Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/44/eabd1143/DC1

Supplementary Materials for

Discovery of a previously unknown biosynthetic capacity of naringenin chalcone synthase by heterologous expression of a tomato gene cluster in yeast

Deze Kong, Sijin Li, Christina D. Smolke*

*Corresponding author. Email: csmolke@stanford.edu

Published 30 October 2020, *Sci. Adv.* **6**, eabd1143 (2020)
DOI: 10.1126/sciadv.abd1143

This PDF file includes:

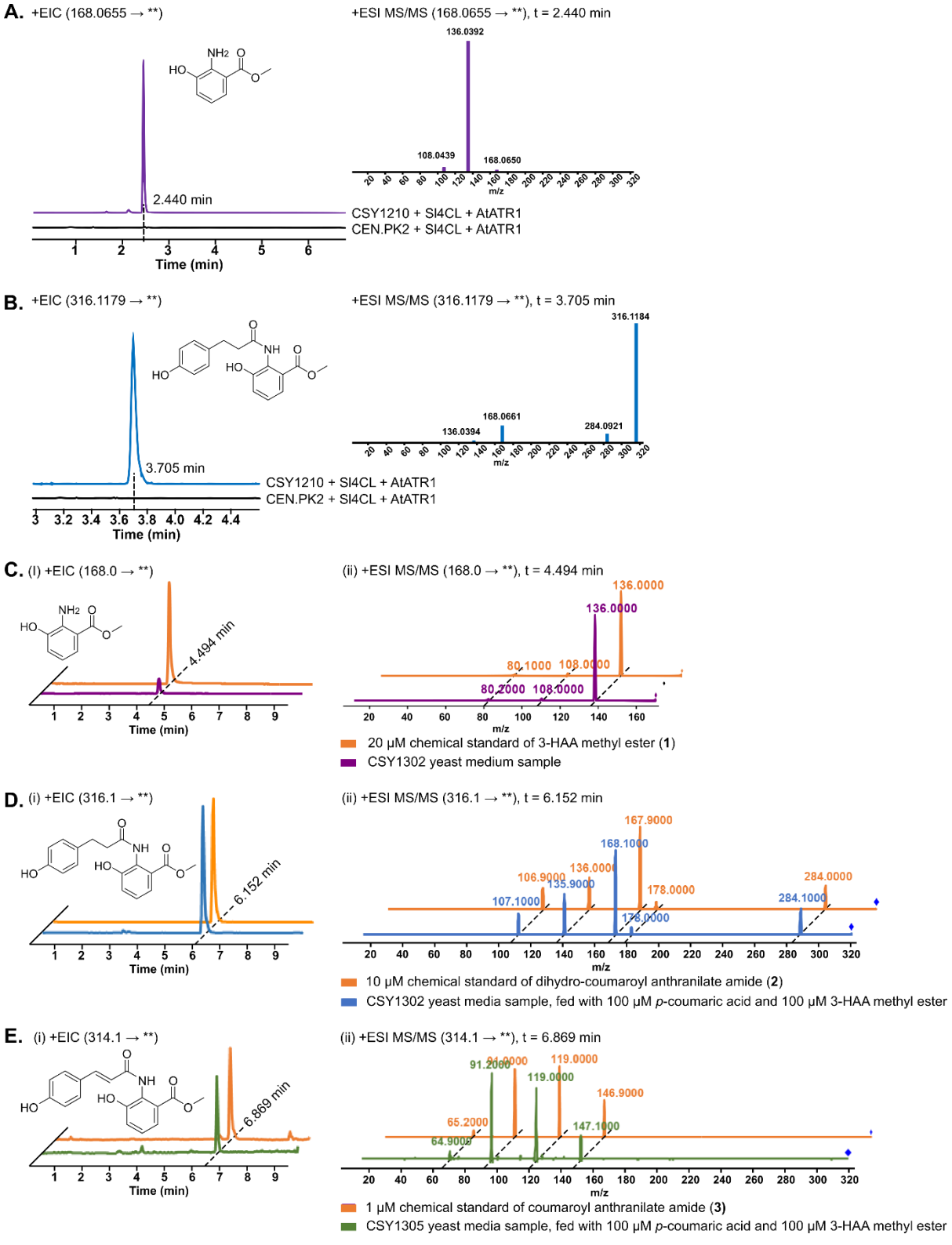
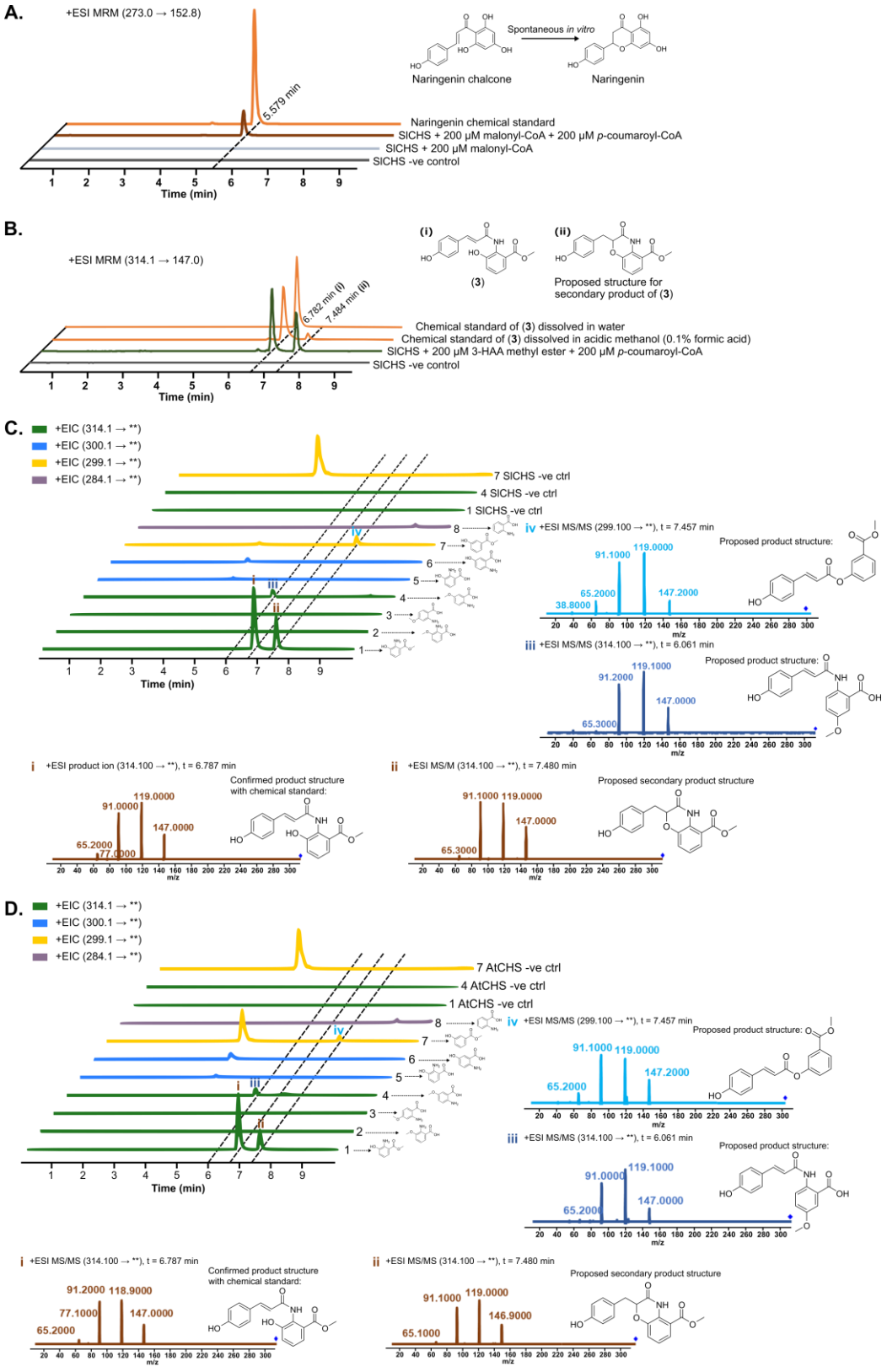


