Evolution of the chalcone–isomerase fold from fatty–acid binding to stereospecific catalysis

Micheline N. Ngaki1*, Gordon V. Louie2*, Ryan N. Philippe2*, Gerard Manning3, Florence Pojer2, Marianne E. Bowman2, Ling Li1, Elise Larsen3, Eve Syrkin Wurtele3 & Joseph P. Noel2

Specialized metabolic enzymes biosynthesize chemicals of ecological importance, often sharing a pedigree with primary metabolic enzymes. However, the lineage of the enzyme chalcone isomerase (CHI) remained unknown. In vascular plants, CHI-catalysed conversion of chalcones to chiral (S)-flavanones is a committed step in the production of plant flavonoids, compounds that contribute to attraction, defence and development. CHI operates near the diffusion limit with stereospecific control. Although associated primarily with plants, the CHI fold occurs in several other eukaryotic lineages and in some bacteria. Here we report crystal structures, ligand-binding properties and in vivo functional characterization of a non-catalytic CHI-fold family from plants. Arabidopsis thaliana contains five actively transcribed genes encoding CHI-fold proteins, three of which additionally encode amino-terminal chloroplast-transit sequences. These three CHI-fold proteins localize to plastids, the site of de novo fatty-acid biosynthesis in plant cells. Furthermore, their expression profiles correlate with those of core fatty-acid biosynthetic enzymes, with maximal expression occurring in seeds and coinciding with increased fatty-acid storage in the developing embryo. In vitro, these proteins are fatty-acid-binding proteins (FAPs). FAP knockout A. thaliana plants show elevated α-linolenic acid levels and marked reproductive defects, including aberrant seed formation. Notably, the FAP discovery defines the adaptive evolution of a stereospecific and catalytically ‘perfected’ enzyme from a non-enzymatic ancestor over a defined period of plant evolution.

CHI (EC 5.5.1.6) catalyses the intramolecular and stereospecific cyclization of chalcones to chiral flavanones by a Michael addition reaction. The origin of CHI has been a mystery, because of the apparent absence of a related protein from primary metabolism. CHI-like homologues in fungi and bacteria lack both key catalytic residues and the chalcone-binding site of bona fide CHI (Fig. 1a–c and Supplementary Fig. 1). Phylogenetic analysis shows that CHI is restricted to vascular plants, and is derived from FAP3. FAP3 forms one of two branches of FAPs (FAP1 and FAP2 form the other) found in derived and basal plants and is probably homologous to CHI-fold proteins from protists and fungi (Supplementary Figs 2 and 3 and Supplementary Files 1 and 2). A. thaliana has five genes (and one pseudogene not shown) encoding CHI-fold proteins: two CHI/CHI-like (CHIL) members, AtCHI (At3g55120) and AtCHIL (At3g63170, At2g26310 and At1g53520, respectively). Pairwise sequence identities of the shared CHI-fold domain range from 10 to 63%. AtCHI is a bona fide enzyme, and mutations of its gene result in plants devoid of flavonoids. The active site of AtCHI retains residues important for catalytic activity, and AtCHI catalyses the in vitro formation of (2S)-naringenin (Fig. 1c). In contrast, AtCHIL has substitutions of several catalytic residues (Supplementary Fig. 3). Likewise, AtFAP1, AtFAP2 and AtFAP3 bear substitutions at nearly all of the critical catalytic positions of AtCHI, and are devoid of catalytic activity.

1Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, Iowa 50011, USA. 2Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. 3Razavi Newman Center for Bioinformatics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

Figure 1 | CHI fold and catalytic reaction. a, Ribbon diagram21 of the AtCHI X-ray crystal structure, colour-coded and labelled according to Jez et al.22. Two nitrate anions associate with catalytic residues in the substrate-binding site. b, Ribbon diagram21 of the MsCHI structure bound to (2S)-naringenin. c, Chalcone is converted to (2S)-flavanone (for example, naringenin) using a combination of electrostatic catalysis and water-mediated charge stabilization during a stereospecific Michael-type addition reaction23,24. Residue numbers for AtCHI appear in parentheses. Catalytic residues are coloured red.

©2012 Macmillan Publishers Limited. All rights reserved
(Supplementary Fig. 3). Studies with green fluorescent protein (GFP) fused to full-length FAP1, FAP2 and FAP3 indicate that these proteins localize in the plastid stroma (Supplementary Fig. 4a–d).

We determined X-ray crystal structures of AtCHI, AtCHIL and the CHI-like domains of AtFAP1 and AtFAP3, and compared these with the previously reported structure of MsCHI. Despite low sequence conservation, all structures align within 2.3 Å root-mean-squared deviation for the backbone atoms of equivalent residues (Supplementary Fig. 5). Ligand-binding pockets reside within the helical layer, bounded by two α-helical segments (helix–turn–helix n6–n7 and helical hairpin n9–n5), β-hairpin β3a–β3b and the face of the core β-sheet, β3a–β3f. Notably, the largest differences in backbone conformations of MsCHI, AtCHI and AtCHIL compared with AtFAP1 and AtFAP3 are confined to the secondary structure elements surrounding the ligand-binding pockets (Supplementary Fig. 5).

