PBS3 and EPS1 complete salicylic acid biosynthesis from isochorismate in Arabidopsis

Authors:

Michael P. Torrens-Spence[1], Anastassia Bobokalonova[1,2], Valentina Carballo[1], Christopher M. Glinkerman[1], Tomáš Pluskal[1], Amber Shen[1,2], Jing-Ke Weng[1,2*]

Affiliations:

[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, Massachusetts 02142, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

*Correspondence: wengj@wi.mit.edu

Salicylic acid (SA) is an important phytohormone mediating both local and systemic defense responses against pathogens in plants[1]. Despite over half a century of research, how SA is biosynthesized in plants remains elusive[2]. In Arabidopsis, the bulk of SA is derived from isochorismate, a key intermediate produced by the isochorismate synthase (ICS)[3,4], which is reminiscent of SA biosynthesis in bacteria. Nevertheless, whereas bacteria employ an isochorismate pyruvate lyase (IPL) that directly converts isochorismate to SA, plants do not contain IPL orthologs. How plants produce SA from isochorismate is unknown. Combining genetic and biochemical approaches, we delineated the SA biosynthetic pathway downstream of isochorismate in Arabidopsis. We show that PBS3, a GH3 acyl adenylase-family enzyme important for pathogen-induced SA accumulation[5], catalyzes ATP- and Mg^{2+}-dependant conjugation of *L*-glutamate primarily to the 8-carboxyl of isochorismate and yields the key SA biosynthetic intermediate isochorismoyl-glutamate A. Moreover, EPS1, a BAHD acyltransferase-family protein with previously implicated role in proper SA accumulation upon pathogen attack[6], harbors a noncanonical active site and an unprecedented isochorismoyl-glutamate A pyruvoyl-glutamate lyase (IPGL) activity that produces SA specifically from the isochorismoyl-glutamate A substrate. Together, PBS3 and EPS1 complete SA biosynthesis from isochorismate in Arabidopsis via a metabolic pathway distinct from that of bacteria[4]. This study closes a major knowledge gap in plant SA biosynthesis and provides new insights for engineering disease resistance in crop plants.

As sessile organisms, plants have evolved intricate immune system to perceive and defend against invasion of a plethora of pathogens. Decades of extensive research has established the role of SA as a pivotal phytohormone that signals plant defense responses in local and distal tissues[1]. In addition to its function in defense signaling, SA also contributes to various aspects of

plant growth and development, including photosynthesis, transpiration, ion uptake and transport, thermogenesis, drought resistance and senescence[7]. SA is biosynthesized from chorismate via two distinct metabolic routes. A minor fraction of SA is produced by 2-hydroxylation of benzoic acid (BA) derived from the general phenylpropanoid metabolism[8]. Previous characterization of the Arabidopsis *sid2* mutant, defective in the *ICS1* gene (*AT1G74710*), revealed that majority, around 90%, of the *de novo* biosynthesized SA upon *Pseudomonas aeruginosa* infection is derived from isochorismate[3]. Whereas SA-producing bacteria employ an IPL that directly converts isochorismate to SA, plants lack IPL orthologs, and previous attempts to identify plant enzyme with *bona fide* IPL activity have not been successful[9], suggesting that plants may use alternative biochemistry to produce SA from isochorismate[2].

Like *sid2*, two additional Arabidopsis mutants, namely *pbs3* and *eps1*, exhibit enhanced susceptibility to bacterial pathogens together with impaired accumulation of SA and SA-derived metabolites, and both encode metabolic enzymes[5,6]. The *pbs3* mutant encodes a GH3 acyl-adenylase-family enzyme (GH3.12, AT5G13320)[5]. The best characterized GH3-family enzyme in plants is JAR1 (GH3.11), which catalyzes ATP-dependent conjugation of isoleucine to jasmonic acid (JA) to yield jasmonyl-isoleucine (JA-Ile), the active form of jasmonate hormone[10]. Although its *in vivo* function remains unknown, PBS3 was shown to contain *in vitro* acyl acid amido synthetase activity, i.e., conjugating *L*-glutamate to 4-substituted BAs[10,11]. Interestingly, SA appears to be a poor substrate of PBS3 and inhibits PBS3 activity *in vitro*[10,11]. On the other hand, the *eps1* mutant reportedly exhibits striking phenotypic resemblance to *pbs3*, including enhanced susceptibility to the bacterial pathogen *Pseudomonas syringae* and impaired *de novo* SA biosynthesis[6]. The *eps1* mutant was mapped to a *BAHD acyltransferase* (*AT5G67160*) of unknown function[6]. Notably, EPS1 harbors an unusual active-site serine

(Ser160) at the residue position corresponding to the highly conserved catalytic histidine in BAHD acyltransferases[6,12], suggesting that EPS1 may contain unconventional catalytic activity different from the canonical acyltransferase activity (Extended Data Fig. 1a, b). Together, these prior studies implicate potential roles of PBS3 and EPS1 in SA metabolism[2].

To resolve the SA biosynthetic pathway in plants, we first took a genetic approach. Two Arabidopsis 2-oxoglutarate/Fe (II)-dependent dioxygenases (AT4G10500 and AT5G24530) were recently characterized as the SA 3-hydroxylase (S3H) and SA 5-hydroxylase (S5H), respectively[13,14]. Collectively, S3H and S5H constitute a major SA catabolic pathway that contributes to the proper maintenance of SA homeostasis in higher plants[14]. The single mutants of *S3H* and *S5H* in Arabidopsis, namely *s3h* and *dmr6*, accumulate elevated levels of free SA and SA-derived metabolites and also exhibit various mild growth phenotypes and enhanced resistance to pathogens associated with hyperactive SA signaling[13,15] (Fig. 1, Fig. 2a, Extended Data Fig. 2, Supplementary Data 1, 2). These phenotypes are greatly exacerbated in the *s3h*/*dmr6* double mutant[14,16] (Fig. 1, Fig. 2a, Extended Data Fig. 2). While the *s3h*/*dmr6* double mutant contain significantly higher levels of free SA and SA-derived metabolites than those of the wild type and the *s3h* and *dmr6* single mutants (Fig. 2a, Extended Data Fig. 2), it also displays a severe stunted-growth phenotype and significantly enhanced resistance against the virulent bacterial pathogen *P. syringae* pv. *tomato* DC3000 (*Pst*DC3000) (Fig. 1). We therefore reasoned that *s3h*/*dmr6* provides an ideal genetic background to study Arabidopsis genes necessary for SA biosynthesis and signaling.

To test this idea, we crossed *s3h*/*dmr6* with the extensively studied SA biosynthetic mutant *sid2* and SA signaling mutant *npr1*[17], and generated the corresponding *s3h*/*dmr6*/*sid2* and *s3h*/*dmr6*/*npr1* triple mutants, respectively. Both triple mutants display a complete rescue of the

stunted-growth phenotype and a partial reversion of the pathogen resistance phenotype of the *s3h*/*dmr6* double mutant (Fig. 1). Consistent with the recovered growth, both the *s3h*/*dmr6/sid2* and *s3h*/*dmr6/npr1* plants also exhibit significantly reduced levels of free SA and SA-derived metabolites compared to those of the *s3h*/*dmr6* plants (Fig. 2a, Extended Data Fig. 2). These observations indicate that hyper-accumulation of SA in the *s3h*/*dmr6* plants not only requires ICS activity, but also NPR1, one of the primary SA receptors in Arabidopsis[18,19]. Upon the initial activation of SA-mediated defense response, NPR1 likely activates a positive feedback loop that further amplifies SA production by upregulating SA biosynthetic genes. These results also suggest that loss-of-function mutations in other genes involved in SA biosynthesis may also alleviate the mutant phenotype of *s3h*/*dmr6*.

To probe the functions of *PBS3* and *EPS1* in the *s3h*/*dmr6* genetic background, we crossed *pbs3* and *eps1* with *s3h*/*dmr6*, and generated the *s3h*/*dmr6/pbs3* and *s3h*/*dmr6/eps1* triple mutants, respectively. Like the *s3h*/*dmr6/sid2* and *s3h*/*dmr6/npr1* triple mutants, the *s3h*/*dmr6/pbs3* and *s3h*/*dmr6/eps1* triple mutants display a complete and nearly complete rescue of the stunted-growth phenotype, respectively (Fig. 1a). The *s3h*/*dmr6/pbs3* and *s3h*/*dmr6/eps1* plants are also less resistant to *Pst*DC3000 compared to the *s3h*/*dmr6* plants (Fig. 1b, c). The *s3h*/*dmr6/pbs3* and *s3h*/*dmr6/eps1* plants accumulate free SA and SA-derived metabolites at levels significantly less than those of the *s3h*/*dmr6* plants (Fig. 2a, Extended Data Fig. 2). These new genetic observations suggest that PBS3 and EPS1 may be directly involved in SA biosynthesis or participate in other auxiliary metabolic processes necessary for proper SA accumulation in Arabidopsis[3].

To define the precise biochemical functions of PBS3 and EPS1 *in vivo*, we performed untargeted metabolomic profiling of three-week-old rosette leaves from wild-type Arabidopsis

and the aforementioned single and multiple mutants using liquid-chromatography high-resolution accurate-mass mass-spectrometry (LC-HRAM-MS) (Methods, Supplementary Data 1, 2). Comparative analysis of the resultant metabolomic datasets between specific mutant backgrounds identified three new metabolite features (compounds **1**-**3**) that accumulate at high levels in the *s3h*/*dmr6* plants but are absent in the *pbs3* and *s3h*/*dmr6/pbs3* plants, relating them to PBS3 function (Fig. 2a, b, c, Extended Data Fig. 3). In-depth analysis of the mass spectra in the context of plausible isochorismate-derived SA biosynthetic pathways led to tentative structural assignment of these metabolites. **1** and **2** were resolved as two regioisomers of isochorismoyl-glutamate with glutamate conjugated to either the 8- or 1-carboxyl of isochorismate, referred to as isochorismoyl-glutamate A and B, respectively (Fig. 2d, e, Extended Data Fig. 4). **3** was identified as *N*-pyruvoyl-*L*-glutamate, a possible degradation product of **1** (Fig. 2d, e, Extended Data Fig. 5a). The identity of **3** was further confirmed by chemical synthesis (Extended Data Fig. 5b, c). Since PBS3 reportedly contains *in vitro* acyl acid amido synthetase activity and prefers *L*-glutamate as its amino acid donor substrate[11], our new findings suggest **1** and **2** may be the *in vivo* enzymatic products of PBS3, which are depleted in the *pbs3* mutant backgrounds. Moreover, **1** is likely further catabolized *in vivo* to yield **3**, and under such reaction scheme, the other co-product would be SA (Fig. 3a). Intriguingly, the *eps1* and *s3h*/*dmr6/eps1* plants contain levels of **1** and **2** orders of magnitude higher than those of the wild-type and *s3h*/*dmr6* plants respectively (Fig. 2a, c, d, Extended Data Fig. 3b, d), suggesting that EPS1 may be involved in the catabolic pathway downstream of **1** and **2**.

We next characterized the biochemical functions of PBS3 and EPS1 using purified recombinant proteins in *in vitro* enzyme assays. Since isochorismate is not commercially available, we devised a pre-assay to enzymatically synthesize isochorismate from chorismate

using recombinant SID2. Starting with 5 mM chorismate, the reaction reaches equilibrium with approximately equal molar ratio of isochorismate and chorismate at the end of the pre-assay (Extended Data Figs. 6a). Consistent with the findings from the Arabidopsis metabolomics experiments, when PBS3 was added to the pre-assay containing approximately equal molar amounts of isochorismate and chorismate, it exhibits specific ATP- and Mg^{2+}-dependent acyl acid amido synthetase activity towards isochorismate with *L*-glutamate as the prefered amino acid substrate, and yields **1** as the main enzymatic product together with **2** as a minor product (Fig. 3a, b, Extended Data Figs. 6, 7). In the absence of isochorismate, PBS3 also exhibits acyl acid amido synthetase activity towards chorismate and produces regio-isomeric chorismoyl-glutamate A and B (Compounds **4** and **5** respectively, Fig. 3c, d, Extended Data Fig. 4), although these products were not detected in Arabidopsis metabolomics datasets. Therefore, PBS3 functions as a substrate- and regio-selective isochorismoyl-glutamate synthase (IGS) both *in vivo* and *in vitro*.