Figure S1. Identification and validation of previously unknown compound production. **(A-B)** Extracted ion chromatograms (EICs) and tandem mass (MS/MS) spectrums detected by qToF-MS are shown for CEN.PK2/CSY1210 + Sl4CL + AtATR1 for detection of 3-HAA methyl ester (**1**) **(A)** and dihydro-coumaroyl anthranilate amide (**2**) **(B)** in CSY1210 + Sl4CL

+ AtATR1. CEN.PK2 is a wild-type yeast strain. CSY1210 expresses *SlCHS*, *SlCYP*, *SlMT1/2/3*. Yeast strains were cultured in synthetic dropout out media (with 2% dextrose) supplemented with 100 μM *p*-coumaric acid for 72 hours at 25°C. Each trace is representative of three biologically independent replicates. **(C-E)** Extracted ion chromatograms (EICs) and tandem mass (MS/MS) spectrums detected by LC-MS/MS are shown for 3-HAA methyl ester (**1**) **(C)**, dihydro-coumaroyl anthranilate amide (**2**) **(D)**, and coumaroyl anthranilate amide (**3**) **(E)** detected from corresponding yeast strain culture medium samples, in comparison to authentic chemical standard. CSY1302 expresses *SlCHS*, *SlCYP*, *SlMT2*. CSY1305 is built from CSY1302 with genotype *TSC13*Δ*::GhECR2*. '**' indicates a thorough MS scan. Each trace of a yeast strain is representative of two biologically independent replicates.