The most intriguing finding from the crystallographic structures is that the ligand-binding pockets of the recombinant FAPs are occupied by fatty acids (Fig. 2a–d). Clearly evident in electron-density maps are two abutting molecules of lauric acid (C12:0) in AtFAP1 (Fig. 2a, b), and one molecule of palmitic acid (C16:0) in AtFAP3 (Fig. 2c, d). The ligand-binding site of the FAPs encompasses a largely buried, non-polar cavity that sequesters the aliphatic chain(s) of the fatty acids, and a conserved Arg–Tyr pair that tethers a fatty-acid carboxylate group (Fig. 2b, d and Supplementary Fig. 6). By comparison, CHI and CHIL possess more sterically restricted pockets owing to the presence of larger aliphatic residues and an inward shift of the secondary structure elements surrounding their ligand-binding clefts (Figs 1 and 2 and Supplementary Fig. 5). The CHI substrate-binding pocket also includes several polar residues that form key hydrogen-bond interactions with the chalcone substrate (Fig. 1c). The corresponding residues in the FAPs are non-polar. AtCHIL possesses a distinct conservation pattern of the active site pocket (Supplementary Fig. 3). Extraction of ligands with ethanol followed by high-performance liquid chromatography–mass spectrometry (HPLC–MS) analyses confirms that all three FAPs (including AtFAP2, which was recalcitrant to crystallization) associate with saturated fatty acids (Fig. 2e). In contrast, with AtCHI and AtCHIL produced under similar conditions, no associated fatty acids were detected analytically or crystallographically. Exogenous fatty acids were not supplied during FAP expression in Escherichia coli, purification or crystallization, so the ligands bound to the FAPs were probably acquired during expression and retained during purification to homogeneity. Indeed, these saturated fatty acids are among the most abundant fatty acids in E. coli.11

There are indications that the true ligands of the FAPs in planta may be different fatty acids. In AtFAP1, the presence of two distinct non-polar tunnels, each occupied by the aliphatic chain of a C12:0 molecule (Fig. 2b and Supplementary Fig. 6a), may indicate a proclivity for longer chain fatty-acid recognition. In AtFAP3, the horseshoe-shaped conformation of C16:0 ports a preference for binding a cis-unsaturated C16 or C18 FA, which typically assumes a bent conformation (Fig. 2d and Supplementary Fig. 6b). Notably, AtFAP1 sequesters exogenously supplied α-linolenic acid (C18:3), which presumably displaces the bound fatty acids acquired during expression in E. coli (Fig. 2e). We further assessed the binding by the FAPs of a series of common fatty acids using thermal-stability assays12 (Supplementary Table 1). AtFAP3 interacts with most fatty acids tested and has maximal relative affinity for C16:0, though C18:0 induces greater thermal stabilization (Supplementary Fig. 7). AtFAP1 preferentially interacts with saturated fatty acids. The strongest stabilizing effect was observed with C18:3 (Supplementary Fig. 8), consistent with the displacement analyses (Fig. 2e) and perhaps a function of AtFAP1 in C18:3 metabolism as indicated also by phenotypic analysis. The thermal stability of AtCHI, included in these analyses as a negative control, was unaffected by the addition of fatty acids (Supplementary Fig. 9).

![Figure 2](image_url)

**Figure 2** | Three-dimensional structure and ligand binding of FAPs. a, Ribbon diagram of AtFAP1 oriented and colour-coded as in Fig. 1. The bound C12:0 molecules are shown as van der Waals spheres where carbon is yellow and oxygen is red. b, Close-up view of the AtFAP1 fatty-acid binding sites with the experimental electron density of each C12:0 shown at 1σ for a SIGMAA-weighted 2Fo − Fo map. c, Ribbon diagram of AtFAP3 rendered as in a, d. Close-up view of the AtFAP3 fatty-acid binding site with the C16:0 ligand shown as in b, e. Analysis of ligands associated with purified AtCHI-fold proteins separated and detected by reversed-phase HPLC–MS. The y-axes represent negative-ion counts for selected masses of anionic forms of fatty acids. The bottom panel depicts binding of C18:3 by His8-tagged AtFAP1 coupled to Ni2+ affinity resin (red) and by the control (blue).
We have obtained, through a variety of approaches, considerable evidence supporting the involvement of AtFAPs in fatty-acid metabolism. Network analysis of transcriptomic data demonstrated maximal expression of AtFAP1 and AtFAP3 in seeds at 6 days after flowering (Supplementary Fig. 10a), coinciding with the accumulation of storage lipids in the developing embryo. AtFAP1 and AtFAP3 co-express with genes encoding fatty-acid biosynthetic enzymes (Supplementary Fig. 10b). Developmental patterns of AtFAP1, AtFAP2 and AtFAP3 expression were characterized in *A. thaliana* lines expressing a β-glucuronidase (GUS) reporter under control of the respective promoters (Supplementary Figs 11, 12 and 13). The FAPs are highly expressed in developing cotyledons, young seedlings, roots, seeds, embryos, macrospores, preanthesis and tapetum. However, although AtFAP2 is expressed throughout the life of the plant, expression of AtFAP1 and especially AtFAP3 is restricted to developing and reproductive tissues.