A previous study of the crystal structure of PBS3 in complex with SA and adenosine monophosphate (AMP) revealed that SA binds to PBS3 in a catalytically nonproductive pose with its carboxyl group pointing away from the catalytic center[10] (Extended Data Fig. 8a). We performed docking simulation of isochorismate and chorismate to the PBS3 structure (PDB: 4EQL)[10]. The 1-carboxyl of isochorismate preferably adopts a similar orientation as the SA carboxyl, which in turn presents the isochorismate 8-carboxyl in a catalytically productive conformation towards the catalytic center, readying it for adenylation (Extended Data Fig. 8b). Although chorismate could be docked to PBS3 active site in a similar orientation as isochorismate, it appears to be a less favored substrate due to the differential placement of its 4-

hydroxyl (Extended Data Fig. 8c). These ligand docking results reveal features of the PBS3 active site consistent with the experimentally observed substrate- and regio-selectivity of PBS3.

We noticed that **1** produced by PBS3 in enzyme assays is unstable and spontaneously decays to produce SA and **3** at a rate of 0.21 pkat (Fig. 3e, Extended Data Fig. 9a, b), whereas **2** is stable under our assay conditions. Interestingly, addition of EPS1 to the PBS3 enzyme assay resulted in rapid turnover of **1** to SA and **3**, with the catalyzed rate four orders of magnitude faster than the spontaneous decay rate of **1** (Fig. 3e, f, Extended Data Fig. 5c, Extended Data Fig. 9c, d). EPS1 displays little to no measurable lyase or hydrolase activity against isochorismate, chorismate, **2**, **3**, **4** and **5**, suggesting its IPGL activity is specific against **1**. Altogether, our *in vitro* enzyme assays further confirmed the chemical identities of **1**-**3**, established the specific acyl acid amido synthetase activity of PBS3 which produces the key SA biosynthetic intermediate **1**, and revealed that EPS1 is an unprecedented IPGL that significantly enhances the rate of SA production from **1**.

Upon pathogen attack, Arabidopsis rapidly activates *de novo* SA biosynthesis in both local and systemic tissues, mostly through the isochorismate-derived pathway[3]. To identify the core Arabidopsis gene set necessary and sufficient for this rapid *de novo* SA biosynthesis, we sought to reconstitute *de novo* SA biosynthesis in *Nicotiana benthamiana* using the Agrobacterium-mediated transient transgenic expression technique[20] (Fig. 4a). Infiltration of *N. benthamiana* leaves with Agrobacterium carrying empty vector did not trigger significant level of *de novo* SA biosynthesis compared to uninfiltrated leaves, thus providing a clean background for this reconstitution exercise. Experimenting with several genes previously implicated in SA metabolism, we discovered that coexpression of *SID2* together with *SID1* (*AT4G39030*), a chloroplast-envelope-localized MATE-family transporter required for pathogen-induced SA

accumulation in Arabidopsis[21], resulted in ectopic accumulation of both chorismate and isochorismate in *N. benthamiana* leaves (Fig. 4b, Extended Data Fig. 10). Furthermore, coexpression of *PmsB*, a bacterial *IPL*[22], together with *SID1* and *SID2* is sufficient to elicit *de novo* SA biosynthesis in *N. benthamiana* leaves (Fig. 3a, Fig. 4b). In comparison, coexpression of *PBS3* together with *SID1* and *SID2* resulted in not only *de novo* production of SA but also a near depletion of both chorismate and isochorismate levels in *N. benthamiana* leaves (Fig. 4b). Moreover, the *PBS3-SID1-SID2* coexpression also led to ectopic accumulation of **1-3**, which were absent in the *PmsB-SID1-SID2* coexpression experiment (Fig. 4c, d, Extended Data Fig. 5c, Extended Data Fig. 7, Supplementary Data 3). On the other hand, coexpression of *EPS1* together with *SID1* and *SID2* did not result in *de novo* production of SA, **1**, **2** and **3** (Fig. 4b, d). When *EPS1* was coexpressed together with *PBS3*, *SID1* and *SID2*, the level of **1** was completely depleted and the level of **2** was also greatly reduced compared to that of the *PBS3-SID1-SID2* coexpression, although the accumulating levels of SA and **3** did not increase further (Fig. 4b, d). Expression of *SID1*, *SID2*, *PmsB*, *PBS3* or *EPS1* alone did not elicit *de novo* SA synthesis nor yield measurable levels of any SA biosynthetic intermediates in *N. benthamiana* leaves (Fig. 4b, c, Extended Data Fig. 10). These reconstitution experiments in *N. benthamiana* further reinforces the critical function of *PBS3* in the plant isochorismate-derived SA biosynthetic pathway, and suggests *PBS3*, *SID1* and *SID2* as a minimum gene set necessary and sufficient to reconstitute *de novo* SA biosynthesis in a heterologous plant host. Although the addition of *EPS1* apparently enhances the turnover rate of **1**, *PBS3-SID1-SID2* coexpression is sufficient to yield SA and its co-product **3** in *N. benthamiana*. This could be attributed to the spontaneous decay of **1** over the course of the Agrobacterium infiltration experiment (5 days), existing IPGL activity in *N.*

benthamiana, and/or *N. benthamiana* endogenous SA production triggered by the initial SA accumulation resulted from the expression of Arabidopsis transgenes.

SA plays a central role in defense signaling of higher plants[2]. In this study, we elucidated the isochorismate-derived SA biosynthetic pathway in Arabidopsis (Fig. 3a). PBS3 is an essential enzyme in this pathway immediately downstream of ICS, responsible for the production of key SA intermediate isochorismoyl-glutamate A. The important role of PBS3 in SA biosynthesis in higher plants is corroborated by a recent phylogenomic study showing that *PBS3* orthologs are widely present in several model flowering plants with predicted emergence predating the monocot/eudicot split[23]. Therefore, it is likely that isochorismoyl-glutamate A may be a conserved SA biosynthetic intermediate in flowering plants. In contrast, EPS1 belongs to a unique clade of BAHD acyltransferases only associated with the Brassicaceae family plants (Extended Data Fig. 1c, Supplementary Data 4, 5), indicating that EPS1 was a recent addition to the SA biosynthetic pathway in Brassicaceae. This evolutionary development could be driven by strong selection pressure to improve SA production kinetics amid pathogen attack in the common ancestor of extant Brassicaceae plants. EPS1 harbors an unusual active site reconfigured from an ancestral BAHD acyltransferase to catalyze the unprecedented IPGL chemistry (Extended Data Fig. 1b), and thus presents an attractive case for future study of molecular evolution that gave rise to novel catalytic function based on ancestral enzyme fold[24]. Although loss of *EPS1* in Arabidopsis led to ectopic hyper-accumulation of isochorismoyl-glutamate A and B and compromised SA production, the *s3h*/*dmr6/eps1* triple mutant still accumulates *N*-pyruvoyl-*L*-glutamate at a level similar to that of the *s3h*/*dmr6* double mutant and higher SA level than the *s3h*/*dmr6/pbs3* triple mutant (Fig. 2a), suggesting spontaneous decay of isochorismoyl-glutamate A likely occurs at absence of EPS1 *in vivo*. On the other hand, non-

Brassicaceae plants may rely on either the spontaneous decay of isochorismoyl-glutamate A or have evolved alternative IPGLs to produce SA. A full understanding of the metabolic pathways underlying SA biosynthesis and catabolism in the plant kingdom will ultimately enable comprehensive approaches to engineer disease resistance in economically important crops.

Acknowledgement

We thank the Bioinformatics and Research Computing and the Metabolite Profiling core Facilities at the Whitehead Institute for assistance with the metabolomics experiments and data analysis. We thank Jen Sheen for providing the *P. syringae* pv. *tomato* DC3000 strain. This work was supported by the Pew Scholar Program in the Biomedical Sciences, the Searle Scholars Program and the National Science Foundation (CHE-1709616).

Author Contributions

M.P.T.S. and J.K.W. designed the research. M.P.T.S. performed all enzyme assay, the reconstitution of SA biosynthesis in *N. benthamiana*. V.C., J.K.W. performed the genetics experiments, metabolomic profiling, and the pathogen inoculation experiment. A.B. and A.S. assisted with various genetic, biochemical and metabolomics experiments. C.M.G. performed chemical synthesis. T.P. performed ligand docking. M.P.T.S. and J.K.W. interpreted the results and wrote the paper.

Competing interests

The authors declare no competing interests.

Figures

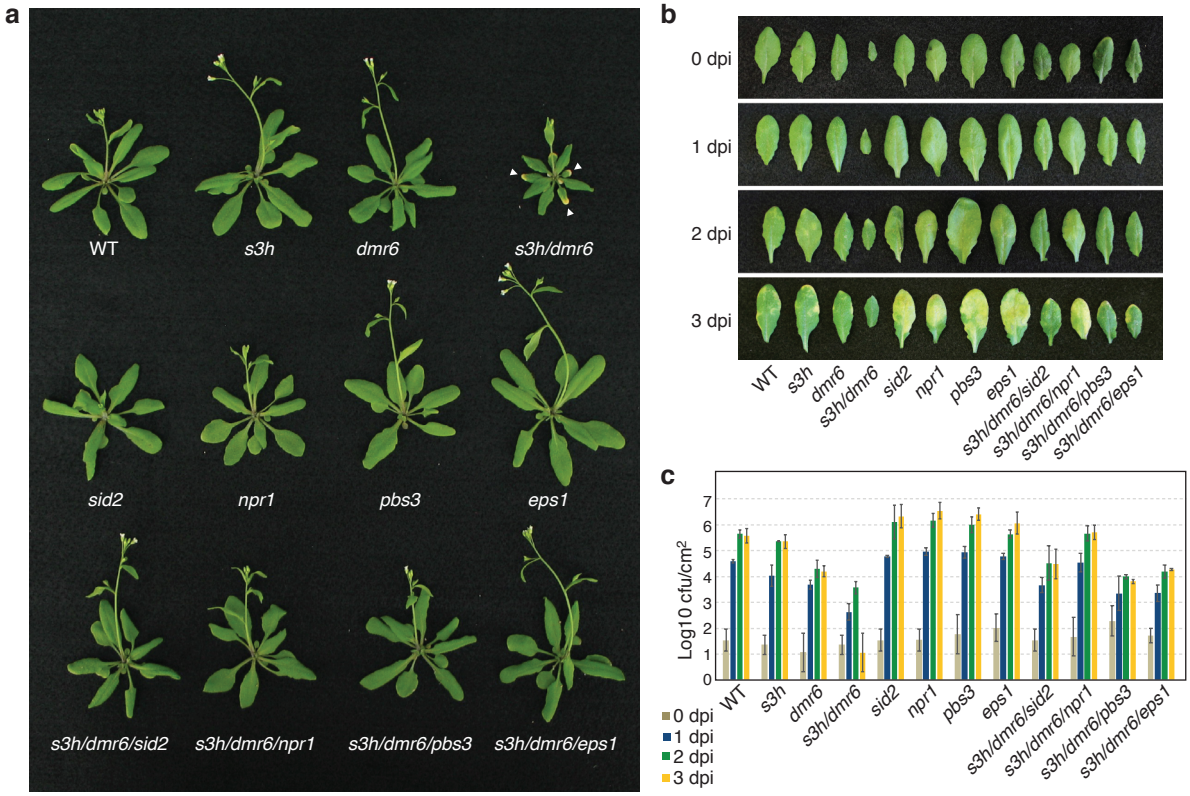


Fig. 1| Characterization of the roles of *PBS3* and *EPS1* in plant growth and defense in an Arabidopsis mutant background that over-accumulates SA. a, Four-week-old green-house-grown Arabidopsis plants. The *s3h/dmr6* double mutant exhibit stunted growth and early senescence (white arrows) phenotypes associated with hyper-accumulation of SA[14], which could be rescued by crossing with the *sid2*, *npr1*, *pbs3* and *eps1* mutants, respectively. **b,** Representative images of Arabidopsis rosette leaves infiltrated with a suspension of the *P. syringae* PstDC3000 strain. Dpi, days post-inoculation. **c,** Quantification of the bacterial growth in Arabidopsis rosette leaves infiltrated with *P. syringae* PstDC3000 strain. Leaves were sampled at 0 through 4 dpi to measure the growth of the bacterial pathogen. The Logarithmic errors (δz) were calculated from the standard deviations of the raw data (n = 3) according to Baird[25]. Cfu, colony-forming unit.