E.

Rxn	Substrate			Enzyme	Product	
	p-Coumaroyl -CoA	3-HAA methyl ester	Malonyl -CoA		(3)	Naringenin
1	+	–	+	SlCHS	0	89.12 ± 0.07
2	+	+	–	SlCHS	100.00 ± 0.05	0
3	+	+	+	SlCHS	14.53 ± 0.01	83.80 ± 0.02
4	+	–	+	–	0	0
5	+	+	–	–	0	0
6	+	–	+	AtCHS	0	100.00 ± 0.04
7	+	+	–	AtCHS	61.61 ± 0.01	0
8	+	+	+	AtCHS	10.53 ± 0.03	96.74 ± 0.01

% Ion count: 100, 80, 60, 40, 20, 0

Figure S2. *In vitro* characterization results for compound production by SlCHS/AtCHS. **(A-B)** Multiple reaction monitoring (MRM) analyses detected by LC-MS/MS are shown to validate the production of naringenin **(A)** and coumaroyl anthranilate amide (**3**) **(B)** under corresponding *in vitro* reaction conditions, in comparison to authentic standards. Each trace is representative of two independent replicates. **(C-D)** Extracted ion chromatograms (EICs) and tandem mass (MS/MS) spectrums by LC-MS/MS are shown for potential condensation products (i-iv) from *in vitro* reaction with 200 μM *p*-coumaroyl-CoA, 200 μM of an anthranilic acid analog, and 4 μg of SlCHS **(C)** or AtCHS **(D)** protein. Anthranilic acid analogues label: (1): 3-HAA methyl ester; (2): 2-amino-3-methoxybenzoic acid; (3): 2-amino-4-methoxybenzoic acid; (4): 2-amino-5-methoxybenzoic acid; (5): 3-HAA; (6): 2-amino-5-hydroxybenzoic acid; (7): 3-hydroxybenzoic methyl ester; (8): anthranilic acid. **(E)** Summary of coumaroyl anthranilate amide (**3**) and naringenin production by SlCHS/AtCHS. Table sign: '+'/ '-' indicates the presence/absence of 200 μM *p*-coumaroyl-CoA, 200 μM 3-HAA methyl ester, 200 μM malonyl-CoA, or 4 μg purified SlCHS protein.+: presence of a substrate, -: absence of a substrate. MRM (314.1 → 147.0) and MRM (273.0 → 152.8) detect the production of (**3**) and naringenin, respectively. The ion counts are normalized by the highest ion count across rxn 1-8 by each column; standard deviation shows the percentage error among two independent replicates. '**' indicates a thorough MS scan.

