Plant Biotechnology 28, 379–385 (2011)
DOI: 10.5511/plantbiotechnology.11.0704b

Original Paper

Metabolomic characterization of the possible involvement of a Cytochrome P450, CYP81F4, in the biosynthesis of indolic glucosinolate in *Arabidopsis*

Kosuke Kai[1,a], Hiroki Takahashi[2], Hirohisa Saga[1], Takumi Ogawa[1], Shigehiko Kanaya[2], Daisaku Ohta[1,*]

[1] Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan;
[2] Graduate School of Information Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
* E-mail: ohtad2g30490@bioinfo.osakafu-u.ac.jp Tel & Fax: +81-72-254-9461

Received June 1, 2011; accepted July 4, 2011(Edited by D. Shibata)

Abstract We studied the biosynthetic role of a cytochrome P450, CYP81F4, in *Arabidopsis*. The *CYP81F4* gene was found within a gene expression network containing the transcription factor gene *MYB34/ATR1* involved in tryptophan metabolisms. Root metabolic profiles in null *cyp81f4* mutant lines were analyzed using an in-house metabolomics scheme based on Fourier transform ion cyclotron resonance mass spectrometry. The *cyp81f4* mutant plants exhibited a build-up of indole-3-yl-methyl glucosinolate with concomitant loss of 1-methoxy-indole-3-yl-methyl glucosinolate. A pathway prediction supported by a detailed glucosinolate analysis indicated that CYP81F4, together with an unidentified methyltransferase, is involved in the formation of the methoxy group in the indole ring to yield 1-methoxy-indole-3-yl-methyl glucosinolate in *Arabidopsis* roots. Current results suggest that 1-methoxy-indole-3-yl-methyl glucosinolate is produced by the activity of CYP81F4 in *Arabidopsis* roots.

Key words: Cytochrome P450, gene coexpression, indolic glucosinolate, metabolomics, van Krevelen diagram.

Plants produce and accumulate a wide variety of secondary metabolites. In such diverse secondary metabolisms, precursor structures are modified through biochemical steps driven by different enzyme classes. These enzyme classes are categorized according to their catalytic principles conferred by specific amino acid sequence motifs, and individual enzymes belonging to each class catalyze their reactions on the basis of the same catalytic principle but with different substrates. Such enzyme classes are ubiquitously found across kingdoms. For example, the *Arabidopsis* genome contains at least 24 flavin-containing monooxygenase genes, 272 cytochrome P450 genes, and more than 20 *S*-adenosylmethionine-dependent methyltransferase genes. Such tentative gene annotations should be confirmed in terms of actual biological functions.

Metabolomics focuses on full biochemical events as an integral outcome of genome-wide gene expression and is thus expected to play a crucial role in gene annotation and curation pipelines (Fukushima et al. 2009; Schauer and Fernie 2006). Current metabolomics studies rely on the systematic comparison of detectable analytes with reference standard compounds on a variety of analytical platforms. These metabolomics studies have demonstrated their robustness in metabolic engineering, process engineering, biomarker finding, and the functional characterization of novel genes. It is of note, however, that reference standards are usually not available to clarify unknown compounds on the basis of analytical chemistry-based evidence. This is also the case for the functional characterization of recombinant enzymes acting on unidentified substrates. Thus, unique analytical methodologies independent of reference standards/substrates are necessary for the characterization of metabolic genes of unknown functions.

In this study, we applied a pathway simulation method on van Krevelen diagrams using the metabolome information from an in-house metabolomics scheme based on Fourier transform ion cyclotron resonance mass

Abbreviations: FT-ICR/MS, Fourier transform ion cyclotron resonance mass spectrometry; 4HOI3M, 4-hydroxy-indole-3-yl-methyl glucosinolate; iGS, indolic glucosinolate; LC/LIT-TOFMS, liquid chromatography/linear ion trap time-of-flight mass spectrometry; MeJA, methyl jasmonate; 1MOI3M, 1-methoxy-indole-3-yl-methyl glucosinolate; 4MOI3M, 4-methoxy-indole-3-yl-methyl glucosinolate; I3M, indole-3-yl-methyl glucosinolate; PCA, principal component analysis; P450, cytochrome P450.

[a] Present address: SYSMEX CORPORATION, Kobe, Hyogo 651-0073, Japan

This article can be found at http://www.jspcmb.jp/
Published online September 10, 2011

spectrometry (FT-ICR/MS). Van Krevelen diagrams have been used to evaluate complex organic materials, such as natural organic matter, coal, and petroleum crude oil, through the use of FT-ICR/MS, which allows for the resolving of all constituents present in samples to determine their possible elemental compositions and molecular formulas (Kim et al. 2003; Wu et al. 2004). Metabolites are plotted on two or three axes according to their hydrogen-to-carbon (H/C), oxygen-to-carbon (O/C), and/or nitrogen-to-carbon (N/C) atomic ratios (Figure 1). Van Krevelen diagrams are thus suited for the holistic comparison of molecular formula modifications by enzyme activities. For example, the reactions of monooxygenase, methyltransferase, and desaturase give rise to distinct changes in the elemental compositions of compounds with +O ($\Delta m/z$=15.99491), +CH_2 ($\Delta m/z$=14.01565), and −H_2 ($\Delta m/z$=2.01565) (Ohta et al. 2010). The metabolic processes able to yield known metabolites could therefore be found through an exhaustive search of possible modifications of elemental compositions (Figure 1).