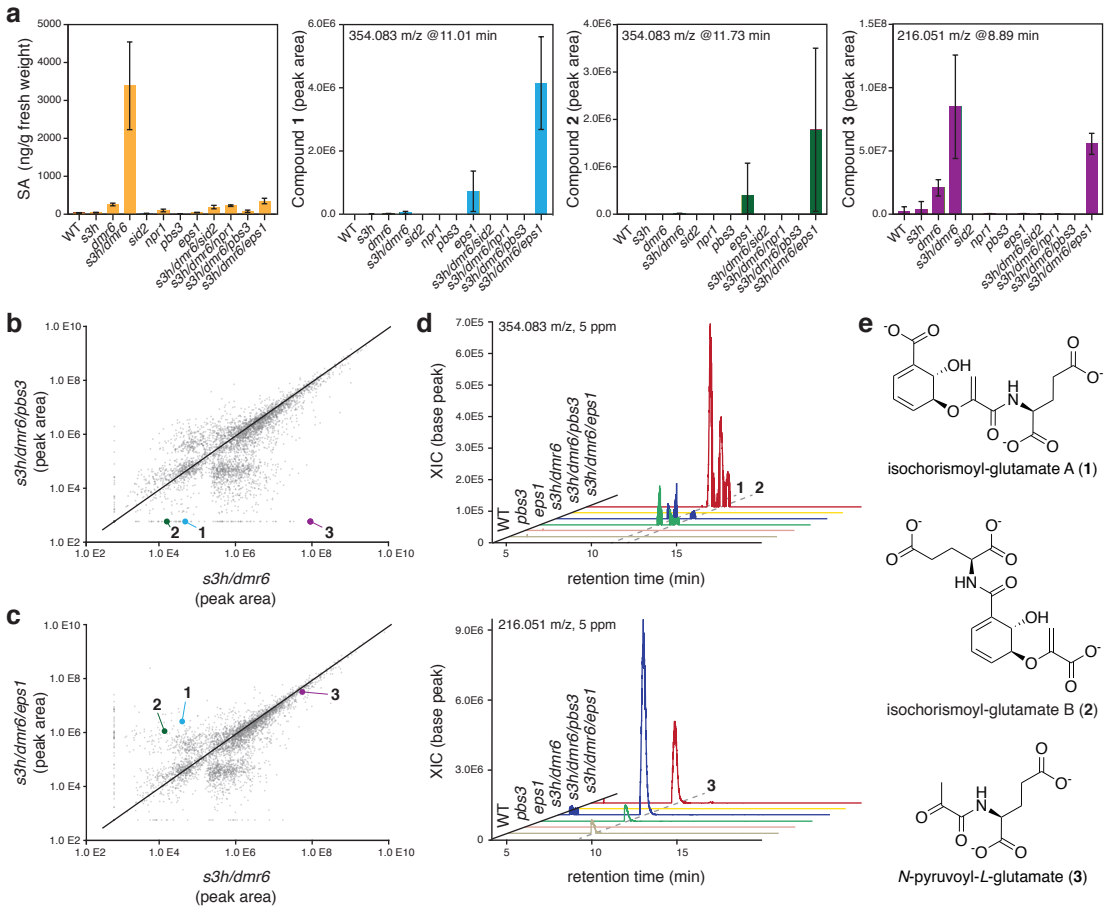


Fig. 2| Key SA biosynthetic intermediates and products revealed by LC-HRAM-MS-based metabolic profiling of Arabidopsis mutants. a, Quantification of SA and three newly identified metabolites (**1-3**) that are differentially accumulated across various Arabidopsis SA biosynthetic and signaling mutants (n = 3). Error bars indicate standard error of the mean (SEM) based on biological triplicates. **b,** Pairwise scatter plot comparing the abundance of global metabolic features of *s3h/dmr6* and *s3h/dmr6/pbs3*. The plot shows the accumulation of **1-3** in *s3h/dmr6*, which are absent in *s3h/dmr6/pbs3*. Data were collected using LC-HRAM-MS under negative mode. Each dot represents a metabolic feature defined by unique m/z and retention time values (n =3). **c,** Pairwise scatter plot comparing the abundance of global metabolic features of *s3h/dmr6* and *s3h/dmr6/eps1*, which shows hyper-accumulation of **1** and **2** in *s3h/dmr6/eps1* compared to *s3h/dmr6* (n =3). **d,** Representative extracted ion chromatograms (XICs) corresponding to **1-3** in select Arabidopsis mutants. The metabolites were resolved by normal-phase chromatography. **e,** Structural assignment of **1-3** (also see Extended Data Fig. 4a, b, 5).

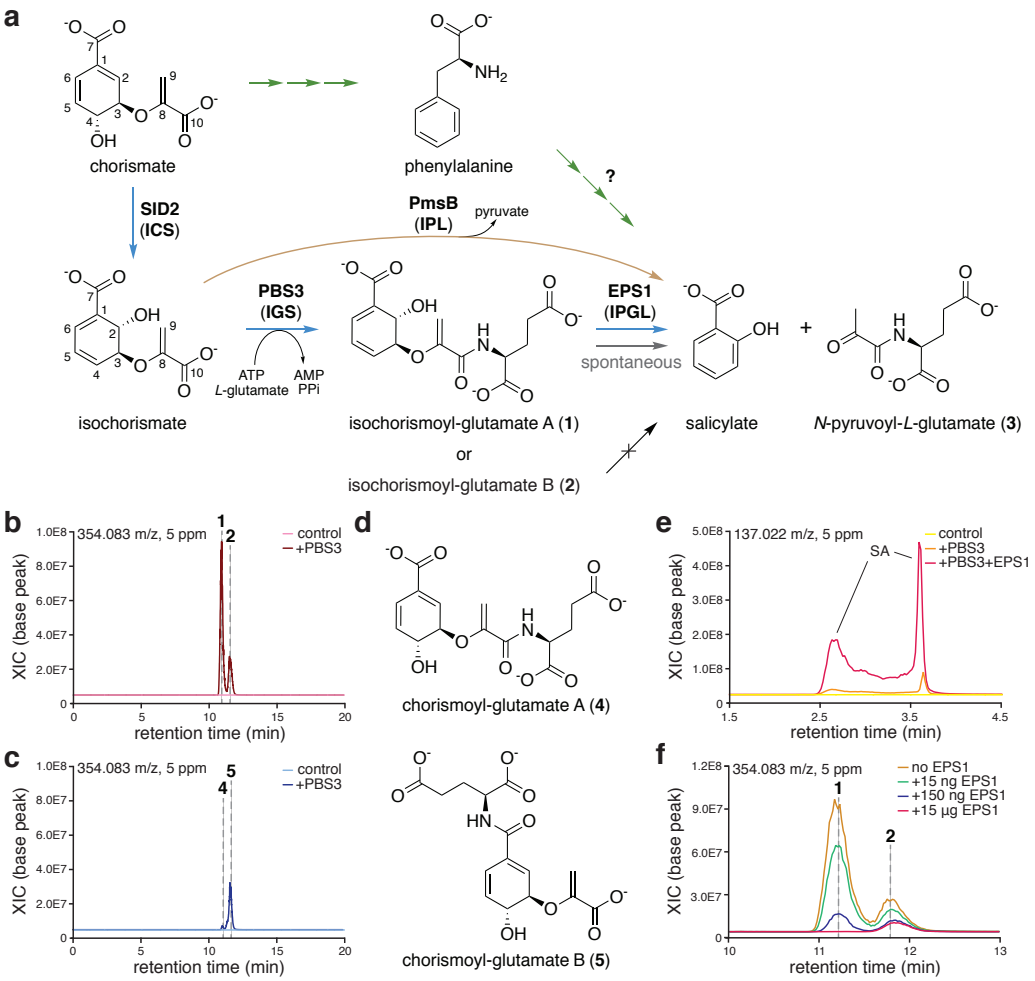


Fig. 3| The biochemical functions of PBS3 and EPS1 in plant SA biosynthesis. a, The SA biosynthetic pathways in higher plants. The isochorismate-derived pathway is highlighted by blue arrows. The minor phenylpropanoid-metabolism-derived pathway is highlighted by green arrows. Some SA-producing bacteria employs an IPL that directly converts isochorismate to SA, which is highlighted by a brown arrow. Enzyme names are in parentheses under the representative protein names. **b,** Extracted ion chromatogram (XICs) showing the *in vitro* enzyme activity of PBS3 that produces **1** and **2** starting from a chorismate-SID2 preassay that contains approximately equal molar isochorismate and chorismate as potential acyl substrates. **c,** PBS3 can also produce **4** and **5** from a assay buffer that contains only chorismate as the acyl substrate. The control traces in **b** and **c** represent enzyme assays performed without adding PBS3. **d**, Structural assignment of **4** and **5** (also see Extended Data Fig. 4c, d). **e,** Representative XICs showing spontaneous decay of **1** that produces SA at the end of PBS3-isochorismate enzyme assay (orange trace). The addition of EPS1 (15 µg) converts all remaining **1** to SA within 5 min (red trace). **f,** Representative XICs showing EPS1 preferably turns over **1** to produce SA and **3** in a EPS1 enzyme concentration-dependent manner (also see Extended Data Fig. 9c, d). The chromatograms shown were resolved using normal-phase chromatography. SA splits into two peaks under such chromatographic condition.

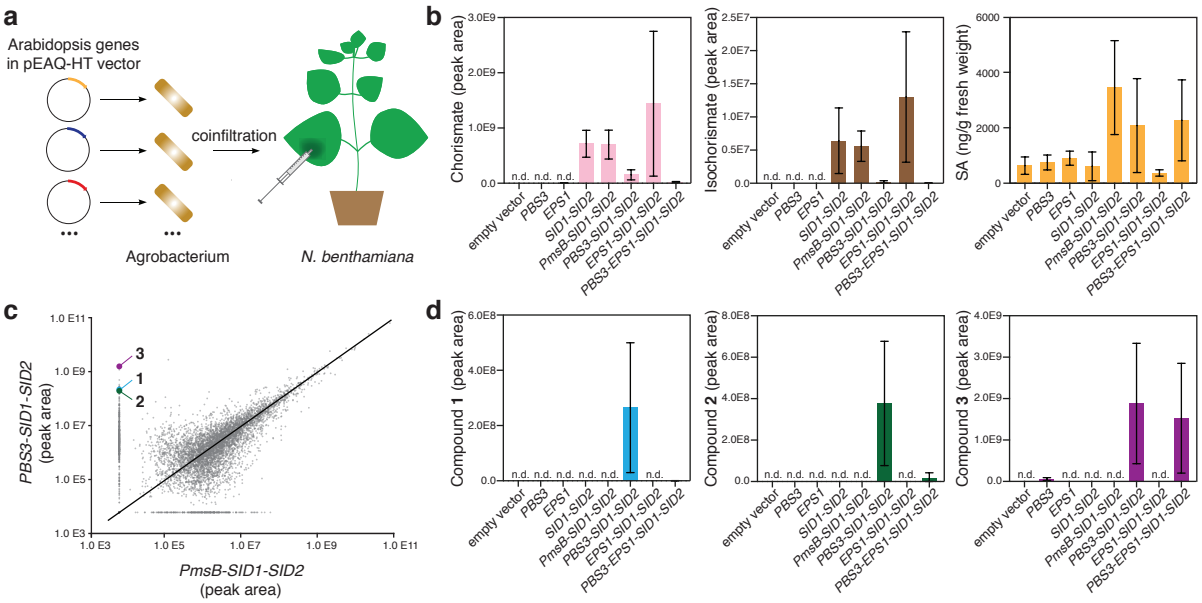


Fig. 4| Reconstitution of *de novo* SA biosynthesis in *N. benthamiana* by Agrobacterium-mediated transient coexpression of transgenes. a, Individual Arabidopsis genes were cloned into the pEAQ-HT vector[20] and separately transformed into Agrobacterium. Combination of the resultant Agrobacterium strains was used to coinfiltrate the leaves of *N. benthamiana* plants. LC-HRAM-MS-based metabolic profiling of leaf extracts was performed 5 days post infiltration. **b,** Quantification of chorismate, isochorismate and SA across various coinfiltration experiments (n = 3). Error bars indicate SEM based on biological triplicates. **c,** Pairwise scatter plot comparing the abundance of global metabolic features of *PBS3-SID1-SID2* conexpression and *PmsB-SID1-SID2* coexpression in *N. benthamiana* leaves. The plot shows hyper-accumulation of **1-3** in *PBS3-SID1-SID2*-coinfiltrated leaves which are absent in the *PmsB-SID1-SID2* coexpression experiment. Data were collected using LC-HRAM-MS under negative mode. Each dot represents a metabolic feature defined by unique m/z and retention time values (n =3). **d,** Quantification of **1-3** across various coinfiltration experiments (n = 3). Error bars indicate SEM based on biological triplicates. n. d., not detected.