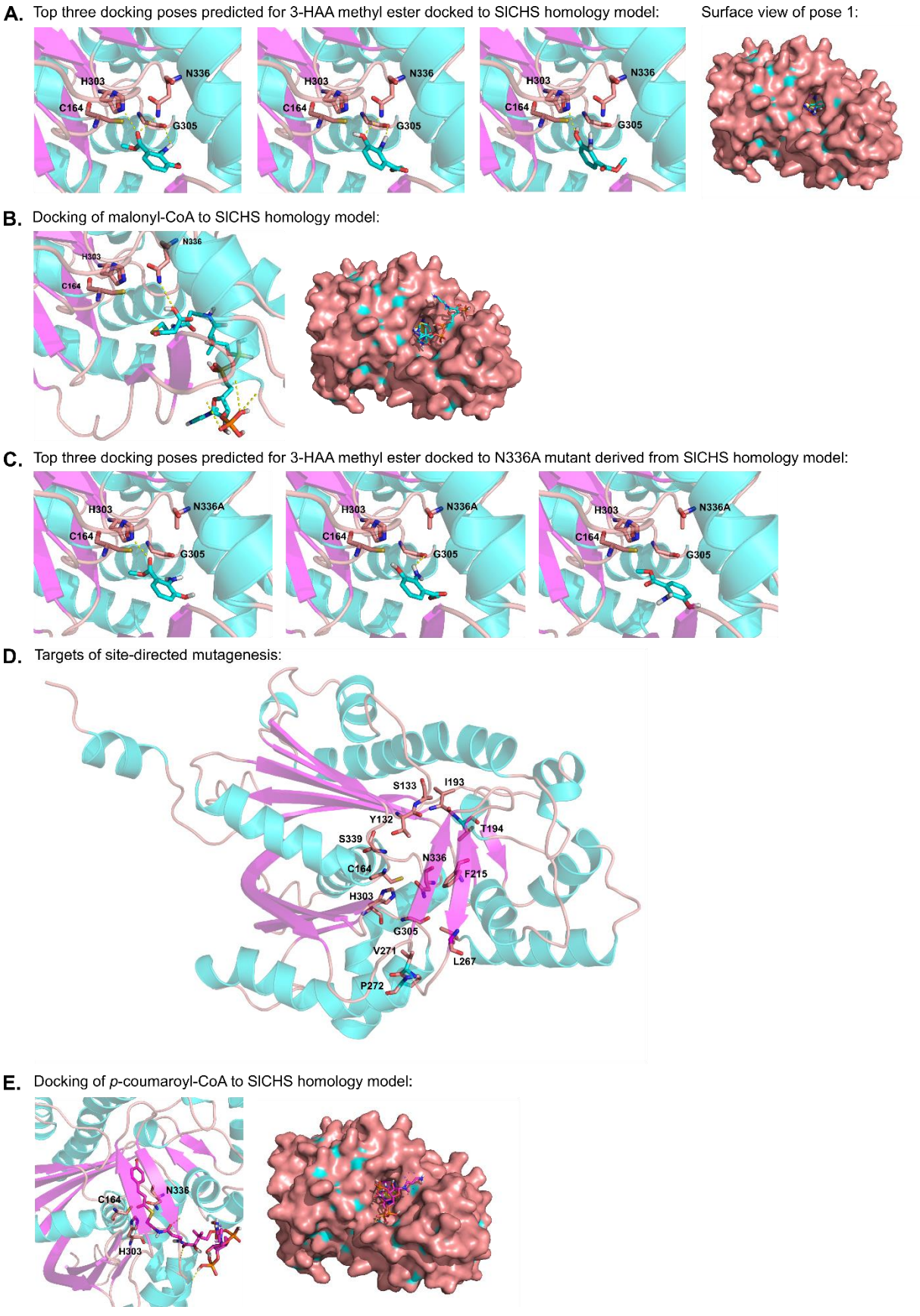


Figure S3. SlCHS homology model and substrate docking simulations. (A) Docking of 3-

HAA methyl ester to SlCHS homology model active site. **(B)** Docking of malonyl-CoA to SlCHS homology model active site. **(C)** Docking of 3-HAA methyl ester to SlCHS N366A mutant model active site. **(D)** Residues chosen for site-directed mutagenesis. **(E)** Docking of *p*-coumaroyl-CoA to SlCHS homology model active site. Dotted line: hydrogen bond interaction.

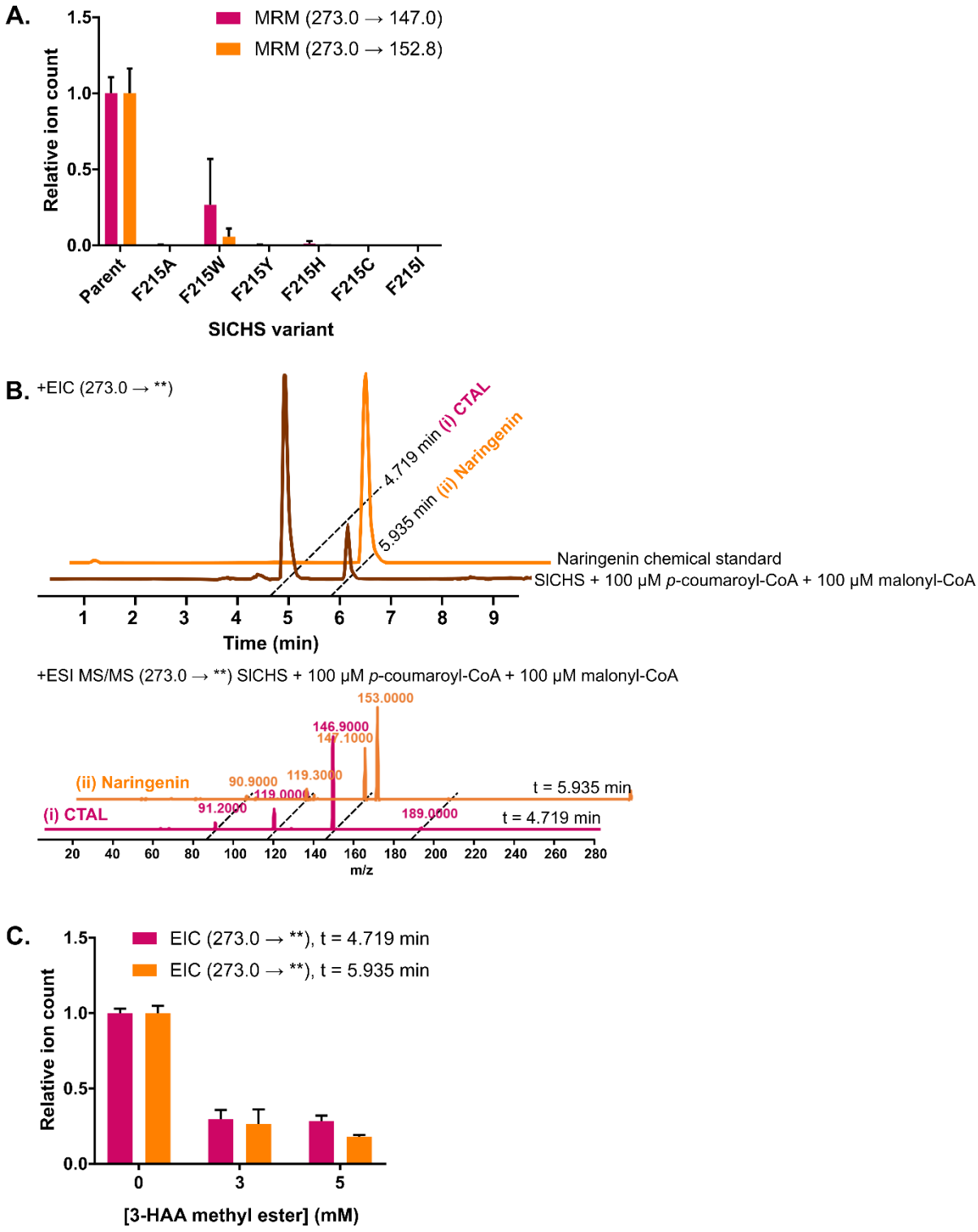
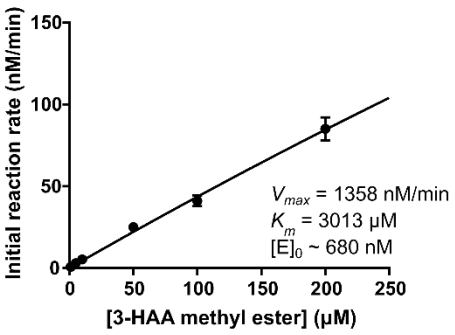


