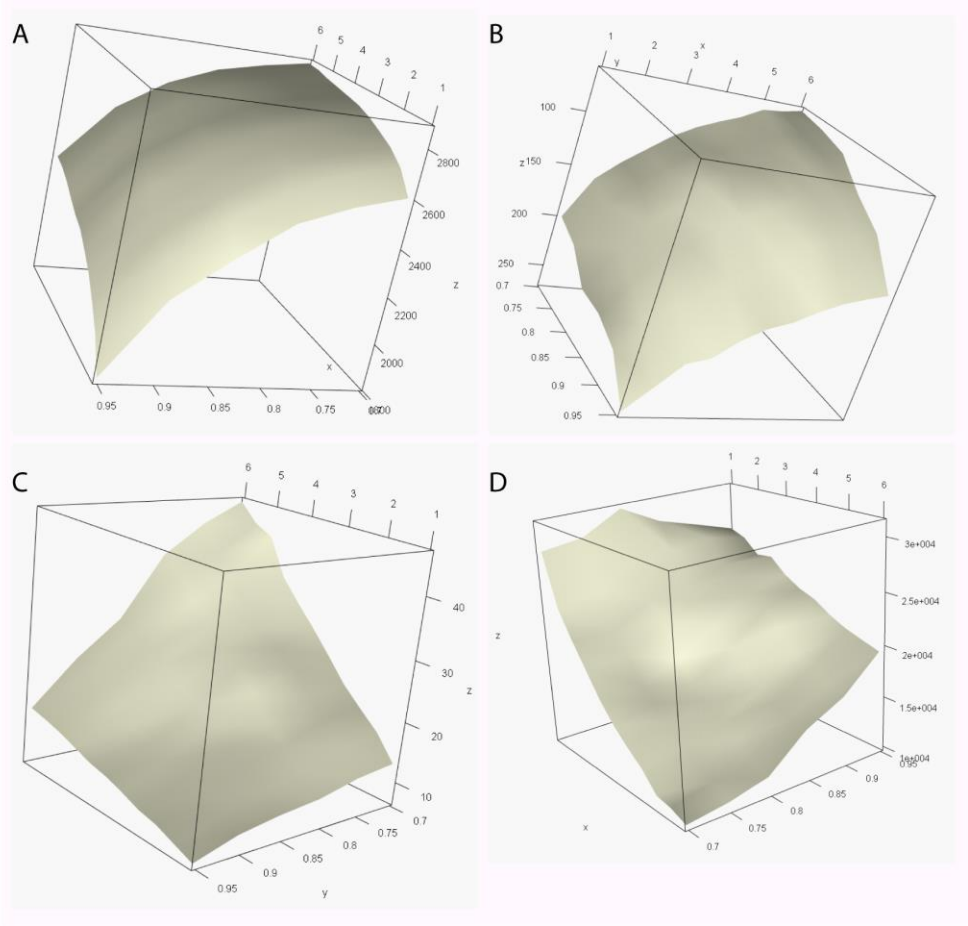
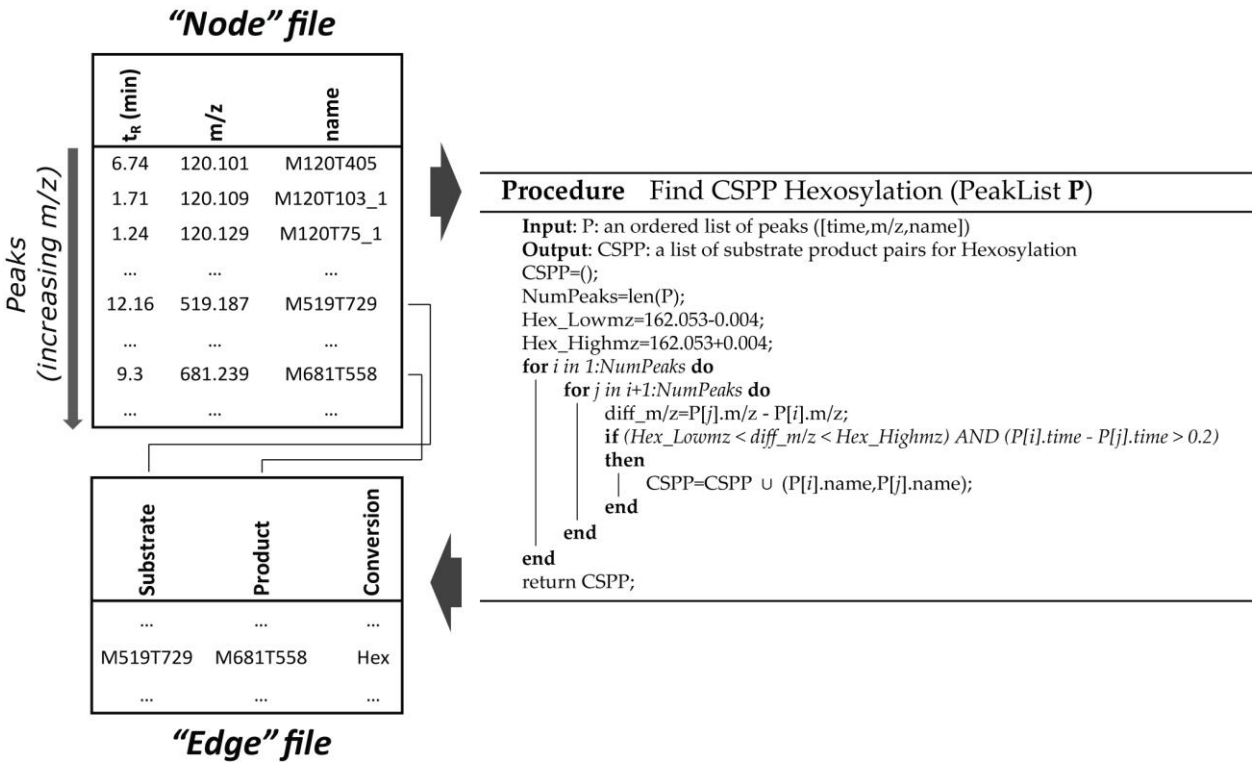
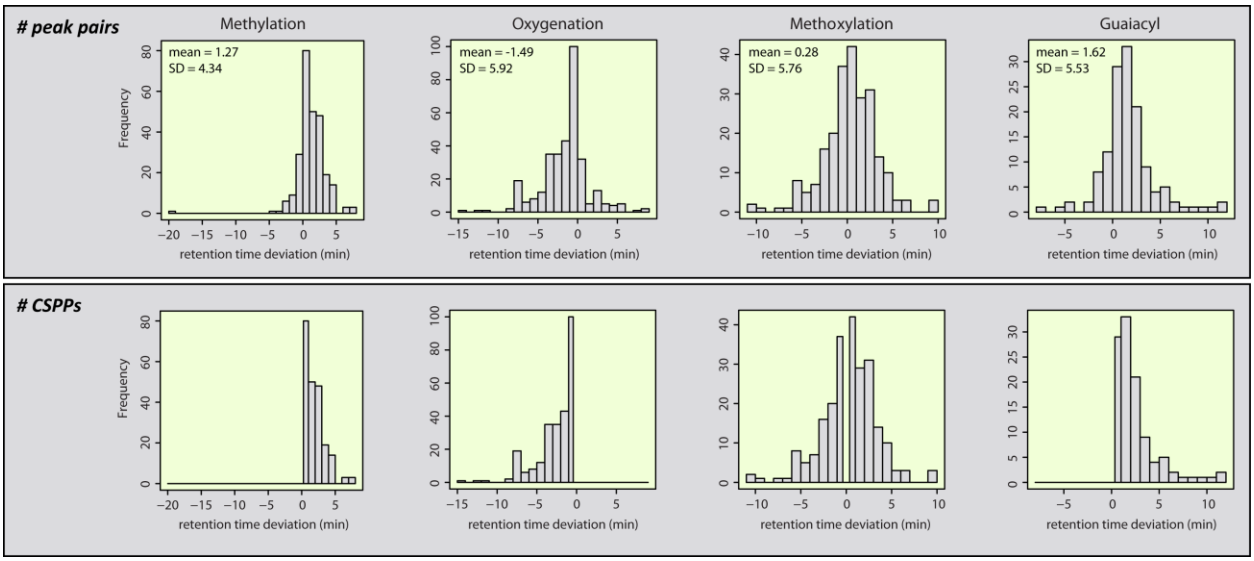
SUPPLEMENTAL FIGURES



Supplemental Figure 1. Peak Grouping Surface Plots. In order to group peaks belonging to the same compound, various settings of the retention time window and the correlation threshold were evaluated on the number of peaks that could be grouped (A), the number of peak groups (B), the average number of assigned peaks per peak group (C) and on an arbitrarily trait representing the co-optimization of both the number of peak groups and the number of assigned peaks to a peak group (D). For the latter trait, to allow that the number of peak groups had a similar impact as the number of assigned peaks, the number of peak groups was multiplied by a correction factor, i.e., the quotient of the average number of assigned peaks and the average number of peak groups. The trait was subsequently obtained by multiplying the corrected number of peak groups with the number of assigned peaks.

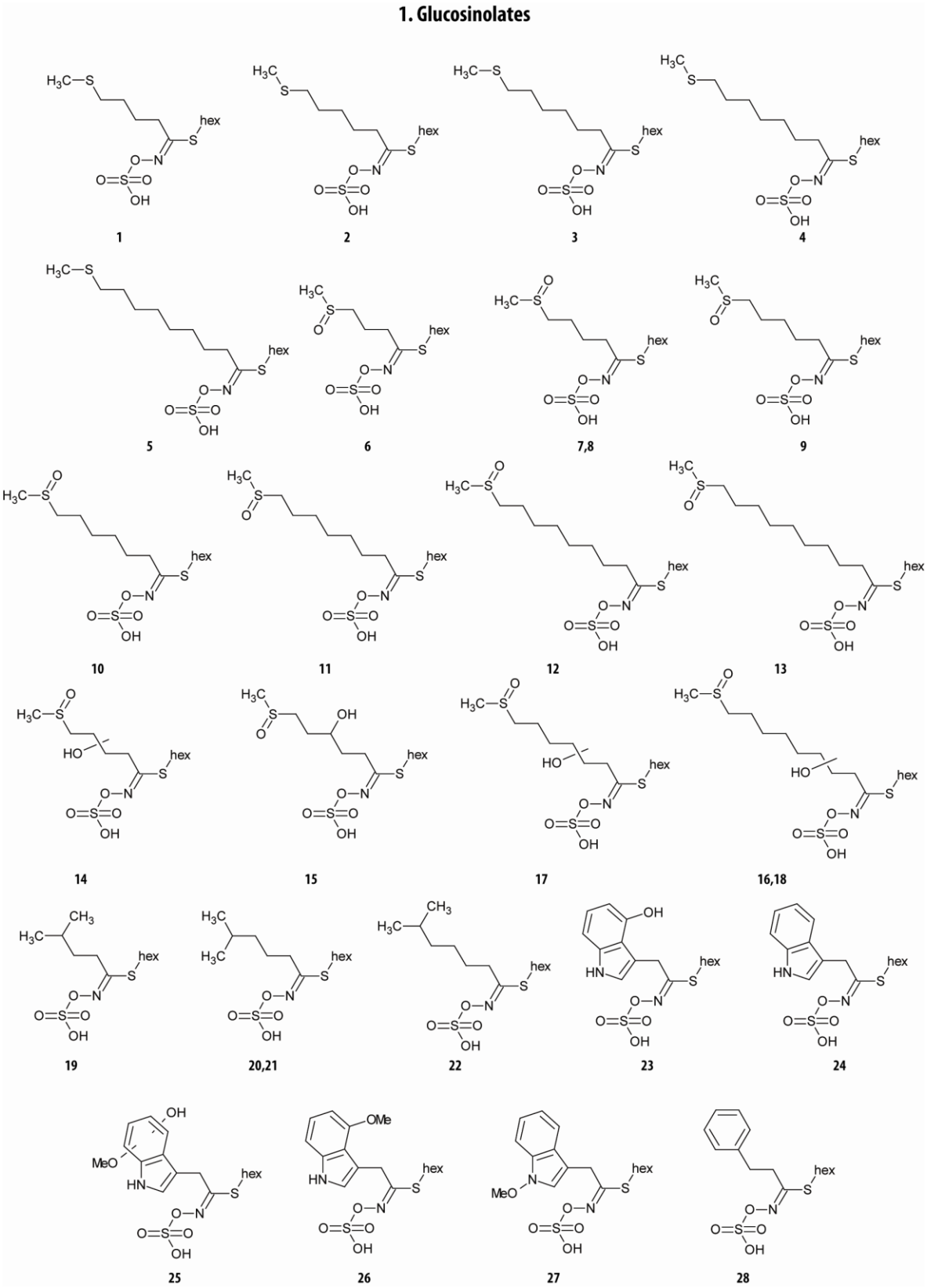


Supplemental Figure 2. CSPP Generation Algorithm. Integrated and aligned chromatogram peaks are ordered with increasing m/z ("node" file) and used for the "edge" file generation (see Methods for explanation). The CSPP algorithm, based on searching peak pairs of which the m/z values differ by the mass expected for e.g., a hexosylation (162.053 Da) and the elution order is in agreement with that expected upon a reversed phase separation – i.e., the hexosylated product eluting earlier – is mentioned in pseudo-code.

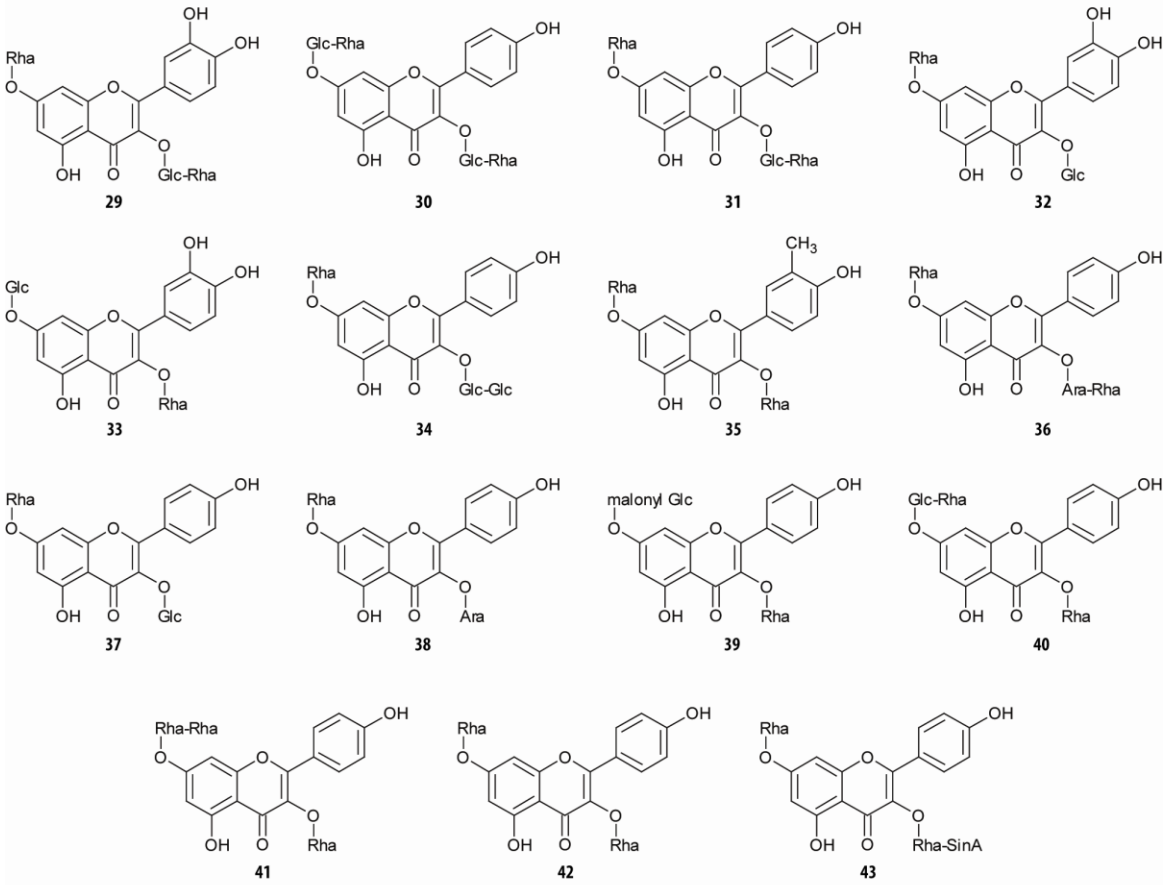


Supplemental Figure 3. Retention Time Difference Distributions. Distribution of the number of peak pairs or CSPPs versus the retention time difference between both peaks of each peak pair or between "substrate" and "product" peak of each CSPP, respectively. For a particular mass difference, CSPPs are obtained by truncating the peak pair-based retention time difference distribution. More specifically, when the "product" is more polar or apolar than the "substrate", only negative or positive retention time differences are allowed, respectively. In any case, a retention time difference of 0 ± 0.2 min is excluded. SD, standard deviation.

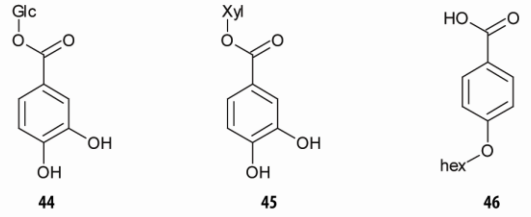
Supplemental Figure 4. Annotated Molecular Structures.