Here, we characterize unidentified metabolic steps in the biosynthesis of indolic glucosinolates (iGS) in *Arabidopsis thaliana*. Glucosinolates are the generic name of a secondary metabolite class with large structural diversity which are mainly produced by plants belonging to the order Brassicales (Brown et al. 2003; Grubb and Abel 2006; Kilbensteuin et al. 2001). In *Arabidopsis*, glucosinolates are divided into two major classes, the previously mentioned iGS along with aliphatic glucosinolates (aGS), which are derived from Trp and Met, respectively. While the core glucosinolate biosynthetic steps have been clearly illustrated in terms of aldoxime formation, glucone formation, and side chain modification (Burow et al. 2010; Halkier and Gershenzon 2006), the individual enzymes/genes involved in biosynthesis *en route* to the diverse glucosinolate structures have not been fully understood. Our metabolic profiling studies demonstrate that *CYP81F4*, together with an unidentified methyltransferase, is involved in the formation of the methoxy group on the indole ring to yield 1-methoxy-indole-3-yl-methyl glucosinolate in the roots of *Arabidopsis*. This is in contrast to the involvement of CYP81F2 in the biosynthesis of 4-methoxy-indole-3-yl-methyl glucosinolate primarily accumulating in the shoots, demonstrating distinct iGS biosynthetic routes operating in different plant organs.

Materials and methods

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia (Col-0) (Lehle Seeds, Round Rock, TX, USA) seedlings were grown at 22°C under

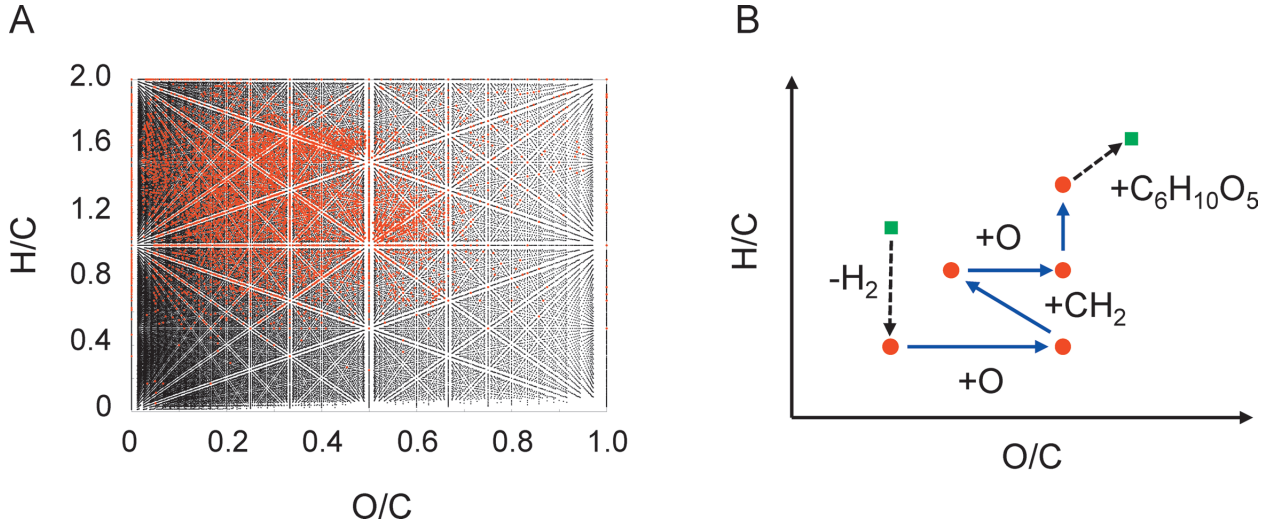

Figure 1. Metabolic activities on the van Krevelen diagram. (A) Metabolites on the van Krevelen diagram. A list of 1,676,252 molecular formulas consisting of C, H, N, and O were calculated to give a molecular weight below 1,000 in light of the combinations of atomic valences. The actual metabolites stored in the KNApSAcK database (50,054 metabolites on March 24, 2010) were also listed to obtain individual atomic ratios. The atomic ratios of the virtual molecules (grey dots) and the actual metabolites (red dots) were illustrated on the 2-D plot. Hydrogen sulfide (H_2S), ammonia (NH_3), phosphoric acid (H_3PO_4), sulfuric acid (H_2SO_4), pyrophosphoric acid ($H_4P_2O_7$), sulfur trioxide (SO_3) in the KNApSAcK database were not included in the plot. (B) Metabolic activities on the van Krevelen diagram. A putative biosynthetic pathway comprised of different enzymes was explained by the changes in the molecular formulas. The enzyme activities such as monooxygenase (+O=15.99491), desaturase (−H_2=2.01565), methyltransferase (+CH_2=14.01565), acetyltransferase (+CH_2CHO=43.018395), and glycosyltransferase (+$C_6H_{10}O_5$=163.0606) give rise to specific changes in the molecular formulas of their substrate molecules. Such modified molecular formulas can be linked to constitute biosynthetic pathways. In ultra-high resolution MS analyses, molecular formulas can be predicted even for ions of unknown identities. Thus, unknown metabolites could be linked to known metabolic processes in terms of molecular formula modification processes. Red circles and green squares represent known and unidentified metabolites, respectively. Blue arrows illustrate established metabolic reactions connected to unidentified enzyme reactions (indicated by dashed arrows).