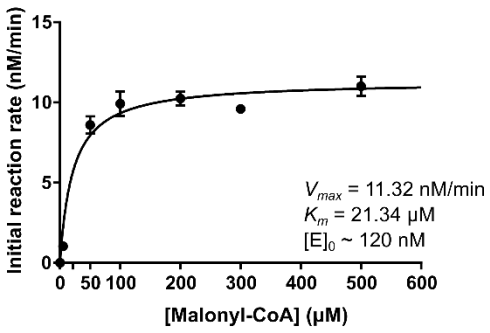
Figure S4. Production of 4-coumaroyltriacetic acid lactone (CTAL) by SlCHS. (A) Relative production of CTAL and naringenin chalcone by F215 mutants in yeast. MRM (273.0 → 147.0) and MRM (273.0 → 152.8) detect the production of CTAL and naringenin chalcone, respectively. **(B)** Extracted ion chromatogram (EIC) is shown for SlCHS *in vitro* reaction sample, in comparison to authentic standard of naringenin. Peaks at t = 4.719 min and t = 5.935

min indicate the production of (i) CTAL and (ii) naringenin, respectively. Tandem mass (MS/MS) spectrums of the peaks (i) and (ii) detected from SlCHS *in vitro* reaction sample are shown. Each trace of a yeast strain is representative of two biologically independent replicates. **(C)** *In vitro* production of CTAL and naringenin in SlCHS canonical activity inhibition assays. EIC (273.0 → **) at t = 4.719 min and EIC (273.0 → **) at t = 5.935 min detect the production of CTAL and naringenin, respectively. Compound production was measured from reaction mixtures fed with 100 µM malonyl-CoA, 200 µM *p*-coumaroyl-CoA and 0, 3, or 5 mM 3-HAA methyl ester inhibitor at the end of the time course (31 min) of a kinetic assay. '**' indicates a thorough MS scan. Data shows the mean of two biologically independent replicates with error bar indicating standard deviation.

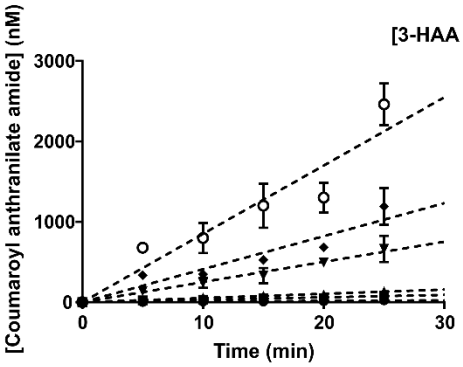
A.



B.

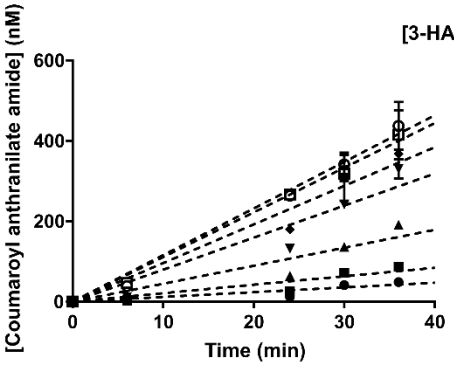


C.