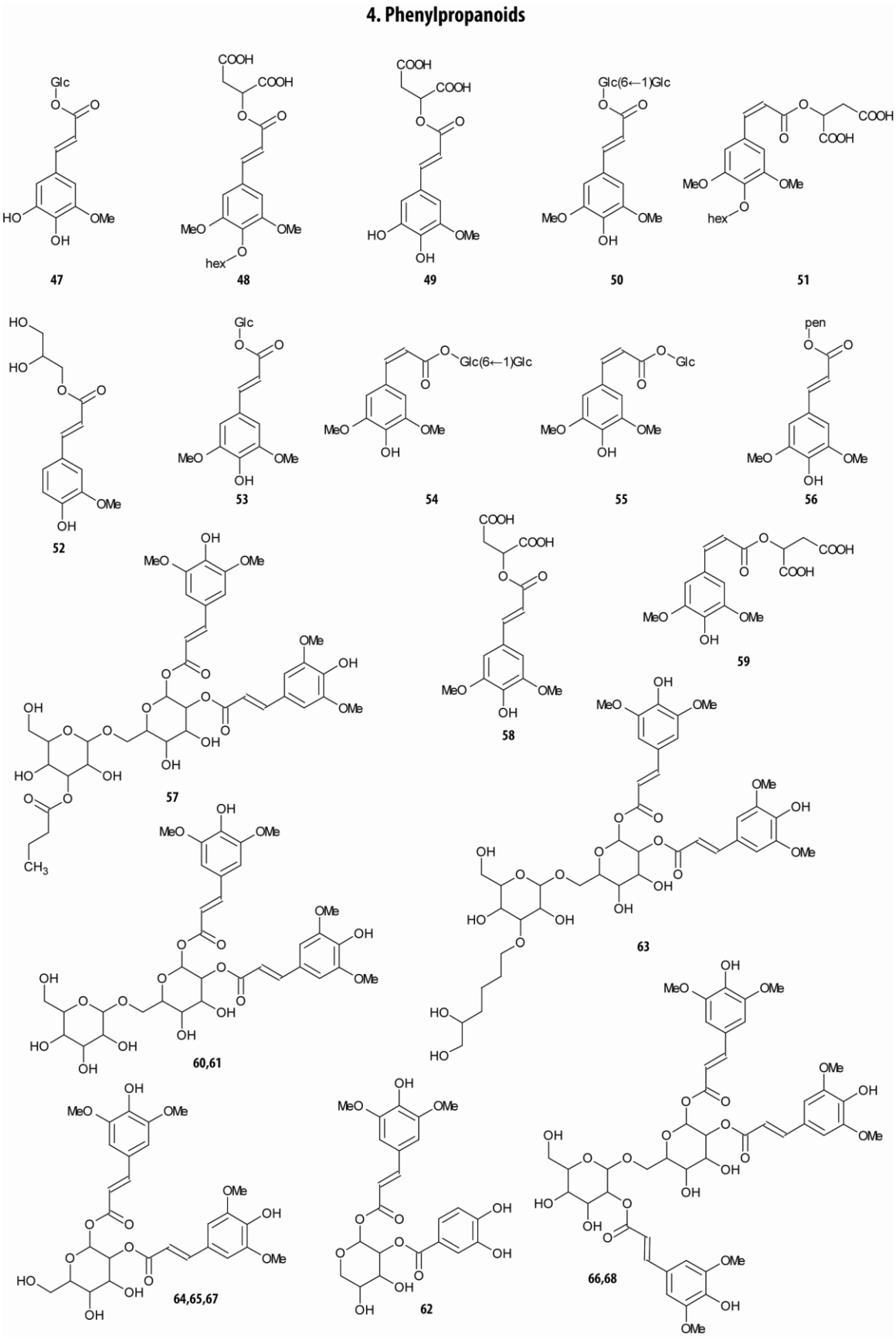


2. Flavonoids

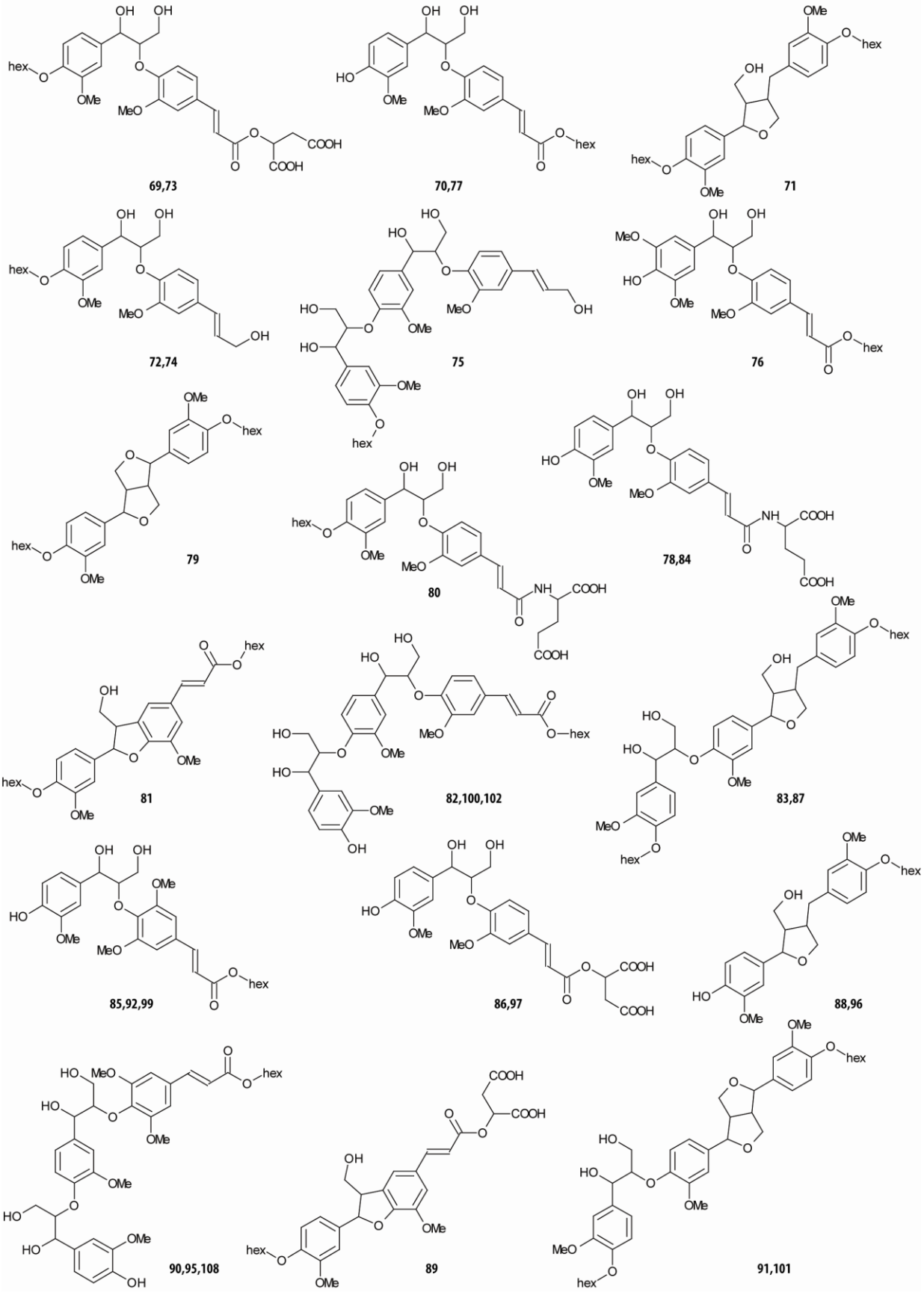


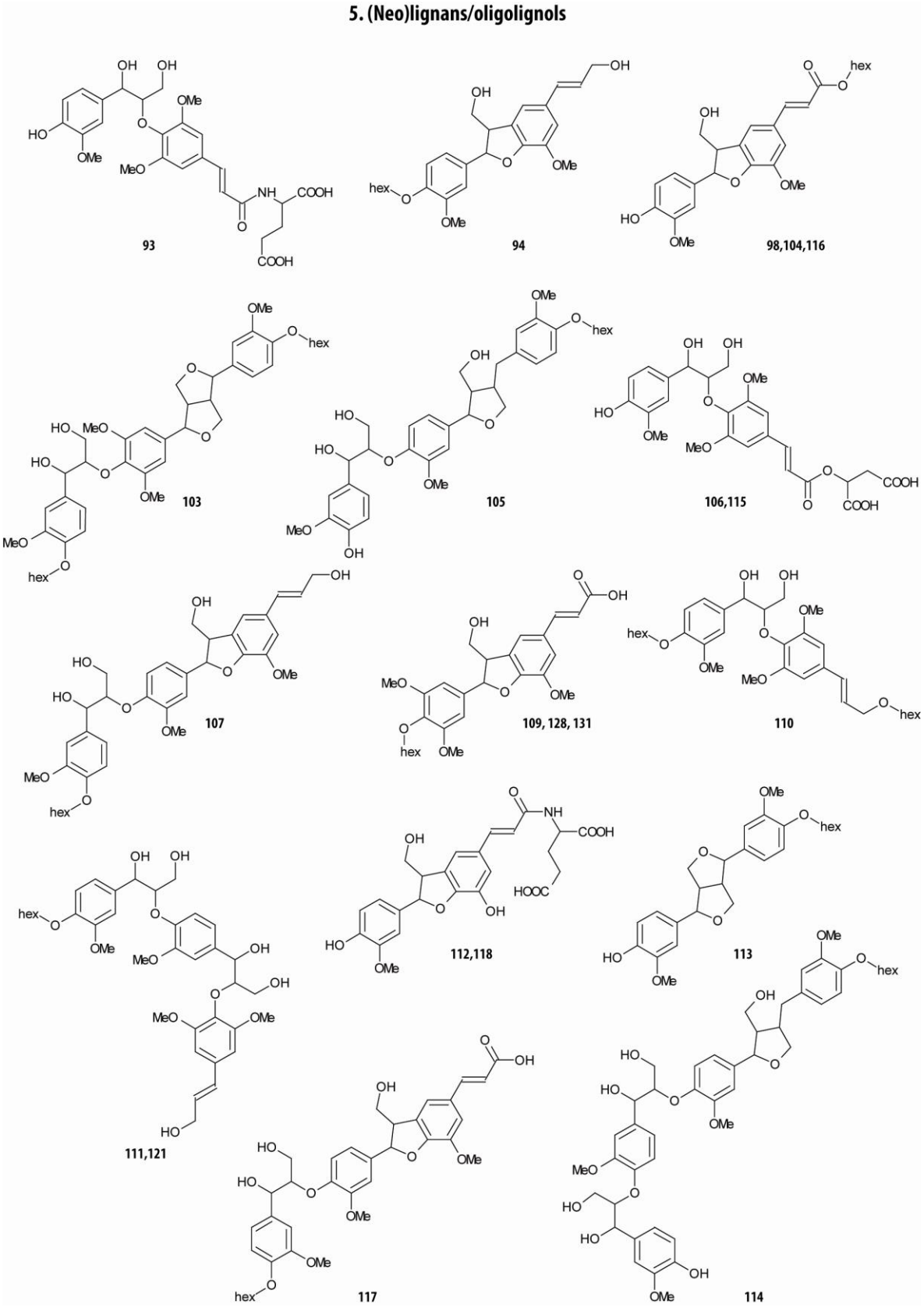
3. Benzenoids



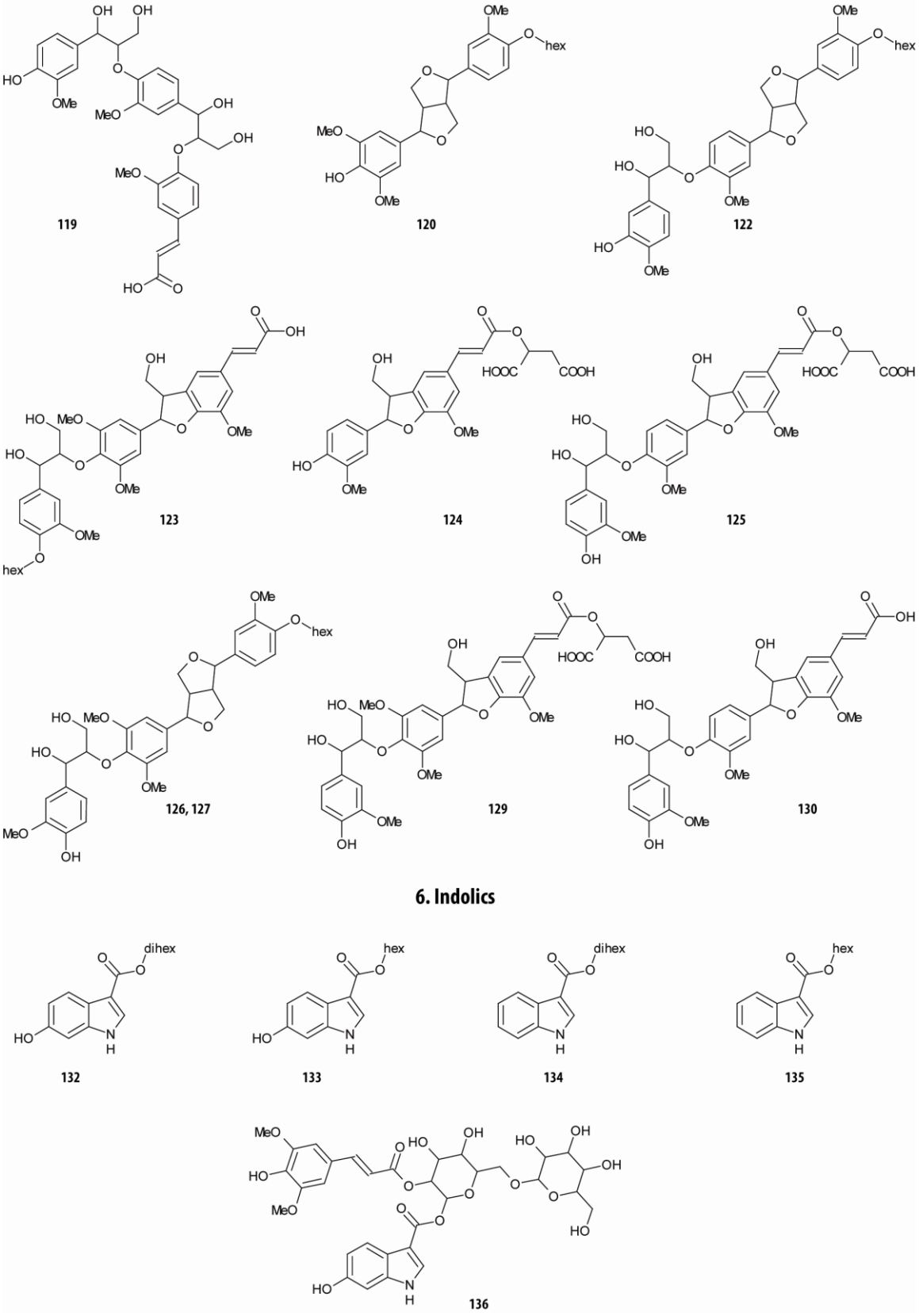

4. Phenylpropanoids
47
48
49
50
51
52
53
54
55
56
57
58
59
60,61
62
63
64,65,67
66,68

5. (Neo)lignans/oligolignols

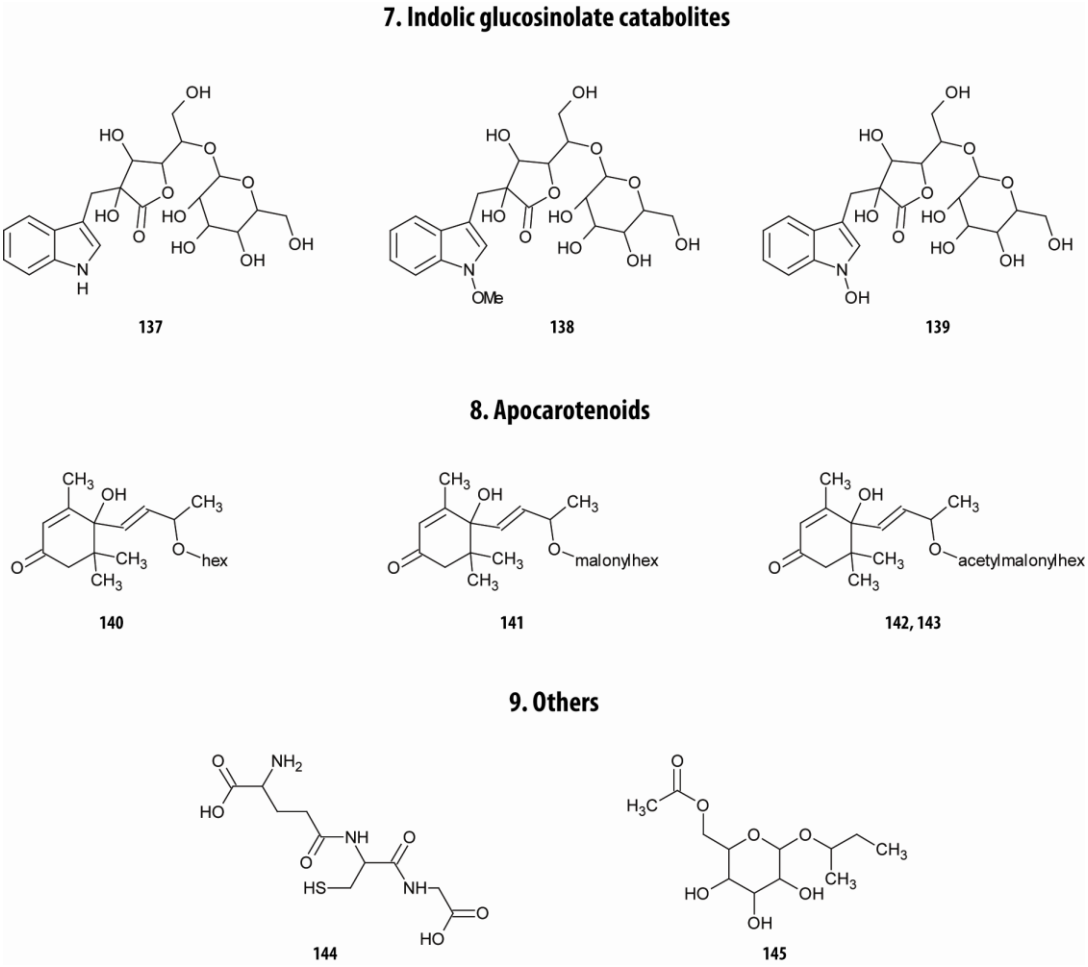



5. (Neo)lignans/oligolignols
93
94
98,104,116
103
105
106,115
107
109, 128, 131
110
111,121
112,118
113
117
114

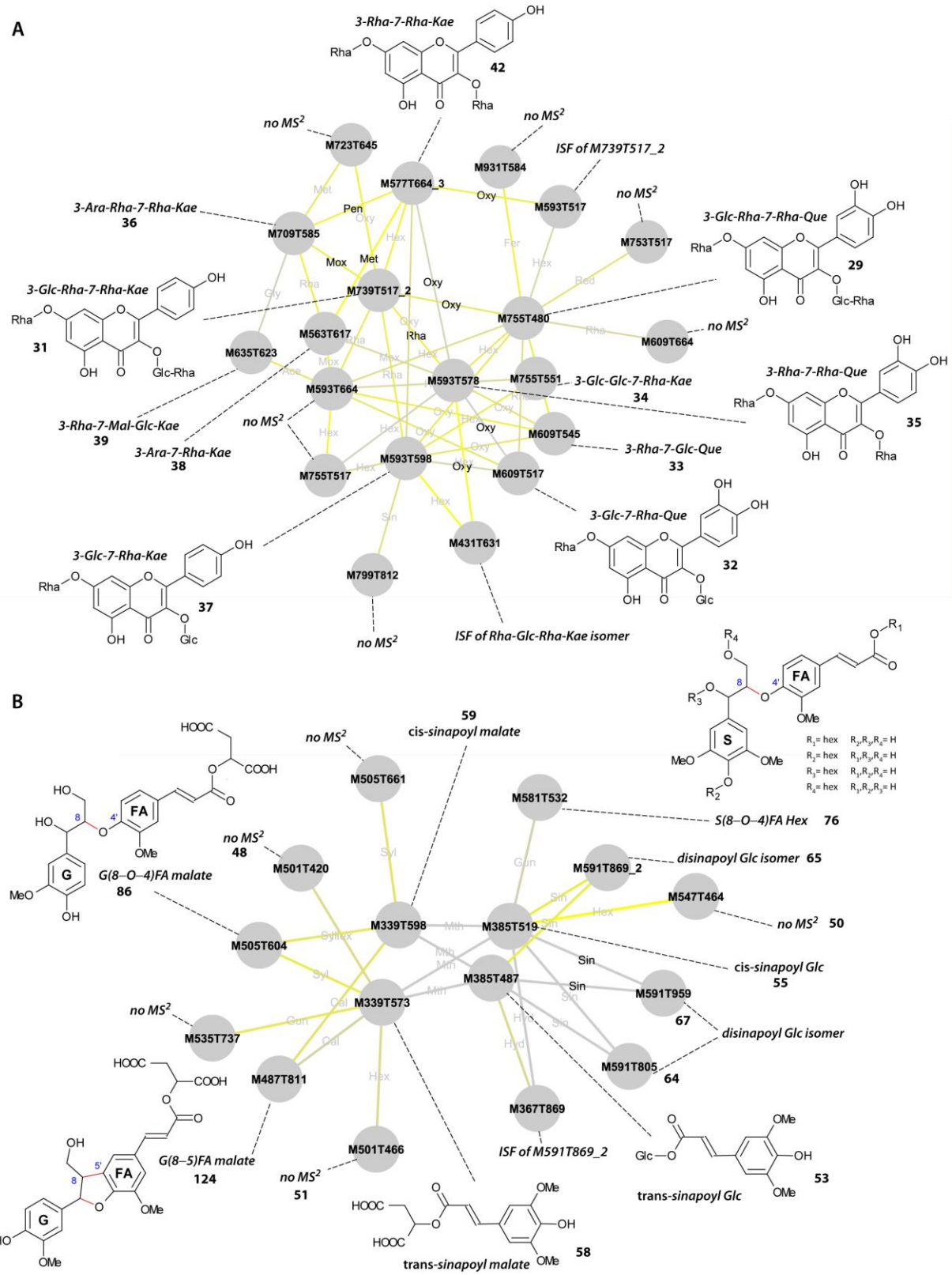
5. (Neo)lignans/oligolignols



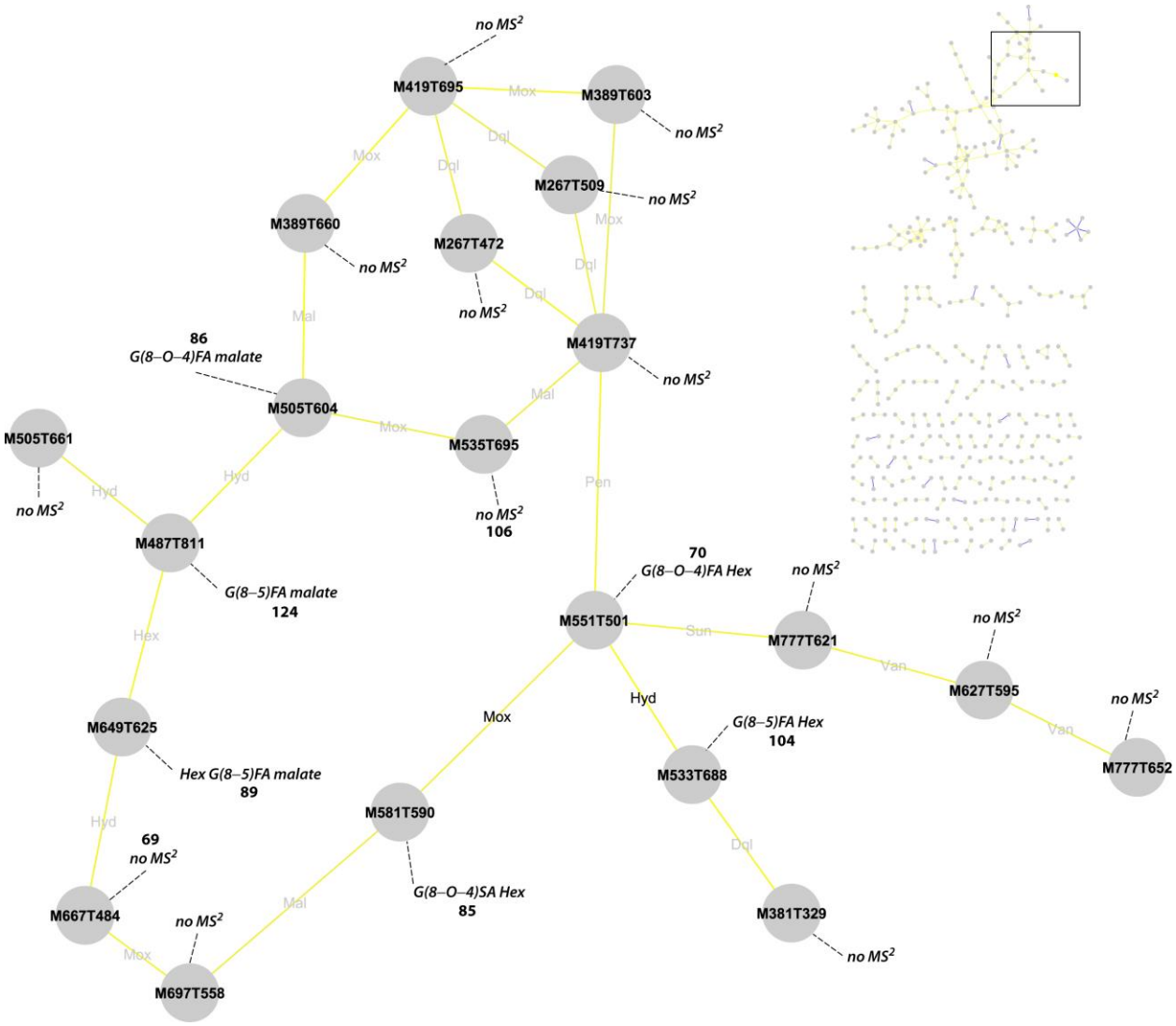
6. Indolics



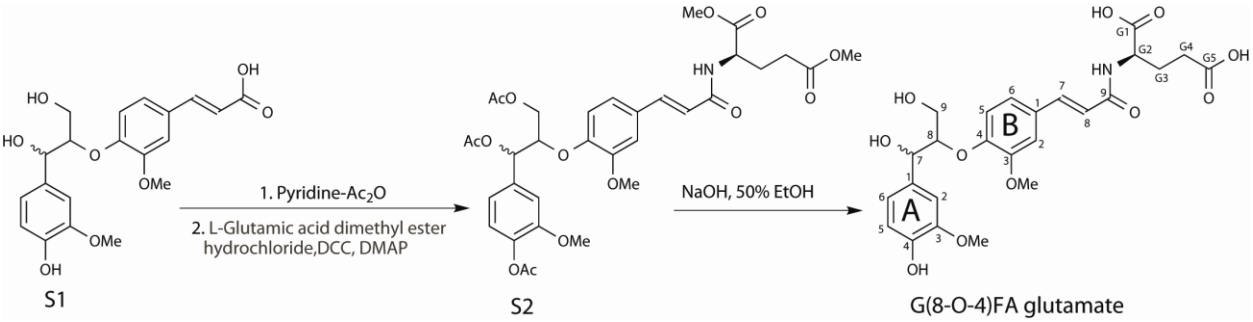
Supplemental Figure 4. Annotated Molecular Structures. See the footnote of Supplemental Data Set 1 for structural elucidation details. Identity, linkage and attachment position of sugars should be interpreted with caution as they cannot be determined by MS^2.


A
3-Rha-7-Rha-Kae
42
no MS2
M723T645
no MS2
M931T584
ISF of M739T517_2
M577T664_3
3-Ara-Rha-7-Rha-Kae
36
M709T585
M593T517
no MS2
M753T517
3-Glc-Rha-7-Rha-Que
29
3-Glc-Rha-7-Rha-Kae
31
M739T517_2
M755T480
M563T617
M635T623
M609T664
no MS2
3-Rha-7-Rha-Que
35
M593T578
M755T551
3-Glc-Glc-7-Rha-Kae
34
M593T664
3-Rha-7-Mal-Glc-Kae
39
no MS2
M609T545
3-Rha-7-Glc-Que
33
3-Ara-7-Rha-Kae
38
M593T598
M755T517
M609T517
3-Glc-7-Rha-Que
32
M431T631
3-Glc-7-Rha-Kae
37
M799T812
no MS2
ISF of Rha-Glc-Rha-Kae isomer
B
59
cis-sinapoyl malate
no MS2
M505T661
M581T532
S(8–O–4)FA Hex
76
no MS2
48
M501T420
M591T869_2
disinapoyl Glc isomer
65
G(8–O–4)FA malate
86
M547T464
M339T598
M385T519
no MS2
50
M505T604
cis-sinapoyl Glc
55
M385T487
M339T573
M591T959
67
no MS2
M535T737
disinapoyl Glc isomer
M591T805
64
M487T811
M367T869
G(8–5)FA malate
124
M501T466
ISF of M591T869_2
no MS2
51
trans-sinapoyl Glc
53
trans-sinapoyl malate
58

Supplemental Figure 5. Partial CSPP Sub-networks of Flavonoids **(A)** and Phenylpropanoids **(B)**. The sub-network for the flavonoids was elaborated starting with one of the more prominent flavonols, i.e., kaempherol 3-*O*-glucosyl-7-*O*-rhamnoside **37**, as "bait" (Veit and Pauli, 1999; Yonekura-Sakakibara et al., 2008) (Supplemental Data Set 1). The phenylpropanoid metabolism sub-network was obtained by using the nodes representing sinapoyl malate (**58** and **59**), sinapoyl glucose (**53** and **55**) and disinapoyl glucose **64** as "baits" (Supplemental Data Set 1). From the 15 flavonols and 22 phenylpropanoid derivatives obtained from the complete sub-networks (Supplemental Data Set 1), only a limited number of compounds are shown to improve the visibility. The phenylpropanoid sub-network also shows oligolignols/(neo)lignans that were highly correlated with sinapoyl malate and/or sinapoyl glucose. Nodes and edges represent chromatogram peaks and (bio)chemical conversions (see Table 1 for conversion types). Node labels are based on the XCMS integration and alignment algorithm. Whenever the similarity between the MS^2 spectra of candidate substrate and product exceeds 0.8, the edge label is black. The color brightness of the edge reflects the Pearson product-moment correlation coefficient between the levels of the CSPP "substrate" and "product" (blue and yellow represent a negative and positive correlation). Based on the MS^n spectra, the identification of hexose, deoxyhexose and pentose residues is not possible. However, up to now, only 3–*O*- and/or 7–*O*-linked glucose (Glc), rhamnose (Rha) and arabinose (Ara) flavonols have been observed in *Arabidopsis* (Yonekura-Sakakibara et al., 2008). Shorthand naming of oligolignols/(neo)lignans is based on (Morreel et al., 2004): units derived from coniferyl alcohol, sinapyl alcohol and ferulic acid are abbreviated as G (guaiacyl unit), S (syringyl unit) and FA, whereas the linkage type is indicated between brackets (see footnote Supplemental Data Set 1). Ara arabinose, Glc glucose, Hex hexose, ISF ion source fragment, Kae kaempherol, Mal-Glc malonyl glucose, Que quercetin, Rha rhamnose.