References

1. Amick Dempsey, D. Salicylic Acid and Disease Resistance in Plants. *CRC Crit. Rev. Plant Sci.* **18**, 547–575 (1999).
2. Chen, Z., Zheng, Z., Huang, J., Lai, Z. & Fan, B. Biosynthesis of salicylic acid in plants. *Plant Signal. Behav.* **4**, 493–496 (2009).
3. Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).
4. Serino, L. *et al.* Structural genes for salicylate biosynthesis from chorismate in Pseudomonas aeruginosa. *Mol. Gen. Genet.* **249**, 217–228 (1995).
5. Nobuta, K. *et al.* The GH3 acyl adenylase family member PBS3 regulates salicylic acid-dependent defense responses in Arabidopsis. *Plant Physiol.* **144**, 1144–1156 (2007).
6. Zheng, Z., Qualley, A., Fan, B., Dudareva, N. & Chen, Z. An important role of a BAHD acyl transferase-like protein in plant innate immunity. *Plant J.* **57**, 1040–1053 (2009).
7. Raskin, I. Role of Salicylic Acid in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 439–463 (1992).
8. León, J., Shulaev, V., Yalpani, N., Lawton, M. A. & Raskin, I. Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 10413–10417 (1995).
9. Zhou, Y., Memelink, J. & Linthorst, H. J. M. An E. coli biosensor for screening of cDNA libraries for isochorismate pyruvate lyase-encoding cDNAs. *Mol. Genet. Genomics* **293**, 1181–1190 (2018).
10. Westfall, C. S. *et al.* Structural basis for prereceptor modulation of plant hormones by GH3 proteins. *Science* **336**, 1708–1711 (2012).

11. Okrent, R. A., Brooks, M. D. & Wildermuth, M. C. Arabidopsis GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. *J. Biol. Chem.* **284**, 9742–9754 (2009).

12. Levsh, O. *et al.* Dynamic Conformational States Dictate Selectivity toward the Native Substrate in a Substrate-Permissive Acyltransferase. *Biochemistry* **55**, 6314–6326 (2016).

13. Zhang, K., Halitschke, R., Yin, C., Liu, C.-J. & Gan, S.-S. Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14807–14812 (2013).

14. Zhang, Y. *et al.* Encodes a Salicylic Acid 5-Hydroxylase That Fine-Tunes Salicylic Acid Homeostasis. *Plant Physiol.* **175**, 1082–1093 (2017).

15. van Damme, M., Huibers, R. P., Elberse, J. & Van den Ackerveken, G. Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *Plant J.* **54**, 785–793 (2008).

16. Zeilmaker, T. *et al.* DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in Arabidopsis. *Plant J.* **81**, 210–222 (2015).

17. Cao, H., Glazebrook, J., Clarke, J. D., Volko, S. & Dong, X. The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88**, 57–63 (1997).

18. Wu, Y. *et al.* The Arabidopsis NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. *Cell Rep.* **1**, 639–647 (2012).

19. Ding, Y. *et al.* Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity. *Cell* **173**, 1454–1467.e15 (2018).

20. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).

21. Nawrath, C., Heck, S., Parinthawong, N. & Métraux, J.-P. EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in Arabidopsis, is a member of the MATE transporter family. *Plant Cell* **14**, 275–286 (2002).

22. Verberne, M. C., Verpoorte, R., Bol, J. F., Mercado-Blanco, J. & Linthorst, H. J. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nat. Biotechnol.* **18**, 779–783 (2000).

23. Okrent, R. A. & Wildermuth, M. C. Evolutionary history of the GH3 family of acyl adenylases in rosids. *Plant Mol. Biol.* **76**, 489–505 (2011).

24. Weng, J.-K. & Noel, J. P. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 309–320 (2012).

25. Baird, D. C. *Experimentation: An Introduction to Measurement Theory and Experiment Design*. (Prentice-Hall, 1995).

Supplementary Information for

PBS3 and EPS1 complete salicylic acid biosynthesis from isochorismate in Arabidopsis

Authors:

Michael P Torrens-Spence[1], Anastassia Bobokalonova[1,2], Valentina Carballo[1], Christopher M Glinkerman[1], Tomáš Pluskal[1], Amber Shen[1,2], Jing-Ke Weng[1,2*]

Affiliations:

[1]Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, Massachusetts 02142, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

*Correspondence: wengj@wi.mit.edu

Methods

Plant materials and growth conditions

Arabidopsis thaliana and *Nicotiana benthamiana* plants were grown at 22 °C in a greenhouse under long-day condition (16-h light/8-h) or in growth chambers under short-day condition (12-h light/12-h) wherever specified. The Arabidopsis T-DNA insertional null mutants *s3h*, *dmr6-2*, *sid2-4*, *npr1-1*, *pbs3-2*, and *eps1-2* in Columbia background were obtained from the Arabidopsis Biological Resource Center (ABRC) under the accessions SALK_059907, GK_249H03, SALK_133146, CS3726, SALK_018225C, and SAIL_734_F07, respectively[1,2]. For simplicity, the mutants were mentioned throughout the paper without their specific allele names called out. *Nicotiana benthamiana* plants were grown at 22 °C in a greenhouse under long-day condition (16-h light/8-h)

Pseudomonas syringae infection assay

P. syringae inoculations were performed according to previously described protocol[3]. Briefly, *Pst* DC3000 was streaked out from a –80 °C glycerol stock onto King's medium B (Protease Peptone #3 20 g/L, glycerol 10 mL/L, K_2HPO_4 1.5 g/L, $MgSO_4 \cdot 7H_2O$ 1.5 g/L, Bacto agar 15 g/L, ph 7.2) with 50mg/mL rifampicin, and grown for 2 days at 30 °C. Fresh bacteria was transferred to a 3 mL liquid culture and grown overnight at 30 °C, a final 200 mL liquid culture was grown overnight until final OD_{600} = 0.8. The culture was pelleted, and diluted in sterile water to a final concentration of 1×10^6 colony-forming unit (cfu)/mL. 6-week-old Arabidopsis leaves were infiltrated with the bacterial suspension on their abaxial side with a 1 mL needleless syringe. At 0, 1, 2, 3 dpi, Infected leaves were harvested and surface sterilized for 1 minute with 70% ethanol and then rinsed with distilled water. Two leaf disks for each genotype and each triplicate were cut with a 0.3 cm^2 cork borer. Leaf disks were placed in a 2.0 mL microfuge tube with 100 µL sterile distilled water, and thoroughly macerated using a plastic pestle. Ground tissue was diluted up to 1 mL with sterile water, and mixed by vortexing. Four serial dilutions were performed and blotted onto King's medium B in 10 µL triplicate drops. After incubation at 30°C for two days, colonies were counted to derive cfu values.

Transient expression of genes in *N. benthamiana*

The open reading frame (ORF) of target genes to be expressed were cloned into pEAQ-HT[4]. Vectors were transformed into *Agrobacterium tumefaciens* strain LBA4404[5]. Bacteria were cultivated at 30 °C in YM Medium (0.4 g/L yeast extract, 10 g/L mannitol, 0.1 g/L NaCl, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.5 g/L $K_2HPO_4 \cdot 3H_2O$) to an OD_{600} between 2.0 and 2.5, washed with 0.5X PBS, then diluted with 0.5X PBS to an OD_{600} of 0.8. For co-infiltration of multiple gene constructs, *A. tumefaciens* strains were mixed in equal concentrations to a final OD_{600} of 0.8. 1 mL of bacterial suspension was infiltrated into leaves of *N. benthamiana* plants between 4 and 6 weeks of age. Leaves were harvested 5 days post infiltration for untargeted metabolomics analysis.

Metabolomic profiling of plant extracts

For untargeted metabolomics analysis, 3-week-old Arabidopsis rosette leaves or *N. benthamiana* leaves 5 days post *A. tumefaciens* infiltration were harvested, suspended in 50% methanol (500 µL per 100 mg fresh weight) and disrupted using a TissueLyser (Qiagen) and zirconia beads. Tissue debris were removed by centrifugation. 2 µL injections of the clarified extracts were separated using a 150 x 2.1 mm ZIC-pHILIC column (5 µm particle size, EMD

Millipore). The chromatographic gradient was run at a flow rate of 0.15 mL/min as follows: 0–20 min, linear gradient from 80% to 20% solvent B; 20–20.5 min, linear gradient from 20% to 80% solvent B; 20.5–28 min, hold at 80% solvent B. Solvent A was 20 mM ammonium carbonate, 0.1% ammonium hydroxide; solvent B was acetonitrile. The column oven and autosampler tray were held at 25 °C and 4 °C, respectively. Data was collected with a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific) for MS^1 scans and dd-MS^2 (data-dependent MS/MS) in both positive and negative polarity modes. The various arabidopsis leaf extracts were also additionally run through a C18 column (Kinetex 2.6-µm, 100-Å, 150 × 3 mm) with a gradient of Solvent A (0.1% formic acid in H_2O) and Solvent B (0.1% formic acid in acetonitrile; 5% B for 2 min, 5-80% B over 40 min, 95% B for 4 min, and 5% B for 5 min; flow rate 0.8 mL/min) before detection by the Q-Exactive mass spectrometer. Full MS resolution, 70,000; mass range, 100–1,250 m/z; dd-MS2 resolution, 17,500; loop count, 5; collision energy, 15–35 eV (stepped); dynamic exclusion, 1 s. The raw data were processed using MZmine 2[6] and further analyzed using metaboanalyst[7].

Chemical synthesis

General methods. All reactions were performed under nitrogen unless otherwise noted. All reagents and solvents were used as supplied without further purification unless otherwise noted. Column chromatography was conducted using Silicycle SiliaFlash P60 SiO_2 (40–63 µm). Analytical TLC was conducting using Millipore SiO_2 60 F_{254} TLC (0.250 mm) plates. HPLC was conducted using a pair of Shimadzu LC-20AP pumps, a Shimadzu CBM-20A communications module, a Shimadzu SPD-20A UV-Vis detector, and a Phenomenex Kinetex 5u C18 100 Å Axia column (150 x 21.2 mm). Melting points were obtained using a Mel-Temp II apparatus and are uncorrected. ^{1}H and ^{13}C NMR spectra were obtained using a Bruker Avance Neo 400 MHz spectrometer equipped with a 5 mm Bruker SmartProbe. IR spectra were obtained using a Bruker Alpha 2 with a Platinum ATR accessory. Mass spectrometric analysis was performed on a JEOL AccuTOF-DART.

Di-*tert*-butyl *N*-pyruvoyl-*L*-glutamate (S2).[8] A vigorously stirred solution of pyruvic acid (0.35 mL, 5.1 mmol) in CH_2Cl_2 (35 mL) at 23 °C was treated sequentially with EDCI (790 mg, 5.1 mmol) and HOBt (690 mg, 5.1 mmol). The reaction mixture was allowed to stir at 23 °C for 5 min, at which time protected glutamate **S1** (1.0 g, 3.4 mmol) was added and the reaction was stirred at 23 °C for an additional 24 h. After 24 h, the reaction mixture was triturated with Et_2O (300 mL) and filtered through Celite. The filtrate was concentrated on a rotary evaporator and the residue was purified by flash chromatography (SiO_2, 20% EtOAc/hexanes) to provide **S2** as a white solid (240 mg, 22%): mp 64–66 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.44 (d, J = 8.4 Hz, 1H), 4.40 (td, J = 8.2, 4.8 Hz, 1H), 2.43 (s, 3H), 2.32–2.19 (m, 2H), 2.18–2.10 (m, 1H), 2.00–1.89 (m, 1H), 1.44 (s, 9H), 1.40 (s, 9H); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 196.2, 171.8, 170.1, 160.0, 82.8, 80.9, 52.3, 31.4, 28.1, 28.0, 27.5, 24.5; IR (film) ν_{max} 3365, 1741, 1714, 1671, 1512, 1367, 1258, 1230, 1144, 844, 620, 606 cm^{-1}; HRMS (DART-TOF) *m/z* 330.1902 ($C_{16}H_{27}NO_6 + H^+$ requires 330.1917).