Homozygous *Atfap1* and *Atfap2* null plants were propagated from transfer DNA (T-DNA) insertion lines (Supplementary Figs 14 and 15). These *Atfap1* and *Atfap2* null plants are indistinguishable from wild-type (WT) plants during vegetative growth (Supplementary Figs 17 and 18), but the *Atfap1* mutants show marked differences during reproductive stages. Specifically, *Atfap1* siliques are shorter than those of WT (~18 mm versus ~20 mm) (Fig. 3a), mutant siliques contain 10–20% abnormal ovules (Fig. 3b), and the number of seeds per siliqua and the yield of viable seeds per plant are reduced in the *Atfap1* null lines (Fig. 3c). The *AtFAP3-RNAi* homozygote lines also present these phenotypic alterations. In addition, the vegetative phenotype of *AtFAP3-RNAi* is altered, demonstrating early bolting, fast growth, increased branching, reduced apical dominance and reduced overall plant height (Supplementary Fig. 19). No *AtFAP3-RNAi* line with more than an approximately 50% reduction in *AtFAP3* RNA (relative to WT) could be identified, nor are any T-DNA-tagged *Atfap3* knockout lines available. These data thus indicate that a functional *AtFAP3* is required for normal plant and embryo development. Double mutants bearing pairwise combinations of homozygous *Atfap1*, *Atfap2*, and *AtFAP3-RNAi* all show additive developmental effects (Supplementary Fig. 18).

Reciprocal crosses between WT and *Atfap1* or *Atfap3-RNAi* lines point to the maternal inheritance of the aberrant reproductive phenotypes, as an elevated rate of ovule abortion is observed only in crosses in which the maternal parent is a null mutant (Supplementary Figs 20–21, Supplementary Tables 2–7). The morphological phenotypes of more than 100 lines of heterozygotic *AtFAP3-RNAi* were analysed; these consistently showed a restoration of the WT vegetative phenotype (Supplementary Fig. 22). However, the aberrant reproductive phenotype of the *AtFAP3-RNAi* mutant was maintained, indicating that phenotypic effects observed in the *AtFAP3-RNAi* lines are due to the specific knockdown of this gene. Analysis of fatty-acid composition reveals that *Atfap1* null lines have elevated total fatty-acid levels both in leaves and seeds relative to WT (Fig. 3d, e, g). These phenotypic alterations in leaves are temperature dependent (apparent only in plants grown at 15 °C and 22 °C) and are due primarily to increased levels of the lower-melting-temperature unsaturated FAs.

![Figure 3](image-url)

**Figure 3** | Phenotypic characterization of *Atfap1* null plants. a, Length of siliques 12 days after flowering; WT siliques are longer than siliques of *Atfap1* nulls (average length = 17.8 mm for WT versus 15.4 mm for *Atfap1*-1; 19.8 mm for WT versus 17.1 mm for *Atfap1*-3 (not shown), *P* < 0.01, *n* = 10 siliques per plant for 10 plants per genotype). b, *Atfap1* null siliques frequently contain abnormal ovules (normal, yellow arrow; aborted, blue arrow; unfertilized, pink arrow). c, *Atfap1* nulls have a greater percentage of abnormal (aborted and unfertilized) ovules than WT (*P* < 0.01 and *n* = 10 siliques per genotype) and low seed yield mass per plant (*Atfap1* average = 1.3 g; WT average = 1.5 g; *P* < 0.01, *n* = 10 plants). Fatty-acid content of leaves, milligrams per gram fresh weight, from plants grown at d, 15 °C, e, 22 °C, f, 26 °C. d, C18:3 and total fatty acids are greater in leaves of *Atfap1* nulls than WT. *e*, C18:3 and total fatty acids are greater in leaves of *Atfap1* nulls than WT. *f*, *Atfap1* nulls and WT have similar fatty-acid content. *g*, In seeds of *Atfap1* mutants, C16:0, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C20:2, C22:1 and total fatty acids (not shown) increase relative to WT. Asterisks indicate statistically significant differences from wild type for both *Atfap1* null lines (*P* < 0.05); *n* = 5 biological replicates, except for *g*, which has three experimental replicates with *n* = 4 biological replicates; error bars, s.d.
particularly C18:3. The fatty-acid content and distribution in leaves are indistinguishable between WT and mutant plants grown at 26 °C (Fig. 3). Similarly, AtFAP3-RNAi plants and Atfap double mutants have altered fatty-acid composition and increased fatty-acid content in leaves and seeds (Supplementary Fig. 23). The temperature dependence of this phenotype reflects a decrease in the more fluid acyl chain component of trienoic fatty acids that accompanies increasing temperature.16,17

We have identified three relatives of CHI in Arabidopsis that appear to be FAPs of undetermined mechanism. Phylogeny and sequence analyses indicate that the CHI fold occurs in most eukaryotic lineages (with the notable exception of animals) and even in bacteria. CHIL, as a divergent relative of FAP3, initially arose in mosses and may have served as the ancestor of enzymatic CHI in vascular plants (Supplementary Figs 2 and 3). This ancestry for plant CHI clarifies the mystery surrounding the function of FAP1 and FAP3, which encode proteins with structural and mechanistic similarities between chalcone synthase (a type-III polylectide synthase) acting one step upstream of CHI in the flavonoid pathway and ketoacyl-synthase III involved in chloroplast and bacterial fatty-acid biosynthesis. Together, CHI and CHS provide the foundation to examine the evolution of more recent metabolic pathways and metabolite classes from widespread biosynthetic pathways of primary metabolism.