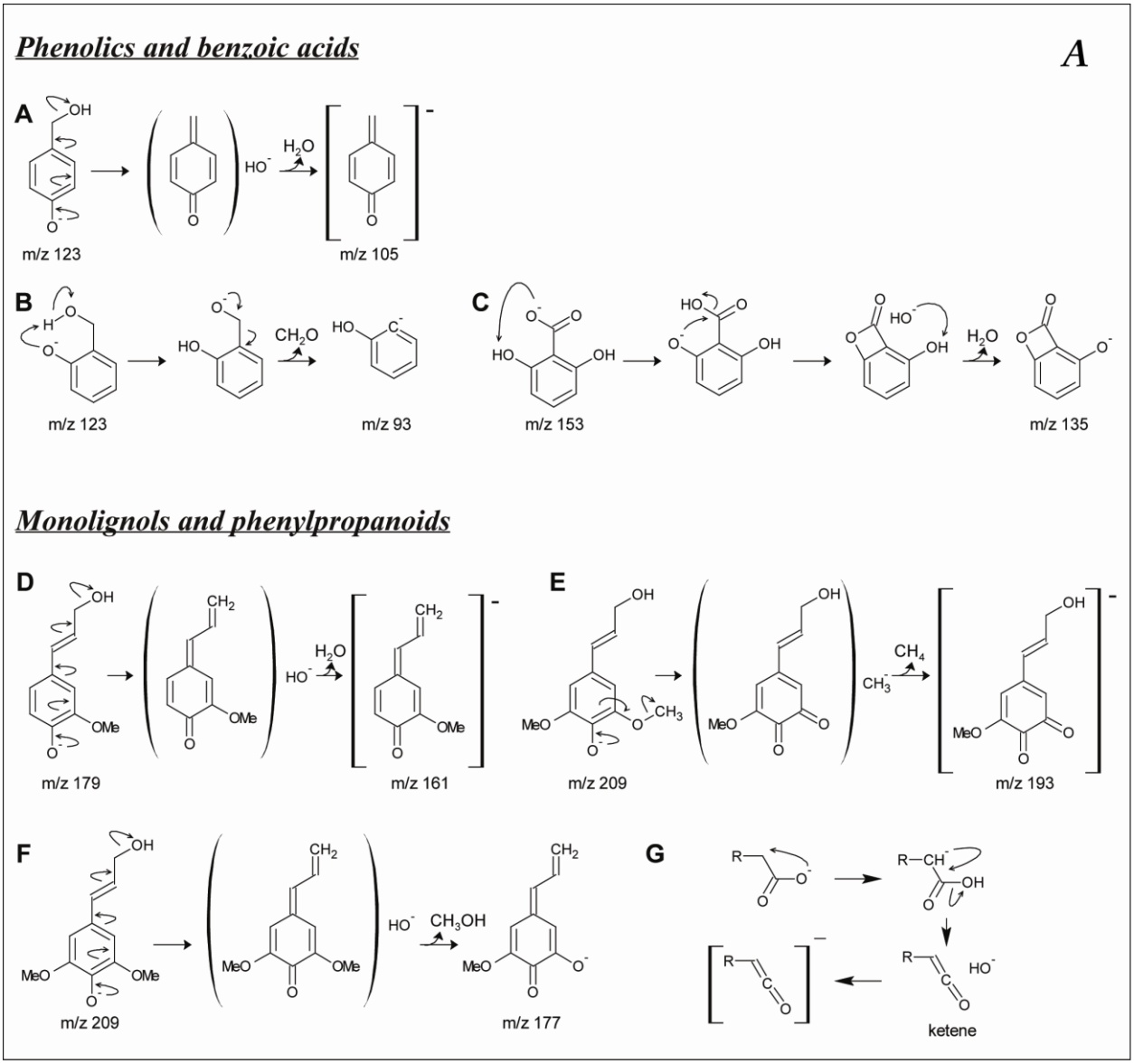


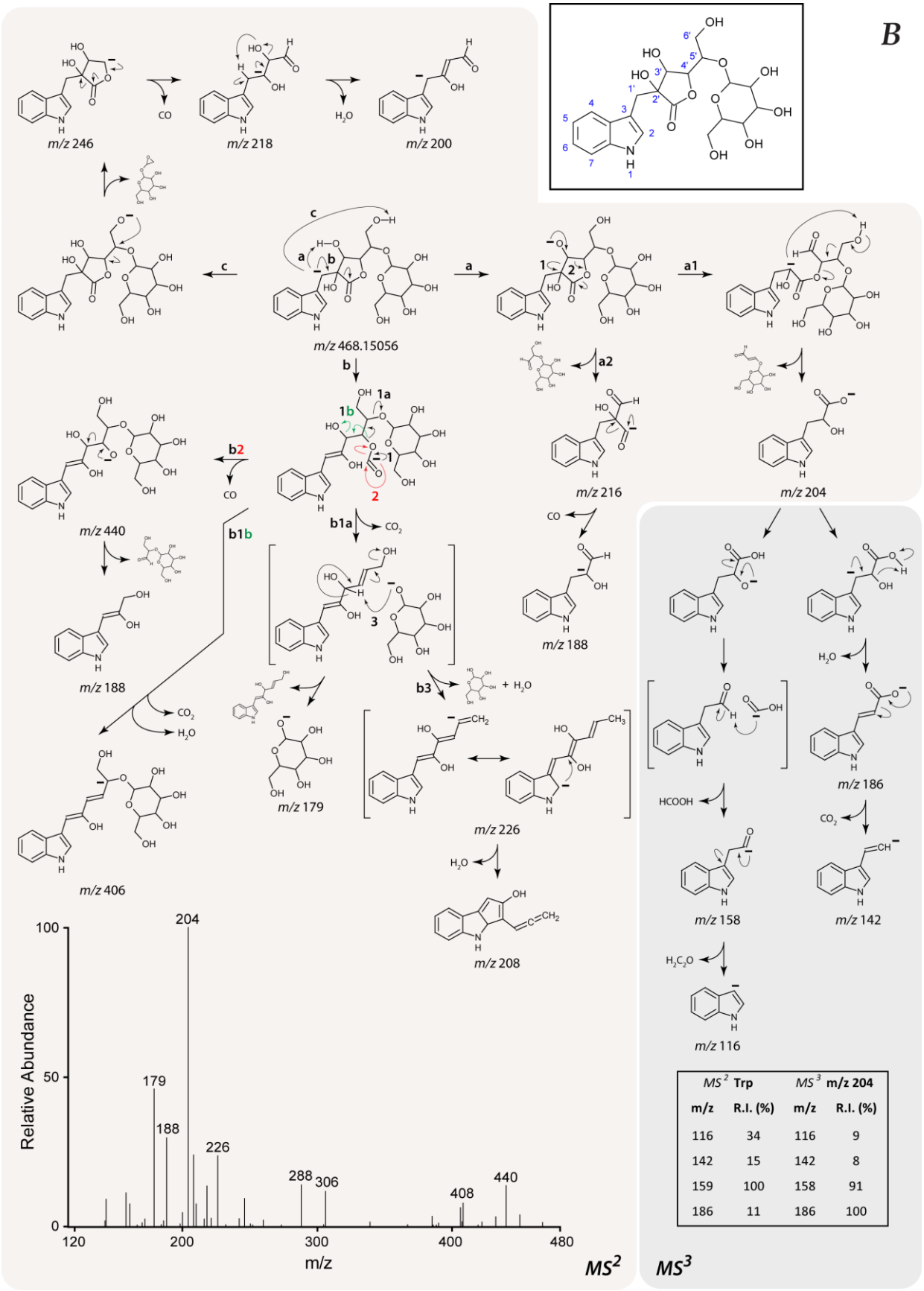
Supplemental Figure 6. CSPP Sub-network of Highly Correlated CSPPs. The sub-network represents mainly the (neo)lignans/oligolignols. An enlarged part of the major cluster is displayed whereas an overview is shown in the upper right corner. Nodes and edges represent chromatogram peaks and (bio)chemical conversions (see Table 1 for conversion types). The darkness of the edge label and the color brightness of the edge reflect the MS^2 spectral similarity and the Pearson product-moment correlation coefficient (see Supplemental Figure 5). Shorthand naming of (neo)lignans/oligolignols is based on (Morreel et al., 2004) (see footnote Supplemental Data Set 1).



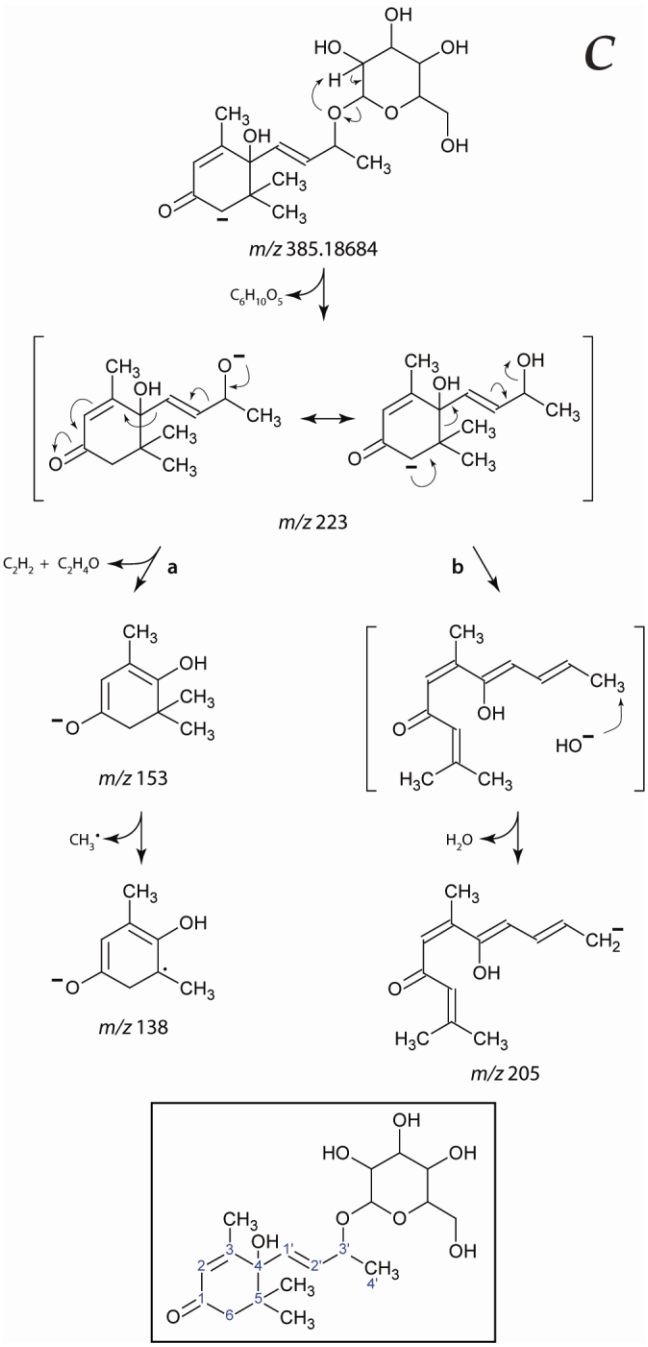
Supplemental Figure 7. Synthesis of G(8–O–4)FA Glu.

Supplemental Figure 8. Gas Phase Fragmentation Pathways of Simple Phenolics and Phenylpropanoids (A), 5´-*O*-β-D-glucosyl Dihydroascorbigen (B) and Corchoionoside C Anions (C).

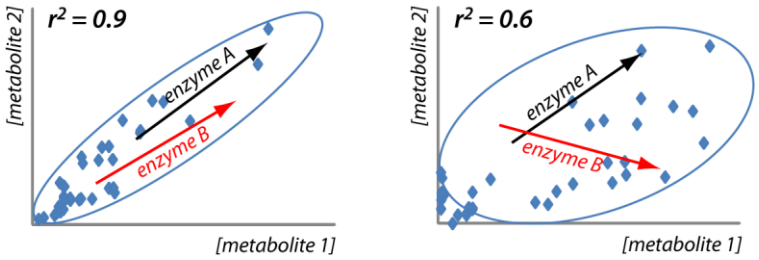




MS^2 Trp		MS^3 m/z 204	
m/z	R.I. (%)	m/z	R.I. (%)
116	34	116	9
142	15	142	8
159	100	158	91
186	11	186	100



Supplemental Figure 8. Gas Phase Fragmentation Pathways of Simple Phenolics and Phenylpropanoids (A), 5′-*O*-β-D-glucosyl Dihydroascorbigen (B) and Corchoionoside C Anions (C).



Supplemental Figure 9. Effect of Shared Control on Correlation. When the levels of two metabolites are controlled by e.g., two enzymes and the co-response on both metabolite levels of a change in enzymatic rate is the same for both enzymes, a high correlation will result (left plot). Otherwise a moderate correlation appears (right plot).

SUPPLEMENTAL TABLES

Supplemental Table 1. MS^2 spectra of simple phenolics

	neutral loss (g/mol)	Hydroxybenzyl alcohols			Hydroxybenzaldehydes		Hydroxybenzoic acids		
		4-hydroxybenzyl alcohol	2-hydroxybenzyl alcohol	4-hydroxy-3-methoxybenzyl alcohol	4-hydroxy-3-methoxy-benzaldehyde	4-hydroxy-3,5-dimethoxy-benzaldehyde	benzoic acid	4-hydroxy-benzoic acid	3-hydroxy-benzoic acid
CE (%)		35	35	30	35	35	35	35	35
[M-H]⁻		123(21)	123(0)	153(100)	151(0)	181(0)	121(0)	137(0)	137(0)
[a]*[M-H-CH_3]⁻*	15			138(31)	136(100)	166(100)			
[M-H-H_2O]⁻	18	105(100)							
[M-H-CH_2O]⁻	30	93(3)	93(100)						

	neutral loss (g/mol)	Hydroxybenzoic acids						
		2,4-dihydroxy-benzoic acid	2,6-dihydroxy-benzoic acid	2,3-dihydroxy-benzoic acid	2,5-dihydroxy-benzoic acid	4-hydroxy-3-methoxybenzoic acid	3-hydroxy-4-methoxybenzoic acid	3,5-dimethoxy-4-hydroxy-benzoic acid
CE (%)		35	28	30	30	30	30	35
[M-H]⁻		153(1)	153(100)	153(23)	153(9)	167(7)	167(10)	197(0)
[a]*[M-H-CH_3]⁻*	15					152(67)	152(100)	182(100)
[M-H-H_2O]⁻	18	135(100)	135(82)					
[M-H-CO_2]⁻	44	109(83)	109(33)	109(100)	109(100)	123(100)	123(65)	153(59)
[M-H-CO_2-CH_3]⁻	59					108(10)	108(5)	

Relative intensity of the daughter ions as compared to the base peak is given between brackets. [a], Elimination of a methyl radical from methoxylated benzene has been described by Reeks et al. (1993). CE, collision energy.

Supplemental Table 2. MS^2 spectra of monolignol-related compounds.

		Hydroxycinnamaldehydes		***Hydroxycinnamyl alcohols***				***Hydroxycinnamic acids***	
	neutral loss (g/mol)	***4-hydroxy-3-methyoxy-cinnamaldehyde***	***4,5-dihydroxy-3-methoxy-cinnamaldehyde***	***cinnamyl alcohol***	***4-hydroxy-3-methoxy-cinnamyl alcohol***	***3,5-dimethoxy-4-hydroxycinnamyl alcohol***	***3,5-dimethoxy-4-hydroxyhydro-cinnamyl alcohol***	***4-hydroxy-hydrocinnamic acid***	***3,4-dihydroxy-hydrocinnamic acid***
CE (%)		35	35	35	35	35	35	35	28
[M-H]⁻		177(5)	193(0)	133(0)	179(0)	209(0)	211(0)	165(0)	181(3)
[a]*[M-H-CH_3]⁻*	15	162(100)	178(100)		164(100)	194(100)	196(100)		
[M-H-CH_4]⁻	16					193(14)			
[b]*[M-H-H_2O]⁻*	18			115(100)		191(33)			
[b]*[M-H-CH_2O]⁻*	30			103(33)					
[M-H-CH_3OH]⁻	32					177(4)			
[M-H-H_2O-CH_3]⁻	33					176(9)			
[M-H-CO_2]⁻	44							121(100)	137(100)
[M-H-CO_2-H_2O]⁻	62								119(5)
[M-H-CO_2-CO]⁻	72							93(16)	

		Hydroxycinnamic acids							
	neutral loss (g/mol)	***cinnamic acid***	***4-hydroxy-cinnamic acid***	***2-hydroxy-cinnamic acid***	***3,4-dihydroxy-cinnamic acid***	***4-hydroxy-3-methoxy-cinnamic acid***	***3-hydroxy-4-methoxy-cinnamic acid***	***4-hydroxy-3,5-dimethoxy-cinnamic acid***	***ferulic acid ethyl ester***
CE (%)		35	30	35	36	32	34	35	35
[M-H]⁻		147(0)	163(3)	163(0)	179(0)	193(1)	193(2)	223(0)	221(0)
[a]*[M-H-CH_3]⁻*	15					178(48)	178(100)	208(100)	206(100)
[M-H-CO_2]⁻	44	103(100)	119(100)	119(100)	135(100)	149(100)	149(4)	179(42)	
[M-H-CO_2-CH_3]⁻	59					134(12)	134(3)	164(26)	

Relative intensity of the daughter ions as compared to the base peak is given between brackets. [a], Elimination of a methyl radical from methoxylated benzene has been described by Reeks et al. (1993). [b], The collision-induced dissociation of ionized primary alcohols has been described by Bowie (1990). CE, collision energy.

SUPPLEMENTAL METHODS

Chemicals

Coniferaldehyde, sinapyl alcohol, dihydrocaffeic acid, *p*-coumaric acid, ferulic acid, 2-hydroxybenzyl alcohol, 2,6-dihydroxybenzoic acid, 2,3-dihydroxybenzoic acid, 3,5-dihydroxybenzoic acid, 2,5-dihydroxybenzoic acid, indole-3-carboxylate, Trp and abscisic acid were purchased from Aldrich (Steinheim, Germany); vanillic acid, isovanillic acid, 4-hydroxybenzyl alcohol, 4-hydroxy-3-methoxybenzyl alcohol, coniferyl alcohol and isoferulic acid were obtained from Acros (Geel, Belgium). Caffeic acid and 4-hydroxy-3,5-dimethoxybenzaldehyde were bought at Janssen (Beerse, Belgium) and Roth (Karlsruhe, Germany), respectively.

Direct Infusion MS^n Analysis of Standards

A 100 μM solution of each standard, flowing at a rate of 10 μl/min, was mixed with a flow of 300 μl/min (water : methanol, 50:50 (v:v), 0.1% acetate) before entering a LCQ Classic ion trap MS (IT-MS) upgraded to a LCQ Deca (Thermo Fisher Scientific, Waltham, MA). With this MS instrument, the standards were more sensitively detected using Atmospheric Pressure Chemical Ionization (APCI) than when using ESI. Analytes were negatively ionized by APCI using the following parameter values: capillary temperature 150°C, vaporizer temperature 350°C, sheath gas 25 (arb), aux gas 3 (arb), source current 5 μA. MS^n analysis was performed by collision induced dissociation (CID) using He as the collision gas. The MS^n spectra were analyzed with Xcalibur vs 1.2.

MS^2 Spectra of Benzenoid and Phenylpropanoid Standards

The gas phase-based fragmentation of negatively ionized molecules is much less understood than that of positively ionized molecules. The main reasons are that anions fragment much more via charge-remote fragmentations, homolytic cleavages, rearrangements and ion-neutral complexes than cations (Bowie, 1990; Eichinger et al., 1994; Cheng and Gross, 2000). However, the various textbook organic chemistry reactions are much more reflected in the gas-phase fragmentation behavior of anions than in those of cations (Born et al., 1997; DePuy, 2000; Gronert, 2001,

2005). Therefore, studies that embark on elucidating the fragmentation pathways of anions increasingly appear in the literature. Among the four major classes of compounds (Supplemental Data Set 1) displayed by the CSPP networks, the low-energy collision-induced dissociation (CID) of the anions from only three compound classes have been extensively studied: glucosinolates (e.g. Fabre et al., 2007; Rochfort et al., 2008; Bialecki et al., 2010; Cataldi et al., 2010), flavonoids (e.g. Justesen, 2000; Fabre et al., 2001; Hughes et al., 2001; Hvattum and Ekeberg, 2003; Cuyckens and Claeys, 2004; Ferreres et al., 2004; Morreel et al., 2006; Yan et al., 2007) and (neo)lignans/oligolignols (e.g. Morreel et al., 2004; Eklund et al., 2008; Ricci et al., 2008; Schmidt et al., 2008; Morreel et al., 2010b; Morreel et al., 2010a; Ricci et al., 2010). Literature data for the fragmentation of the fourth major class displayed by the CSPP networks, benzenoids and phenylpropanoid anions, are available but their gas-phase fragmentation behavior upon low-energy CID has not been comprehensively analyzed. Therefore, various benzenoids and phenylpropanoids were subjected to negative ionization low-energy CID in an IT-MS instrument. Although both charge-driven, i.e., fragmentations that start from the most acidic site (Thevis et al., 2003), and charge-remote reactions might be responsible for the fragmentations, the former type will occur whenever possible (Cheng and Gross, 2000). Therefore charge-driven fragmentation pathways are proposed in this study. In the absence of a carboxylic acid function, ionization of benzenoids and phenylpropanoids will lead to a phenoxide anion from which the charge-driven CID pathways will proceed. In case both a carboxylic acid and a phenolic function are present, the carboxylic acid function will mainly take up the charge as it is more acidic in the gas-phase as compared to the phenolic function (e.g., 348.2 and 349.2 kcal/mol for acetic acid and phenol, respectively) (Harrison, 1992). In that case, the acid function will be mainly responsible for the fragmentation initiation.