continuous light (100 μE m^{-2} s^{-1}) as described previously (Morikawa et al. 2006). For glucosinolate analysis and RNA isolation, *Arabidopsis* seedlings were grown under sterile conditions on germination medium (GM) [1×MS salt, 1% (w/v) sucrose, pH 5.7] plate containing 0.8% (w/v) agar. The seedlings were then transferred to a GM liquid culture and shaken at 120 rpm at 22°C under continuous light (100 μE m^{-2} s^{-1}).

RT-PCR analyses

Plant T-DNA insertion lines with regard to *CYP81F4* (At4g37410: SALK_024438, SM_3_20372) and *ATR1/MYB34* (At5g60890: WiscDsLox424F3) were obtained from the Arabidopsis Biological Resource Center (http://signal.salk.edu/cgi-bin/tdnaexpress) (Alonso et al. 2003). Gene-specific primer sets together with a T-DNA border primer LBa1 (5′-TGGTTCACGTAGTGGGCCATCG-3′) and a transposon-specific primer Spm32 (5′-TACGAATAAGAGCGTCCATTTTAGAGTGA-3') were used for the mutant characterization. Gene specific primer sets were prepared for *CYP81F4* (81F4-Fw 5′-GGATCCCATATGTTTAACTATGTGATTAT-3′; 81F4-Rv 5′-TCTAGACTAAACTTTCGTGTAGGCCG-3′) and *ATR1* (ATR1-Fw, 5′-ATGGTGAGGACACCATGTTGCAAA-3′; ATR1-Rv 5′-TCAGACAAAGACTCCAACCATATTGTC-3′). The same gene specific primer sets for *CYP81F4* (81F4-Fw and 81F4-Rv) and *ATR1* (ATR1-Fw and ATR1-Rv) were used for the amplification of the corresponding gene transcripts. The *actin* gene (Morikawa et al. 2006), used as the internal control, was amplified under the same PCR conditions using a primer pair of Act-F (5′-ATGGCTGATGGTGAAGACATTC-3′) and Act-R (5′-TCAGAAGCACTTCCTGTGAAC-3′).

Metabolomics

Samples were analyzed by direct infusion using an IonSpec Explorer FT-ICR/MS with a 7-tesla magnet (IonSpec Inc., Lake Forest, CA, USA) (Oikawa et al. 2006). Under our analytical conditions, the resolving power ($m/\Delta m_{50\%}$) was approximately 100,000 (for m/z=400 with 1 s of ion accumulation time), and the mass accuracy was below 2 ppm. In the negative mode analyses, disaccharide (*m/z* 341.10894; $[M-H]^-$), 8-methoxythiooctyl glucosinolate (*m/z* 476.10882; $[M-H]^-$), flavonoid diglycoside (*m/z* 609.14611; $[M-H]^-$), and flavonoid triglycoside (*m/z* 755.20402; $[M-H]^-$) were used as endogenous mass calibrants for correcting analytical fluctuations. We prepared samples in triplicate (as a minimum) for each analysis with 10 successive spectral scans.

Metabolome information

The metabolome data were subjected to a KNApSAcK database search (Oikawa et al. 2006; Shinbo et al. 2006). In the assignment of elemental compositions to *m/z* values, we set the *m/z* error within a 2 ppm tolerance limit. The initial lists of elemental compositions were further narrowed down by incorporating the Seven Golden Rules proposed by Kind and Fiehn (2007). A total of 18,078 possible candidates of elemental compositions were initially listed from 353 ions detected in the extracts of wild-type and *cyp81f4* lines. After applying the restriction rule, we obtained 8,156 elemental compositions that were used for the calculation of H/C and O/C ratios for the analysis on a van Krevelen diagram. To search for potential substrates on this diagram, we prepared a list of substrate and product pairs stored in the KNApSAcK data base which contains different enzyme classes involved in the 6,982 actual enzyme reactions (Supplementary Table 1).