METHODS SUMMARY

Structural biology. Crystal structures of the CHI domains of AtCHI, AtCHIL, AtFAP1 and AtFAP3 were determined as described in Methods and summarized in Supplementary Table 8. In vitro biochemistry. CHI activity was assayed. Lipids bound to the purified AtFAP proteins were extracted with ethanol and analysed by reverse-phase HPLC–MS. The effects of fatty acids on the melting temperature of the AtFAP proteins were measured using a Thermofluor-based assay.12 Phylogenetic and sequence analyses. Homologues of AtCHI were identified by iterative PSI-BLAST and profile hidden Markov model. Alignments were curated using protein structure superpositions and homologous groups determined manually. Informatics. Co-expression patterns of A. thaliana genes encoding FAPs and genes associated with fatty-acid biosynthesis were evaluated with data compiled from 72 microarray experiments using MetOmGraph (http://www.metnetdb.org).13 Plant growth and mutant analyses. Plants were grown in randomised block designs at 22 °C under 16 h light/8 h dark. Fatty acids were measured 5 h after onset of light. Temperature effects were observed after 10 day growth at the specified temperatures. Homozygous Atfap1-1, Atfap2-3, Atfap2-1 and Atfap3-2 mutant alleles (Supplementary Fig. 14 and Supplementary Table 3) were generated from SALK_130560, SALK_039829, SAIL_171_C12 and SAIL_616_D09 (Arabidopsis Biological Resource Center) stocks, respectively. Atfap3 RNAi lines (Supplementary Fig. 16 and Supplementary Table 9) were generated from A. thaliana Col-0 plants, transformed with Agrobacterium tumefaciens strain GV3101 containing vector CATMA1a44560 (Nottingham Arabidopsis Stock Centre).10 Atfap RNA was evaluated by quantitative RT–PCR (Supplementary Figs 15 and 16). Transgenic A. thaliana lines were generated with AtFAP1, -2 or -3 promoters or promoter + CTP and/or + CDS fused to GUS and GFP. Fatty-acid quantification in plants. Fatty acids were extracted from plant tissues using barium hydroxide hydrolysis, modified by methylation and analysed by gas chromatography–mass spectrometry (GC–MS). Monodecanoic acid was used as an internal standard and a reference fatty-acid mixture was used for calibration and retention time determination (http://www.plantmetabolomics.org).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 April 2011; accepted 1 March 2012.

Published online 13 May 2012.

1. Hartmann, T. From waste products to ecoc hemicals: fifty years research of plant secondary metabolism. Phytochemistry 68, 2831–2846 (2007).
2. Ferrer, J.-L., Auch, M. B., Stewart, C. J. & Noel, J. P. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol. Biochem. 46, 356–370 (2008).
3. Peer, W. A. & Murphy, A. S. Flavonoids and auxin transport: modulators or regulators? Trends Plant Sci. 21, 556–563 (2007).
4. Bednar, R. A. & Hadcock, J. R. Purification and characterization of chalcone isomerases from soybeans. J. Biol. Chem. 263, 5982–5988 (1988).
5. Jez, M. & Noel, J. P. Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. J. Biol. Chem. 277, 1361–1369 (2002).
6. Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. Biochemistry 15, 5631–5640 (1976).
7. Jez, M. J., Bowman, M. E. & Noel, J. P. Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry 41, 5168–5176 (2002).
8. Jez, M. J., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. Nature Struct. Biol. 7, 786–791 (2000).
9. Gensheimer, M. S. & Mushagin, A. Chalcone isomerase family and fold: no longer unique to plants. Protein Sci. 13, 540–544 (2004).
10. Shirley, B. W., Hanley, S. & Goodman, H. M. Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent tests mutants. Plant Cell 4, 334–347 (1992).
11. Oursel, D. et al. Identification and relative quantification of fatty acids in Escherichia coli membranes by gas chromatography/mass spectrometry. Rapid Commun. Mass Spectrom. 21, 3229–3233 (2007).
12. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols 9, 2212–2221 (2007).
13. Mentzen, W. I. & Wurtele, E. S. Regulation of aromaticity. Arabidopsis Plant Biol. 8, 99 (2008).
14. Mentzen, W. I., Peng, J., Ransom, N., Nikolau, B. J. & Wurtele, E. S. Articulation of three core metabolic processes in Arabidopsis: fatty acid biosynthesis, leucine catabolism and starch metabolism. Plant Biol. 8, 76–90 (2008).
15. Ruuska, S. A., Girke, T., Benning, C. & Ohlrogge, J. B. Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14, 1191–1206 (2002).
16. Falcone, D. L., Ogas, J. P. & Somerville, C. R. Regulation of membrane fatty acid composition by temperature in mutants of Arabidopsis with alterations in membrane lipid composition. Plant Biol. 4, 17–31 (2004).
17. Murakami, Y., Tsuyama, Y., Kobayashi, Y., Kodama, H. & Iba, K. Trienoic fatty acids and plant tolerance of high temperature. Science 287, 476–479 (2000).
18. Austin, M. S. & Noel, J. P. The chalcone synthase superfamily of type III polylectide synthases. Nat. Prod. Rep. 20, 79–110 (2003).
19. Hilpon, P. et al. Versatile gene-specific tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. Genet. Res. 14, 2176–2197 (2004).
20. Bonaventure, G., Salas, J. J., Pollard, M. R. & Ohlrogge, J. B. Disruption of the FATTY gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell 15, 1020–1033 (2003).
21. Delano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).
22. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. We thank A. Perera, B. Nikolau, H. Iarsan and J. Peng for technical training (to M.N.N.), J. Peng for the fap1-1 homozygote mutant line, D. Nettleton and H. Wang for statistical analysis of initial seed fatty-acid data, and Eric Scheff for assistance in Bayesian phylogenetic analysis. This research was supported in part by a Fulbright fellowship (to M.N.N.). This material is based in part upon work supported by the Beckman Foundation award under award number MCB-0645794 (to J.P.N.), EEC-0813570 (to E.S.W.), MCB-0591170 (to E.S.W.), and by National Cancer Institute award number CA14195 (to G.M.). The Plant Sciences Institute at Iowa State University (to E.S.W.). J.P.N. is an investigator with the Howard Hughes Medical Institute. Portions of this research were conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory, on behalf of the US Department of Energy, Office of Basic Energy Sciences. We thank the staff at the Advanced Light Source for assistance with X-ray data collection. Any opinions, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Contributions. M.N.N. experimentally characterized the FAP genes in planta. M.E.B., R.N.P. and E.L. expressed, purified and crystallized proteins. L.L. designed genetics experiments and constructs. G.V.L. and F.P. performed fatty-acid binding analyses and solved the X-ray crystal structures. R.N.P. performed thermal-shift assays of fatty-acid binding. G.M. and E.L. performed phylogenetic and sequence analyses. J.P.N. designed the biochemical experiments; E.S.W. designed the bioinformatics and functional genomics experiments. The manuscript was written by R.N.P., G.M., E.L., M.N.N., J.P.N., G.M. and E.S.W.