The MS^2 product ions of various hydroxybenzoic acid, hydroxybenzyl alcohol and hydroxybenzaldehyde anions are listed in Supplemental Table 1. Upon fragmentation in the negative mode, alcohols may lose water and/or formaldehyde (Bowie, 1990). Both neutral losses are also observed with hydroxybenzyl alcohols (Supplemental Table 1), but their importance depends on the relative position of the phenol and the alcohol. When the phenol is in the *para* position of the hydroxymethyl functionality, water loss is favored (Supplemental Figure 8A.A), whereas formaldehyde loss is favored when the phenol is in the *ortho*-position (Supplemental Figure 8A.B). In case of a phenol in the *para* position of the hydroxymethyl, water loss is

initiated when the phenoxide anion of 4-hydroxybenzyl alcohol is converted to a quinone methide with the subsequent elimination of a hydroxide anion. Prior to the dissociation of this neutral/anion complex, the hydroxide anion will abstract a proton from the quinone methide (Bowie, 1990). The spectrum of 2-hydroxybenzylalcohol is clearly dominated by a so-called *ortho* effect (Supplemental Figure 8A.B): the phenoxide anion abstracts a proton from the alcohol function via a six-membered cyclic transition state in a McLafferty-type rearrangement (Grossert et al., 2006). The resulting alkoxide anion then undergoes a 1,2-elimination (Bowie, 1990).

Ionization of hydroxybenzoic acids yields carboxylate anions. The spectra of these anions are characterized by a neutral loss of 44 g/mol due to decarboxylation (Supplemental Table 1), which was also observed by Levsel et al. (2007). Nevertheless, decarboxylation was hardly observed for benzoic acid or monohydroxybenzoic acids. This agrees with the conclusion of Bandu et al. (2004) that multiple electron-withdrawing groups should be present on the benzene ring before decarboxylation occurs. Interestingly, a product ion arising from the loss of water was only observed for 2,6- and 2,4-dihydroxybenzoic acid. A reaction mechanism for this water loss due to an *ortho* effect is proposed in Supplemental Figure 8A.C. Following the proton transfer from the C_2 hydroxyl group to the carboxylate anion in a McLafferty-type rearrangement involving a six-membered cyclic transition state (Bandu et al., 2006; Grossert et al., 2006), an internal nucleophilic acyl substitution occurs between the C_2 oxyanion and the carboxylic acid in which a hydroxide anion is expelled (Attygalle et al., 2006). Water elimination is then mediated by an anion/neutral complex (Bowie, 1990). Because of the absence of a product peak associated with water loss in the mass spectra of other positional isomers of dihydroxybenzoic acid, this reaction can only proceed if the proton can be taken from a hydroxyl function at C_4 or C_6; due to the electron-withdrawing carbonyl function attached to C_1, hydroxyl functions at C_4 or C_6 are more acidic than those at the C_3 or C_5 position.

In the absence of a carboxylic acid function, the major loss observed in the spectra of monolignols and monolignol-related compounds was the charge-remote elimination of a neutral methyl radical (Supplemental Table 2). The most elaborate MS^2 spectrum was observed for 4-hydroxy-3,5-dimethoxycinnamyl alcohol. Besides methyl radical loss, a major MS^2 peak corresponding with dehydration was observed. This dehydration likely proceeds by a reaction mechanism (Supplemental Figure 8A.D) that is similar to that described for hydroxybenzyl

alcohols (Supplemental Figure 8A.A). The absence of water loss in the spectrum of 4-hydroxy-3,5-dimethoxyhydrocinnamyl alcohol supports the involvement of the aliphatic double bond in the loss of water in 4-hydroxy-3,5-dimethoxycinnamyl alcohol. The loss of water is observed when both *ortho* positions of the phenol group are methoxylated and not when only one *ortho* position is methoxylated, proving that the substitution of the aromatic ring influences the stability of the anion. Minor abundant product ions in the spectrum of 4-hydroxy-3,5-dimethoxycinnamyl alcohol originate from the loss of methanol, methane and the combined elimination of water and a methyl radical. Methanol loss likely occurs by a S_N2-type mechanism in which the expelled hydroxide anion acts as the nucleophile (Supplemental Figure 8A.F). The proposed fragmentation mechanism for the loss of methane is given in Supplemental Figure 8A.E.

Decarboxylation was the main fragmentation pathway of the hydroxycinnamic acids (Supplemental Table 2). Although carboxylate anions are reported to loose water upon conversion to their enolate anions (Bowie, 1990) (Supplemental Figure 8A.G) with subsequent ketene formation and the elimination of a hydroxide anion, this pathway was not observed in the IT-MS obtained fragmentation spectra of hydroxycinnamic acids. Likely, such a pathway is more favored under high-energy CID, although it has been sporadically suggested to occur under low-energy CID (Kanawati et al., 2007; Kanawati et al., 2008; Kanawati and Schmitt-Kopplin, 2010). Methyl radical loss of 4-hydroxy-3-methoxycinnamic acid was less pronounced than for 3-hydroxy-4-methoxycinnamic acid owing to the greater degree of radical delocalization in the latter. Finally, the specific loss of 62 Da in the spectrum of dihydrocaffeic acid indicates that a flexible side chain is necessary for this fragmentation. As the corresponding m/z 119 ion is also formed upon MS^3 of the m/z 137 first product ion, this loss represents a combined decarboxylation and dehydration. The additional water loss likely occurs by a charge-remote process involving the *ortho*-dihydroxybenzene moiety of the compound.

In general, the major fragmentation reactions observed for all of these phenolics were decarboxylation when a carboxylic acid was present and a demethylation when an aromatic methoxygroup was present. In the absence of *ortho* effects, water and formaldehyde eliminations were associated with aliphatic alcoholgroups. No dissociation of the phenolic function itself was observed. Hydroxycinnamaldehydes did not show specific fragmentation mechanisms.

Structural Elucidation of MS^n Spectra

Structural elucidation was based on knowledge obtained from the curated CSPP network (see Results) and further supported by interpretation of the MS^n spectra whenever possible. In metabolomics, a full identification (based on purification of the unknown compound followed by NMR analysis, or by spiking a purchased or synthesized standard) of all molecules is not possible. Therefore, two other levels of structural elucidation, i.e., structural annotation and structural characterization, have been defined (Sumner et al., 2007). Below, a structural annotation (Sumner et al., 2007) is obtained whenever MS^n information was used in addition to CSPP network information (see Supplemental Data Set 1 for information obtained from the CSPP network). In the absence of MS^n data, a structural characterization, based solely on the information from the CSPP network, is performed (Sumner et al., 2007). Below, the structural annotations and characterizations are described for the 145 compounds of which the structure could be predicted. The MS^n elucidation approach is more elaborately explained and referenced for the first representative compounds from each structural type.

Glucosinolates

1. 4-methylthiobutyl Gluc

The chemical formula of the anion of compound **1** was $C_{12}H_{22}O_9NS_3$ (m/z 420.04612). The base peak in its MS^2 spectrum appeared at m/z 259, a first product ion which is characteristic for glucosinolates and that is derived from the common moiety in glucosinolates consisting of a glucose moiety attached to a sulphated thiohydroximate. This ion at m/z 259 is formed via a rearrangement and represents a sulphated glucose moiety (Rochfort et al., 2008; Bialecki et al., 2010; Cataldi et al., 2010). Further support was obtained from the MS^3 spectrum of the m/z 259 ion that was identical to the one published previously (Rochfort et al., 2008). Other first product ions associated with the common glucose moiety in glucosinolates were observed as well at m/z 291, 275, 241 and 195 (Fabre et al., 2007; Bialecki et al., 2010; Cataldi et al., 2010). First product ions associated with the variable side-chain of glucosinolates (Fabre et al., 2007) were observed at m/z 340 (-80 Da, sulphite loss), 242 (-178 Da, loss of thio-glucose fragment), 224 (-196 Da, thio-glucose loss) and 178 (-242 Da, combined loss of glucose and sulphite). Based on the chemical formula, this compound is a saturated, methionine-derived glucosinolate. Therefore, compound **1** is 4-methylthiobutyl glucosinolate (Gluc) or glucoerucin. Matching to MassBank

returned a hit and score of 4 and 0.42, respectively. The low score arose from the difference in the applied collision energy to obtain the library spectrum and the one used in current study.

2. 5-methylthiopentyl Gluc

The ion of compound **2** had $C_{13}H_{24}O_9NS_3$ (m/z 434.06176) as chemical formula and was associated with 4-methylthiobutyl Gluc **1** in the CSPP networks. Its MS^2 spectrum was very similar to that of 4-methylthiobutyl Gluc **1**. However, the first product ions associated with the variable side-chain were all shifted by 14 amu in the MS^2 spectrum of compound **2** as compared to those in the MS^2 spectrum of 4-methylthiobutyl Gluc **1**. Consequently, this compound was elucidated as 5-methylthiopentyl glucosinolate or glucoberteroin. Further support was obtained from its retention time: compound **2** eluted 2.4 min later than 4-methylthiobutyl Gluc **1**, but 2.4 min earlier than the next member in this homologous series, i.e., 6-methylthiobutyl Gluc **3**.

3. 6-methylthiohexyl Gluc

Also for this compound with chemical formula $C_{14}H_{26}O_9NS_3$ (m/z 448.07705), a similar MS^2 spectrum was obtained as that of 5-methylthiopentyl Gluc **2**. Again, the main differences were the m/z values of the first product ions associated with the variable side-chain; all were shifted by 14 amu as compared to those in the MS^2 spectrum of 5-methylthiopentyl Gluc **2**. Compound **3** was characterized as 6-methylthiohexyl glucosinolate or glucolesquerellin.

4. 7-methylthioheptyl Gluc

The anion of **4** had $C_{15}H_{28}O_9NS_3$ (m/z 462.09239) as chemical formula and its MS^2 spectrum was very similar to that of 4-methylthiobutyl Gluc **1**, mainly differing by the first product ions associated with the variable side-chain: they were all shifted with 42 amu as compared to the corresponding m/z peaks in the MS^2 spectrum of 4-methylthiobutyl Gluc **1**. Therefore, this compound was annotated as 7-methylthioheptyl glucosinolate. Matching to MassBank rendered a score and hit of 0.84 and 7, respectively.

5. 8-methylthiooctyl Gluc

Again, the MS^2 spectrum of the anion of **5**, having a chemical formula equal to $C_{16}H_{30}O_9NS_3$ (m/z 476.10812), was very similar to that of compounds **1** and **4**, essentially differing by a shift of 56 amu and 14 amu, respectively, for all first product ions representing the variable side-chain. Compound **5** is 8-methylthiooctyl glucosinolate. A score and hit of 0.58 and 7 were obtained when matching to MassBank.

6. 3-methylsulfinylpropyl Gluc

The ion of compound **6** had a chemical formula of $C_{11}H_{20}O_{10}NS_3$ (m/z 422.02544). Together with first product ions at m/z 259, 275 and 291 in its MS^2 spectrum, this compound is a methionine-derived glucosinolate. Opposite to the MS^2 spectra of the methylthioalkyl glucosinolates (compounds **1**, **4** and **5**), the first product ion at m/z 358 was the base peak rather than the m/z 259 ion. This base peak arises from a loss of 64 Da which is derived from a methylsulfinyl endgroup (Fabre et al., 2007). The methylsulfinylalkyl side-chain looses also a methyl radical yielding the first product ion at m/z 407. Therefore, compound **6** is 3-methylsulfinylpropyl glucosinolate or glucoiberin. Matching to MassBank returned a score and hit of 0.63 and 9.

7. 4-methylsulfinylbutyl Gluc

The MS^2 spectrum of the anion of **7**, having a chemical formula of $C_{12}H_{22}O_{10}NS_3$ (m/z 436.04079), was similar to that of 3-methylsulfinylpropyl Gluc **6**, but the base peak was shifted with 14 amu, leading to the annotation of compound **7** as 4-methylsulfinylbutyl glucosinolate or glucoraphanin. A score and hit of 0.72 and 6 were obtained when matching to MassBank.

8. 4-methylsulfinylbutyl Gluc

The chemical formula ($C_{12}H_{22}O_{10}NS_3$, m/z 436.04080) and the MS^2 spectrum of the anion of **8** were identical to 4-methylsulfinylbutyl Gluc **7** and, thus, compound **8** is characterized as an isomer. Matching to MassBank yielded a score and hit of 0.72 and 6.

9. 5-methylsulfinylpentyl Gluc

The ion of compound **9** had $C_{13}H_{24}O_{10}NS_3$ (m/z 450.05654) as chemical formula and its MS^2 spectrum was similar to that of compounds **6**, **7** and **8**. However, compared to the MS^2 spectrum of compounds **7** and **8**, the base peak and the first product ion arising from methyl radical loss were shifted with 14 amu. Compound **9** is annotated as 5-methylsulfinylpentyl glucosinolate or glucoalyssin. A score and hit of 0.79 and 5 were obtained when matching to MassBank.

10. 6-methylsulfinylhexyl Gluc

With a chemical formula of $C_{14}H_{26}O_{10}NS_3$ (m/z 464.07213) and a MS^2 spectrum similar to those of compounds **6**, **7**, **8** and **9** except for the m/z shifts of the base peak and of the first product ion associated with methyl radical loss, this compound is annotated as 6-methylsulfinylhexyl glucosinolate or glucohesperin. Matching to the MassBank did not yield any result. Searching for

glucohesperin showed that it was not present in the database. Structure prediction with MetFrag (Wolf et al., 2010) using the PubChem database explained the variable side-chain-characteristic first product ions at m/z 449 and 400, but could not explain any of the glucose moiety-based first product ions that are characteristic for the CID spectra of glucosinolates. This underscores the weakness of current MS^2 structural prediction programs in the case of negative ions (Heinonen et al., 2008).

11. 7-methylsulfinylheptyl Gluc

The chemical formula of the anion was $C_{15}H_{28}O_{10}NS_3$ (m/z 478.08779) and the MS^2 spectrum indicates that this compound is also a methylsulfinylalkyl glucosinolate. More specifically, compound **11** was annotated as 7-methylsulfinylheptyl glucosinolate or glucoibarin. Matching to MassBank returned a score and hit of 0.77 and 7.

12. 8-methylsulfinyloctyl Gluc

Based on the MS^2 spectrum and the chemical formula ($C_{16}H_{30}O_{10}NS_3$, m/z 492.10339), this compound belonged also to the methylsulfinylalkyl glucosinolates. Compound **12** was annotated as 8-methylsulfinyloctyl Gluc or glucohirsutin. Comparison to the corresponding MassBank spectrum rendered a score and hit of 0.71 and 7.

13. 9-methylsulfinylnonyl Gluc

The chemical formula ($C_{17}H_{32}O_{10}NS_3$, m/z 506.11965) and the MS^2 spectrum were representative for methylsulfinylalkyl glucosinolates, and aided the structural annotation of compound **13** as 9-methylsulfinylnonyl glucosinolate or glucoarabin. Glucoarabin was not traced in the MassBank database.

14. hydroxy-4-(methylsulfinyl)-butyl Gluc

The ion of this compound had a chemical formula of $C_{12}H_{22}O_{11}NS_3$ (m/z 452.03595) and was closely associated with the methylsulfinylalkyl glucosinolates in the CSPP networks. Its structure was characterized as hydroxy-4-(methylsulfinyl)-butyl glucosinolate, yet no MS^2 spectrum was available for confirmation.

15. 3-hydroxy-5-(methylsulfinyl)-pentyl Gluc

Based on its close association with hydroxy-4-(methylsulfinyl)-butyl Gluc **14** in the CSPP networks and its anion's chemical formula of $C_{13}H_{24}O_{11}NS_3$ (m/z 466.05157), compound **15** was characterized as hydroxy-5-(methylsulfinyl)-pentyl glucosinolate. In the plant kingdom, a similar

compound has been purified in which the hydroxyl group was attached to the 3-position (Fabre et al., 2007).

16. hydroxy-8-(methylsulfinyl)-octyl Gluc

The ion of compound **16** had a chemical formula of $C_{16}H_{30}O_{11}NS_3$ (m/z 508.09845) and was the first compound in the hydroxy-(methylsulfinyl)-alkyl glucosinolate series for which MS^n spectra were obtained, supporting as well the structural characterizations of compounds **14**, **15** and **17**. The MS^2 spectrum of compound **16** showed first product ions at m/z 259 and 291 reminiscent of the common glucosinolate moiety, yet the base peak was observed at m/z 444 owing to a loss of 64 Da that is characteristic for a methylsulfinyl endgroup. Based on its MS^2 spectrum, the additional oxygen in its chemical formula as compared to the chemical formula of the corresponding methylsulfinyloctyl glucosinolate, should be present as a hydroxyl function on the variable side-chain of the glucosinolate. Therefore, compound **16** was annotated as hydroxy-8-(methylsulfinyl)-octyl glucosinolate. No hit with the same chemical formula was obtained when matching to MassBank.

17. hydroxy-6-(methylsulfinyl)-hexyl Gluc

Based on the chemical formula, $C_{14}H_{26}O_{11}NS_3$ (m/z 480.06348), and its close association with the methylsulfinylalkyl glucosinolates in the CSPP networks, this compound was structurally characterized as hydroxy-6-(methylsulfinyl)-hexyl glucosinolate. No MS^2 spectrum was obtained to verify this structure.

18. hydroxy-8-(methylsulfinyl)-octyl Gluc

The ion of compound **18** had the same chemical formula ($C_{16}H_{30}O_{11}NS_3$, m/z 508.09845) as hydroxy-8-(methylsulfinyl)-octyl Gluc **16**. Also its MS^2 spectrum was similar except that m/z 291 was the base peak rather than the m/z 444 first product ion. In addition, a new first product ion at m/z 391 appeared. The differences in the fragmentation pattern of the glucosinolate common moiety as compared to those of other glucosinolates arises from the effect of the additional hydroxyl function, suggesting that the hydroxyl function is attached close to the glucose thiohydroximate moiety. No hit with the same chemical formula was obtained when matching to MassBank.

19. 3-methylbutyl Gluc

The chemical formula of anion **19** was $C_{12}H_{22}O_9NS_2$ (m/z 388.07390). Its MS^2 spectrum was similar to those of the methylthioalkyl glucosinolates, i.e., dominated by first product ions due to fragmentations of the common moiety. The absence of fragmentations typical for the variable side-chain moiety suggested that no functional groups were present on the variable side-chain. Therefore, compound **19** is a leucine-derived glucosinolate and was annotated as 3-methylbutyl glucosinolate. No hit with the same chemical formula was obtained when matching to MassBank.

20. 4-methylpentyl Gluc

The ion of compound **20** had $C_{13}H_{24}O_9NS_2$ (m/z 402.08964) as chemical formula and its MS^2 spectrum was similar to that of 3-methylbutyl Gluc **19** except that the first product ion due to sulphite loss (-80 Da) was shifted by 14 amu. Therefore, this compound was annotated as 4-methylpentyl glucosinolate. No hit with the same chemical formula was obtained when matching to MassBank.

21. methylpentyl Gluc

Compound **21** had the same chemical formula (ion at m/z 402.08973, $C_{13}H_{24}O_9NS_2$) as 4-methylpentyl Gluc **20**, but no MS^2 spectrum was obtained. Compound **21** was characterized as methylpentyl glucosinolate.

22. 5-methylhexyl Gluc

Based on its anion's chemical formula ($C_{14}H_{26}O_9NS_2$, m/z 416.10553) and its MS^2 spectrum that was similar to the other methylalkyl glucosinolates except for the m/z shift of the first product ion resulting from sulphite loss (-80 Da), compound **22** was annotated as 5-methylhexyl glucosinolate. No hit with the same chemical formula was obtained when matching to MassBank.

23. 4-hydroxyglucobrassicin

The ion of compound **23** had a chemical formula of $C_{16}H_{19}O_{10}N_2S_2$ (m/z 463.04780) in which the additional nitrogen, as compared to the chemical formulae of the glucosinolates discussed above, should be present in the variable side-chain. Its characterization as a glucosinolate was indeed confirmed by the first product ions at m/z 259 and 275 that result from rearrangement reactions at the glucose thiohydroximate moiety. However, the most abundant first product ions were observed at m/z 285 and 267 due to losses of a thioglucose fragment and thioglucose (Fabre et

al., 2007), respectively. Subsequent losses of HCN and sulphite from the latter first product ion lead to the first product ions observed at m/z 240 and 160. This was confirmed in the MS^3 spectrum of the m/z 267 first product ion. The latter gas phase fragmentation reactions are typical for indolic glucosinolates (Fabre et al., 2007) and compound **23** was annotated as 4-hydroxyindol-3-ylmethyl glucosinolate or 4-hydroxyglucobrassicin. No hit with the same chemical formula was obtained when matching to MassBank.

24. glucobrassicin

The chemical formula of this anion was $C_{16}H_{19}O_9N_2S_2$ (m/z 447.05346) and its MS^2 spectrum showed the same type of first product ions as the spectrum of 4-hydroxyglucobrassicin **23**, yet the most abundant MS^2 ions were observed at m/z 259 and 275 and also the other first product ions typical for the glucosinolate common moiety were observed, i.e., at m/z 195, 241 and 291. However, indicative for indolic glucosinolates were the ions at m/z 269 and 251 due to losses of a thioglucose fragment and thioglucose, and, from the latter ion, the further HCN and sulphite losses yielding the first product ions at m/z 244 and 144. Therefore, this compound was annotated as indol-3-ylmethyl glucosinolate or glucobrassicin. Matching to MassBank yielded a score and hit of 0.60 and 10, respectively.

25. hydroxy-methoxyglucobrassicin

The ion of compound **25** had $C_{17}H_{21}O_{11}N_2S_2$ (m/z 493.05834) as chemical formula and its MS^2 spectrum showed the m/z 315 ion (loss of thioglucose fragment) as a very abundant peak, characteristic for indolic glucosinolates, yet the base peak was at m/z 259. This compound was annotated as hydroxy-methoxyglucobrassicin. No hit with the same chemical formula was obtained when matching to MassBank.

26. 4-methoxyglucobrassicin

The chemical formula of the ion of compound **26** was $C_{17}H_{21}O_{10}N_2S_2$ (m/z 477.06391). The most abundant MS^2 ions were observed at m/z 259, 275 and 291, characteristic for the glucose thiohydroximate moiety. However, first product ions resulting from the losses of a thioglucose fragment and of thioglucose, typical for the indolic moiety were observed as well at m/z 299 and 281. The latter ion fragmented by HCN loss yielding the m/z 254 first product ion. This compound was annotated as 4-methoxyindol-3-ylmethyl glucosinolate or 4-

methoxyglucobrassicin. Matching to MassBank rendered a score and hit of 0.87 and 8, respectively.

27. neoglucobrassicin

The ion of this compound had the same chemical formula ($C_{17}H_{21}O_{10}N_2S_2$, m/z 477.06398) as 4-methoxyglucobrassicin **26**. Its MS^2 spectrum showed first product ions at m/z 259, reminiscent of the common moiety in glucosinolates, and at m/z 299 due to the loss of a thioglucose fragment typical for indolic glucosinolates. However, the base peak was observed at m/z 446 resulting from methoxyl radical loss. Such a homolytic fragmentation would be expected when the methoxyl function is linked to the indole nitrogen. Therefore, this compound is 1-methoxyindol-3-ylmethyl glucosinolate or neoglucobrassicin. No hit with the same chemical formula was obtained when matching to MassBank.

28. 2-phenylethyl Gluc

The anion with chemical formula of $C_{15}H_{20}O_9NS_2$ (m/z 422.05842) had a MS^2 spectrum similar to that of the methylthioalkyl and the methylalkyl glucosinolates, i.e., in which the fragmentations typical for the glucose thiohydroximate moiety prevailed. Based on the high RDB value (ring and double bonds=6.5) as compared to that of other glucosinolates, this compound was annotated as 2-phenylethyl glucosinolate or gluconasturtiin. No hit with the same chemical formula was obtained when matching to MassBank.