[3-HAA methyl ester]	Gradient (nM/min)
1 μM	0.82 ± 0.05
5 μM	3.05 ± 0.18
10 μM	5.26 ± 0.17
50 μM	25.12 ± 0.73
100 μM	41.09 ± 3.28
200 μM	85.00 ± 7.09

D.



[3-HAA methyl ester]	Gradient (nM/min)
0.4 mM	1.18 ± 0.17
0.8 mM	2.11 ± 0.28
1.6 mM	4.47 ± 0.54
3 mM	7.96 ± 0.78
5 mM	9.61 ± 0.96
10 mM	11.59 ± 0.38
15 mM	11.10 ± 0.28

E. Kinetic model used to simulate progress curve data fitting in DynaFit:

E + M → EM : *k1*
EM + M → EM2 : *k2*
EM2 + M → EM3 : *k3*
EM3 + M → EM4 : *k4*
EM3 → E + P : *kp*

E: enzyme concentration, 108 nM
M: malonyl-CoA concentration, varied from 5 μM to 100 μM
P: product concentration (nM) is fitted by each dataset

[3-HAA methyl ester] (mM)	[Malonyl-CoA] (μM)	Reaction rate* (nM/min)	Relative RMS (%)**
	5	0.7589	9.14624
	50	2.544	5.68088
0	100	3.066	7.32803
	5	0.0944	4.76973
	50	1.526	3.7826
3	100	1.969	5.27058
	5	0.0523	17.4288
	50	0.6571	8.45193
5	100	0.8697	4.9313

*Reaction rate is obtained by fitting first derivative of progress curve data obtained from DynaFit to equation *M(1-exp(-ax))* in MATLAB 2017a, in which '*M*' represents the reaction rate.
**Relative RMS (%) is obtained from DynaFit progress curve fitting for each dataset.

Figure S5. Additional SlCHS *in vitro* kinetic data. (A) Kinetic characterization of SlCHS amide formation at low substrate concentration range. Reaction rates were measured at 1, 5, 10, 50, 100 and 200 μM 3-HAA methyl ester. **(B)** Kinetic characterization of SlCHS synthesis of naringenin chalcone (canonical activity). Reaction rates were measured at 5, 50, 100, 200, 300 and 500 μM malonyl-CoA. **(A-B)** Kinetic curve is fitted by GraphPad Prism 7 Michaelis-Menten nonlinear regression, data shows slope estimated from progress curve by GraphPad Prism 7 with error bar indicating the relative error in slope estimation. **(C-D)** Progress curve data for coumaroyl anthranilate amide (**3**) synthesis kinetic shown in **(A) (C)** or Fig. 5A **(D)**. Slope (reaction rate) is calculated by GraphPad Prism 7 linear regression tool, data shows the mean of two independent replicates with error bar indicating the standard deviation. **(E)** Kinetic model used for progress curve data fitting by DynaFit for SlCHS canonical activity inhibition assays. Refer to Table S2 for curve fitting results. Table: DynaFit progress curve fitting results for SlCHS canonical activity inhibition assays.