Author Information. Coordinates and structure factors are deposited in Protein Data Bank under accession numbers 4DOO (AtFIAP1). 4DOK (AtCHIL). 4DL0 (AtFAP3) and 4DQO (AtFAP1). Reprints and information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.S.W. (mash@astate.edu) or J.P.N. (noel@salish.edu).
**METHODS**

Phylogenetic and sequence analyses. Homologues of AtCHI were gathered by PSI-BLAST from the National Center for Biotechnology Information NR database, followed by iterative profile hidden Markov model building and searches against public protein, expressed sequence tag and genome sequence databases. Gene models were built with GeneWise and refined manually. Protein sequences in each class were aligned with MUSCLE and manually edited, and classes were aligned by ClustalW and further edited based on the structural superposition of the available structures. An alignment of all available proteol and plant proteins and selected fungal and bacterial proteins is attached (Supplementary File 1). A consensus phylogenetic tree from 30,000 trees sampled from 1000 generations run of MrBayes (Supplementary Fig. 2 and Supplementary File 2). A maximum likelihood tree building using PhyML produced a highly similar branching pattern supporting the overall family classification, though several terminal branches were not conserved (not shown).

Expression and purification of *A. thaliana* CHI-fold proteins. The coding sequence for each *A. thaliana* protein’s CHI-fold domain was inserted between the NcoI and BamHI sites of the expression vector pHIS8, which, under the control of a T7 promoter, yields the target protein fused to a thrombin-cleavable amino-terminal His6-tag31. Proteins were heterologously expressed in E. coli BL21 (DE3) (Novagen) expression host. E. coli cells were grown at 37 °C in Terrific Broth to an attenuation (600 nm) of 1.5, induced with 1 mM isopropyl-1-thiogalactoside (IPTG), then grown for an additional 6 h at 22 °C. Bacterial cells were collected by centrifugation, re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0; 0.5 M NaCl; 20 mM imidazole; 1% Tween20; 10% glycerol; and 20 mM 2-mercaptoethanol) and lysed by sonication. Proteins were isolated from E. coli lysates by affinity chromatography with Ni- affinity-Ni (Ni2+-NTA) agarose, and eluted with lysis buffer containing 0.25 M imidazole. Partly purified proteins were treated with thrombin for cleavage of the His6-tag, then further purified by gel-exclusion chromatography with a Superdex 200 HR26/60 column (Pharmacia Biotechnology). Selenomethionine-substituted AtCHIL was generated from E. coli cultures grown in M9 minimal medium supplemented with an amino acid cocktail including L-selenomethionine (Sigma-Aldrich), and were otherwise handled as described above for unlabelled protein.

Analysis of fatty-acid binding by *A. thaliana* CHI-fold proteins. Bound ligands were extracted from protein samples of AtFAP1, AtFAP2 and AtFAP3 (1–2 mg in 20 mM ammonium bicarbonate) by addition of ethanolic to a final concentration of 80%. After incubation at –20 °C for 3 days, the treated samples were centrifuged (16,000g at 4 °C) to remove denatured protein and other particulates, and the supernatant was evaporated under vacuum. Residual material was dissolved in 200 μl 3-propanol, which was then passed through a 0.45 μm nylon filter in preparation for further characterization by HPLC–MS. The analysis of binding of exogenously provided α,ω-linolenic acid by AtFAP1 used the aforementioned amino-terminal His6-tagged form of the protein coupled to Ni-affinity resin. Immobilized protein (2 mg) was incubated in buffer (12.5 mM Tris-HCl, pH 8.0, 50 mM NaCl) with 0.5 mM α,ω-linolenic acid, and subsequently washed thoroughly with buffer including 5% ethanol. The AtFAP1 protein was released with 0.25 M acetic acid, then processed for extraction of bound ligand as described above. Control samples were treated identically, except that the AtFAP1 protein was omitted.