Flavonoids

29. 3-Glc(2←1)Rha-7-Rha-Que

The ion of compound **29** had $C_{33}H_{39}O_{20}$ (m/z 755.20427) as chemical formula. Its MS^2 spectrum was dominated by the loss of 146 Da due to expelling a deoxyhexose residue yielding the ion at m/z 609. MS^3 fragmentation of this first product ion lead to second product ions at m/z 463 and 447 indicating that a second deoxyhexose and a hexose were lost, respectively. The occurrence of both m/z 463 and 447 revealed that the deoxyhexose and the hexose were attached to two different sites on the aglycone. The aglycone was represented by second product ions at m/z 300 and 301. The former represents the aglycone radical anion resulting from homolytic cleavage of the O-glycosidic bond (Hvattum and Ekeberg, 2003). MS^4 fragmentation of the aglycone yielded third product ions due to the loss of CO or both CO and CO_2, indicative for a flavonol whereas the third product ions at m/z 151 and 179 arose from a Retro Diels-Alder (RDA) cleavage of the

C ring. Both ions are annotated as the $^{1,3}A^-$ and $^{1,2}A^-$ ions of quercetin in Fabre et al. (2001). Other groups have also pinpointed the typical RDA cleavages occurring during gas phase fragmentations of flavonoids (Justesen, 2000; Hughes et al., 2001). Because all glycosides were linked to the 3- and/or 7-position, because diglycosides were more prominent on the 3-position than on the 7-position and because all deoxyhexoses and hexoses were rhamnose and glucose, respectively, in a thorough study of Arabidopsis flavonoids (Yonekura-Sakakibara et al., 2008), compound **29** was annotated as 3-rhamnosyl(1→2)glucoside-7-rhamnoside-quercetin. The rhamnose is 1→2 linked to glucose (neohesperidose) rather than 1→6 as in rutinose because the cross-ring cleavage $^{0,2}X_0^-$ first and second product ions at m/z 489 are observed (Cuyckens and Claeys, 2004). Furthermore, the interglycosidic bond in a neohesperidose moiety is much more fragile than that in a rutinose moiety, rendering information on the sugar sequence for the former (Ferreres et al., 2004; Yan et al., 2007). Matching to MassBank yielded a score and hit of 0.80 and 9.

30. 2-Glc-2-Rha-Kae

The chemical formula of this anion was $C_{39}H_{49}O_{24}$ (m/z 901.26343). The MS^2 spectrum showed a loss of 308 Da as major fragmentation pathway leading to the ion at m/z 593. This is due to the loss of a disaccharide composed of a deoxyhexose, likely rhamnose, and a hexose, likely glucose. Because no first product ion due to interglycosidic cleavage was observed, a 1→6 rather than a 1→2 glycosidic linkage might be inferred (Cuyckens and Claeys, 2004; Ferreres et al., 2004; Yan et al., 2007). MS^3 of the first product ion at m/z 593 yields then a second product ion at m/z 447 due to the loss of a second deoxyhexose. Subsequent dissociation of this ion eliminates another hexose and yields the second product ions at m/z 285 and 284, representing the aglycone. This aglycone eliminates CO yielding the ion at m/z 257, whereas the second product ion at m/z 179 can be annotated as the $^{1,2}A^-$ ion arising from RDA cleavage of a flavonol C ring (Fabre et al., 2001). Therefore, this compound was characterized as kaempherol to which two disaccharides, each comprising a rhamnosyl and a glucosyl unit, are linked. No hit with the same chemical formula was obtained when matching to MassBank. Using MetFrag, 20 biological compounds with the same chemical formula were downloaded from the Pubchem database of which 5 were predicted via *in silico* fragmentation to match equally likely the MS^2 spectrum (containing only one first product ion at m/z 593, see above) of compound **30**: two of those 5 compounds were kaempherol glycosides and two others were apigenin glycosides.

31. 3-Glc(2←1)Rha-7-Rha-Kae

Anion **31** had $C_{33}H_{39}O_{19}$ (m/z 739.20820) as chemical formula. The MS^2 and MS^3 spectra were similar to those of 3-Glc(2←1)Rha-7-Rha-Que **29** except that the product ion for the aglycone was detected 16 amu lower, i.e., at m/z 285. The CO and combined CO and CO_2 losses observed in the MS^4 spectrum were also at m/z values 16 amu lower than the corresponding losses in the MS^4 spectrum of 3-Glc(2←1)Rha-7-Rha-Que **29**. Therefore, this compound is 3-rhamnosyl(1→2)glucoside-7-rhamnoside-kaempherol. A score and hit of 0.77 and 1 was obtained when matching to MassBank.

32. 3-Glc-7-Rha-Que

The chemical formula of this compound's anion was $C_{27}H_{29}O_{16}$ (m/z 609.14553). The MS^2 base peak at m/z 463 indicated a deoxyhexose (rhamnose) loss. However, also a hexose (glucose) loss was evident from the first product ion at m/z 447. Both sugars are attached to different positions of the aglycone. A first product ion representing a quercetin aglycone was observed at m/z 301. This was confirmed by MS^3 fragmentation of the m/z 463 first product ion rendering both the aglycone anion and the aglycone radical anion at m/z 301 and 300, respectively (Hvattum and Ekeberg, 2003). Likely, the rhamnose and glucose moieties are linked to the 7-*O*- and 3-*O*-positions because glycosidic cleavage in negative ionization mode occurs presumably more readily at the 7-*O*-position since the reverse is true in positive ionization mode (Cuyckens and Claeys, 2004). This was indeed verified for some flavonol di-*O*-glycosides from *Farsetia aegyptia* (Shahat et al., 2005). Compound **32** was annotated as 3-glucosyl-7-rhamnosyl-quercetin. Matched to MassBank, a score and hit of 0.33 and 2 was obtained with a library spectrum recorded on a quadrupole-time-of-flight MS.

33. 3-Rha-7-Glc-Que

The chemical formula ($C_{27}H_{29}O_{16}$, m/z 609.14527) of the anion was the same as for 3-Glc-7-Rha-Que **32**. The MS^2 spectrum was similar as well. The major difference was that hexose loss was more facile than rhamnose loss based on the abundance of the corresponding first product ions at m/z 447 and 463, respectively. Based on the same reasoning as for 3-Glc-7-Rha-Que **32**, this compound was annotated as 3-rhamnosyl-7-glucosyl-quercetin. MassBank matching returned a score and hit of 0.72 and 7 with a library spectrum of 3-glucosyl-7-rhamnoside-quercetin (see compound **32**). However, 3-rhamnosyl-7-glucoside-quercetin was not present in the database.

34. 3-Glc(6←1)Glc-7-Rha-Kae

The anion had $C_{33}H_{39}O_{20}$ (m/z 755.20409) as chemical formula. The base peak in its MS^2 spectrum was at m/z 609 due to rhamnose loss. A minor peak was observed at m/z 432 resulting from the elimination of a dihexoside. Thus, the rhamnose is likely attached to the 7-O-position, whereas the dihexoside is 3-O-linked (Cuyckens and Claeys, 2004; Shahat et al., 2005). MS^3 fragmentation of the m/z 609 first product ion rendered a second product ion at m/z 285, representing the kaempherol aglycone. This was verified by MS^4 fragmentation of the second product ion at m/z 285, which was characterized by CO loss (m/z 257), CO_2 loss (m/z 241), combined CO/CO_2 loss (m/z 213), 2CO loss (m/z 229), C_2H_2O loss (m/z 243) and by the RDA-derived $^{1,3}A^-$ ion at m/z 151 (Fabre et al., 2001; Hughes et al., 2001). As no interglycosidic cleavage was observed in the MS^3 spectrum, the dihexoside is 1→6 linked and represents a gentiobiose moiety (Ferreres et al., 2004; Yan et al., 2007). Therefore, this compound was annotated as 3-glucosyl-(6←1)-glucosyl-7-rhamnosyl-kaempherol. No hit with the same chemical formula and the same aglycone was obtained when matching to MassBank. Using MetFrag, 110 biological compounds were retrieved with an equal chemical formula of which 22, via *in silico* fragmentation, equally likely matched with the MS^2 spectrum (containing one first product ion) of compound **34**: they comprised a diverse set of flavonol and flavone aglycone structures.

35. 3-Rha-7-Rha-Que

The chemical formula for the anion was $C_{27}H_{29}O_{15}$ (m/z 593.15133). The MS^2 base peak appeared at m/z 447 indicating a rhamnose loss. Further MS^3 fragmentation of the MS^2 base peak lead to a second rhamnose loss yielding the peaks at m/z 300 and 301 corresponding to a quercetin radical anion and a quercetin anion, respectively. The aglycone structure was verified by MS^4 fragmentation. This compound was annotated as 3-rhamnosyl-7-rhamnosyl-quercetin. A score and hit value of 0.58 and 2 were obtained when matching to MassBank.

36. 3-Ara(2←1)Rha-7-Rha-Kae

This anion had a chemical formula of $C_{32}H_{37}O_{18}$ (m/z 709.19849). The MS^2 spectrum had a base peak at m/z 563 owing to rhamnose loss (-146 Da). Further MS^3 fragmentation rendered a base peak at m/z 417 due to a second rhamnose loss. A further pentose loss (likely arabinose) in the MS^3 spectrum explained the kaempherol anion peak at m/z 285. The MS^3 ion at m/z 417 expelled a water molecule yielding the second product ion at m/z 399. Together with the

abundant second product ion representing the kaempherol radical anion at m/z 284, this suggested a rhamnose 1→2 linked to arabinose (Hvattum and Ekeberg, 2003; Ferreres et al., 2004; Yan et al., 2007). The latter disaccharide moiety cleaved off leading to the low abundant radical anion at m/z 430 in the MS^2 spectrum. Therefore, this compound was structurally elucidated as 3-arabinosyl-(2←1)-rhamnosyl-7-rhamnosyl-kaempherol. No hit with the same chemical formula was obtained when matching to MassBank. Using MetFrag, 18 compounds with the same chemical formula were retrieved from the Pubchem database of which 5, via *in silico* fragmentation, equally likely matched the MS^2 (containing one peak) spectrum of compound **36**: 3 kaempherol and 2 apigenin glycosides. Among the kaempherol glycosides, 3-arabinosyl-(2←1)-rhamnosyl-7-rhamnosyl-kaempherol was present.

37. 3-Glc-7-Rha-Kae

As chemical formula $C_{27}H_{29}O_{15}$ (m/z 593.15055) was obtained for this anion. The MS^2 spectrum was characterized by glucose and rhamnose loss leading to the first product ions at m/z 431 and 447, the latter ion being the most abundant (Cuyckens and Claeys, 2004; Shahat et al., 2005). The spectrum showed also the aglycone peak at m/z 285. MS^3 fragmentation of the m/z 447 ion rendered second product ions typical for kaempherol (see description of MS^n spectra from compound **34**). This compound was annotated as 3-glucosyl-7-rhamnosyl-kaempherol. The best fit when matching to MassBank was obtained for 3-rhamnosyl-7-rhamnosyl-quercetin (score=0.74, hit=4). However, the second best fit was for 3-glucosyl-7-rhamnosyl-kaempherol (score=0.48, hit=5).

38. 3-Rha-7-Ara-Kae

This anion had $C_{26}H_{27}O_{14}$ (m/z 563.14006) as chemical formula. Its MS^2 spectrum was dominated by the base peak at m/z 431 and the less abundant ion at m/z 417 due to pentose (arabinose) and rhamnose losses. The aglycone anion at m/z 285 was reminiscent of a kaempherol. Based on Cuyckens and Claeys (2004) and on Shahat et al. (2005), this compound was annotated as 3-rhamnosyl-7-arabinosyl-kaempherol. No hit with the same chemical formula was obtained when matching to MassBank. Entering the MS^2 data in MetFrag returned 137 biological compounds with the same chemical formula as compound **38** from the Pubchem database. Following *in silico* fragmentation, the best hit was obtained for the flavone 7-xylosyl(1→4)rhamnosyl-scutellarein (score=1, # explained peaks=4). Among the next 4 best hits (score=0.985, # explained peaks=3) were the flavonols 3-arabinosyl-7-rhamnosyl-kaempherol, 3-

rhamnosyl-7-arabinosyl-kaempherol, 3-rhamnosyl-4'-arabinosyl-kaempherol and the flavone 6-xylosyl-7-rhamnosyl-scutellarein.

39. 3-Rha-7-Mal-Glc-Kae

The chemical formula for this anion was $C_{29}H_{31}O_{16}$ (m/z 635.16151). Upon MS^2 fragmentation, the loss of rhamnose and malonylglucose yield the first product ions at m/z 489 and 431. The first product ion at m/z 285 suggested a kaempherol structure as aglycone. As the m/z 431 ion was the most abundant (Cuyckens and Claeys, 2004; Shahat et al., 2005), this compound was annotated as 3-rhamnosyl-7-malonylglucosyl-kaempherol. No hit with the same chemical formula was obtained when matching to MassBank. No biological compounds with this chemical formula could be retrieved from either the Pubchem or the ChemSpider databases using MetFrag.

40. 3-Rha-7-Rha(4←1)Glc-Kae

The anion had a chemical formula of $C_{33}H_{39}O_{19}$ (m/z 739.20920). The MS^2 base peak was observed at m/z 431 due to the combined loss of glucose and rhamnose. Although a very minor peak at m/z 577 was observed resulting from glucose loss, the major loss as a disaccharide moiety indicates that no 1→2 linkage was involved (Ferreres et al., 2004; Yan et al., 2007) and that this disaccharide moiety was present at the 7-position (Cuyckens and Claeys, 2004; Shahat et al., 2005). A rhamnose loss from the 3-position renders the MS^2 ion at m/z 593. The aglycone was represented by the first product ion at m/z 285. In the MS^3 spectrum of the first product ion at m/z 431, the flavonoid-specific RDA cleavage $^{1,3}A^-$ and $^{1,2}A^-$ ions at m/z 151 and 179 were observed (Fabre et al., 2001; Hughes et al., 2001). Therefore, the aglycone was annotated as a kaempherol and the compound as 3-rhamnosyl-7-rhamnosyl-(4←1)-glucosyl-kaempherol. Matching to MassBank retrieved 3-rhamnosyl-(2←1)-glucosyl-7-rhamnosyl-kaempherol (score=0.77, hit=2). Our proposed structure was not present in this database.

41. 3-Rha(4←1)Rha-7-Rha-Kae

The chemical formula for this anion was $C_{33}H_{39}O_{18}$ (m/z 723.21490). No MS^2 spectrum was obtained for this low abundant compound, but this compound was connected to 3-Ara(2←1)Rha-7-Rha-Kae **36** via a "methylation" conversion in the CSPP networks. Furthermore, the levels of both compounds were highly correlated. The possible candidate structure, i.e., 3-rhamnosyl-

(4←1)-rhamnosyl-7-rhamnosyl-kaempherol, has been previously observed in the plant kingdom (Rasoanaivo et al., 1990).

42. 3-Rha-7-Rha-Kae

The anion's chemical formula was $C_{27}H_{29}O_{14}$ (m/z 577.15646). MS^2 fragmentation yielded a base peak at m/z 431 (rhamnose loss) and a minor peak at m/z 285 representing the aglycone following two rhamnose losses. The MS^3 spectrum of the m/z 431 ion yielded ions at m/z 151 and 179 corresponding to the flavonoid RDA cleavage $^{1,3}A^-$ and $^{1,2}A^-$ ions (Justesen, 2000; Yonekura-Sakakibara et al., 2008), thus, pinpointing to a kaempherol aglycone. Both rhamnoses, connected in a disaccharide moiety via a 2←1 linkage, could have been attached to the 3-O-position or the rhamnoses were separately attached at two different positions on the kaempherol moiety. As this compound was one of the three most abundant, already identified (Veit and Pauli, 1999), flavonols in leaves, this compound was annotated as 3-rhamnosyl-7-rhamnosyl-kaempherol or kaempferitrin. Matching to MassBank returned a score and hit of 0.52 and 2.

43. 3-sinapoyl-Rha-7-Rha-Kae

The chemical formula of the anion was $C_{38}H_{39}O_{18}$ (m/z 783.21356). This compound was connected to the previous compound, i.e., 3-Rha-7-Rha-Kae **42**, via a "sinapoylation" conversion in the CSPP networks. Rhamnose elimination was the main fragmentation pathway rendering the MS^2 base peak at m/z 637. Smaller peaks at m/z 577 and 431 were also visible, the first derived from cleavage of the sinapoyl ester bond, the second due to the loss of a second rhamnose moiety. This was further verified by MS^3 fragmentation of the m/z 637 ion. In both the MS^2 and MS^3 spectra, the ion at m/z 285 indicated that kaempherol was the aglycone part. Therefore, this compound was identified as 3-sinapoyl-rhamnosyl-7-rhamnosyl-kaempherol. No hit with the same chemical formula was obtained when matching to MassBank. Eight hits were returned from the Pubchem database using MetFrag. *In silico* fragmentation provided non-zero scores for only three of them; all were flavone glycosides bearing a hydroxycinnamic acid further supporting our proposed structure for compound **43**.

Opposite to the observation for glucosinolate anions, MetFrag predicted reasonable structures that were all close to the true structure for all the flavonol glycoside anions. This seemed to be mainly due to the accurate prediction of interglycosidic cleavages and the use of the parent ion molecular weight for searching the Pubchem database.

Benzenoids

44. protocatechoyl Glc

The chemical formula for the ion was $C_{13}H_{15}O_9$ (m/z 315.07229). The MS^2 base peak at m/z 153 resulted from a hexose loss (-162 Da) and the MS^3 spectrum showed that decarboxylation, yielding the ion at m/z 109, was the main fragmentation pathway for the m/z 153 ion. MS^2 ions at m/z 225 and 165, resulting from hexose cross-ring cleavages, indicated that the hexose was connected as a hexoside with a free reducing end (Carroll et al., 1995; Mulroney et al., 1999; March and Stadey, 2005) or via an ester bond (Vanholme et al., 2012). The aglycone is a dihydroxybenzoic acid. In *Arabidopsis*, both 3,4-dihydroxybenzoic acid (protocatechuic acid) and 2,4-dihydroxybenzoic acid (homogentisic acid) occur. However, in the case of a homogentisic acid moiety, the MS^3 spectrum of this moiety should show a base peak due to water loss (Supplemental Table 1 and Supplemental Methods), which is not observed here. Therefore, protocatechoyl glucose was proposed as structure for this molecule. Further support for the ester bond was provided by searching the complementary pairs of ions associated with the two characteristic cleavages that esters undergo (Debrauwer et al., 1992; Fournier et al., 1993; Fournier et al., 1995; Stroobant et al., 1995). The first charge-remote cleavage type produces a carboxylate anion and a neutral which remain initially together in an ion-dipole complex. The complex can then dissociate or might be preceded by a proton transfer between the carboxylate anion and the neutral, hence, leading to the loss of a neutral carboxylic acid. From the resulting complementary pair of ions, only the carboxylate ion is observed at m/z 153. A second cleavage type, which occurs to a lesser extent, produces a neutral ketene and an alkoxide anion that remain together in an ion-dipole complex. Again the complex can dissociate as such or might be preceded by a proton transfer yielding a ynolate ion and the neutral alcohol. This complementary pair of ions was observed at m/z 135 (ynolate ion) and 179 (alkoxide ion). No hit with the same chemical formula was obtained when matching to MassBank. However, when using MetFrag, 23 biological compounds were retrieved from the Pubchem database and, following *in silico* fragmentation, the best hit returned our proposed structure for compound **44**.

45. protocatechoyl Xyl

The ion's chemical formula was $C_{12}H_{13}O_8$ (m/z 285.06195). The MS^2 spectrum was very similar to that of protocatechoyl Glc, but the MS^2 base peak (m/z 153) was due to a pentose loss instead of a hexose loss. This compound was characterized as protocatechoyl xylose. No hit with the

same chemical formula was obtained when matching to MassBank. Using MetFrag, Pubchem returned 47 biological compounds with the same chemical formula. After *in silico* fragmentation, the best hit returned our proposed structure.

46. *p*-hydroxybenzoic acid hex

This ion had $C_{13}H_{15}O_8$ (m/z 299.07748) as chemical formula. The MS^2 spectrum showed a base peak at m/z 137 resulting from hexose loss and MS^3-induced decarboxylation yielded the peak at m/z 93. Fortunately, decarboxylation of the aglycone could be pinpointed in the MS^3 spectrum using a linear ion trap despite that this fragmentation pathway did not lead to a stable peak in the MS^2 spectrum of the corresponding aglycone standard using a 3D ion trap (Supplemental Methods, Supplemental Table 1). This compound was annotated as *p*-hydroxybenzoic acid hexoside because no ester bond was evident based on its two types of characteristic cleavages and the hexose cross-ring cleavages (see protocatechoyl Glc **44**). No hit with the same chemical formula was obtained when matching to MassBank. Using MetFrag, 84 biological compounds were retrieved from Pubchem. Following *in silico* fragmentation, the three best hits (score=1, # explained peaks=4/5) were the *o*-, *m*- and *p*-hydroxybenzoic acid hexosides.

Phenylpropanoid derivates

47. 5-hydroxyferuloyl Glc

The ion's chemical formula was $C_{16}H_{19}O_{10}$ (m/z 371.09825). Elimination of a hexose moiety lead to the base peak at m/z 209. First product ions at m/z 251, 281 and 311 due to hexose cross-ring cleavages indicated that the hexose was linked in an ester bond (Vanholme et al., 2012) or that it was a hexoside with a free reducing end (Carroll et al., 1995). An ester bond is confirmed by its characteristic second type of cleavage in which an ynolate ion was formed at m/z 191. Further dissociation yielded the first product ion at m/z 176 via methyl radical loss. MS^3 fragmentation of the base peak at m/z 209 showed that methyl radical loss was the major fragmentation pathway (ion at m/z 194), but also decarboxylation (m/z 165) and a combined methyl radical loss / decarboxylation (m/z 150) were observed. As these losses are typical for hydroxycinnamic acids, e.g. MS^2 spectrum of 4-hydroxy-3-methoxy-cinnamic (ferulic) and 4-hydroxy-3,5-dimethoxy cinnamic (sinapic) acid in Supplemental Table 2 (see also Supplemental Methods), and because decarboxylation only occurred following hexose loss in the MS^2 spectrum, this compound was annotated as 5-hydroxyferuloyl glucose. No hit with the same

chemical formula was obtained when matching to MassBank. From the Pubchem, 20 biological compounds having the same chemical formula were retrieved using MetFrag. None of them was a hydroxycinnamic acid.

48. sinapoyl malate hex

This ion had $C_{21}H_{25}O_{14}$ (m/z 501.12423) as chemical formula and was connected via a "hexosylation" conversion to *trans*-sinapoyl malate **58**. The levels of both compounds were also highly correlated across biological replicates. Its MS^2 spectrum showed first product ions resulting from either hexose (m/z 339) or malate (m/z 385) loss. The loss of both moieties lead to the first product ion at m/z 223 representing a sinapate ion. Therefore, this compound was characterized as sinapoyl malate hexoside.

49. 5-hydroxyferuloyl malate

The ion's chemical formula was $C_{14}H_{13}O_9$ (m/z 325.05674). Although no MS^2 spectrum was obtained for this compound, a MS^2 spectrum was recorded for an in-source fragment resulting from the loss of 116 Da which corresponds with a malate moiety. The MS^2 spectrum of the in-source fragment was identical to the MS^3 spectrum of 5-hydroxyferuloyl Glc **47**. Therefore, this compound was elucidated as 5-hydroxyferuloyl malate. No hit with the same chemical formula was obtained when matching to MassBank. Using MetFrag, the MS^2 spectrum of the in-source-produced 5-hydroxyferulic acid was analyzed. From the Pubchem database, 430 biological compounds having the same chemical formula were retrieved. However, the *in silico* obtained spectrum for 5-hydroxyferulic acid was not even included in the 150 best hits. In negative ionization mode, MetFrag rendered good results for smaller phenolics, i.e., the benzenoids, but was not efficient anymore for explaining the, more complicated, gas phase fragmentation behavior of larger phenolics such as the phenylpropanoids.