Indolic glucosinolate analyses

Arabidopsis seedlings were transferred to a GM liquid culture and shaken at 120 rpm at 22°C under continuous light (100 μE m^{-2} s^{-1}). *Arabidopsis* roots were extracted using hot methanol/water (70 : 30, v/v). indole-3-yl-methyl glucosinolate (I3M), 1-methoxy-indole-3-yl-methyl glucosinolate (1MOI3M), and 4-methoxy-indole-3-yl-methyl glucosinolate (4MOI3M) were analyzed using a LC/LIT-TOFMS (Nano Frontier LD, Hitachi High-Technologies Corp., Tokyo, Japan). Metabolites were separated using a Cadenza CD-C18 column (2×150 mm, Imtakt Corp., Kyoto, Japan) in a linear gradient elution using solvent A (H_2O containing 0.1% [v/v] formic acid) and solvent B (acetonitrile containing 0.1% [v/v] formic acid). The elution was kept from 0 min to 5 min at an A : B ratio of 95 : 5 and then linearly shifted to an A : B ratio of 45 : 55 in 30 min. On the chromatogram, I3M, 4MOI3M and 1MOI3M were distinguished from each other based on their elution times of 13.3 min, 17.0 min and 19.3 min, respectively. 4MOI3M and 1MOI3M with the same *m/z* values yielded unique patterns of product ion formations upon MS/MS analyses, as previously reported (Cataldi et al. 2007; Rochfort et al. 2008) (Supplementary Figure 1). The endogenous amounts of I3M, 4MOI3M, and 1MOI3M were compared according to their ion peak areas detected in the ion-selective mode of the mass chromatograms.

Results and discussion

Indolic glucosinolates in Arabidopsis

Three major iGSs accumulate in *Arabidopsis*. Indole-3-yl-methyl glucosinolate (I3M) is converted to 4-methoxy-indole-3-yl-methyl glucosinolate (4MOI3M) and 1-methoxy-indole-3-yl-methyl glucosinolate (1MOI3M) (Figure 2). Pfalz et al. (2009) have determined, through metabolic quantitative trait loci analysis and recombinant enzyme characterization, that *CYP81F2* is in charge of the conversion of I3M to 4MOI3M via 4-hydroxy-indole-3-yl-methyl glucosinolate (4HOI3M). 4MOI3M is involved in the defense systems against both herbivory feeding and infection by microbial pathogens (Bednarek et al. 2009; Clay et al. 2009). On the other hand, the enzymes in charge of the hydroxylation and subsequent methyltransfer reactions to yield 1MOI3M remain unknown, while genes belonging to the CYP81F subfamily are strong candidates for catalyzing iGS biosynthetic steps yet to be clarified (Figure 2).

CYP81F family genes in Arabidopsis

The P450 gene subfamily CYP81F in *Arabidopsis* is composed of four member genes: *CYP81F1* (At4g37430), *CYP81F2* (At5g57220), *CYP81F3* (At4g37400), and

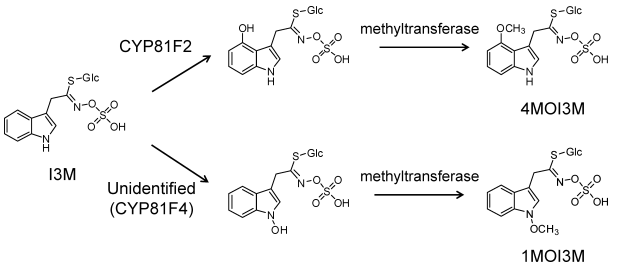

Figure 2. Proposed biosynthesis pathway of methoxyindole glucosinolates in *Arabidopsis*. Indolic glucosinolate biosynthetic routes forming the methoxy groups via a couple of hydroxylase and methyltransferase reactions. CYP81F2 is in charge of the hydroxylation of I3M at the 4 position of the indole ring (Pfalz et al. 2009). *CYP81F4* is suggested to be involved in the hydroxylation of the 1 position of the indole ring.

CYP81F4 (At4g37410). The primary structure of CYP81F1 is 68%, 66%, and 63% identical to those of CYP81F2, CYP81F3, and CYP81F4, respectively. These P450 genes exhibited distinct expression manners (Arabidopsis eFP browser, http://www.bar.utoronto.ca/efp/development/; Genevestigator, https://www.genevestigator.com/gv/index.jsp), suggesting possible metabolic roles in the specified organs. Under our experimental conditions, *CYP81F4* was characterized by its higher expression in the roots (Figure 3A), while plant organs other than roots might also accumulate *CYP81F4* transcript at lower levels. Gene co-regulation analyses (Obayashi et al. 2007; Ogata et al. 2009) indicated that *CYP81F4* is present within a gene coexpression module containing *ATR1/MYB34*, an MYB-like transcription factor. Gene coexpression analyses provide a wealth of information for studying the metabolic functions of both well-characterized proteins and genes that have unidentified biological roles (Hirai 2009). It has been demonstrated that coexpressed glucosinolate biosynthetic genes are in fact under the control of specific transcription factors (Gigolashvili et al. 2007; Hirai 2009). *ATR1/MYB34* is known to modulate the expression of the *ASA1* gene that encodes the alpha subunit of anthranilate synthase, which catalyzes the rate-limiting step of tryptophan synthesis (Bender and Fink 1998; Celenza et al. 2005; Grubb and Abel 2006). The loss-of-function allele suggests that *ATR1* also functions as a control point to regulate iGS homeostasis (Bender and Fink 1998; Grubb and Abel 2006). Coexpressed genes include *PYK10* (At3g09260), *NAI2* (At3g15950), and jacalin lectin family protein encoding genes (At1g76790, At2g39310, At3g16460, At3g16450), involved in the functioning of the ER body and glucosinolate degradation (Nagano et al. 2008; Wittstocka and Burow 2010). These observations prompted us to clarify a possible metabolic role of *CYP81F4* in the iGS biosynthetic pathway derived from Trp.