Extracts were analysed for fatty-acid content by HPLC-MS on an Agilent 1100 Series LC-MSD instrument with electrospray-ionization eluent introduction into an XCT ion trap mass spectrometer (Agilent). Chromatographic separations employed an Agilent Zorbax Eclipse XDB-C18 (4.6 mm × 150 mm, 5 mm particle size) reversed-phase column run at a flow rate of 0.5 ml min−1, and a linear gradient with initial and final mobile phases consisting of 95% water:5% acetonitrile (0.1% formic acid and 5% water:95% acetonitrile:0.1% formic acid, respectively. The identities of fatty-acid components were established by mass determination and comparison of chromatographic retention times with authentic fatty-acid standards (Sigma-Aldrich).

Thermal-shift binding assay. The shift in melting temperature for AtFAP proteins observed under increasing concentrations of fatty acids was measured using a Thermofluor-type assay similar to one previously published40. Protein melting temperatures were assayed using the LightCycler-480 System II (Roche), using the following program: 30s at 20 °C, ramp up to 85 °C at 0.06 °C s−1, 30s at 20 °C, excitation wavelength 483 nm, emission wavelength 588 nm. Using SYPRO Orange (Sigma), an environmentally sensitive dye that interacts with hydrophobic amino-acid residues, the melting of a protein can be observed as a fluorescence increase when the hydrophobic core residues of a protein fold are exposed to the environment in a denatured protein. SYPRO Orange dye can interact with the unfolded protein and cause an increase in fluorescence. The maximum or minimum of a first derivative curve of the fluorescence profile will indicate the melting temperature (Tm). The melting temperature of a protein in the absence of any additional compound provides a baseline melting temperature for the protein (Tm0). Compounds can be screened for binding interactions with the protein of interest. When a ligand binds a protein, a change in Gibbs free energy occurs and may cause a change in observed Tm for that protein. ΔTm can be measured for a collection of putative ligands and the strength of binding interactions can be ranked and compared for compounds with similar physicochemical properties, providing a relative measure of binding affinities.

A quantitative analysis of these relative binding affinities for AtFAPs versus the collection of fatty acids shown in Supplementary Table 1 (12 different fatty acids) could be obtained by a simple dose–response analysis of ΔTm with respect to fatty-acid concentration. Five micromoles of each protein and 10% SYPRO Orange (from 5000× stock solution) were mixed with from 10−3 to 10−11 M of the 12 different fatty acids, incubated at room temperature for 20 min, then assayed for protein melting temperature. Mean ΔTm (Tm – Tm0) was plotted against fatty-acid concentration for each FA/FAP combination using data from four replicate experiments. A dose–response: stimulation (three parameters) standard (Hill) slope nonlinear regression was fit to the data, using this model: \[ y = y_m + \frac{(Y_m - y_m)}{1 + (10^E C)^{-H}} \]. Calculated values of EC50 ± s.e.m. were reported as relative measures of binding constants, whereas maximum ΔTm was reported as a qualitative measure of free-energy change in the closed state.

Crystallization and structure determination of *A. thaliana* CHI-fold proteins. Crystals of AtCHI-fold proteins were grown by vapour diffusion at 4 °C from 1:1 mixtures of protein solution (10–15 mg ml−1) in 12.5 mM Tris-HCl, pH 7.5, 50 mM NaCl) and reservoir solution. The reservoir solution contained 28% polyethylene glycol (PEG) 8000, 0.3 M magnesium nitrate, 2 mM diithiothreitol (DTT) and 100 mM HEPES-Na+ (pH 7.0) for AtCHI; 28% PEG 8000, 0.2 M calcium acetate, 2 mM DTT and 100 mM TAPS-Na+ (pH 8.5) for AtCHIL; 19% PEG 3350, 0.3 M potassium chloride, 2 mM DTT and 100 mM TAPS-Na+ (pH 8.5) for AtFAP1; and 7–9% PEG 8000, 0.2 M calcium acetate, 2 mM DTT and 100 mM PIPES-Na+ (pH 6.5) for AtFAP5. Crystals grow typically occurred over a period of 2–10 days, and was sometimes expedited through seeding with finely crushed microcrystals. Crystals of AtFAP1 were soaked overnight in mother liquor supplemented with 1 mM K3PtCl6.

Crystals were flash frozen by immersion in liquid nitrogen after a brief incubation in a cryoprotection solution (consisting of reservoir solution supplemented with 17–20% ethylene glycol). X-ray diffraction data were collected from frozen crystals at the FIP beamline of the European Synchrotron Radiation Facility, beamlines 8.2.1 and 8.2.2 of the Advanced Light Source, Lawrence Berkeley National Laboratory, or beamlines 1-5 and 9-1 of the Stanford Synchrotron Radiation Laboratory. Diffraction intensities were measured on ADSC Quantum or MarResearch charge-coupled device detectors, and were indexed, integrated and scaled with MOSFLM and SCALA, HKL2000, or XDS and XSCALE programs.

Structural class solutions for the CHI-fold proteins were obtained through either single/multiple-wavelength anomalous dispersion or molecular-replacement analyses, as detailed in Supplementary Table 8. For single/multiple-wavelength anomalous dispersion analyses, the location of anomalous scatterers and initial phase estimates were determined with the program SOLVE, and preliminary structural models were automatically built with the program RESOLVE. For molecular-replacement analyses, Molrep was used, and where necessary, search coordinate-sets were constructed through homology modelling with Modeller. ARRP/WARP was used for automated rebuilding of initial structure models for AtCHIL and AtFAP3. Subsequent structural refinements used REFMAC4 and CNS. X-PLOR was used for graphical map inspection and manual rebuilding of atomic models. Programs from the CCP4 suite were used for all other crystallographic calculations. Structural superspositions were calculated with SSM. The identity of two ordered small-molecules bound in the active site of AtCHI was inferred to be nitrate because of the high concentration (0.3 M) of magnesium nitrate in the crystallization medium. From the X-ray crystallographic analysis, the shape and level of the electron density associated with these small molecules are also consistent with nitrate.