***N*-pyruvoyl-*L*-glutamate (3)**.[9] A vigorously stirred suspension of **S2** (50 mg, 0.15 mmol) in THF (1.2 mL) at 23 °C was treated with 3 N HCl(aq) (3.0 mL). The reaction mixture was allowed to stir at 23 °C for 24 h, then concentrated under a stream of N_2. The resulting solid residue was purified by preparative reverse-phase HPLC (CH_3CN-0.1% TFA/H_2O-0.1% TFA 5:95 to 95:5 over 30 min, 7 mL/min, R_t = 10.1 min) to provide **3** as a hygroscopic colorless film (16 mg, 48%): ^{1}H NMR (DMSO-d_6, 400 MHz) δ 12.47 (bs, 2H), 8.68 (d, J = 8.1 Hz, 1H), 4.23

(ddd, J = 9.6, 8.1, 4.7 Hz, 1H), 2.34 (s, 3H), 2.26 (t, J = 7.4 Hz, 2H), 2.06 (dtd, J = 15.5, 7.1, 4.7 Hz, 1H), 1.91 (ddt, J = 14.0, 9.7, 7.1 Hz, 1H); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 196.7, 173.9, 172.4, 161.2, 51.4, 30.1, 25.6, 24.9; IR (film) ν_{max} 2926, 1719, 1637, 1524, 1411, 1358, 1165, 969, 611 cm^{-1}; HRMS (DART-TOF) *m/z* 218.0651 ($C_8H_{11}NO_6$ + H^+ requires 218.0665).

Recombinant protein production and purification

E. coli cultures transformed with expression vectors were grown at 37 °C in TB media to an optical density of 1.5, induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG), and allowed to grow for an additional 16 h at 18 °C. *E. coli* cells were harvested by centrifugation, washed with phosphate buffered saline, resuspended in 100 mL of lysis buffer [(50 mM phosphate buffer pH 6.2, 0.5 M NaCl, 20 mM imidazole, 10 mM sodium ascorbate, 2 mM ethylenediaminetetraacetic acid (EDTA) and 2 mM dithiothreitol (DTT)], and lysed by sonication (Branson sonifier 450). The resulting crude protein lysate was clarified by centrifugation prior to Qiagen Ni-NTA gravity flow chromatographic purification. After loading the clarified lysate, the protein bound Ni-NTA resin was washed with 10 column volumes of lysis buffer and eluted with 1 column volume of elution buffer (50 mM phosphate buffer pH 6.2, 0.5 M NaCl, 200 mM imidazole, 10 mM sodium ascorbate, 2 mM EDTA and 2 mM DTT). Purified recombinant enzymes were concentrated and washed multiple times with storage buffer (25 mM phosphate buffer pH 6.2, 50 mM NaCl, 10mM sodium ascorbate, and 2 mM DTT) using an amicon ultra-15 centrifugal filter unit with ultracel-30 membrane (EMD Millipore) to achieve a final protein concentration of 1 mg/mL. The purity of the recombinant proteins was evaluated by ImageJ[10] densitometric analysis as compared to a bovine serum albumin standard.

Enzyme assays

Initial isochorismate synthase activity titration assays were performed in 100 µL reactions composed of 25 mM Tris and 5 mM chorismate. Reactions were started with the addition of 0.5 µg recombinant SID2, incubated at 25 °C for various time points, and quenched with the addition of 100 µL MeOH. To establish PBS3 activity *in vitro*, isochorismate was first generated enzymatically by adding 2 µg of SID2 to a 100 µL reaction containing 50 mM Tris, 1 mM *L*-glutamate, 5mM chorismate, 5 mM ATP, 5 mM $MgCl_2$ and incubating for 5 minutes at 25 degrees. After incubation, 8 µg of PBS3 was added to each reaction mixture, incubated at 25 °C for various time points and then quenched with 100 µL MeOH. The ATP- and Mg^{2+} dependence of PBS3 was tested with identical reactions excluding ATP and $MgCl_2$, respectively. 1 mM *L*-alanine, *L*-arginine, *L*-asparagine, *L*-aspartic acid, *L*-cysteine, *L*-glutamine, glycine, *L*-histidine, *L*-isoleucine, *L*-leucine, *L*-lysine, *L*-phenylalanine, *L*-proline, *L*-serine, *L*-threonine, *L*-tryptophan, *L*-tyrosine and *L*-valine was substituted for the *L*-glutamate in separate assays to assess the proteinogenic *L*-amino acid substrate selectivity of recombinant PBS3. The relative amino acid preference was determined by measuring AMP production. The production of various isochorismoyl-amino acid conjugates was also verified by analyzing the MS data. To characterize the biochemical function of EPS1, isochorismoyl-glutamate conjugates were first produced enzymatically in 100 µL reaction systems containing 50 mM Tris, 1 mM *L*-glutamate, 5mM chorismate, 5 mM ATP, and 5 mM $MgCl_2$ with addition of 2 µg of SID2 and 8 µg of PBS3 incubated for 20 minutes at 25 °C. Various amounts of recombinant EPS1 and were added to the reactions, incubated for an additional 5 minutes at 25 °C prior to quenching with an equal volume of MeOH. Enzyme assays were analyzed by LC-HRAM-MS using ZIC-pHILIC column following the same procedure as described above.

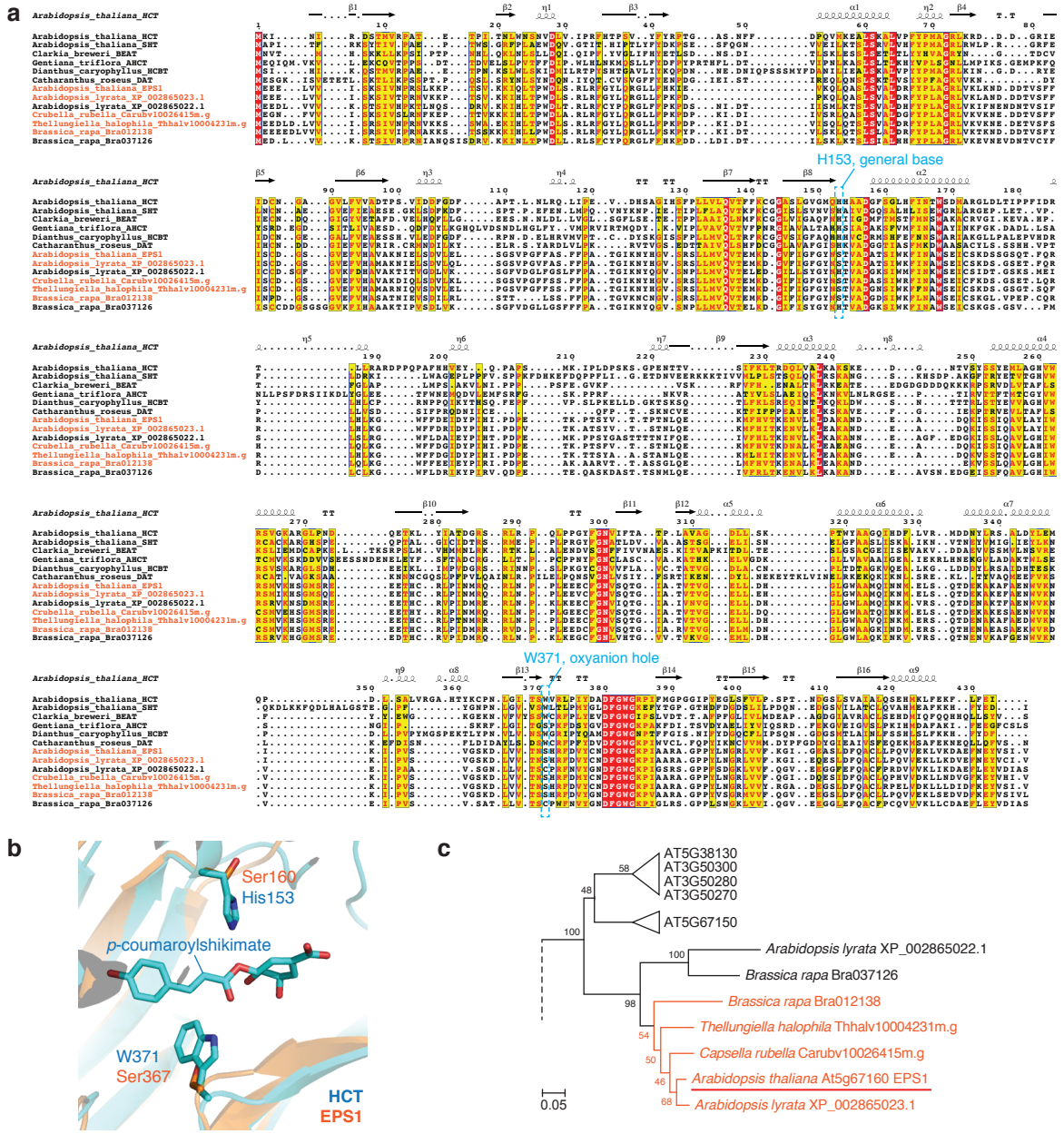
Sequence alignment, phylogenetic analysis and homology modeling

Sequence alignment shown in Extended Data Fig. 1a was generated using T-coffee[11], and displayed using ESPript 3.0[12]. The Neighbor-Joining (NJ) tree shown in Extended Data Fig. 1a and Supplementary Data 4 was inferred using MEGA 7[13], based on sequence alignment built using MUSCLE[14] (Supplementary Data 5). The bootstrap percentages were based on 1000 replications. The EPS1 structural homology model was generated using the SWISS-MODEL[15] server with Arabidopsis HCT (PDB: 5KJU) as the template.

Ligand docking

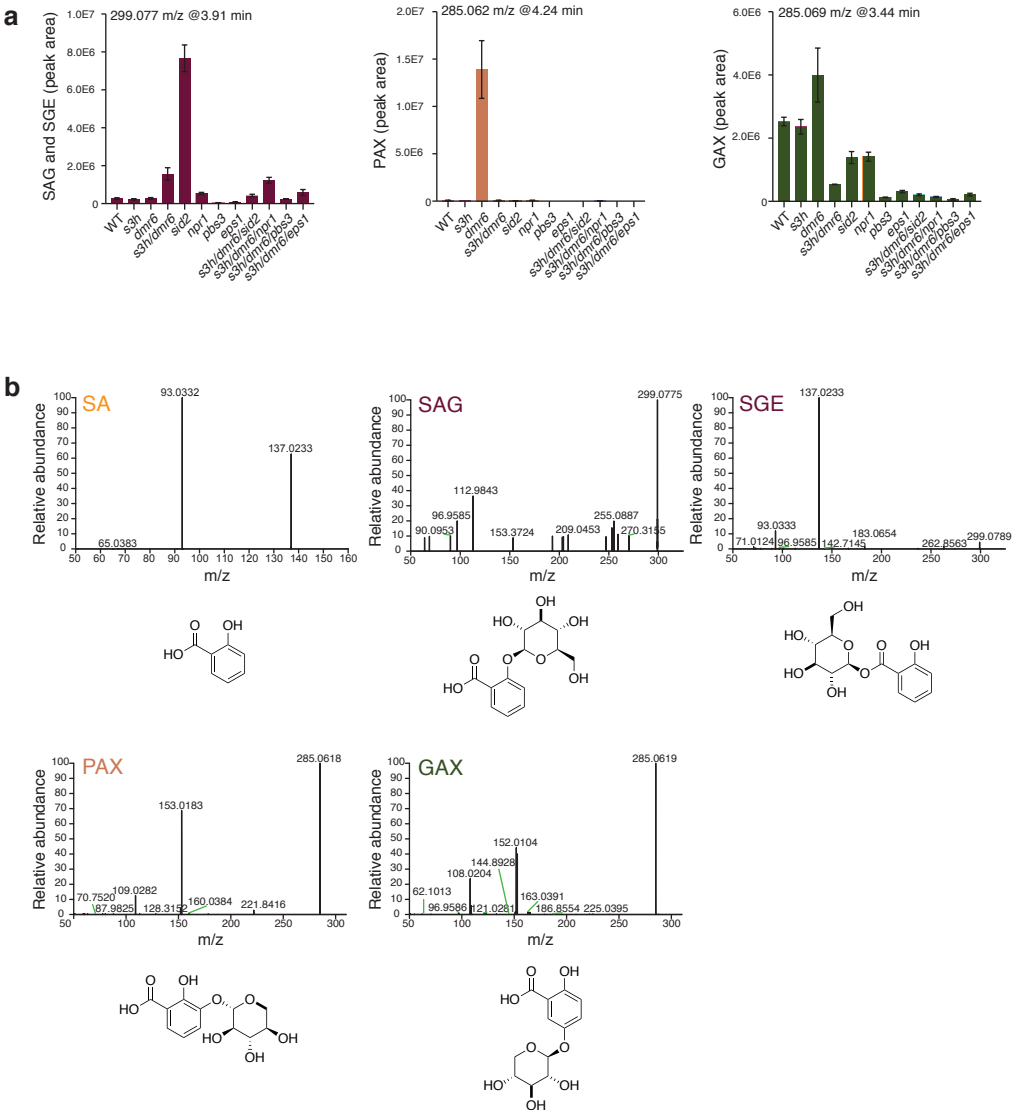
Docking simulation of chorismate and isochorismate into the active site of Arabidopsis PBS3 crystal structure (PDB: 4EQL)[16] was performed using AutoDock Vina 1.1.2[17]. The bound ligands in 4EQL were removed before docking. The docking simulation generated top nine possible poses for each of the docked ligands, which were ranked by the predicted affinity (kcal/mol). The top poses for isochorismate and chorismate were shown in Extended Data Fig. 8b and c respectively. Molecular graphics were rendered with PyMOL 2.0.7 (Schrödinger).