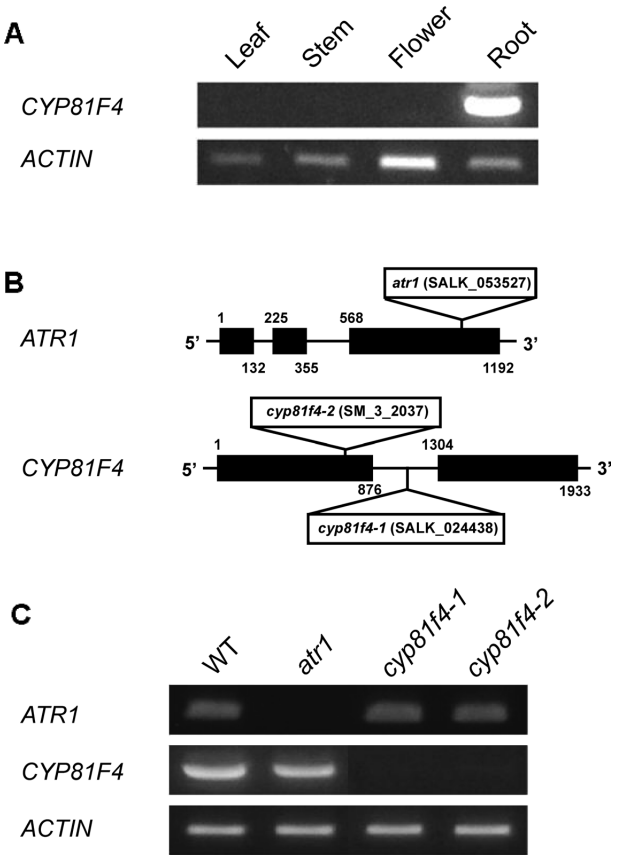

Figure 3. T-DNA and transposon insertion mutants of *ATR1* and *CYP81F4*. (A) Expression levels of *CYP81F4* in *Arabidopsis* plants. (B) The positions of the T-DNA and transposon insertion in the *atr1* (WiscDsLox424F3), the *cyp81f4-1* (SALK_024438), and the *cyp81f4-2* (SM_3_20372) lines. (C) Expression levels of the *ATR1* and *CYP81F4* genes in the mutant lines.

Metabolomics with CYP81F4 T-DNA insertion mutants

For metabolome analyses, we selected *Arabidopsis* mutant lines carrying gene disruption events within *ATR1*/*MYB34* and *CYP81F4* (Figure 3B). In this study, we used two T-DNA insertion lines (At4g37410: SALK_024438, SM_3_20372). The *CYP81F4* mutant alleles overlap a gene of unknown function At4g37409 on the database. The At4g37409 (a coding sequence for 131 amino acids) annotation is likely to be an artifactual annotation. Putative EST clones assigned to At4g37409 contained unusually long 3′-untranslated regions that could be ascribed to the *CYP81F4* transcript in the reverse direction. In these mutant lines, transcripts of the target genes disappeared (Figure 3C). The *CYP81F4* transcript level was apparently unaffected by the *ATR1* disruption, though it was expected, based on gene coexpression analyses. We do not exclude the possible involvement of *ATR1*/*MYB34* in the transcriptional regulation of CYP81F4, but our results indicated that *ATR1*/*MYB34* was not essential for the expression of *CYP81F4*. As shown in Figure 3A, *CYP81F4* was highly

expressed in the roots, and we therefore focused on the root metabolomes of the *cyp81f4* lines (*cyp81f4-1* and *cyp81f4-2*). The metabolome data were composed of mass-to-charge ratio (*m/z*) values with ion intensities using an FT-ICR/MS metabolomics scheme (Oikawa et al. 2006). Among a variety of MS principles, FT-ICR/MS offers the best performance in terms of the mass resolution as it allows for precise molecular formula prediction (Guan et al. 1996; Marshall et al. 1998), a key step for metabolite identification. Principal component analysis (PCA) demonstrated that the root metabolomes formed in the *cyp81f4* lines clearly differed from those in wild-type plants (Figure 4A), with the PC1 (59.2%) and PC2 (16.7%) separation shown in Supplementary Table 2. Several metabolite candidates were assigned to some of the analytes contributing to the *cyp81f4*-type metabolome, including an analyte A (*m/z* 447.0541, $[M-H]^-$; $C_{16}H_{20}N_2O_9S_2$; H/C=1.25 and O/C=0.56), of which ion intensity levels were 7- and 9.6-fold higher in the *cyp81f4-1* and *cyp81f4-2* lines, respectively. While it was suggested that the analyte A was, in fact, I3M as a result of a KNApSAcK database search (Shinbo et al. 2006), a number of unidentified metabolites are listed (Supplementary Table 2). On the other hand, we cannot rule out the possibility that these unknown metabolites was also related to CYP81F4-dependent reactions. In general, one of the biggest challenges in metabolomics is due to the difficulties in assigning chemical identification to every analyte, and supportive information is required for narrowing down possible substrate candidates, thereby generating working hypotheses for further studies including recombinant enzyme characterization together with genetic complementation.