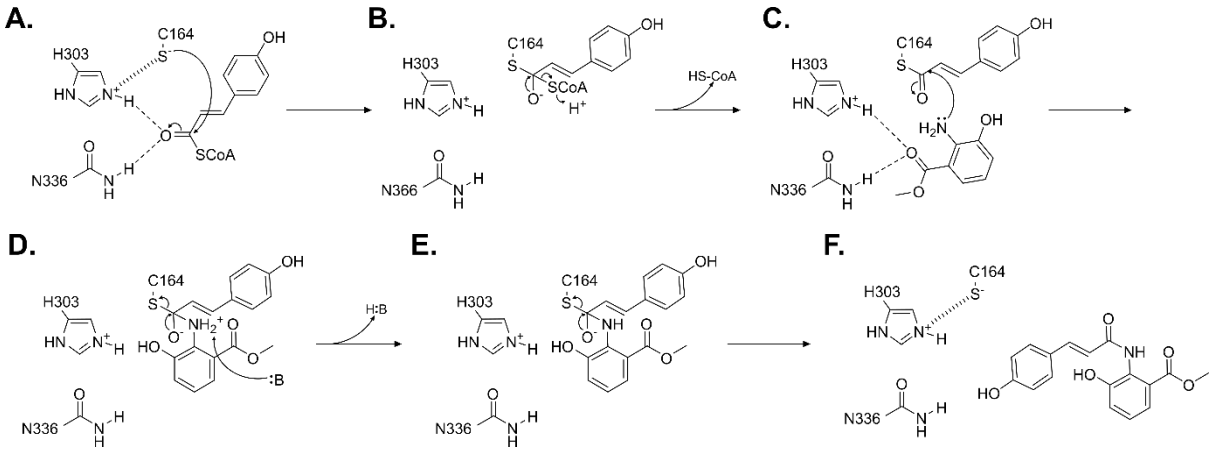


Figure S6. Proposed SlCHS catalytic mechanism for amide formation. (A) Activated cysteine (C164) initiating nucleophilic attack to the carbonyl group on *p*-coumaroyl-CoA. **(B)** Loading of *p*-coumaroyl-CoA onto C164. **(C)** Amine group of 3-HAA methyl ester attacking the carbonyl group. **(D)** Formation of the amide bond and deprotonation of the amine group by a general base. **(E-F)** Release of the final product. Dashed line: ionic interaction; dotted line: hydrogen bond interaction. C164-H303-N336 is the catalytic triad of SlCHS.

Table S1. Summary of compound production from combinatorial expression of tomato cluster genes. Table sign: '+'/ '-' indicates the presence/absence of a gene or a compound. MRM (168.0 → 136.0) and MRM (316.1 → 168.0) detect the production of 3-HAA methyl ester (**1**) and dihydro-coumaroyl anthranilate amide (**2**), respectively. Results for each test group is representative of three biologically independent replicates.

Group	Fed substrate	Genes			Product	
	p-Coumaric acid	*SlCHS*	*SlMT2*	*Sl4CL*	(1)	(2)
1	+	+	+	+	+	+
2	-	+	+	+	+	-
3	+	+	+	-	+	-
4	+	-	+	+	+	-
5	+	+	-	+	-	-
6	-	+	-	-	-	-
7	-	-	+	-	+	-
8	-	-	-	+	-	-

Table S2. Parameters fitted to kinetic curve.

A. Without inhibition coefficients.

[3-HAA methyl ester] (mM)	*n*	V_{max} (nM/min)	K_m (μM)	Adjusted R-square
0	**1**	**3.64**	**20.13**	**0.9988**
0	1.2	3.28	14.88	0.9941
0	1.5	3.014	11.08	0.985
3	1	3.267	62.78	0.9855
3	1.5	2.368	34.07	0.9994
3	**1.7**	**2.259**	**32.43**	**1**
5	1	1.463	65.83	0.991
5	1.2	1.211	44.91	0.9973
5	**1.5**	**1.057**	**35.88**	**1**

B. With competitive inhibition coefficient (K_c)*.

*V_{max} was fixed at 3.64 nM/min and K_m fixed at 20.13 μM for the fitting.

[3-HAA methyl ester] (mM)	*n*	K_c (μM)	Adjusted R-square	RMSE
3	**1**	**1040**	**0.9892**	**0.1036**
3	1.1	651.6	0.982	0.1338
3	1.2	426.2	0.9656	0.1852
3	1.3	286.7	0.9424	0.2398
5	**1**	**379.3**	**0.9669**	**0.07916**
5	1.1	272.4	0.948	0.09921
5	1.2	196.8	0.925	0.1192
5	1.3	142.8	0.899	0.1382

C. With uncompetitive inhibition coefficient (K_u)*.

*V_{max} was fixed at 3.64 nM/min and K_m fixed at 20.13 μM for the fitting.

[3-HAA methyl ester] (mM)	*n*	K_u (μM)	Adjusted R-square	RMSE
3	**1**	**3682**	**0.8932**	**0.8932**
3	1.1	2647	0.8948	0.8948
3	1.3	1585	0.8963	0.8963
3	1.5	1053	0.8971	0.8971
5	1	1298	0.733	0.2247
5	1.1	1092	0.7377	0.2227

5	1.3	789.6	0.7449	0.2197
5	**1.5**	**583.5**	**0.7504**	**0.2173**

D. With competitive and uncompetitive inhibition coefficients (K_c, K_u)*.

*V_{max} was fixed at 3.64 nM/min and K_m fixed at 20.13 μM for the fitting.

[3-HAA methyl ester] (mM)	*n*	K_c (μM)	K_u (μM)	Adjusted R-square	RMSE
3	1	1216	25880	0.9855	0.1203
3	1.1	1056	7267	0.9904	0.0981
3	1.3	774.3	2742	0.9966	0.05784
3	1.5	549.9	1551	0.9994	0.02398
3	**1.7**	**377**	**1006**	**1**	**0.00302**
5	1	654.8	3497	0.994	0.04129
5	1.1	618.6	2320	0.9965	0.03166
5	1.3	470.9	1354	0.9993	0.01434
5	**1.5**	**341.4**	**897.1**	**1**	**0.0001239**

n: Hill coefficient for cooperativity approximation

K_c: competitive inhibition coefficient

K_u: uncompetitive inhibition coefficient

Adjusted R-squared: R-square adjusted based on number of independent variables as a parameter against over-fitting.

RMSE: Root mean square error.

Table S3. Plasmids and yeast strains constructed in this study.