Extended Data Figures

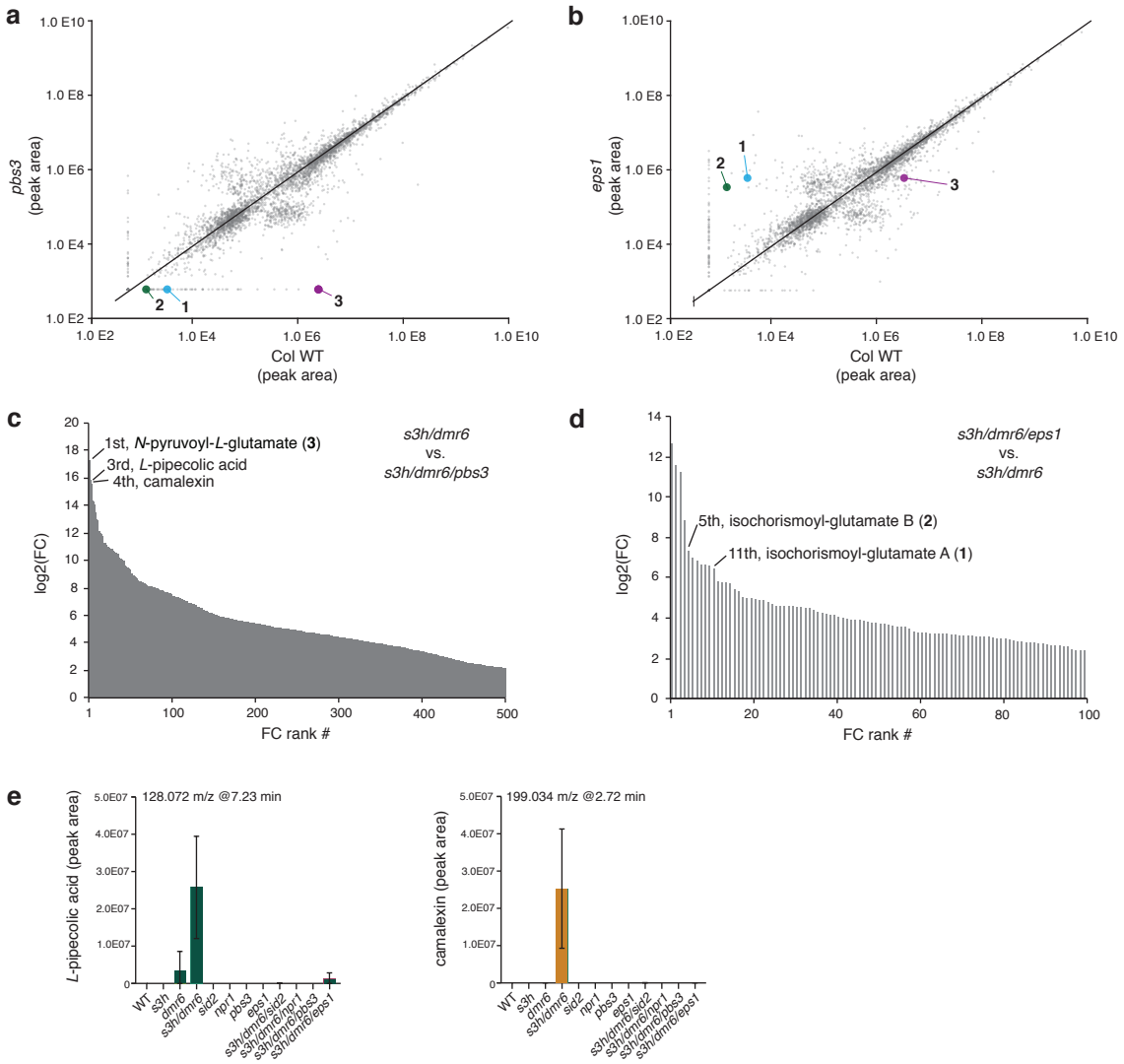


Extended Data Fig. 1| EPS1 represents a unique clade of BAHD acyltransferase family proteins associated with the Brassicaceae family plants. a, Sequence alignment of EPS1 and its orthologs from other Brassicaceae species (orange, also illustrated in Extended Data Fig. 1c) together with several other related and well characterized BAHDs[18,19]. The two catalytic residue positions important for canonical BAHD acyltransferase catalytic function are highlighted in blue dotted boxes. H153 (Arabidopsis HCT numbering) serves as a general base that deprotonates the hydroxyl or amine group of the acyl acceptor substrate, priming it for nucleophilic attack on the carbonyl carbon of the acyl CoA substrate[20], and is ubiquitously conserved among canonical BAHDs. The indole nitrogen of Trp371 serves as an oxyanion hole

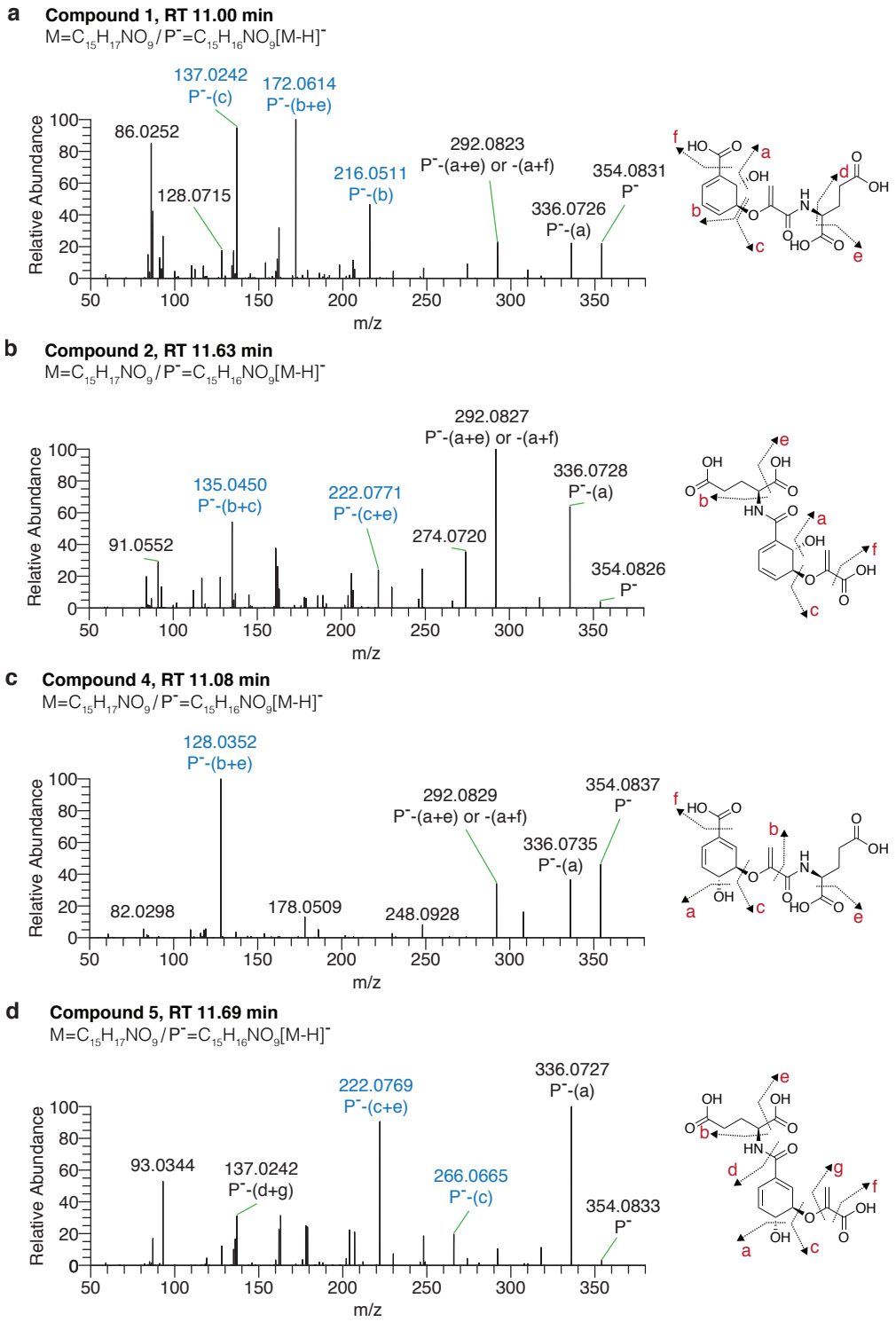
that stabilizes the negative charge on the tetrahedral intermediate formed after the nucleophilic attack[20]. Trp371 is conserved among the majority of BAHDs. **b**, Overlay of the active sites of Arabidopsis HCT (in complex with the product *p*-coumaroylshikimate, blue, PDB: 5KJU) and the EPS1 homology model (orange). EPS1 contains serine substitutions at the two catalytic residues His153 and Trp371 (AtHCT numbering) highly conserved among canonical BAHDs. **c**, Neighbor-joining phylogenetic analysis of BAHDs from multiple reference genomes suggests that EPS1 belongs to a unique clade of BAHDs restricted in the Brassicaceae family (orange). The EPS1 orthologous clade is sister to three other BAHD clades also restricted in Brassicaceae. Bootstrap values (based on 1000 replicates) are indicated at the tree nodes. The scale measures evolutionary distances in substitutions per amino acid. The full tree is presented in Supplementary Data 4 and the sequence alignment is presented in Supplementary Data 5.