50. sinapoyl gentiobiose

The chemical formula of this ion was $C_{23}H_{31}O_{15}$ (m/z 547.16645). Although no MS^2 spectrum was obtained for this compound, the CSPP network showed an association with *cis*-sinapoyl Glc **55** via a "hexosylation" reaction. Furthermore, levels of both compounds were very highly correlated across biological replicates. Therefore, this compound was characterized as a sinapoyl diglycoside. Because one or more sinapic acid moieties esterified to gentiobiose [glucosyl-

(1→6)-glucose] has been observed in other *Brassicaceae* (Supplemental Data Set 1), this compound was more narrowly defined as sinapoyl gentiobiose.

51. sinapoyl malate hex

This ion had the same chemical formula as sinapoyl malate hex **48** and was also via a "hexosylation" conversion linked to *trans*-sinapoyl malate **58** in the CSPP networks. Additionally, the levels of the latter compound were highly correlated with those of compound **51** across biological replicates. Its MS^2 spectrum was dominated by malate loss (m/z 385) and MS^3 fragmentation of the m/z 385 first product ion yielded a sinapate ion due to hexose loss. Therefore, this compound is a sinapoyl malate hexoside isomer.

52. feruloyl glycerol

The ion had $C_{13}H_{15}O_6$ (m/z 267.08784) as chemical formula. The MS^2 spectrum showed a base peak at m/z 149 and smaller peaks at m/z 134, 178 and 193. The first three peaks likely arise from the m/z 193 ion via decarboxylation, a combined decarboxylation and methyl radical loss, and a methyl radical loss, respectively. These losses are typically for the hydroxycinnamic acids and indicate that a ferulic acid moiety is present (see Supplemental Methods and Supplemental Table 2). The 74 Da loss leading to the first product ion at m/z 193 corresponds with an esterified glycerol moiety. Indeed, the second ester-specific ketene-producing cleavage yielded the first product ynolate ion at m/z 175. A rearrangement followed by a decarboxylation rendered the peak at m/z 223, whereas the peak at m/z 192 resulted from homolytic cleavage of the ester bond (Bowie, 1990). Therefore, this compound was characterized as feruloyl glycerol. No hit with the same chemical formula was obtained when matching to MassBank.

53. *trans*-sinapoyl Glc

The ion's chemical formula was $C_{17}H_{21}O_{10}$ (m/z 385.11406). The MS^2 spectrum showed a base peak at m/z 223 due to hexose loss (-162 Da) which, upon MS^3 fragmentation, yielded second product ions at m/z 208, 179 and 164 due to methyl radical loss, decarboxylation and a combined methyl radical loss and decarboxylation. Thus, this MS^3 spectrum shows the typical collision-induced dissociation fingerprint of sinapic acid (Supplemental Methods, Supplemental Table 2). The observation of first product ions at m/z 325, 295 and 265 arise from hexose cross-ring cleavages and occur whenever the hexose is connected as a hexoside with a free reducing end (Carroll et al., 1995) or when the hexose is linked as an ester (Vanholme et al., 2012). Since

decarboxylation only occurred after hexose loss and the second ester-characteristic cleavage produced the MS^2 ynolate ion at m/z 205, this compound was characterized as sinapoyl glucose. Matching to MassBank rendered a score and hit of 0.51 and 5.

54. sinapoyl gentiobiose

The ion had the same chemical formula as sinapoyl gentiobiose **50**. It was connected to sinapoyl malate hexoside **48** via a "malate-hexoside transesterification" conversion in the CSPP networks and the levels of the CSPP substrate and product peaks were highly correlated across biological replicates. Its MS^2 spectrum was dominated by the first product ion at m/z 223 representing a sinapate ion. Other major first product ions were observed at m/z 385 (-162 Da, dehydrated hexose loss), m/z 367 (-180 Da, hexose loss), m/z 349 (-198 Da, combined loss of hexose and water), m/z 323 (-224 Da, sinapic acid loss), m/z 289 (-258 Da, combined loss of hexose, water and two molecules of formaldehyde). The latter first product ion arises from hexose cross-ring cleavages. Also the less abundant first product ions at m/z 325 and 295 arise from hexose cross-ring cleavage occurring on the m/z 385 first product ion. As already mentioned, cross-ring cleavages can occur when the hexose is linked as an ester (Vanholme et al., 2012). Consequently, the MS^2 data suggest a disaccharide ester-linked to sinapic acid. For the same reasoning as described for compound **50**, compound **54** is defined as a sinapoyl gentiobiose isomer.

55. *cis*-sinapoyl Glc

The same chemical formula and similar MS^n spectra were obtained for this compound as for *trans*-sinapoyl Glc **53**. Because this compound eluted later and was much less abundant, it was characterized as the *cis* isomer of sinapoyl Glc, i.e., *cis*-sinapoyl glucose. The MassBank score and hit value were 0.48 and 5.

56. sinapoyl pen

The chemical formula of this ion was $C_{16}H_{19}O_9$ (m/z 355.10357). No MS^2 spectrum was obtained, but the compound was a CSPP substrate for a "dihydroxybenzoylation" conversion to the CSPP product dihydroxybenzoyl sinapoyl pen **62**. Also the levels of both compounds were highly correlated across biological replicates. Thus, this compound was annotated as sinapoyl pentose.

57. disinapoyl butanoyl gentiobiose

The ion had $C_{38}H_{47}O_{20}$ (m/z 823.26695) as chemical formula. No MS^2 spectrum was recorded, but the compound was connected with trisinapoyl gentiobiose **66** via a "dihydroxybenzoylation" in the CSPP networks. Both compound levels were highly correlated across biological replicates, strongly suggesting that compound **57** was a sinapoyl gentiobiose derivate as well. A putative candidate is disinapoyl butanoyl gentiobiose.

58. *trans*-sinapoyl malate

The chemical formula of this ion was $C_{15}H_{15}O_9$ (m/z 339.07220). The MS^2 spectrum showed a base peak at m/z 223 due to malate loss (-116 Da) and MS^3 fragmentation of this first product ion rendered second product ions at m/z 208, 179 and 164 which are characteristic for sinapic acid (Supplemental Table 2). The second ketene-producing cleavage was not observed for this ester, because the ester cleavage reaction leading to the sinapate anion is too favorable. Besides classical charge-remote ester cleavage (Stroobant et al., 1995), the neighboring carboxylic acid functions on the malate moiety facilitate proton transfer to the ester, enhancing a cleavage reaction similar to a β-keto acid decarboxylation in solution chemistry. Therefore, this compound is sinapoyl malate. Matching to MassBank returned a score and hit of 0.71 and 1.

59. *cis*-sinapoyl malate

The chemical formula and MS^n spectra were the same as for *trans*-sinapoyl malate **58** and, because compound **59** had lower levels and was eluting later, it was characterized as *cis*-sinapoyl malate. Matching to MassBank returned a score and hit of 0.71 and 1.

60. disinapoyl gentiobiose

The ion's chemical formula was $C_{34}H_{41}O_{19}$ (m/z 753.22528). It's MS^2 spectrum was dominated by a peak at m/z 591 resulting from hexose loss (-162 Da). No neutral losses of -60, -90 and/or -120 Da were observed, indicating that this hexose was connected via its reducing end in a glycosidic rather than an ester bond (Carroll et al., 1995; March and Stadey, 2005; Vanholme et al., 2012). This ion dissociated further by expelling a sinapic acid residue leading to the second product ion at m/z 367. A second hexose loss (-144 Da) lead to the second product ion at m/z 223; the latter ion being reminiscent of a second sinapic acid moiety in the molecule. Second product ions at m/z 307 and 277 (-60 and -90 Da losses) indicated that this hexose was linked via an ester bond. This was further confirmed by the second product ion at m/z 349 representing the ynolate ion due to the second ketene-producing cleavage typical for esters. Consequently, this

compound was characterized as disinapoyl gentiobiose. Disinapoyl gentiobiose has been observed in other *Brassicaceae* in which both sinapoyl moieties were connected to the 1-O- and 2-O positions of gentiobiose (Supplemental Data Set 1). No hit with the same chemical formula was obtained when matching to MassBank.

61. disinapoyl gentiobiose

This compound had the same chemical formula and similar MS^n spectra as disinapoyl gentiobiose **60** and is, consequently, an isomer. This compound was characterized as disinapoyl gentiobiose. No hit with the same chemical formula was obtained when matching to MassBank.

62. dihydroxybenzoyl sinapoyl pen

The chemical formula of this ion was $C_{23}H_{23}O_{12}$ (m/z 491.11868). The MS^2 spectrum showed two sets of complementary peaks, i.e., at m/z 337 and m/z 153 and at m/z 267 and m/z 223, suggesting that there were two ester bonds involved. Upon CID, ester bonds are subjected to a charge-remote cleavage producing a carboxylate anion and a neutral that remain in an ion-neutral complex (Debrauwer et al., 1992; Stroobant et al., 1995). Following dissociation of the ion-neutral complex, the carboxylate ions provide the peaks at m/z 153 and m/z 223 representing dihydroxybenzoate and sinapate. Alternatively, before complex dissociation, the carboxylate ion can abstract a proton from the neutral of which the ion is visible by peaks at m/z 337 and m/z 267 representing the sinapoyl pentose and the dihydroxybenzoyl pentose moiety, respectively. Therefore, this compound was presumably dihydroxybenzoyl sinapoyl pentose. Further support was derived from the remaining MS^2 first product ions. The second ester-characteristic ketene-producing cleavage is observed for the sinapoyl moiety rendering ions at m/z 285 and 205 representing the alkoxide and ynolate ions, respectively. Although this second cleavage type is not observed for the dihydroxybenzoyl moiety, the latter can initiate a rearrangement converting dihydroxybenzoyl sinapoyl pentose into (iso)vanilloyl 5-hydroxyferuloyl pentose. When the dihydroxybenzoyl moiety is deprotonated, the resulting phenoxide ion is stabilized by the ortho-hydroxy group. The phenoxide anion can then attack one of the methoxy groups on the sinapoyl system converting it into a 5-hydroxyferuloyl system in a similar S_N2 reaction as observed for the methanol loss upon CID of sinapyl alcohol (3,5-dihydroxy-4-methoxycinnamyl alcohol; see Supplemental Methods, Supplemental Table 2 and Supplemental Figure 8A). This rearrangement is followed by charge-remote ester cleavages, providing the base peak at m/z 323 (vanillic acid loss, -168 Da), the ion at m/z 209 (5-hydroxyferulate ion), or the ketene-producing cleavage

providing the sinapic acid-derived ynolate ion at m/z 191. The latter first product ion can lose a methyl radical yielding the first product ion at m/z 176. The rearrangement can only occur when the dihydroxybenzoyl and sinapoyl moieties are close together, e.g., connected to the 1-O- and 2-O-positions of the pentose ring. The first product ion at m/z 233 likely resulted from pentose cross-ring cleavage of the m/z 323 ion which represented the sinapoyl pentose moiety. Noticeable, assuming the rearrangement occurs in the reverse direction, i.e., producing dihydroxybenzoyl sinapoyl pentose from (iso)vanilloyl 5-hydroxyferuloyl pentose, all MS^n peaks could be explained in exactly the same way. Therefore, it cannot be excluded that this compound is (iso)vanilloyl 5-hydroxyferuloyl pentose. When matching to MassBank, no hit with the same chemical formula as that of dihydroxybenzoyl sinapoyl pentose was obtained.

63. disinapoyl hexanetriol dihex

The ion's chemical formula was $C_{40}H_{53}O_{21}$ (m/z 869.31077). This compound was connected to disinapoyl butanoyl gentiobiose **57** via a "malate-hexose transesterification" conversion in the CSPP networks. However, this CSPP connection is based on the mass difference and does not necessarily visualize a true biochemical transesterification. Nonetheless, as the levels of both compounds were highly correlated across biological replicates, they are likely structurally similar. In the full MS, the detection of a peak at m/z 753 might originate from the in-source loss of hexanetriol (-116 Da). Therefore, this compound was annotated as disinapoyl hexanetriol dihexos-e/-ide.

64. disinapoyl Glc

The chemical formula of this ion was $C_{28}H_{31}O_{14}$ (m/z 591.17154). Its MS^2 spectrum was identical to the MS^3 spectrum of the first product ion at m/z 591 of disinapoyl gentiobiose **60**. Therefore, this compound was characterized as disinapoyl glucose. When matching to MassBank, no hit with the same chemical formula as compound **64** was obtained.

65. disinapoyl Glc

The chemical formula and MS^2 spectrum were identical to that of disinapoyl Glc **64**. Therefore compound **65** was characterized as a disinapoyl glucose isomer. When matching to MassBank, no hit with the same chemical formula as compound **65** was obtained.

66. trisinapoyl gentiobiose

This ion had $C_{45}H_{51}O_{23}$ (m/z 959.28407) as chemical formula. Its MS^2 spectrum showed a base peak at m/z 797 due to hexose loss (-162 Da). Further MS^3 fragmentation of this first product ion eliminated a sinapic acid-derived unit, rendering the peaks at m/z 591 and 573 corresponding with the two cleavages typical for esters (Debrauwer et al., 1992; Fournier et al., 1993; Fournier et al., 1995). The first product ions at m/z 501 and 349 are due to hexose cross-ring cleavage and the loss of a second sinapic acid-derived unit. This compound was characterized as trisinapoyl gentiobiose. When matching to MassBank, no hit with the same chemical formula as compound **66** was obtained.

67. disinapoyl Glc

The chemical formula and MS^2 spectrum were identical to that of disinapoyl Glc **64**. Therefore compound **67** was characterized as a disinapoyl glucose isomer. When matching to MassBank, no hit with the same chemical formula as compound **67** was obtained.

68. trisinapoyl gentiobiose

The chemical formula was the same as obtained for trisinapoyl gentiobiose **66**. No MS^2 spectrum was obtained, but it was connected to disinapoyl gentiobiose **60** via a "sinapic acid derivatization" conversion in the CSPP networks. Levels of both compounds were also correlated across biological replicates. Therefore, this compound was annotated as a trisinapoyl gentiobiose isomer.

(Neo)Lignans/Oligolignols

69. hex G(8–O–4)FA malate

The ion's chemical formula was $C_{30}H_{35}O_{17}$ (m/z 667.18775). No MS^2 spectrum was obtained, but it was connected to G(8–O–4)FA malate **86** via a "hexosylation" conversion in the CSPP networks. Levels of both compounds were highly correlated across biological replicates. Therefore, this compound was characterized as the hexoside of guaiacylglycerol 8–O–4 feruloyl malate ether.

70. G(8–O–4)FA hex

The ion's chemical formula was $C_{26}H_{31}O_{13}$ (m/z 551.17649). Upon CID, hexose loss (m/z 389, -162 Da) was the dominating fragmentation pathway. The MS^3 spectrum of this base peak showed two small neutral losses, i.e., water loss (-18 Da) and a combined water / formaldehyde

loss (-48 Da), leading to the ion at m/z 371 and 341. These are characteristic for 1,4-propanediol or 1,3-propanediol moieties and, thus, for a dibenzylbutanediol or β-aryl ether (neo)lignan/oligolignol (Eklund et al., 2008; Morreel et al., 2010a). In the case of a β-aryl ether, cleavage of the 8–O–4-linkage provides information on the composing units (Morreel et al., 2010a; Morreel et al., 2010b). The ions at m/z 195 and 193 represent a guaiacyl unit connected to a unit derived from ferulic acid. The former ion fragments further to m/z 165 due to formaldehyde loss, whereas the latter ion shows the typical fragmentation pattern (m/z 178, 149 and 134) of ferulic acid (Supplemental Table 2). Therefore, this compound was characterized as guaiacylglycerol 8–O–4 ferulic acid ether hexoside. The shorthand name for the aglycone, G(8–O–4)FA, is based on Morreel et al. (2004) and is explained in the legend of Supplemental Data Set 1. When matching to MassBank, no hit with the same chemical formula as compound **70** was obtained, but the CID spectrum has been described previously (Vanholme et al., 2012). Using MetFrag, 51 biological compounds having the same chemical formula were retrieved from the Pubchem database. None of them was a β-aryl ether.

71. lariciresinol dihex

This ion had $C_{32}H_{43}O_{16}$ (m/z 683.25571) as chemical formula. Dissociation of a hexose lead to the MS^2 peak at m/z 521. Further MS^3 fragmentation yielded a peak at m/z 359 indicating a second hexose loss. The second product ion at m/z 329 indicated a further formaldehyde loss and is typical for lariciresinol (Morreel et al., 2010a), i.e., pinoresinol or G(8–8)G in which one of the tetrahydrofuran rings is reduced. Its structural characterization was confirmed by its connection to pinoresinol dihex **79** via a "reduction" conversion. The levels of both compounds were very highly correlated across biological replicates. When matching to MassBank, no hit with the same chemical formula as compound **71** was obtained. Using MetFrag, 51 biological compounds having the same chemical formula were retrieved from the Pubchem database. None of them was a resinol-derived compound. Because lariciresinol hexoside, represented by the first product ion at m/z 521, was expected to be present in the Pubchem database, the MS^3 spectrum was entered into MetFrag. Both second product ions from lariciresinol hexoside were recognized leading to a perfect match.

72. G(8–O–4)G hex

The ion's chemical formula was $C_{28}H_{37}O_{14}$ (m/z 597.21896). Its MS^2 spectrum showed the loss of an acetate adduct (-60 Da) rendering the peak at m/z 537. Another peak at m/z 375 indicated

that this compound was hexosylated. Furthermore, a first product ion at m/z 327 was found due to the combined loss of water and formaldehyde (-48 Da) from the aglycone (m/z 375). This is characteristic for β-aryl ethers (Morreel et al., 2010a). Therefore, this compound is guaiacylglycerol 8–O–4 coniferyl ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **72** was obtained. Using MetFrag, 57 biological compounds having the same chemical formula were retrieved from the Pubchem database. Two hexosides isomers of G(*t*8–O–4)G hexoside were present among the 10 best hits. The MS^3 spectrum of the first product ion at m/z 375 represented the fragmentation of the aglycone. When the accurate mass of this aglycone was entered into MetFrag, 388 hits with the Pubchem database were returned. However, β-aryl ethers were not among the top 50 hits as only three of the four fragment ions (m/z 327, m/z 195 and m/z 179, but not m/z 165) were recognized.

73. hex G(8–O–4)FA malate

The ion's chemical formula ($C_{30}H_{35}O_{17}$; m/z 667.18877) was the same as for hex G(8–O–4)FA malate **69**. Its MS^2 spectrum showed a peak at m/z 551 due to malate loss. Following further MS^3 dissociation of the first product ion at m/z 551, a peak appeared at m/z 389 due to hexose loss and also at m/z 341 due to the combined loss of water and formaldehyde from the aglycone (m/z 389) which is typical for β-aryl ethers. Therefore, this compound is an isomer of guaiacylglycerol 8–O–4 feruloyl malate ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **73** was obtained. No MetFrag search was performed because the aglycone is not present in the Pubchem database.

74. G(t8–O–4)G hex

The ion had $C_{26}H_{33}O_{12}$ (m/z 537.19774) as chemical formula. Its MS^2 spectrum was dominated by a peak at m/z 375 due to hexose loss. The latter ion was further fragmented to the ions at m/z 327, 195 and 179 in the MS^3 spectrum. These are due to a combined water/formaldehyde loss, characteristic for the 8–O–4-linkage in (neo)lignans/oligolignols and cleavage of this linkage resulting in second product ions representing each of the units in this dimer (Morreel et al., 2010a). The fragmentations in this MS^3 spectrum are well-documented and the aglycone-based MS^2 spectrum has been published before (Morreel et al., 2010a). This compound is the *threo* isomer of guaiacylglycerol 8–O–4 coniferyl ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **74** was obtained. MetFrag-based structural elucidation rendered the same result as for G(*t*8–O–4)G hex **72**.

75. G(8–O–4)G(8–O–4)G hex

This ion had $C_{36}H_{45}O_{16}$ (m/z 733.27194) as chemical formula. No MS^2 spectrum was obtained, but the compound was connected to G(t8–O–4)G hex **74** via a "G unit addition" conversion. Furthermore, both compound levels were very highly correlated across biological replicates. Therefore, this compound was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 coniferyl ether hexoside.

76. S(8–O–4)FA hex

This ion had $C_{27}H_{33}O_{14}$ (m/z 581.18709) as chemical formula. Upon CID, a base peak at m/z 419 was observed due to hexose loss. MS^3 fragmentation yielded the type I small neutral losses at m/z 401 (water loss) and 371 (combined water/formaldehyde loss) that are characteristic for β-aryl ethers (Morreel et al., 2010a). The type II cleavage of the linkage reveals the type of units involved (Morreel et al., 2010a): peaks at m/z 225 and 193 indicate a syringyl unit coupled to a unit derived from ferulic acid. Therefore, this compound is syringylglycerol 8–O–4 ferulic acid ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **76** was obtained. The aglycone was not present in the Pubchem database, rendering MetFrag-based structural verification redundant.

77. G(8–O–4)FA hex

The ion had the same chemical formula ($C_{26}H_{31}O_{13}$, m/z 551.17679) and very similar MS^n spectra as G(8–O–4)FA hex **70**. The additional MS^3 peak at m/z 195 confirmed that this compound contains a guaiacyl unit and is an isomer of guaiacylglycerol 8–O–4 ferulic acid ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **77** was obtained. As the aglycone is not present in the Pubchem database, no MetFrag-based annotation was performed.

78. G(8–O–4)FA Glu

The chemical formula of this ion was $C_{25}H_{28}O_{11}N$ (m/z 518.16698). No MS^2 spectrum was recorded, but the compound was connected to G(8–5)FA Glu **112** via a "hydration" conversion. The levels of both compounds were very highly correlated across biological replicates. This compound was annotated as guaiacylglycerol 8–O–4 feruloyl glutamic acid ether. As the aglycone is not present in the Pubchem database, no MetFrag-based annotation was performed. However, as glutamate derivates of (neo)lignans/oligolignols were not expected, the standard

was synthesized allowing the authentication of these compound structures. Therefore, this proved that (neo)lignans/oligolignols can be derivatized with glutamate.

79. pinoresinol dihex

The ion's chemical formula was $C_{32}H_{41}O_{16}$ (m/z 681.24042). Upon CID, a hexose loss was observed leading to the ion at m/z 519. MS^3 fragmentation of this first product ion rendered the peak at m/z 357 due to a second hexose loss. The MS^4 spectrum of this second product ion was identical to the MS^2 spectrum of pinoresinol or G(8–8)G (Ye et al., 2005; Guo et al., 2007; Eklund et al., 2008; Ricci et al., 2008; Morreel et al., 2010a; Hanhineva et al., 2012). Therefore, this compound is pinoresinol dihexoside. When matching to MassBank, no hit with the same chemical formula as compound **79** was obtained. Using MetFrag, 37 molecules with the same chemical formula were retrieved from the Pubchem database of which pinoresinol diglucoside was among the best hits upon *in silico* fragmentation. To determine whether MetFrag would be suitable to determine the resinol aglycone structure, the MS^4 spectrum of the m/z 357 second product ion (representing pinoresinol) was entered. Almost 1000 biological molecules were retrieved from the PubChem database and, upon *in silico* fragmentation, pinoresinol was among the top 25 best hits. All fragment ions were recognized by the software.

80. hex G(8–5)FA Glu

The chemical formula of this ion was $C_{31}H_{36}O_{15}N$ (m/z 662.20968). Its MS^2 spectrum was dominated by a peak at m/z 500 due to hexose loss. When subjected to MS^3 fragmentation, water and formaldehyde losses rendered the peaks at m/z 482 and 470. The second product ion at m/z 371 was formed by glutamic acid loss (-129 Da). It also fragmented further by water and formaldehyde which explains the peaks at m/z 353 and 341. These losses are typical for 8–5-linked neolignans/oligolignols (Morreel et al., 2010a). Taking the chemical formula into account, this compound was characterized as dihydroconiferyl alcohol 8–5 feruloyl glutamic acid hexoside. When matching to MassBank, no hit with the same chemical formula as compound **80** was obtained. Using MetFrag, 4 biological compounds having the same chemical formula were retrieved from the Pubchem database. None of them was a phenylcoumaran.