Plant growth. This study used WT *A. thaliana* Columbia (Col-0 and Ws), and mutant and transgenic lines derived from them. Seeds (approximately five per pot) were sown on soil, in flats containing either 21 or 32 pots; genotypes were distributed in a completely randomized design. After incubation for 3 days at 4 °C to break dormancy, flats were moved to a growth-chamber under long-day conditions (16:8 hours light-dark cycle) at 22 °C and 75% relative humidity; after a week pots were thinned such that each pot contained two plants. For additional plant morphological analyses and for seed collection, plants were grown in the greenhouse at 20–23 °C under continuous illumination (170 μmol m−2 s−1) in soil.
treated with granular Marathon (Olympic Horticultural Products) to protect against insect damage.

In the specified experiments, after 3 weeks of growth at 22 °C, plants were further grown at 15, 22 and 26 °C until senescence (control plants were grown at 22 °C for this period). Genotypes were distributed in a randomized design. Fatty-acid determinations were made after expanded rosette leaves harvested after 10 days of growth at 15, 22 and 26 °C. Growth and morphological phenotype of plant lines were observed every 2 days throughout development, and any differences between WT and mutants were recorded.

**Molecular constructs, transgenic A. thaliana lines and microscopy.** To evaluate subcellular location of the proteins, we generated promoter::target-GUS/GFP fusion constructs using promoters of different lengths for FAP1 and FAP3, and promoter::CDS-GUS/GFP fusion for FAP2. The primers used for amplification were as follows: FAP1 promoter 1 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATGGGATCAGCAAT-3'), FAP1 promoter 2 (5'-GGGGACCACTTTGTACAGGAAACGGCAGGCTTCATGGATGGGATCAGCAAT-3'), FAP2 promoter (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCAGACTGAACAACTCAGGATATAATACCTCC-3'), FAP3 promoter 1 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCAGACTGAACAACTCAGGATATAATACCTCC-3'), FAP3 promoter 2 (5'-GGGGACCACTTTGTACAGGAAACGGCAGGCTTCATGGATGGGATCAGCAAT-3').

To evaluate the spatial distribution of FAP expression in *plants*, we generated promoter::GUS/GFP fusion constructs using promoters of two different lengths. The primers used for amplification were as follows: FAP1 promoter 1 (5'-GGGGACGAAGTTTGTACAAAAAAGCAGGCTTCATGGATGGGATCAGCAAT-3'), FAP1 promoter 2 (5'-GGGGACCACTTTGTACAGGAAACGGCAGGCTTCATGGATGGGATCAGCAAT-3'), FAP2 promoter (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCAGACTGAACAACTCAGGATATAATACCTCC-3'), FAP3 promoter 1 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCAGACTGAACAACTCAGGATATAATACCTCC-3'), FAP3 promoter 2 (5'-GGGGACCACTTTGTACAGGAAACGGCAGGCTTCATGGATGGGATCAGCAAT-3'). Five independent transgenic lines in the T2 generation (five individual plants from each transgenic line at each stage of development) were evaluated for GFP location by confocal microscopy.

**Methods.** To evaluate the putative transgenic lines, four plants from 10 independent lines for each construct were analysed by PCR using the attB-adapted primers (5'-GGGGACGAAGTTTGTACAAAAAAGCAGGCTTCATGGATGGGATCAGCAAT-3', forward primer, 5'-GGGGACCACTTTGTACAGGAAACGGCAGGCTTCATGGATGGGATCAGCAAT-3', reverse primer). The primers used for amplification of the open reading frame (ORF) of FAP1 and 3 was introduced into pEarleyGate 103 and analyzed by PCR. Representative transgenic lines in the T2 generation were evaluated by immunoblotting and for GFP location by confocal microscopy.

**DNA and RNA isolation from plant tissue and RT–PCR.** Genomic DNA was isolated from leaves using a cetyl trimethylammonium bromide extraction protocol. Tissue was frozen and ground to a powder in liquid nitrogen. Fatty-acid analysis from plant tissue. Fatty acids were extracted using a barium hydroxide hydrolysis protocol. Total RNA was isolated from leaves by the method of Bradford.

**Fatty-acid analysis from plant tissue.** Fatty acids were extracted using a barium hydroxide hydrolysis protocol. Recently expanded rosette leaves were collected and placed in liquid nitrogen. Each replicate comprised leaves from a single plant. Approximately 0.1 g fresh weight of leaves or 5 mg of seeds was used for each
extraction. Tissues were frozen in liquid nitrogen upon collection, and placed in a pre-cooled tissue homogenizer. Twenty microlitres of internal standard (nonadecanoic acid, C19:0, 2 mg ml⁻¹ dissolved in chloroform) and 1 ml of barium hydroxide were added, and the mixture was further homogenized according to the procedure in plant metabolomics.org (http://www.plantmetabolomics.org). The homogenate was transferred to a glass tube and 550 μl of 1,4-dioxane (Aldrich catalogue number 123-91-1) added. The tube was tightly capped and incubated for 24 h at 110 °C. The solution was then acidified with six drops of 0.6 M HCl hexane (2 × 3 ml) was added, and the mixture was centrifuged. The hexane layer was transferred to a new tube, dried under nitrogen gas, methylated with 2 ml of HCl-methanol for 1 h at 80 °C, followed by two extractions in hexane (2 ml each). Samples were dried under nitrogen gas, acetylated with 1 ml of acetonitrile and 70 μl of bis-(trimethylsilyl)trifluoroacetamide for 20 min at 60 °C. The solvent was evaporated under nitrogen gas and 200 μl (for leaf samples) or 1.5 ml (for seed samples) of chloroform was added.