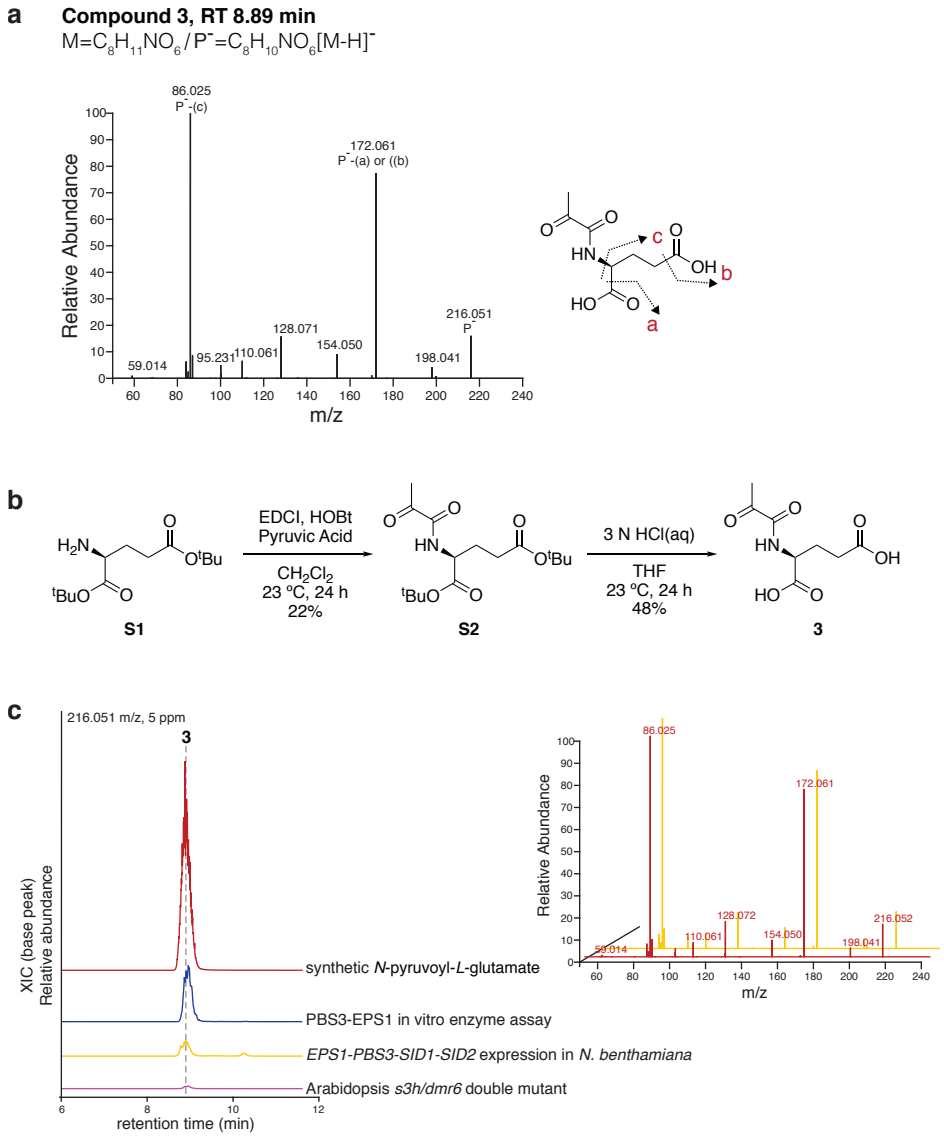
Extended Data Fig. 2| Levels of SA-derived metabolites in various Arabidopsis mutants. a, Relative quantification of four major SA-derived metabolites that are differentially accumulated across various Arabidopsis SA biosynthetic and signaling mutants (n = 3). Error bars indicate standard error of the mean (SEM) based on biological triplicates. SAG, salicylic acid 2-*O*-*β*-D-glucoside; SGE, salicylic acid glucose ester; PAX, o-pyrocatechuic acid 5-*O*-*β*-D-xyloside; GAX, gentisic acid 5-*O*-*β*-D-xyloside. The three metabolites were chromatographically separated on a C18 column, measured by LC-HRAM-MS and data processed by MZmine2. **b,** The measured MS^2 spectra of SA, SAG, SGE, PAX and GAX and their chemical structures.



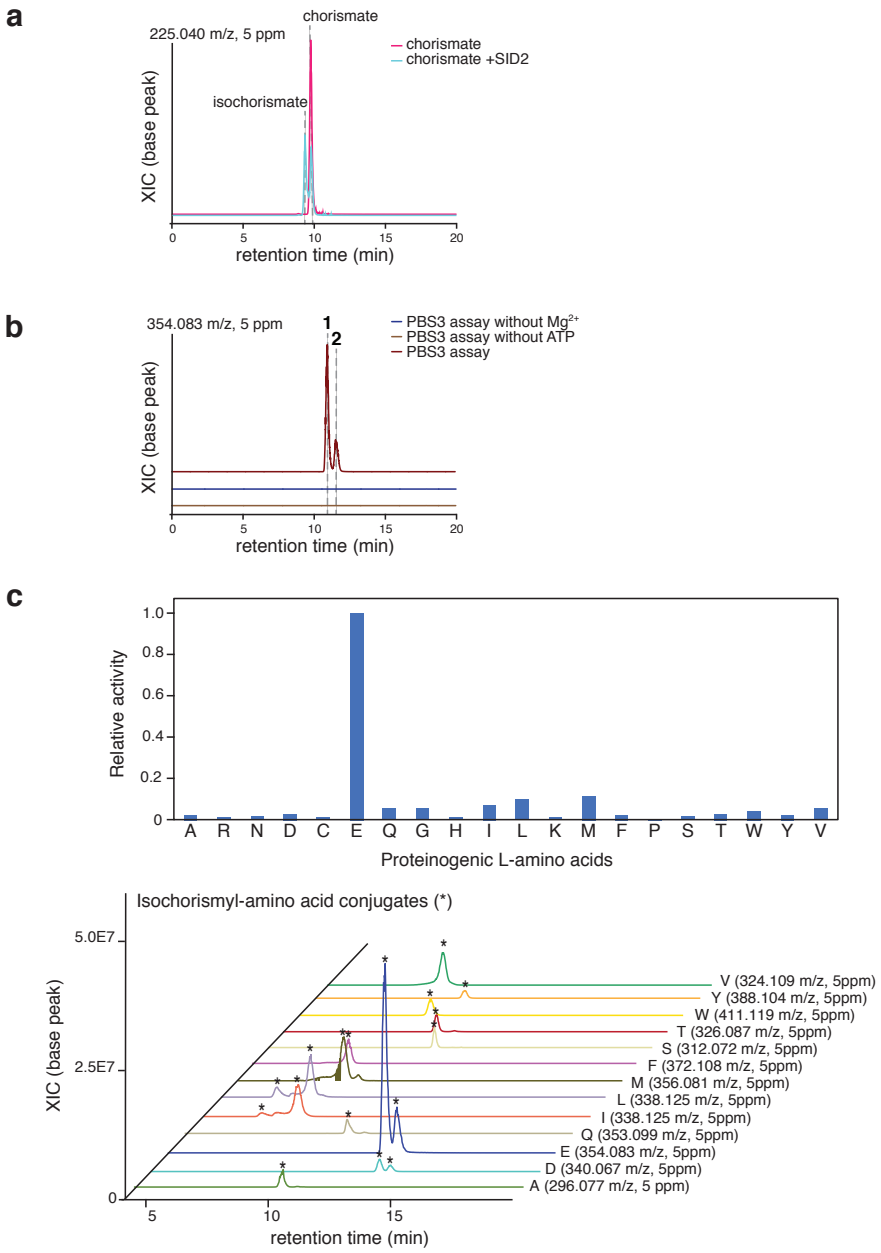
Extended Data Fig. 3| Analysis of differentially accumulated metabolic features across Arabidopsis SA-biosynthetic mutants. a, Pairwise scatter plot comparing the abundance of global metabolic features of wild type and *pbs3*. The plot shows the accumulation of **1-3** in wild type, which are absent in *pbs3*. Data were collected using LC-HRAM-MS under negative mode. Each dot represents a metabolic feature defined by unique m/z and retention time values (n =3). **c,** Pairwise scatter plot comparing the abundance of global metabolic features of wild type and *eps1*, which shows hyper-accumulation of **1** and **2** in *eps1* compared to wild type (n =3). **b,** Ranking of metabolic features that are differentially accumulated between *s3h/dmr6* and *s3h/dmr6/pbs3* based on Log2 fold change. **b,** Ranking of metabolic features that are differentially accumulated between *s3h/dmr6/eps1* and *s3h/dmr6* based on Log2 fold change. **e,** Relative quantification of *L*-pipecolic acid and camelexin that are differentially accumulated across various Arabidopsis SA biosynthetic and signaling mutants (n = 3). Error bars indicate standard error of the mean (SEM) based on biological triplicates.



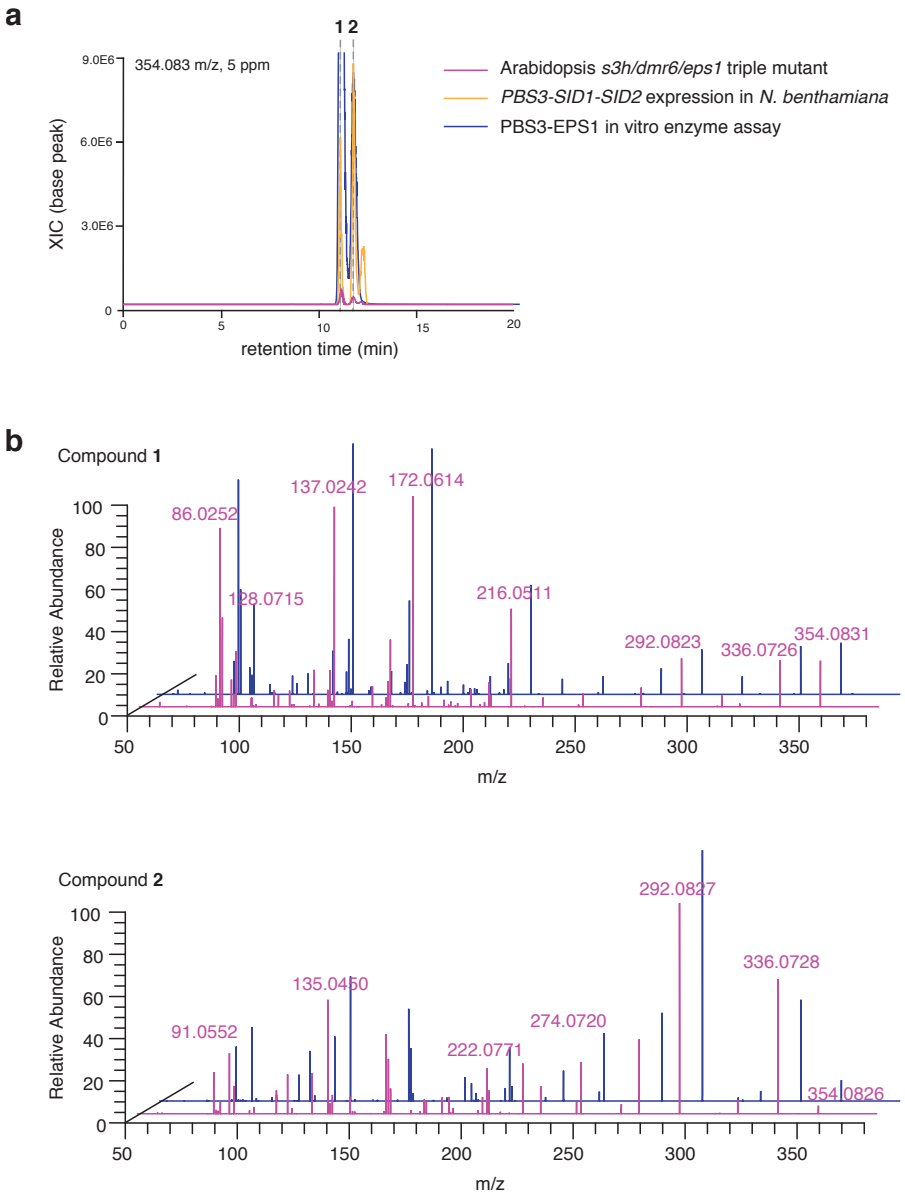
Extended Data Fig. 4| Structural elucidation of compounds 1, 2, 4 and 5. MS^2 fragmentation patterns of 1, 2, 4 and 5 (**a-d**). The proposed fragmentation sites are shown. P^-, protonated precursor ion. The unique fragments associated with each compound that helped to determine the identify of the specific structural isomer are highlighted in blue.