81. G(8–5)FA dihex

The chemical formula of this ion was $C_{32}H_{39}O_{17}$ (m/z 695.21981). No MS^2 spectrum was recorded, but the compound was connected to G(8–5)FA hex **104** in the CSPP network via a

"hexosylation" reaction. Levels of both compounds were very highly correlated across biological replicates. Therefore, this compound was annotated as dihydroconiferyl alcohol 8–5 ferulic acid dihexoside.

82. G(8–O–4)G(8–O–4)FA hex

This ion had $C_{36}H_{43}O_{17}$ (m/z 747.25195) as chemical formula. This compound was linked to G(8–O–4)FA hex **70** via a "G unit addition" in the CSPP network. Moreover, the abundances of both compounds were highly correlated. Its MS^2 spectrum indicated the presence of a hexose moiety (m/z 585, -162 Da). Therefore, this compound was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 ferulic acid ether hexoside.

83. G(8–O–4)lariciresinol dihex

The chemical formula of the ion was $C_{42}H_{55}O_{20}$ (m/z 879.33274). The abundance was too low to obtain a MS^2 spectrum. However, the compound was connected to lariciresinol dihex **71** via a "G unit addition" conversion in the CSPP networks and both compound levels were highly correlated across biological replicates. This compound was annotated as guaiacylglycerol 8–O–4 lariciresinol ether dihexoside.

84. G(8–O–4)FA Glu

This compound had the same chemical formula and showed the same CSPP network connections as G(8–O–4)FA Glu**78**. The abundances of both compounds were mutually highly correlated across biological samples.

85. G(8–O–4)SA hex

The chemical formula of this ion was $C_{27}H_{33}O_{14}$ (m/z 581.18713). The base peak in the MS^2 spectrum was at m/z 419 indicating a hexose loss (-162 Da). MS^3 fragmentation of this MS^2 base peak yielded the small neutral losses characteristic for β-aryl ether neolignan/oligolignols (Morreel et al., 2010a), i.e., the so-called type I fragmentations: water (-18 Da) and a combined water/formaldehyde (-48 Da) loss leading to ions at m/z 401 and 471. The type II fragmentations start by cleavage of the 8–O–4-linkage providing product ions representing the connecting units. These ions, i.e., at m/z 223 and 195 indicated that a guaiacyl unit was coupled to a unit derived from sinapic acid. This compound was characterized as guaiacylglycerol 8–O–4 sinapic acid ester hexoside. When matching to MassBank, no hit with the same chemical formula as

compound **85** was obtained. Using MetFrag, 27 biological compounds having the same chemical formula were retrieved from the Pubchem database. None of them was a β-aryl ether.

86. G(8–O–4)FA malate

The ion's chemical formula was $C_{24}H_{25}O_{12}$ (m/z 505.13495). The MS^2 base peak was at m/z 389 due to malate loss (-116 Da). Subjecting the ion to MS^3 fragmentation showed peaks at m/z 341 (-48 Da loss, type I cleavage of β-aryl ethers) and at m/z 193 and 195 (due to β-aryl ether-characteristic type II cleavages) corresponding with a unit derived from ferulic acid and a guaiacyl unit. Therefore, this compound was characterized as guaiacylglycerol 8–O–4 feruloyl malate ether. When matching to MassBank, no hit with the same chemical formula as compound **86** was obtained.

87. G(8–O–4)lariciresinol dihex

The chemical formula was the same at that of G(8–O–4)lariciresinol dihex **83**. The abundance was too low to obtain a MS^2 spectrum. However, the compound was connected to lariciresinol dihex **71** via a "G unit addition" conversion in the CSPP networks and both compound levels were highly correlated across biological replicates. This compound was annotated as another guaiacylglycerol 8–O–4 lariciresinol ether dihexoside isomer.

88. lariciresinol hex

This ion had $C_{26}H_{33}O_{11}$ (m/z 521.20286) as chemical formula. The MS^2 spectrum indicated a hexose loss (-162 Da) rendering the base peak at m/z 359. Its MS^3 spectrum was dominated by formaldehyde loss leading to the peak at m/z 329. This spectrum is typical for the CID spectrum of lariciresinol, i.e., reduced pinoresinol [G(8–8)G] (Eklund et al., 2008; Morreel et al., 2010a; Hanhineva et al., 2012). Therefore, this compound was characterized as lariciresinol hexoside. When matching to MassBank, no hit with the same chemical formula as compound **88** was obtained. Using MetFrag, 107 compounds were returned from the Pubchem database of which, upon *in silico* fragmentation, the best hit was our proposed structure.

89. hex G(8–5)FA malate

The chemical formula of this ion was $C_{30}H_{33}O_{16}$ (m/z 649.17719). Elimination of malate lead to the MS^2 base peak at m/z 533 which, upon MS^3 fragmentation, expelled a hexose group leading to the second product ion at m/z 371. This aglycone underwent losses of 18 and 30 Da representing water and formaldehyde and providing the second product ions at m/z 353 and 341.

Both losses are characteristic for phenylcoumaran neolignans/oligolignols (Morreel et al., 2010a; 2010b). Therefore, this compound is dihydroconiferyl alcohol 8–5 feruloyl malate hexoside. When matching to MassBank, no hit with the same chemical formula as compound **89** was obtained. No structural elucidation via MetFrag was performed as G(8–5)FA is not present in the PubChem database.

90. G(8–O–4)G(8–O–4)SA hex

The ion had $C_{37}H_{45}O_{18}$ (m/z 777.26334) as chemical formula. No MS^2 fragmentation occurred, but the compound was connected to G(8–O–4)FA hex **70** via a "S unit addition" conversion in the CSPP network. The levels of both compounds are highly correlated across biological replicates. Although this would suggest the compound to be S(8–O–4)G(8–O–4)FA hex, the 8–O–4 coupling of a sinapyl alcohol radical to a oligolignol radical is not favored owing to oxidation potential differences (Syrjänen and Brunow, 1998). Therefore, this compound was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 sinapic acid ether hexoside.

91. G(8–O–4)pinoresinol dihex

The ion's chemical formula was $C_{42}H_{53}O_{20}$ (m/z 877.31705). No MS^2 fragmentation was recorded, but the compound was connected to pinoresinol dihex **79** via a "G unit addition" conversion in the CSPP network. Levels of both compounds were correlated across biological replicates and, thus, this compound was annotated as guaiacylglycerol 8–O–4 pinoresinol ether dihexoside.

92. G(8–O–4)SA hex

The chemical formula of this ion was $C_{27}H_{33}O_{14}$ (m/z 581.18785). Its MS^2 spectrum showed a base peak at m/z 419 due to hexose loss. Upon MS^3, the latter ion expelled water and formaldehyde (-48 Da) yielding the second product ion at m/z 371. In addition to this β-aryl ether-characteristic type I fragmentation, type II fragmentations lead to the ions at m/z 223 and m/z 195 representing the units derived from sinapic acid (supported by the observation of a second product ion at m/z 208 due to methyl radical loss from the m/z 223 ion) and from coniferyl alcohol (this ion fragmented further by formaldehyde loss yielding the second product ion at m/z 165). Therefore, this compound was characterized as guaiacylglycerol 8–O–4 sinapic acid ether hexoside. When matching to MassBank, no hit with the same chemical formula as

compound **92** was obtained. MetFrag-based structural characterization was not performed as G(8–O–4)SA is not present in the PubChem database.

93. G(8–O–4)SA Glu

The ion had $C_{26}H_{30}O_{12}N$ (m/z 548.17812) as chemical formula, but no MS^2 spectrum was obtained. In the CSPP network, the compound was linked to G(8–O–4)FA Glu **78** via a "methoxylation" conversion. Furthermore, the levels of both compounds were correlated across biological replicates. Therefore, this compound is annotated as guaiacylglycerol 8–O–4 sinapoyl glutamic acid ether.

94. G(8–5)G hex

The chemical formula of the acetate adduct of this compound was $C_{28}H_{35}O_{13}$ (m/z 579.20835). MS^2 fragmentation yielded the deprotonated compound at m/z 519. Other second product ions at m/z 357 (due to hexose loss) and at m/z 339 and 327 (losses of water and formaldehyde) – type I fragmentations typically observed in the spectrum of phenylcoumarans – pointed to a hexoside of a phenylcoumaran. The type II ion at m/z 221 indicated the presence of a guaiacyl unit and, thus this compound was characterized as (8–5)-dehydrodiconiferyl alcohol hexoside. When matching to MassBank, no hit with the same chemical formula as compound **94** was obtained. To observe the extent that the aglycone could be characterized using MetFrag, the MS^3 spectrum of the m/z 357 first product ion was entered. 988 compounds with the same chemical formula were retrieved from the PubChem database, yet G(8–5)G was not among the 50 best hits upon *in silico* fragmentation. Only four out of the five second product ions (m/z 339, 327, 221 and 203 but not m/z 191) could be explained by MetFrag. Furthermore, MetFrag returned G(8–8)G as a better hit rendering explanations for all of the second product ions.

Although MetFrag recognized most of the product ions upon CID of any of the guaiacyl dimers, i.e., G(8–O–4)G, G(8–8)G and G(8–5)G, the large number of structural isomers (belonging to various biochemical classes) from the PubChem database that were equally likely good hits, was very confusing. Additionally, the MetFrag-proposed fragments were often not in agreement with previously published gas phase fragmentation reactions for these compounds (Eklund et al., 2008; Morreel et al., 2010a; Morreel et al., 2010b). Because of this lack of specificity and the absence of most of the (neo)lignan/oligolignol core structures in the PubChem database, MetFrag-based structural elucidation was not considered anymore for the remainder of these compounds.

95. G(8–O–4)G(8–O–4)SA hex

The ion's chemical formula was $C_{37}H_{45}O_{18}$ (m/z 777.26288). No MS^2 spectrum was obtained, but the compound was connected via a "S unit addition" conversion to G(8–O–4)FA hex **70** in the CSPP network. The levels of both compounds were very highly correlated across biological replicates. In agreement with the same reasoning as mentioned for compound **90**, compound **95** was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 sinapic acid ether hexoside.

96. lariciresinol hex

This ion had the same chemical formula and similar MS^n spectra as those of lariciresinol hex **88**. Therefore, this compound was characterized as a lariciresinol hexoside isomer. When matching to MassBank, no hit with the same chemical formula as compound **96** was obtained.

97. G(8–O–4)FA malate

The chemical formula of this ion was the same as the one of G(8–O–4)FA malate **86** and also the MS^n spectra were similar. Consequently, this compound was characterized as an isomer of guaiacylglycerol 8–O–4 feruloyl malate ether. When matching to MassBank, no hit with the same chemical formula as compound **97** was obtained.

98. G(8–5)FA hex

This ion had $C_{26}H_{29}O_{12}$ (m/z 533.16673) as chemical formula. No MS^2 spectrum was recorded, but a connection with G(8–O–4)FA hex **70** via a "hydration" conversion was observed in the CSPP network. Levels of both compounds were highly correlated and this compound was annotated as dihydroconiferylalcohol 8–5 ferulic acid hexoside or glycosmisic acid hexoside.

99. G(8–O–4)SA hex

The chemical formula of this ion was the same as that of G(8–O–4)SA hex **92**. The MS^n spectra were very similar to those of G(8–O–4)SA hex **92**. Thus, this compound was characterized as a guaiacylglycerol 8–O–4 sinapic acid ether hexoside isomer. When matching to MassBank, no hit with the same chemical formula as compound **99** was obtained.

100. G(8–O–4)G(8–O–4)FA hex

The ion had $C_{36}H_{43}O_{17}$ (m/z 747.25195) as chemical formula. No MS^2 spectrum was recorded, but a connection with G(8–O–4)FA hex **70** via a "G unit addition" conversion was observed in

the CSPP network. Levels of both compounds were correlated and this compound was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 ferulic acid ether hexoside.

101. G(8–O–4)pinoresinol dihex

The ion had $C_{42}H_{53}O_{20}$ (m/z 877.31685) as chemical formula. No MS^2 spectrum was recorded, but a connection with pinoresinol dihex **79** via a "G unit addition" conversion was observed in the CSPP network. Levels of both compounds were correlated and this compound was annotated as guaiacylglycerol 8–O–4 pinoresinol ether dihexoside.

102. G(8–O–4)G(8–O–4)FA hex

The chemical formula of this ion was the same as for G(8–O–4)G(8–O–4)FA hex **100** ($C_{36}H_{43}O_{17}$, m/z 747.25195). No MS^2 spectrum was obtained for this precursor ion, but an in-source hexose elimination rendered the ion at m/z 585.19801 ($C_{30}H_{33}O_{12}$) in the full MS spectrum. The MS^2 spectrum obtained for the latter ion showed ions at m/z 567 and 535 due to water loss (-18 Da) and the combined loss of water and formaldehyde (-48 Da). Both dissociations are typical type I fragmentations of β-aryl ethers (Morreel et al., 2010a; Morreel et al., 2010b; Hanhineva et al., 2012). Type II fragmentations were evident as well (Morreel et al., 2010a; Morreel et al., 2010b; Hanhineva et al., 2012). A neutral loss of 196 Da, indicating the presence of a guaiacylglycerol moiety, lead to the ion at m/z 389 representing a dimeric moiety. The m/z 341 ion could be explained by another β-aryl ether-specific type I-associated combined water/formaldehyde loss. The ion at m/z 193 suggests the presence of a unit derived from ferulic acid. Therefore, this compound was characterized as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 ferulic acid ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **102** was obtained.

103. G(8–O–4)S(8–8)G dihex

The ion had $C_{36}H_{43}O_{17}$ (m/z 747.25195) as chemical formula. No MS^2 spectrum was obtained but the compound was connected via a "S unit addition" conversion to G(8–8)G dihex or pinoresinol dihex **79** in the CSPP network. The levels of both compounds were highly correlated across biological replicates. Although this would suggest the 8–O–4-coupling of a sinapyl alcohol radical to a radical from pinoresinol leading to the S(8–O–4)G(8–8)G aglycone, this reaction is unfavorable (Syrjänen and Brunow, 1998). Therefore, the most logical structure to be annotated was guaiacylglycerol 8–O–4 medioresinol ether dihexoside or buddlenol E dihexoside.

104. G(8–5)FA hex

The ion had the same chemical formula ($C_{26}H_{29}O_{12}$, m/z 533.16580) as G(8–5)FA hex **98**. The MS^2 spectrum showed a base peak at m/z 371 due to hexose loss (-162 Da). MS^3 dissociation of the latter first product ion yielded the typical type I-associated neutral losses (water loss, formaldehyde loss and a combined water/methyl radical loss) of a phenylcoumaran neolignan/oligolignol (Morreel et al., 2010a; Morreel et al., 2010b) providing the second product ions at m/z 353, 341 and 338. A decarboxylation was evident from the peak at m/z 327. Type II-associated peaks were observed at m/z 235 and 191 (Morreel et al., 2010a) resulting from cleavage of the phenylcoumaran into its composing units. Therefore, this compound was characterized as dihydroconiferylalcohol 8–5 ferulic acid hexoside or glycosmisic acid hexoside. When matching to MassBank, no hit with the same chemical formula as compound **104** was obtained.

105. G(8–O–4)lariciresinol hex

The chemical formula of this ion was $C_{36}H_{45}O_{15}$ (m/z 717.27758). Upon CID, the type I fragmentations of a β-aryl ether were observed as first product ions at m/z 699 and 669 (Morreel et al., 2010a; Morreel et al., 2010b). Hexose loss resulted in the first product ion at m/z 555. The MS^3 spectrum of the latter ion showed also a β-aryl ether-associated type I peak at m/z 507. The peak at m/z 195 is due to a β-aryl ether-associated type II cleavage and indicated the presence of a guaiacylglycerol 8–O–4 ether moiety. A second product ion at m/z 329 results from the further formaldehyde loss from a lariciresinol moiety, which has been previously observed to be a dominating pathway upon CID of lariciresinol (Morreel et al., 2010a). Therefore, this compound was characterized as guaiacylglycerol 8–O–4 lariciresinol ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **105** was obtained.

106.G(8–O–4)SA malate

This ion had $C_{25}H_{27}O_{13}$ (m/z 535.14570) as chemical formula. The MS^2 spectrum was dominated by a base peak at m/z 419 due to malate loss. Further MS^3 fragmentation of this ion yielded a β-aryl ether type I pathway-associated peak at m/z 371 (Morreel et al., 2010a; 2010b). Furthermore, type II fragmentation of the β-aryl ether linkage rendered ions at m/z 223 and 195 representing moieties derived from sinapic acid and guaiacylglycerol. Therefore, this compound was characterized as guaiacylglycerol 8–O–4 sinapoyl malate ether. When matching to MassBank, no hit with the same chemical formula as compound **106** was obtained.

107. G(8–O–4)G(8–5)G hex

The chemical formula of the acetate adduct of this compound was $C_{38}H_{47}O_{17}$ (m/z 775.28366). This compound was connected to G(8–5)G hex **94** via a "G unit addition" conversion in the CSPP network. Both compound levels were highly correlated across biological replicates. Its MS^2 spectrum was dominated by the combined loss of the acetate and the hexoside yielding the first product ion at m/z 553. Further MS^3 fragmentation of this first product ion lead to second product ions at m/z 535, 523 and 505 due to the loss of water, formaldehyde and the combined loss of water and formaldehyde, respectively. These second product ions are characteristic type I fragmentations of β-aryl ethers (Morreel et al., 2010a). Thus, this compound was annotated as guaiacylglycerol 8–O–4 dehydrodiconiferyl alcohol ether hexoside.

108. G(8–O–4)G(8–O–4)SA hex

This ion's chemical formula was $C_{37}H_{45}O_{18}$ (m/z 777.26346). No MS^2 spectrum was obtained, but the compound was connected to G(8–O–4)FA hex **70** via a "S unit addition" conversion in the CSPP network. Both compound levels were correlated across biological replicates. However, as mentioned above (see compounds **95** and **103**), a sinapyl alcohol radical will not readily couple via an 8–O–4-linkage to a guaiacyl-derived phenolic function. Therefore, this compound should be annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 sinapic acid ether hexoside.

109. S(8–5)FA hex

The chemical formula of this ion was $C_{27}H_{31}O_{13}$ (m/z 563.17746). No MS^2 spectrum was obtained, but the compound was connected to G(8–5)FA hex **98** via a "methoxylation" conversion in the CSPP network. Both compound levels were correlated across biological replicates. Thus, this compound was annotated as dihydrosinapyl alcohol 8–5 ferulic acid hexoside.

110. G(8–O–4)S(8–8)G dihex

The ion had $C_{43}H_{55}O_{21}$ (m/z 907.32659) as chemical formula. Its MS^2 spectrum showed a peak at m/z 745 indicative for a hexose loss (-162 Da). No further MS^n information was obtained, but the compound was connected to pinoresinol dihex **79** via a "S unit addition" conversion. Both compound levels were very highly correlated across biological replicates. Because of the resilience of the sinapyl alcohol radical to form an 8–O–4-linkage to a guaiacyl phenolic function

(see compounds **95** and **103**), this compound should be annotated as guaiacylglycerol 8–O–4 medioresinol ester dihexoside. When matching to MassBank, no hit with the same chemical formula as compound **110** was obtained.

111. G(8–O–4)G(8–O–4)S hex

This ion had as chemical formula $C_{37}H_{47}O_{17}$ (m/z 763.28446). No MS^2 spectrum was obtained, but the compound was connected to G(*t*8–O–4)G hex **74** via a "S unit addition" conversion in the CSPP network. Both compound levels were highly correlated across biological replicates. Because a sinapyl alcohol radical does not readily form an 8–O–4-linkage to a guaiacyl phenolic function (see compounds **95** and **103**), this compound was annotated as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 sinapyl alcohol ether hexoside.

112. G(8–5)FA Glu

The chemical formula of this ion was $C_{25}H_{26}O_{10}N$ (m/z 500.15606). MS^2 dissociation lead to the m/z 371 ion due to a neutral loss of 129 Da, indicating the presence of a glutamic acid-derived moiety. Fragmentation of the first product ion at m/z 371 rendered the first product ions at m/z 353, 341, 327, 235 and 191 in a similar way as observed upon MS^3 fragmentation of the m/z 371 ion in the MS^2 spectrum of G(8–5)FA hex **104**. This suggested dehydroconiferyl alcohol 8–5 feruloyl glutamic acid as structure for compound **112**. The remaining first product ions provided further evidence. The phenylcoumaran-associated type II cleavage (Morreel et al., 2010a) produced the first product ion at m/z 364 representing a moiety derived from the feruloyl glutamic acid unit of compound **112**. Interestingly, as has been shown earlier for dicarboxylic acids (Kanawati and Schmitt-Kopplin, 2010), a combined loss of water and carbondioxide from the glutamate moiety yielding the first product ion at m/z 438 (-62 Da) was observed. Therefore, this compound was characterized as dehydroconiferyl alcohol 8–5 feruloyl glutamic acid. When matching to MassBank, no hit with the same chemical formula as compound **112** was obtained.

113. pinoresinol hex

The chemical formula of this ion was $C_{26}H_{31}O_{11}$ (m/z 519.18705). The MS^2 spectrum was dominated by the ion at m/z 357 due to hexose loss. MS^3 fragmentation of this base peak yielded a spectrum identical to the MS^2 spectrum observed for pinoresinol (Morreel et al., 2010a). Therefore, this compound is G(8–8)G hexoside or pinoresinol hexoside. When matching to MassBank, no hit with the same chemical formula as compound **113** was obtained.

114. G(8–O–4)G(8–O–4)lariciresinol hex

This ion had $C_{46}H_{57}O_{19}$ (m/z 913.35306) as chemical formula. The MS^2 base peak at m/z 751 arose from hexose loss. A water loss (-18 Da) and a combined water/formaldehyde loss (-48 Da), i.e., the type I fragmentations of a β-aryl ether (Morreel et al., 2010a; Morreel et al., 2010b), occurred both from the precursor ion (yielding the first product ions at m/z 895 and 865) as well as from the MS^2 base peak (leading to the first product ions at m/z 733 and 703). The first product ion at m/z 555 is formed by expelling a G unit from the MS^2 ion at m/z 751; a type II fragmentation known to occur in β-aryl ethers. The further structural characterization was solely based on the CSPP network in which this compound was linked to G(8–O–4)lariciresinol hex **105** via a "G unit addition" conversion. Both compound levels were very highly correlated across biological replicates. Compound **114** was characterized as guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 lariciresinol ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **114** was obtained.

115. G(8–O–4)SA malate

The ion of this compound had $C_{25}H_{27}O_{13}$ (m/z 535.14509) as chemical formula. No MS^2 spectrum was obtained, but the compound was connected to *trans*-sinapoyl malate **58** via a "G unit addition" conversion in the CSPP network. Both compound levels were highly correlated across biological replicates. Therefore, this compound was annotated as guaiacylglycerol 8–O–4 sinapoyl malate ester, i.e., an isomer of compound **106**.

116. G(8–5)FA hex

The chemical formula of this ion was identical and its MS^2 spectrum highly similar to those of G(8–5)FA hex **104**. Therefore, this compound was characterized as another dihydroconiferylalcohol 8–5 ferulic acid hexoside or glycosmisic acid hexoside isomer. When matching to MassBank, no hit with the same chemical formula as compound **116** was obtained.