Metabolic pathway prediction

To clarify the metabolic origin of the analyte A, we studied a pathway simulation on a van Krevelen diagram using the metabolome information from the FT-ICR/MS (Figure 4B). Our target gene (*CYP81F4*) encodes a putative P450 monooxygenase. The relationship between the substrate and the product of CYP81F4 can be explained in terms of the elemental composition change of one oxygen atom on the van Krevelen diagram. Furthermore, the CYP81F2 enzyme activity is expected to be involved in the modification of iGS. Metabolite candidates related to the analyte A were specified on a van Krevelen diagram (Figure 4C). An analyte B (*m/z* 477.0649, $[M-H]^-$; $C_{17}H_{22}N_2O_{10}S_2$; H/C=1.29 and O/C=0.59) was explained by the addition of a methoxy group (CH_2O) to the analyte A, which could be ascribed to a monooxygenase reaction coupled with a methyltransferase reaction. The level of the analyte B decreased in the mutant lines (0.43- and 0.63-fold in the *cyp81f4-1* and *cyp81f4-2* lines, respectively), and it was 37th (score: −0.76) in the PC1 contribution list (not

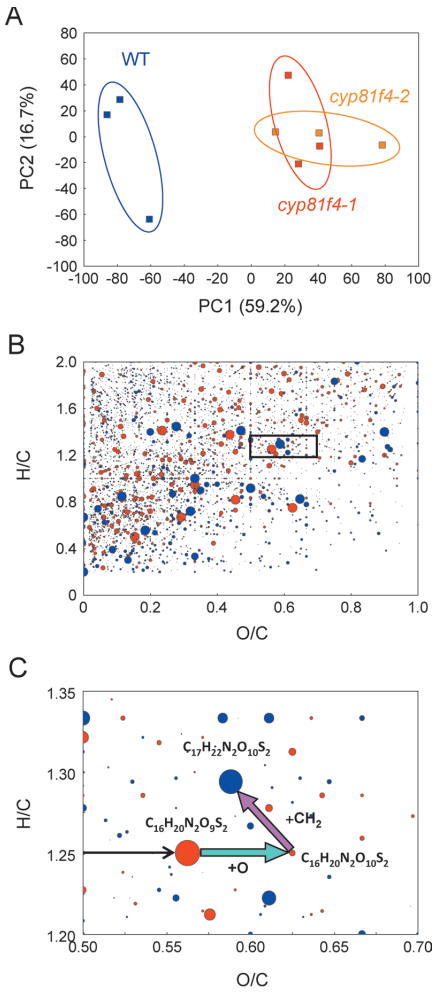

Figure 4. Identification of CYP81F4 functions on a van Krevelen diagram. Root extracts (triplicates) were subjected to FT-ICR/MS analyses (negative mode). The ultrahigh-resolution MS data were converted to metabolome information using the Dr.DMASS program (Oikawa et al. 2006). (A) Metabolome differentiation by PCA. PC1 (59.2%) contributed to separating the metabolomes formed in wild-type plants (blue) and the *cyp81f4* mutants (red and yellow). Contributions of the metabolites (PCA loading data) are shown in Supplementary Table 2. (B) Metabolome analyses on the van Krevelen diagram. Allowing 2 ppm *m/z* errors, molecular formulas (8,156 possibilities) were predicted for the analytes (353 independent *m/z* values) detected using the root extracts from wild-type plants and *cyp81f4-1* plants. The ratios of O/C and H/C were calculated from the predicted molecular formulas and plotted on a van Krevelen diagram. Ion intensities of analytes were compared between wild-type plants and *cyp81f4-1* plants. Stronger ion intensities in wild-type and *cyp81f4-1* were shown in blue and red circles, respectively. The sizes of the circles reflect the intensity values. (C) A group of metabolites indicated by the black rectangle in Figure 4B were magnified. The accumulation level of the analyte A (shown by the red circle; $C_{16}H_{20}N_2O_9S_2$; H/C=1.25 and O/C=0.56) was 7-fold higher in the *cyp81f4-1* plants, and the accumulation of the analyte B (indicated by the blue circle; $C_{17}H_{22}N_2O_{10}S_2$; H/C=1.29 and O/C=0.59) was 0.43 in *cyp81f4-1* plants compared with wild-type plants. The difference in the elemental compositions between these analytes was CH_2O, corresponding to a methoxy group formation. The analyte C ($C_{16}H_{20}N_2O_{10}S_2$; H/C=1.25 and O/C=0.63), a hydroxylated form of the analyte A, was detected only in the *cyp81f4-1* plants (indicated by the small red circle). Putative reactions of monooxygenase and methyltransferase are indicated by blue and pink arrows, respectively.

