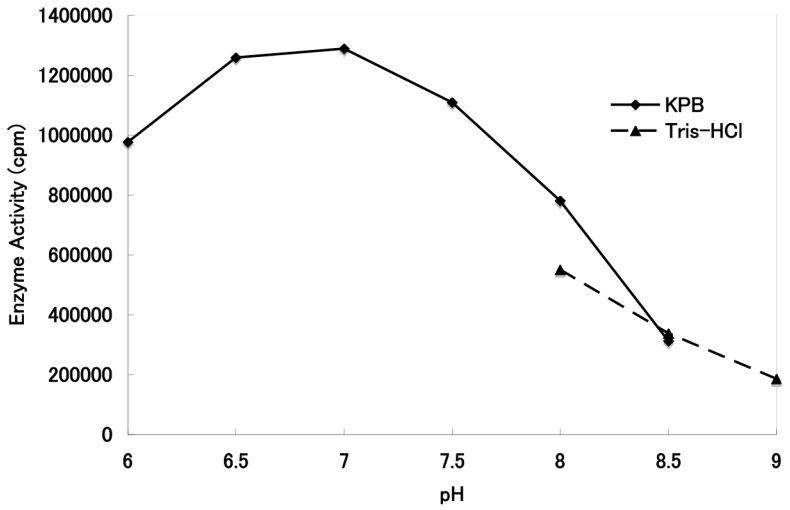

Fig. 4 The pH dependence of enzyme activity of A. *arborescens* PCS.

Three-dimensional Homology Modeling. The model was produced by the SWISS-MODEL package (http://expasy.ch/spdbv/) provided by the Swiss-PDB-Viewer program (Guex, N., and Peitsch, M. C. *Electrophoresis* **18**, 2714-2723, 1977). A standard homology modeling procedure was applied based on the sequence homology of residues 6-403 of *A. arborescens* PCS and the X-ray crystal structures of CHS including *M. sativa* CHS (1BI5A.pdb, 1BQ6A.pdb, 1CGKA.pdb, 1CGZA.pdb, 1CHWA.pdb, 1CHWB.pdb, 1CMLA.pdb), *M. sativa* CHS C164A mutant (1D6FA.pdb), *M. sativa* CHS N336A mutant (1D6HA.pdb), *M. sativa* CHS H303Q mutant (1D6IA.pdb, 1D6IB.pdb), *M. sativa* CHS G256A mutant (1I86A.pdb), *M. sativa* CHS G256V mutant (1I88A.pdb, 1I88B.pdb), *M. sativa* CHS G256L mutant (1I89A.pdb, 1I89B.pdb), *M. sativa* CHS G256F mutant (1I8BA.pdb, 1I8BB.pdb), and *M. sativa* CHS F215F mutant (1JWXA.pdb). The corresponding Ramachandran plot was also created with Swiss PDB-Viewer software to confirm that the majority of residues grouped in the energetically allowed regions. Calculation of cavity volumes (Connolly's surface volumes) was then performed with CASTP program (http://cast.engr.uic.edu/cast/).

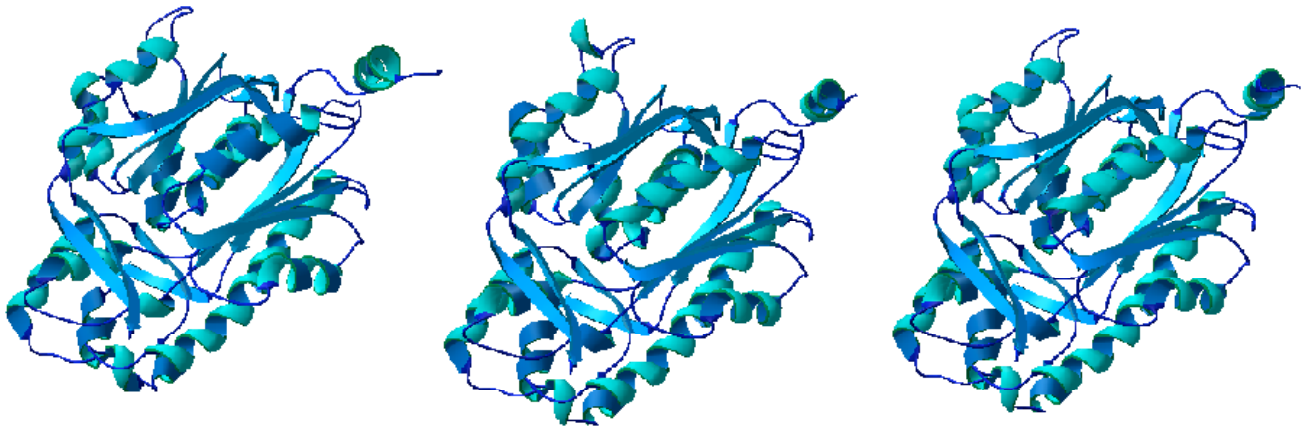

Fig. 5 Ribbon display of (from left to right) (A) crystal structure of *M. sativa* CHS (1CGKA.pdb), (B) three-dimensional structure of *A. arborescens* PCS, and (C) *R. palmatum* ALS predicted by Swiss-PDB-Viewer software.