Plasmid	Construct summary	Marker(s)	Expression organism
pCS4544	P_{TEF1p}-*Sl4CL*-T_{CYC1}	TRP, AmpR	*S. cerevisiae*
pCS4545	P_{TEF1p}-*AtATR1*-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4546	P_{GPD}-*SlCHS*-T_{ADH1}	LEU, AmpR	*S. cerevisiae*
pCS4547	P_{PYK1}-*SlMT1*-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4548	P_{PYK1}-*SlMT2*-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4549	P_{PYK1}-*SlMT3*-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4550	P_{GPD}-*SlCHS*-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4551	P_{GPD}-*SlCHS*_C164A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4552	P_{GPD}-*SlCHS*_H303A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4553	P_{GPD}-*SlCHS*_N336A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4554	P_{GPD}-*SlCHS*_F215A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4555	P_{GPD}-*SlCHS*_S133A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4556	P_{GPD}-*SlCHS*_S339A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4557	P_{GPD}-*SlCHS*_T132A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4558	P_{GPD}-*SlCHS*_I193A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4559	P_{GPD}-*SlCHS*_T194-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4560	P_{GPD}-*SlCHS*_L267A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4561	P_{GPD}-*SlCHS*_L271A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4562	P_{GPD}-*SlCHS*_L272A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4563	P_{GPD}-*SlCHS*_G305A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4564	P_{GPD}-*SlCHS*_D270A-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4565	P_{GPD}-*SlCHS*_D270E-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4566	P_{GPD}-*SlCHS*_D270N-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4567	P_{GPD}-*SlCHS*_C164S-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4568	P_{GPD}-*SlCHS*_S339T-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4569	P_{GPD}-*SlCHS*_F215W-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4570	P_{GPD}-*SlCHS*_F215Y-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4571	P_{GPD}-*SlCHS*_F215H-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4572	P_{GPD}-*SlCHS*_F215C-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4573	P_{GPD}-*SlCHS*_F215I-T_{CYC1}	URA, AmpR	*S. cerevisiae*
pCS4574	pET28 w/ *At4CL*	KanR	*E. coli*
pCS4575	pET28 w/ *SlCHS*	KanR	*E. coli*
pCS4576	pET28 w/ *AtCHS*	KanR	*E. coli*
pCS4577	pCS3410 w/ *TSC13*_gRNA1	G418	*S. cerevisiae*

pCS4578	pCS3410 w/ *TSC13*_gRNA2	G418	*S. cerevisiae*
pCS4579	pCS3410 w/ *SlCHS*_gRNA	G418	*S. cerevisiae*

Strain	Genotype
CEN.PK2-1D	MATα ura3-52, trp1-289; leu2-3/112, his3Δ1, MAL2-8C, SUC2
CSY1210	CEN.PK2-1D; pYES1L-P_{GPD}-*SlCHS*-T_{ADH1}, P_{TEF1}-*SlMT1*-T_{CYC1}, P_{PYK1}-*SlMT2*-T_{MFA1}, P_{PGK1}-*SlMT3*-T_{PHO5}, P_{HXT7}- *SlCYP*-T_{PGK1}
CSY1301	CEN.PK2-1D; YMR206WΔ:: P_{GPD}-*SlCHS*-T_{ADH1}, P_{TEF1}-*Sl4CL*-T_{CYC1}, HIS5, P_{PGK1}-*SlMT1*-T_{PHO5}
CSY1302	CEN.PK2-1D; YMR206WΔ:: P_{GPD}-*SlCHS*-T_{ADH1}, P_{TEF1}-*Sl4CL*-T_{CYC1}, HIS5, P_{PGK1}-*SlMT2*-T_{PHO5}
CSY1303	CEN.PK2-1D; YMR206WΔ:: P_{GPD}-*SlCHS*-T_{ADH1}, P_{TEF1}-*Sl4CL*-T_{CYC1}, HIS5, P_{PGK1}-*SlMT3*-T_{PHO5}
CSY1304	CSY1302; *TSC13*Δ559-942
CSY1305	CSY1302; *TSC13*Δ:: *GhECR*
CSY1306	CSY1302; *TSC13*Δ:: *MdECR*
CSY1307	CSY1305; *SlCHS*Δ

Table S4. MRM transitions and detection parameters developed in this study.

Compound	MRM transition	Fragmentor	CollisionEnergy (V)
3-HAA methyl ester (**1**)	168.0 → 136.0	100	10
Dihydro-coumaroyl anthranilate amide (**2**)	316.1 → 168.0	100	10
Coumaroyl anthranilate amide (**3**)	314.1 → 147.0	100	20/40
Naringenin chalcone	273.0 → 152.8	100	10
Naringenin	273.0 → 152.8	100	10
4-coumaroyltriacetic acid lactone (CTAL)	273.0 → 147.0	100	40