Chloroform-dissolved material was subjected to fatty-acid methyl ester analysis using a Model 6890 series gas chromatograph (Agilent) equipped with a Mass Detector 5973 (Agilent) and an HP-1 silica capillary column (30 m × 0.32 mm, inner diameter), using helium as a carrier gas. The injector and detector temperature was 300 °C. The initial oven temperature was 100 °C for 2 min, increased to 240 °C at a rate of 5 °C min⁻¹, and maintained at 240 °C for 5 min. The total running time was 40 min. AMDIS software version 2.65 was used for peak integration; peaks were identified through the Agilent NIST05 mass spectra libraries (http://www.nist.gov/srd/nist1a.htm). Fatty-acid methyl esters were quantified in each sample in accordance with the amount of nonadecanoic acid internal standard added.

Statistical and bioinformatics analysis. All experiments were conducted a minimum of three times. For each experiment, plants were collected and analysed in randomized block design. Data are presented as mean ± s.d. We compared two sets of independent samples using Student’s t-test (two-tailed) with assumption of equal variances and P < 0.05 was considered significant.

Pearson correlation coefficients of A. thaliana transcript accumulation across changes in plant genotype, environment and development were calculated using MetaOmGraph (http://www.metnetdb.org) and a data set consisting of 951 chips from 72 public microarray experiments. 23–31

23. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. Genome Res. 14, 988–995 (2004).
24. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).
25. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948 (2007).
26. Hueslenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755 (2001).
27. Guindon, S., Lefhec, F., Duroux, P. & Gascuel, O. PHYML Online – a web server for fast maximum-likelihood-based phylogenetic inference. Nucleic Acids Res. 33, W557–W559 (2005).
28. Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A. & Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry 39, 890–902 (2000).
29. Doublie, S. Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. 276, 523–530 (1997).
30. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols 9, 2212–2221 (2007).
31. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EMBCB Newsletter on Protein Crystallography No. 26 (1992).
32. Evans, P. R. Scaling and assessment of data quality. Acta Crystallogr. D 62, 72–82 (2005).
33. Otwonowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
34. Kabasch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 26, 795–800 (1993).
35. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. Acta Crystallogr. D 55, 849–861 (1999).
36. Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr. D 59, 38–44 (2003).
37. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. J. Appl. Cryst. 30, 1022–1025 (1997).
38. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993).
39. Perrakas, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. Nature Struct. Biol. 6, 458–463 (1999).
40. Brunger, A. T. & Warren, G. L. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D 54, 905–921 (1998).
41. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D 60, 2256–2266 (2004).
42. Ma, H., Stage, C. & Roach, D. Invitrogen. Gateway technology: a universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Catalog nos. 12535-019 and 12535-027. http://wolfsen.hju.ac.il/expRESSION/gatewayman.pdf (2003).
43. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743 (1998).
44. Lohar, D. P., Shuller, K., Buzas, D. M., Gresshoff, P. M. & Stiller, J. Transformation of Lotus japonicum using the herbicide resistance bar gene as a selectable marker. J. Exp. Bot. 52, 1691–1702 (2001).
45. Li, L., Iarslan, H., James, M. G., Myers, A. M. & Wurtele, E. S. Genome wide co-expression among the starch debranching enzyme genes AtISA1, AtISA2, and AtISA3 in Arabidopsis thaliana. J. Exp. Bot. 58, 3323–3342 (2007).
46. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. 45, 615–629 (2006).
47. Jefferson, R. A. Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Rep. 5, 387–405 (1987).
48. Doyle, J. J. & Doyle, J. L. Isolation of plant DNA from fresh tissue. Focus 12, 13–15 (1990).
49. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Rev. Biochem. 72, 248–254 (1976).
50. Markham, J. E., Li, J., Caoon, E. B. & Jaworski, J. G. Separation and identification of major plant sphingolipid classes from leaves. J. Biol. Chem. 281, 22684–22694 (2006).
51. Moon, S. & Nikolau, B. J. Plantmetabolomics Platform2: fatty acids. http://www.plantmetabolomics.org (2009).
52. Perera, M. A. D. N., Dietrich, C. R., Meeley, R., Schnable, P. S. & Nikolau, B. J. in Advanced Research on Plant Lipids (eds Murata, M. et al.) 225–228 (Kluwer Academic, 2002).
53. Fatland, B. L., Nikolau, B. J. & Wurtele, E. S. Reverse genetic characterization of cytosolic acetyl-CoA generation by ATP-citrate lyase in Arabidopsis. Plant Cell 17, 182–203 (2005).
54. Davies, N. A. The New Automated Mass Spectrometry Deconvolution and Identification System (AMIDS) (IASIS, 1998).