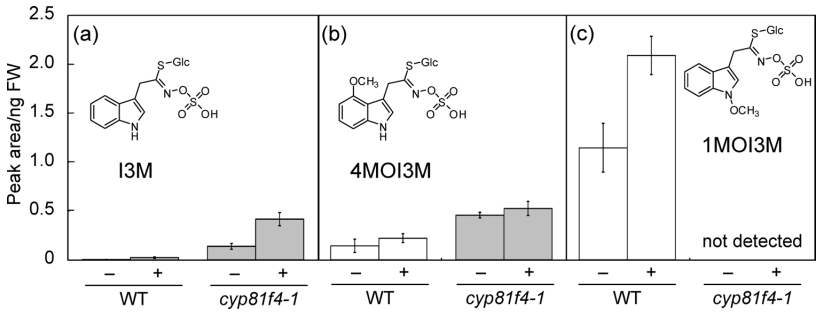

Figure 5. Glucosinolate levels in the roots of wild-type and *cyp81f4-1* plants. Three days after germination, wild-type and the *cyp81f4-1* seedlings were transferred to a GM liquid culture. Ten days after the transfer, seedlings were treated with 250 μM MeJA for 12 h. The roots were extracted with hot-methanol and analyzed using an LC/LIT-TOFMS in the negative mode. The accumulation levels of I3M, 4MOI3M (retention time = 17 min), and 1MOI3M (retention time = 19.3 min) were estimated from the peak areas from the ions of *m/z* 447 (I3M) and *m/z* 477 (4MOI3M and 1MOI3M). 4MOI3M and 1MOI3M were distinguished according to the MS/MS fragmentation patterns (data not shown). The 250 mM MeJA treatment (+) and mock treatment with 0.25% EtOH (–) were carried out in triplicate experiments. Bars indicate standard errors (n = 3).

shown) to form the *cyp81f4*-metabolome. In addition, we were able to detect the presence of an analyte C (*m/z* 463.0487, $[M-H]^-$; $C_{16}H_{20}N_2O_{10}S_2$; H/C = 1.25 and O/C = 0.63) corresponding to the possible product of the CYP81F4 hydroxylase reaction (a hydroxylated form of I3M), supporting the possibility that the analyte B is synthesized via a methoxy group formation reaction (Figure 4C). A KNApSAcK database search indicated that the analyte B could be either 4MOI3M or 1-methoxy-indole-3-yl-methyl glucosinolate (1MOI3M).

Targeted analyses of indolic glucosinolates

LC/LIT-TOFMS analyses distinguished 1MOI3M from 4MOI3M in terms of the different elution times and MS/MS fragmentation patterns (Cataldi et al. 2007; Rochfort et al. 2008) (Supplementary Figure 1). Figure 5 compares the accumulation levels of I3M, 1MOI3M, and 4MOI3M between the roots of wild-type plants and *cyp81f4-1* plants. In wild-type plant roots, the amounts of I3M and 4MOI3M were significantly lower than that of 1MOI3M. The 1MOI3M level in wild-type roots was elevated approximately two-fold by 250 μM methyl jasmonate (MeJA) treatment. This treatment also stimulated the accumulation of I3M, while the 4MOI3M level was not significantly altered. In the roots of the *cyp81f4-1* plants, the 1MOI3M accumulation was not detectable, indicating that *CYP81F4* was involved in the formation of 1MOI3M. The transcriptomics data at AtGenExpress (http://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp) indicate that MeJA treatments induce the expression of *CYP81F4*, supporting the idea that *CYP81F4* contributed to the elevated 1MOI3M levels in the MeJA treated plants (Figure 5). It has been reported that 1MOI3M is the major IG accumulating in the roots (Petersen et al. 2002). In the *cyp81f4-1* line, 1MOI3M was not detectable, but both 4MOI3M and I3M levels were slightly elevated, suggesting that CYP81F4 was not responsible for 4MOI3M biosynthesis from I3M, but was instead involved in the 1MOI3M biosynthesis in the roots (Figure 2). In other words, other *CYP81F* family genes might not be involved in the biosynthesis of 1MOI3M. The catalytic functions of CYP81F1 and CYP81F3 are not clear, though the metabolic profiling results suggest that these enzymes are involved in the production of 4MOI3M in the shoots.