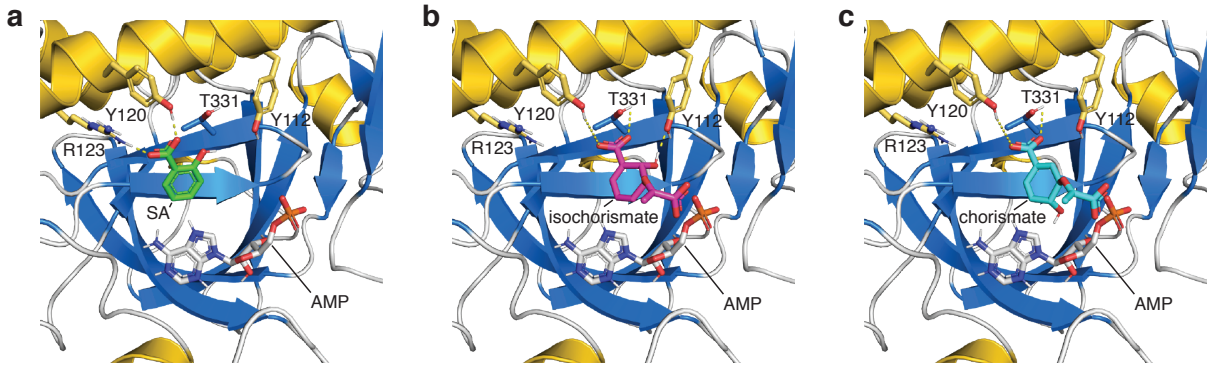
Extended Data Fig. 5| Structural elucidation of compound 3. a, MS^2 fragmentation patterns of 3. The proposed fragmentation sites are shown. P^-, protonated precursor ion. **b,** Organic synthesis scheme for **3**. **c**, Overlay of extracted ion chromatogram (XICs) showing the detection of **3** from LC-HRAM-MS analysis of synthetic **3**, representative PBS3-EPS1 enzyme assay, *N. benthamiana* co-expressing *EPS1-PBS3-SID1-SID2* and the Arabidopsis *s3h/dmr6* double mutant. The MS^2 spectrum match between synthetic **3** and **3** from *N. benthamiana* co-expressing *EPS1-PBS3-SID1-SID2* is shown.



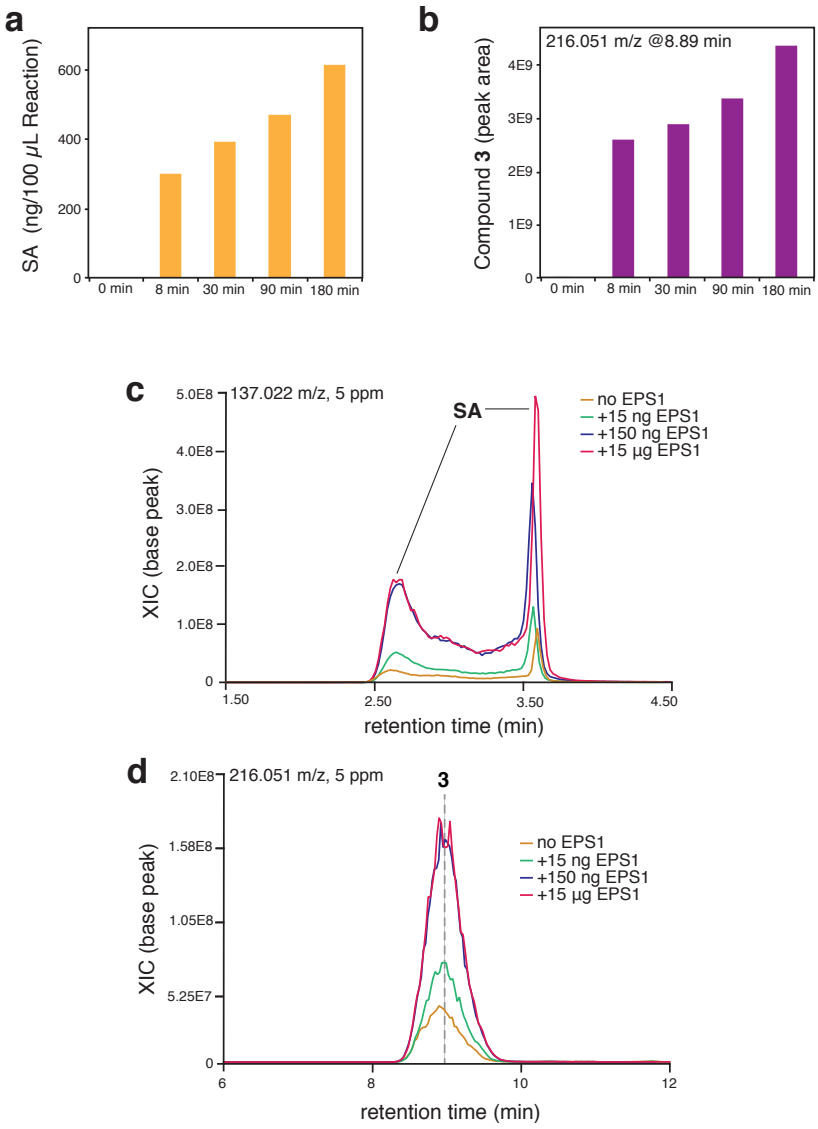
Extended Data Fig. 6| Characterization of the biochemical activity of PBS3 *in vitro*. a, A preassay using SID2 to enzymatically generate isochorismate. 2 µg of SID2 was added a 100 µL reaction containing 50 mM Tris, 1 mM L-glutamate, 5mM chorismate, 5 mM ATP, 5 mM $MgCl_2$. Incubation for 5 minutes at 25 degrees resulted in the production of isochorismate. The reaction also reaches equilibrium that gives approximately equal molar of isochorismate and chorismate. **b,** PBS3 activity requires ATP and Mg^{2+}. **c,** Amino acid preference of PBS3. Individual assays were carried out using isochorismate as the acyl substrate and 1 mM of each of the 20 proteinogenic amino acids as the amino acid substrate. The relative activity was determined based on AMP production. The production of isochorismoyl-amino acid conjugates were also confirmed by LC-HRAM-MS analysis, which are denoted by asterisks.



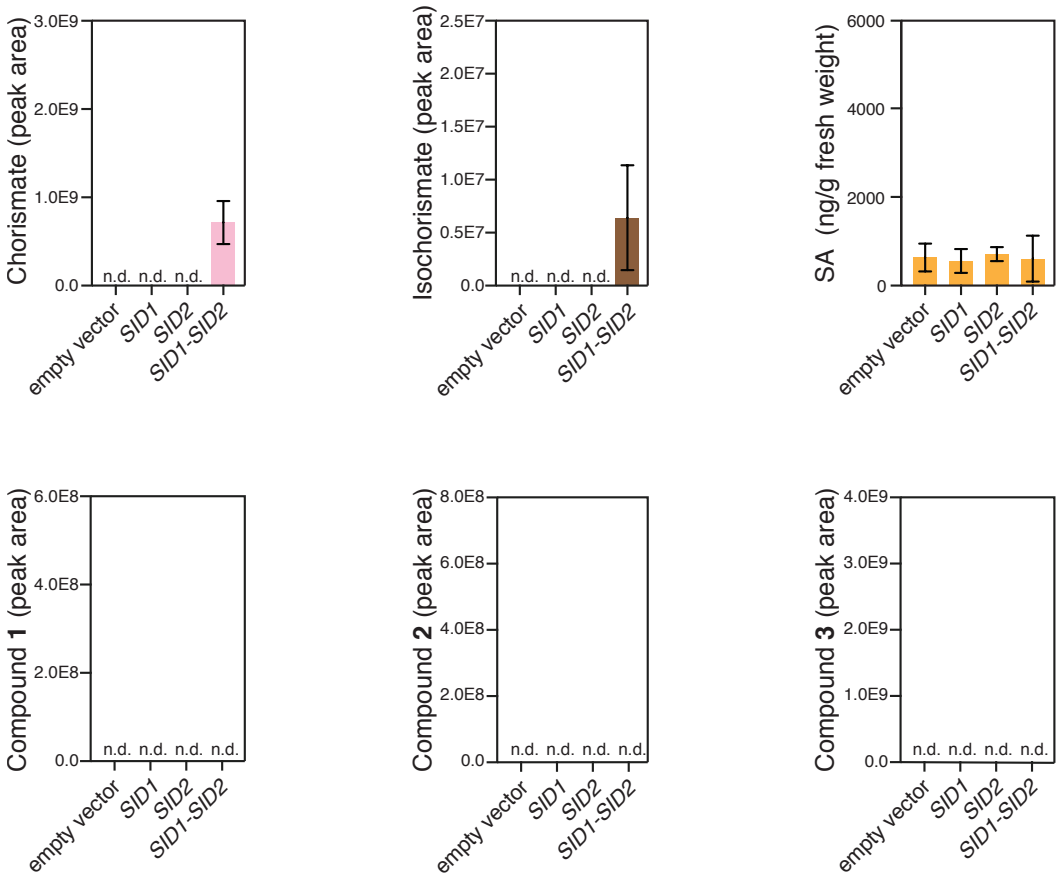
Extended Data Fig. 7| Compound 1 and 2 identity verification across Arabidopsis, *N. benthamiana* and in vitro enzyme assay experiments. a, Overlay of extracted ion chromatogram (XICs) showing the detection of **1** and **2** from LC-HRAM-MS analyses of representative PBS3-EPS1 enzyme assay, *N. benthamiana* co-expressing *PBS3-SID1-SID2* and the Arabidopsis *s3h/dmr6/eps1* triple mutant. **b,** The MS^2 spectra of **1** and **2** measured from the Arabidopsis *s3h/dmr6/eps1* triple mutant sample and the PBS3-EPS1 enzyme assay are overlaid.



Extended Data Fig. 8| Simulated docking of isochorismate and chorismate to the PBS3 crystal structure. a, The crystal structure of PBS3 in complex with SA and AMP (PDB: 4EQL) showing the catalytically nonproductive binding mode of SA in the active site. **b,** The top simulated binding mode of isochorismate in the active site with the calculated binding affinity of -6.9 kcal/mol. **c,** The top simulated binding mode of chorismate in the active site with the calculated binding affinity of -6.8 kcal/mol. The 0.1 kcal/mol decrease in binding affinity compared to isochorismate is likely due to the differential placement of the 4-hydorxyl in chorismate. In contrast, the 2-hydroxyl engages hydrogen bonding with the *p*-hydroxyl of Y112 (**b**). The 4-hydroxyl of chorismate may also impede ATP binding. Active-site residues potentially involved in substrate binding are labeled.



Extended Data Fig. 9| EPS1 enhances the rate of SA production from 1. a, SA production in PBS3 enzyme assays from spontaneous decay of **1**, with an estimated rate of 0.21 pkat. **b,** Production of **3** in PBS3 enzyme assays from spontaneous decay of **1**. **c,** Representative extracted ion chromatograms (XICs) showing EPS1-dependent production of SA from **1**. The chromatograms shown were were resolved using normal-phase chromatography. SA splits into two peaks under such chromatographic condition. **d,** Representative XICs showing EPS1-dependent production of **3** from **1**. Also see Fig. 3f.



Extended Data Fig. 10| Quantification of chorismate, isochorismate, SA and compound 1-3 in *SID1*, *SID2* and *SID1-SID2* infiltration experiments in *N. benthamiana*. Error bars indicate SEM based on biological triplicates (n = 3). n. d., not detected.

Supplementary references

1. Alonso, J. M. Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. *Science* **301**, 653–657 (2003).
2. Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. & Weisshaar, B. GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res.* **40**, D1211–5 (2012).
3. Katagiri, F., Thilmony, R. & He, S. Y. The Arabidopsis Thaliana-Pseudomonas Syringae Interaction. *Arabidopsis Book* **1**, e0039 (2002).
4. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).
5. Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J. & Schilperoort, R. A. A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. *Nature* **303**, 179 (1983).
6. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, (2010).
7. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).
8. Koji, O. *et al.* Amino acid derivative. *Patent* (2008).
9. Wehbe, J., Rolland, V., Fruchier, A., Roumestant, M.-L. & Martinez, J. Enantioselective synthesis of new 4-substituted glutamic acid derivatives. *Tetrahedron: Asymmetry* **15**, 851–858 (2004).
10. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
11. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, W13–7 (2011).
12. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–4 (2014).
13. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
14. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).
15. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
16. Westfall, C. S. *et al.* Structural basis for prereceptor modulation of plant hormones by GH3 proteins. *Science* **336**, 1708–1711 (2012).
17. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
18. D'Auria, J. C. Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340 (2006).
19. Gou, J.-Y., Yu, X.-H. & Liu, C.-J. A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18855–

18860 (2009).
20. Levsh, O. *et al.* Dynamic Conformational States Dictate Selectivity toward the Native Substrate in a Substrate-Permissive Acyltransferase. *Biochemistry* **55**, 6314–6326 (2016).