Involvement of CYP81F4 in the new step of indolic glucosinolate biosynthesis

Our results suggest that CYP81F4 is an essential hydroxylase in the roots of *Arabidopsis* that couples to an unidentified methyltransferase to produce 1MOI3M through methoxy group formation at the indole ring (Figure 2). Further characterization of *CYP81F4* will be accomplished through recombinant enzyme characterization and genetic complementation of the gene disruption events. Also, it remains to be determined how 1MOI3M and its degradation products contribute to the defense mechanisms or possible interactions with other organisms, including insects and microorganisms (Bednarek et al. 2009; Clay et al. 2009). Currently, we are performing metabolomics approaches directed to the root metabolisms possibly involved in the interaction of plants and other organisms are currently being performed.

Acknowledgements

This work was supported by New Energy and Industrial Technology Development Organization (as part of the project called Development of Fundamental Technologies for Controlling the Process of Material Production of Plants). This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (grant no. 21580418).

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. (2003)

Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301: 653–657

Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, et al. (2009) A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323: 101–106

Bender J, Fink GR (1998) A Myb homologue, ATR1, activates tryptophan gene expression in Arabidopsis. *Proc Natl Acad Sci USA* 95: 5655–5660

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Photochemistry* 62: 471–481

Burow M, Halkier BA, Kliebenstein DJ (2010) Regulatory networks of glucosinolates shape *Arabidopsis thaliana* fitness. *Curr Opin Plant Biol* 13: 348–353

Cataldi, TRI, Rubino A, Lelario F, Bufo SA (2007) Naturally occurring glucosinolates in plant extracts of rocket salad (*Eruca sativa* L.) identified by liquid chromatography coupled with negative ion electrospray ionization and quadrupole ion-trap mass spectrometry. *Rap Commun Mass Spectr* 21: 2374–2388

Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J (2005) The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiol* 137: 253–262

Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM (2009) Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* 323: 95–101

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2009) Integrated omics approaches in plant systems biology. *Curr Opin Chem Biol* 13: 532–538

Gigolashvili T, Berger B, Mock HP, Müller C, Weisshaar B, Flügge UI (2007) The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J* 50: 886–901

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. *Trends Plant Sci* 11: 89–100

Guan SH, Marshall, AG, Scheppele SE (1996) Resolution and chemical formula identification of aromatic hydrocarbons and aromatic compounds containing sulfur, nitrogen, or oxygen in petroleum distillates and refinery streams. *Anal Chem* 68: 46–71

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* 57: 303–333

Hirai MY (2009) A robust omics-based approach for the identification of glucosinolate biosynthetic genes. *Phytochem Rev* 8: 15–23

Kim S, Kramer RW, Hatcher PG (2003) Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal Chem* 75: 5336–5344

Kind T, Fiehn O (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform* 8: 105

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiol* 126: 811–825

Marshall AG, Hendrickson CL, Jackson GS (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 17: 1–35

Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D, et al. (2006) Cytochrome P450 CYP710A encodes the sterol C22-desaturase in Arabidopsis and tomato. *Plant Cell* 18: 1008–1022

Nagano AJ, Fukao Y, Fujiwara M, Nishimura M, Hara-Nishimura I (2008) Antagonistic jacalin-related lectins regulate the size of ER body-type beta-glucosidase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol* 49: 969–980

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of coexpressed genes and cis elements for identifying coregulated gene groups in Arabidopsis. *Nucleic Acids Res* 35: D863–D869

Ogata Y, Sakurai N, Aoki K, Suzuki H, Okazaki K, Saito K, Shibata D (2009) KAGIANA: an excel-based tool for retrieving summary information on Arabidopsis genes. *Plant Cell Physiol* 50: 173–177

Ohta D, Kanaya S, Suzuki H (2010) Application of Fourier-transform ion cyclotron resonance mass spectrometry to metabolic profiling and metabolite identification. *Curr Opin Biotechnol* 21: 35–44

Oikawa A, Nakamura Y, Ogura T, Kimura A, Suzuki H, Sakurai N, Shinbo Y, Shibata D, Kanaya S, Ohta D (2006) Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. *Plant Physiol* 142: 398–413

Petersen BL, Chen S, Hansen CH, Olsen CE, Halkier BA (2002) Composition and content of glucosinolates in developing *Arabidopsis thaliana*. *Planta* 214: 562–571

Pfalz M, Vogel H, Kroymann J (2009) The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in Arabidopsis. *Plant Cell* 21: 985–999

Rochfort SJ, Trenerry VC, Imsic M, Panozzo J, Jones R (2008) Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MS^n fragmentation. *Phytochem* 69: 1671–1679

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci* 11: 508–516

Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, Asahi H, Kurokawa K, Arita M, Saito K, Ohta D, Shibata D, Kanaya S (2006) KNApSAcK: a comprehensive species-metabolite relationship database. *Biotechnol Agric Forest* 57: 165–184

Wittstocka U, Burow M (2010) Glucosinolate Breakdown in Arabidopsis: Mechanism, Regulation and Biological Significance. *The Arabidopsis Book* 8: 1–14

Wu Z, Rodgers RP, Marshall AG (2004) Two- and three-dimensional van Krevelen diagrams: a graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband fourier transform ion cyclotron resonance mass measurements. *Anal Chem* 76: 2511–2516