117. G(8–O–4)G(8–5)FA hex

This ion had $C_{36}H_{41}O_{16}$ (m/z 729.24154) as chemical formula. It's MS^2 spectrum was dominated by the peak at m/z 567 resulting from hexose loss. The MS^3 spectrum of this MS^2 base peak showed ions at m/z 549 and 519 resulting from β-aryl ether characteristic type I cleavages leading to water (-18 Da) and a combined water/formaldehyde (-48 Da) loss (Morreel et al., 2010a; Morreel et al., 2010b). The type II fragmentations yielded ions at m/z 195 and 371

representing a guaiacylglycerol moiety and a dimeric moiety. The latter first product ion lost water and formaldehyde rendering the peaks at m/z 353 and 341 and can, thus, be pinpointed as a phenylcoumaran (Morreel et al., 2010a). Therefore, this compound was characterized as guaiacylglycerol 8–O–4 glycosmisic acid ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **117** was obtained.

118. G(8–5)FA Glu

The ion had the same chemical formula and almost identical MS^n spectra as G(8–5)FA Glu **112** and is, thus, an isomer of dehydroconiferyl alcohol 8–5 feruloyl glutamic acid. When matching to MassBank, no hit with the same chemical formula as compound **118** was obtained.

119. G(8–O–4)G(8–O–4)FA

The chemical formula and the MS^2 spectrum were the same as obtained for the in-source fragment ion of G(8–O–4)G(8–O–4)FA hex **102**. Therefore, this compound is an isomer of guaiacylglycerol 8–O–4 guaiacylglycerol ether 8–O–4 ferulic acid ether. When matching to MassBank, no hit with the same chemical formula as compound **119** was obtained.

120. G(8–8)S hex

The ion had $C_{27}H_{33}O_{12}$ (m/z 549.19770) as chemical formula. The base peak in its MS^2 spectrum was the ion at m/z 387 resulting from a hexose loss. When this ion was subjected to MS^3 fragmentation, the second product ions at m/z 372 and 341 arising from methyl radical (-15 Da) and formic acid (-46 Da) losses were those expected from the type I fragmentation of resinol lignans (Morreel et al., 2010a; see also references mentioned for pinoresinol dihex **79**). The type II cleavages of the resinol structure itself provided the ions at m/z 181, 166, 151 and 136, indicating that this compound is medioresinol hexoside. When matching to MassBank, no hit with the same chemical formula as compound **120** was obtained.

121. G(8–O–4)G(8–O–4)S hex

The chemical formula of this ion was $C_{37}H_{47}O_{17}$ (m/z 763.28385). No MS^2 spectrum was obtained, but this compound was connected via a "S unit addition" conversion to G(*t*8–O–4)G hex **74** in the CSPP network. Moreover, both compound levels were very highly correlated across biological replicates. Because 8–O–4-coupling of a sinapyl alcohol radical to the phenolic function of a G unit is not favored (Syrjänen and Brunow, 1998), only guaiacylglycerol 8–O–4

guaiacylglycerol ether 8–O–4 sinapyl alcohol ether hexoside is possible as annotation for this (neo)lignan/oligolignol.

122. G(8–O–4)pinoresinol hex

The ion's chemical formula was $C_{36}H_{43}O_{15}$ (m/z 715.26152). The compound was connected to the acetate adduct of G(8–5)G hex **94** via the spurious "vanillyl alcohol condensation" conversion in the CSPP network. The addition of vanillyl alcohol does not occur in *Arabidopsis* to our knowledge, but the levels of both peaks were highly correlated across biological replicates. In addition, compound **122** was also connected via a "G unit addition" to the deprotonated form of G(8–5)G hex **94**. Its MS^2 spectrum clearly indicated that it was a hexoside (first product ion at m/z 553, -162 Da loss). The most straightforward annotation for this compound was G(8–O–4)G(8–5)G hex, yet the MS^3 spectrum of this compound showed the characteristic type I fragmentations for a β-aryl ether (-18, -30 and -48 Da losses yielding the second product ions at m/z 535, 523 and 505; Morreel et al., 2010a). Furthermore, the β-aryl ether-specific type II fragmentations yielded the second product ions at m/z 357, 343 and 327 which are typical for a resinol moiety (Morreel et al., 2010b). Therefore, G(8–O–4)pinoresinol hex was taken as annotation for compound **122**.

123. G(8–O–4)S(8–5)FA hex

The chemical formula of this ion was $C_{37}H_{43}O_{17}$ (m/z 759.25082). The base peak at m/z 597 in its MS^2 spectrum was derived from a hexose loss. Further MS^3 fragmentation of this first product ion showed the typical type I cleavages of a β-aryl ether (m/z 579 and 549) (Morreel et al., 2010a; see structural elucidation of e.g. compounds **70**, **74**, **76**, **85**), but also a decarboxylation was evident from the ion at m/z 553. The characteristic type II cleavages of a β-aryl ether yielded the ions at m/z 195 and m/z 401 representing a guaiacylglycerol moiety and the remaining dimeric moiety. Further dissociation of m/z 401 lead to the second product ions at m/z 383 and 371 due to water and formaldehyde losses and correspond with the type I cleavages of a phenylcoumaran. As this dimeric moiety (m/z 401) underwent also a decarboxylation (m/z 357), compound **123** was characterized as guaiacylglycerol 8–O–4 dehydrosinapyl alcohol ether 8–5 ferulic acid hexoside. When matching to MassBank, no hit with the same chemical formula as compound **123** was obtained.

124. G(8–5)FA malate

This ion had $C_{24}H_{23}O_{11}$ (m/z 487.12414) as chemical formula. The MS^2 spectrum showed a base peak at m/z 371 due to malate loss (-116 Da). MS^3 fragmentation of the MS^2 base peak yielded a spectrum identical to the MS^3 spectrum described for G(8–5)FA hex **104**. Therefore, this compound was characterized as dihydroconiferylalcohol 8–5 feruloyl malate or glycosmisoyl malate. When matching to MassBank, no hit with the same chemical formula as compound **124**was obtained.

125. G(8–O–4)G(8–5)FA malate

The chemical formula of this ion was $C_{34}H_{35}O_{15}$ (m/z 683.19815). Its MS^2 spectrum was dominated by malate loss (-116 Da) leading to the peak at m/z 567. Upon MS^3 dissociation, the β-aryl ether-characteristic type I losses of 18 and 48 Da were observed at m/z 549 and 519 (Morreel et al., 2010a; see structural elucidation of e.g. compounds **70**, **74**, **76**, **85**). A β-aryl ether-associated type II cleavage explained the origin of the ion at m/z 371. Other second product ions at m/z 353 and 341 corresponded with the phenylcoumaran-associated type I losses occurring from the dimeric moiety represented by the m/z 371 ion. Therefore, this compound was characterized as guaiacylglycerol 8–O–4 glycosmisoyl malate ether. When matching to MassBank, no hit with the same chemical formula as compound **125** was obtained.

126. G(8–O–4)S(8–8)G hex

This ion had $C_{37}H_{45}O_{16}$ (m/z 745.27271) as chemical formula. No MS^2 spectrum was recorded, but this compound was connected to G(8–8)G hex or pinoresinol hex **113** via a "S unit addition" conversion in the CSPP network. The levels of both compounds were correlated across biological replicates. Although the most logical structure would be S(8–O–4)G(8–8)G hex, the 8–O–4-radical radical coupling of sinapyl alcohol to a guaiacyl phenolic endgroup of an oligolignol/(neo)lignan is not favored (Syrjänen and Brunow, 1998). Therefore, this compound was annotated as guaiacylglycerol 8–O–4 medioresinol ether hexoside.

127. G(8–O–4)S(8–8)G hex

This ion had $C_{37}H_{45}O_{16}$ (m/z 745.27233) as chemical formula. The MS^2 base peak was observed at m/z 583 and resulted from hexose loss. The MS^3 spectrum of this first product ion was showing type I losses (at m/z 565 and 535) and type II cleavages (at m/z 387 and 195) typical for a β-aryl ether (Morreel et al., 2010a; see structural elucidation of e.g. compounds **70**, **74**, **76**, **85**). The quartet of peaks at m/z 387, 373, 357 and 343 are typical for a resinol-containing oligolignol

(Morreel et al., 2010b) and, thus, this compound was characterized as guaiacylglycerol 8–O–4 medioresinol ether hexoside. When matching to MassBank, no hit with the same chemical formula as compound **127** was obtained.

128. S(8–5)FA hex

The chemical formula of this ion was $C_{27}H_{31}O_{13}$ (m/z 563.17473). No MS^2 spectrum was recorded, but this compound was connected to G(8–5)FA hex **104** via a "methoxylation" conversion in the CSPP network. The levels of both compounds were highly correlated across biological replicates and this compound was annotated as dihydrosinapyl alcohol 8–5 ferulic acid hexoside.

129. G(8–O–4)S(8–5)FA malate

This ion had $C_{35}H_{37}O_{16}$ (m/z 713.20870) as chemical formula. It's MS^n spectra were similar to those of G(8–O–4)G(8–5)FA malate **125**, but the m/z values of all MS^n peaks were increased with 30 amu. Therefore, this compound was characterized as guaiacylglycerol 8–O–4 dihydrosinapyl alcohol 8–5 feruloyl malate ether. When matching to MassBank, no hit with the same chemical formula as compound **129** was obtained.

130. G(8–O–4)G(8–5)FA

The chemical formula of this ion was $C_{30}H_{31}O_{11}$ (m/z 567.18674). No MS^2 spectrum was generated, but the compound was connected in the CSPP network to the ion of G(8–5)FA malate **124** representing the dilignol core structure resulting from an in-source malate loss. Both compounds were connected via a "G unit addition" conversion and their levels were correlated across biological replicates. This compound was annotated as guaiacylglycerol 8–O–4 glycosmisic acid ether.

131. S(8–5)FA hex

The ion had $C_{27}H_{31}O_{13}$ (m/z 463.17432) as chemical formula. No MS^2 spectrum was recorded, but this compound was connected to G(8–5)FA hex **104** via a "methoxylation" conversion in the CSPP network. The levels of both compounds were highly correlated across biological replicates and this compound was annotated as dihydrosinapyl alcohol 8–5 ferulic acid hexoside.

Indolics

132. 6-hydroxyindole-3-carboxylate dihex

This ion had $C_{21}H_{26}O_{13}N$ (m/z 500.14021) as chemical formula. Its MS^2 spectrum had a base peak at m/z 338 due to hexose loss. MS^3 fragmentation of the latter ion yielded a peak at m/z 176 due to a second hexose loss. The second product ion at m/z 132 arose due to a decarboxylation occurring from the m/z 176 ion and indicated that the aglycone was hydroxyindole-3-carboxylate. In the MS^2 spectrum, abundant peaks were present due to hexose cross-ring cleavages (ions at m/z 440, 410 and 380) reminding of an ester-linked hexose. However, also in the MS^3 spectrum, although much less abundant, these hexose cross-ring cleavages were evident (ions at m/z 278, 248 and 218) despite that no second carboxyl acid function was available. This indicates that one of the two hexoses is esterified whereas the other one is present as a phenolic hexoside. Therefore, this compound is 6-hydroxyindole-3-oyl hexose hexoside. The position of the hydroxyl group was derived from its connection to 6-hydroxyindole-3-carboxylate hex **133** in the CSPP network. When matching to MassBank, no hit with the same chemical formula as compound **132** was obtained. Using MetFrag, 8 biological compounds having the same chemical formula were retrieved from the Pubchem database. None of the returned hits was an indolic compound.

133. 6-hydroxyindole-3-carboxylate hex

The ion's chemical formula was $C_{15}H_{16}O_8N$ (m/z 338.08829). The MS^2 spectrum showed a base peak at m/z 176 due to hexose loss (-162 Da). Because no cross-ring cleavages were observed, this hexose is attached as a hexoside. The MS^3 spectrum of the m/z 176 first product ion rendered a second product ion at m/z 132 due to a decarboxylation. This MS^3 spectrum was similar to that of indole-3-carboxylate hex **135** except for the 16 amu shift to higher m/z values of first and second product ions. This compound was more precisely characterized as 6-hydroxyindole-3-carboxylate hexoside because this compound has been previously observed in Arabidopsis leaf extracts. When matching to MassBank, no hit with the same chemical formula as compound **133** was obtained. Via MetFrag, 54 structural isomers were returned from the PubChem database. After *in silico* fragmentation, two methoxyindolic compounds were found among the five best hits.

134. indole-3-carboxylate dihex

The chemical formula of this ion was $C_{21}H_{26}O_{12}N$ (m/z 484.14531). No MS^2 spectrum was recorded, but this compound was connected to indole-3-carboxylate hex **135** via a "hexosylation" conversion in the CSPP network. The levels of both compounds were highly

correlated across biological replicates and this compound was annotated as indole-3-carboxylate dihexoside.

135. indole-3-carboxylate hex

This ion had $C_{15}H_{16}O_7N$ (m/z 322.09348) as chemical formula. The major peak (m/z 160) in its MS^2 spectrum originates from hexose loss (-162 Da). Additional first product ions at m/z 262, 232 and 202, corresponding with losses of 60, 90 and 120 Da, respectively, result from hexose cross-ring cleavages and occur whenever the reducing end of the hexose is free (Carroll et al., 1995) or when the hexose is linked via an ester bond (Vanholme et al., 2010; see also explanation MS^n spectra of protocatechoyl Glc **44**). Evidence that the hexose is linked in an ester bond arises as well from the observation of a decarboxylation occurring upon MS^3 fragmentation of the aglycone (second product ion at m/z 116) which does not occur from the hexosylated compound. In addition, an ester bond might as well fragment via formation of a ketene neutral (Debrauwer et al., 1992; see other references in the explanation for protocatechoyl Glc **44**). This yields the hexose-associated second product ion at m/z 179. Based on the chemical formula of the aglycone, a nitrogen-containing aromatic molecule is expected. As the ynolate ion is not observed in the MS^2 spectrum (see explanation for protocatechoyl Glc **44**), this suggests that the ester function is directly connected to the aromatic system. Therefore, this compound was characterized as indole-3-oyl hexose and the aglycone structure was confirmed by MS^n analysis of an indole-3-oyl standard. When matching to MassBank, no hit with the same chemical formula as compound **135** was obtained. With MetFrag, 157 structural isomers were downloaded from the PubChem database. Following *in silico* fragmentation, our proposed structure belonged to the four best hits. All of the product ions could be explained by the software.

136. 6-hydroxyindole-3-carboxylate dihex sinA

The chemical formula of this ion was $C_{32}H_{36}O_{17}N$ (m/z 706.19914). The MS^2 base peak at m/z 544 resulted from hexose loss (-162 Da). An additional loss of the sinapic acid moiety as a ketene (-206 Da) yielded the first product ion at m/z 338. Ester cleavage of the precursor ion produced the complementary peaks at m/z 223, representing sinapate, and 482 originating from sinapic acid loss. The ion at m/z 367 corresponded with the sinapoyl hexose moiety. The MS^3 spectrum of the first product ion at m/z 544 rendered the complementary ions at m/z 367 and 176 representing sinapoyl hexose and 6-hydroxyindole-3-carboxylate. The presence of 6-hydroxyindole-3-carboxylate was further supported in the MS^3 fragmentation of the first product

ion at m/z 338 which rendered ions at m/z 176 and 132 due to a hexose loss and a subsequent decarboxylation from the 6-hydroxyindole-3-carboxylate aglycone. Furthermore, the presence of a sinapoyl hexose moiety was proven by MS^4 dissociation of the second product ion at m/z 367, yielding the third product ion at m/z 223. Therefore, this compound was characterized as 6-hydroxyindole-3-oyl sinapoyl dihexoside. When matching to MassBank, no hit with the same chemical formula as compound **136** was obtained. Using MetFrag, two structural isomers were returned from the PubChem database, but none was retained as a valuable candidate after *in silico* fragmentation.

Indolic glucosinolate catabolites

137. 5'-Glc-dihydroascorbigen

The ion had $C_{21}H_{26}O_{11}N$ (m/z 468.15056) as chemical formula. Its MS^2 spectrum was not similar to any of the compounds described up to now, yet a clear connection with the indolics was evident from the CSPP network. The compound was linked with indole-3-carboxylate dihex **134** via an "oxygenation" conversion. Both compound levels were highly correlated across biological replicates. Searching indolics with the same chemical formula in the CAS database yielded one candidate, i.e., 5'-*O*-β-D-glucosyl dihydroascorbigen, of which the structure is expected to follow similar dissociation channels in the gas phase as those represented by the MS^2 spectrum of our unknown compound **137**. An outline of the various gas phase fragmentation reactions is shown in Supplemental Figure 8B.

Although both charge-driven and charge-remote pathways can occur during the gas phase fragmentations of negative ions, the former type of pathways will prevail whenever possible (Thevis et al., 2003). Therefore, in the explanation of the MS^n dissociations of compound **137** (Supplemental Figure 8B), charge-remote pathways were considered whenever no charge-driven pathway could be deduced. Arguably, the most acidic site is the $C_{1'}$-position which is allylic to the indole moiety. Following deprotonation, a proton transfer from the 3´-OH group via a five-center intermediate (Supplemental Figure 8B, pathway **a**) or from the 3´-OH group via an eight-center intermediate (Supplemental Figure 8B, pathway **c**) can be envisaged. Alternatively, an E1cb-like reaction might open the lactone ring (Supplemental Figure 8B, pathway **b**). Subsequent fragmentation along pathway **a** also opens up the lactone ring. Dependent whether the negative charge then ends up at the $C_{2'}$-position via cleavage of the 3´–4´-linkage or leads to

the cleavage of the 2´–3´-linkage, a 2´-hydroxyindolylpropanyl aldehyde (m/z 188) or 2´-hydroxyindolylpropanoic acid (m/z 204) anion is formed. The latter structure resembles that of Trp leading to a similar fragmentation behavior (Supplemental Figure 8B, table insert). Interestingly, because ammonia is a less electronegative species than water, the peak at m/z 186 (ammonia loss) in the MS^2 spectrum of Trp is much smaller than the same peak in the MS^3 spectrum of compound **137** (due to water loss in the latter case). Furthermore, the peak at m/z 158 is due to decarboxylation in the case of Trp fragmentation but formic acid loss upon dissociation of the 2'-hydroxyindolylpropanoyl anion. Formic acid loss is characteristic for α-hydroxy acids (Bandu et al., 2006).

Pathway **b** explains the ions at m/z 440, 406, 226, 208 and 179. The E1cb reaction opens the lactone ring via carbon monoxide loss (m/z 440) followed by cleavage of the 3´–4´-bond, hence, providing a second pathway leading to the ion at m/z 188. Decarbonylation from the ester group in an α-hydroxy ester moiety has been previously suggested to occur in negative ionization tandem-in-space MS/MS (Mancel et al., 2004). However, instead of decarbonylation, the E1cb reaction might as well lead to decarboxylation of the lactone ring. Concomitantly with the decarboxylation, desaturation of the 3´–4´-bond or the 4´–5´-bond will occur. Whereas the former desaturation leads to water loss providing the ion at m/z 406, desaturation of the 4´–5´-bond will result in the elimination of a glucose anion (m/z 179). However, as the glucose anion is transiently withhold in an ion-neutral complex (Bowie, 1990), complex dissociation might be preceded by a proton transfer from the 3´-position to the glucose anion and the subsequent elimination of water from the indolyl-containing anion. This yields the highly conjugated anion at m/z 226. Expelling another water molecule renders then the peak at m/z 208.

Following initial proton transfer from the 6´-OH group (pathway c), the resulting alkoxide anion could attack the C6´-position in an internal S_N2 reaction, cleaving the 5´–6´-bond and, thus, providing the indole-bearing anion at m/z 246 and a neutral epoxide. A further decarbonylation from the lactone ring delivers the ion at m/z 218. Internal S_N2 reactions leading to epoxide formation have been previously described to occur in the gas phase (Binkley et al., 1996; Mancel et al., 2004). Finally, a charge-remote glycosidic bond cleavage explains the loss of 162 Da yielding the peak at m/z 306 (Carroll et al., 1995), whereas a subsequent water loss produces the peak at m/z 288 (Mulroney et al., 1999). Because the most abundant product ions agreed with the proposed structure, this compound was characterized as 5'-*O*-β-D-glucosyl

dihydroascorbigen. When matching to MassBank, no hit with the same chemical formula as compound **137** was obtained. Via MetFrag, 9 structural isomers were returned from the PubChem database. After *in silico* fragmentation, one indolic compound was found among the retained hits, yet its aglycone was not dihydroascorbigen. However, a high resolution tandem-in-space MS/MS spectrum has been obtained previously (Montaut and Bleeker, 2010) and was highly similar to the tandem-in-time MS^2 spectrum obtained in our study (S. Montaut, personal communication).

138. 5´-Glc-dihydroneoascorbigen

This ion had $C_{22}H_{28}O_{12}N$ (m/z 498.16129) as chemical formula. Its MS^n spectra were similar to those of 5´-Glc-dihydroascorbigen **137** except that most of the product ions appeared at m/z values that were 30 Th higher. Therefore, this compound was characterized as 5´-*O*-β-D-glucosyl dihydroneoascorbigen. When matching to MassBank, no hit with the same chemical formula as compound **138** was obtained. Via MetFrag, 12 structural isomers were returned from the PubChem database. After *in silico* fragmentation, three indolic compounds were found among the retained hits, yet their aglycones were not similar to dihydroneoascorbigen.

139. hydroxy-dihydroascorbigen hex

The chemical formula of this ion was $C_{21}H_{26}O_{12}N$ (m/z 484.14515). Its MS^2 spectrum was characterized by the loss of hexose (-162 Da) yielding the base peak at m/z 322. MS^3 fragmentation of the latter first product ion yielded second product ions at m/z 188 and 204 that were reminiscent of the 2´-hydroxyindolylpropanyl aldehyde or 2´-hydroxyindolylpropanoic acid anions that are produced upon MS^2 fragmentation of 5´-Glc-dihydroascorbigen **137**. Therefore, compound **139** was structurally characterized as hydroxyl-dihydroascorbigen hex.

Apocarotenoids

140. corchoionoside C

This ion had $C_{19}H_{29}O_8$ (m/z 385.18684) as chemical formula. The MS^2 spectrum of compound **140** showed a hexose loss leading to the ion at m/z 223. Possible biological candidate compound classes containing representatives with the chemical formula of the aglycone in the CAS database, were the jasmonates and the apocarotenoids. The similarity between the MS^2 spectrum of abscissic acid (an apocarotenoid) and the MS^3 spectrum of the aglycone of compound **140** suggested the latter to be an apocarotenoid. In both cases, the major product ion was observed at

m/z 153 corresponding with the apocarotenoid ring structure. Blumenol A glucoside or corchoionoside C has already been observed in the *Brassicaceae* (Cutillo et al., 2005) and had the correct chemical formula. Indeed, all MS^n fragmentations could be explained based on the anion of this candidate structure. In case of the base peak, the presumed charge-driven mechanism starting from deprotonation of the 6-position is given in Supplemental Figure 8C (pathway **a**). A proton transfer from the 3´-OH function, succeeded by ethyne elimination, explains the m/z 153 ion production. This reaction is driven by the neutral loss of two highly electronegative species, i.e., ethyne and acetaldehyde, and the resonance-stabilized product ion. Such an ion might readily loose a methyl radical as both 5-linked methyl groups are allylic to the conjugated double bonds. Indeed, this is verified by the observation of a base peak at m/z 138 in the MS^4 spectrum of the m/z 153 second product ion. An alternative fragmentation pathway occurs when charge delocalization leads to the elimination of the 3´-OH function (pathway **b**, Supplemental Figure 8C). This results in a dehydration when an allylic proton is abstracted from the 3´-position via a hydroxide ion-neutral complex. Therefore, this compound was characterized as corchoionoside C. When matching to MassBank, no hit with the same chemical formula as compound **140** was obtained. Via MetFrag, 68 structural isomers were returned from the PubChem database. After *in silico* fragmentation, the four best hits in which all four first product ions (m/z 153, 161, 205 and 223) could be explained, were all apocarotenoids. Among these four, a stereomer of corchoionoside C (or 6S,9R blumenol A hexoside) was included, i.e., 6S,9S-roseoside.

141. blumenol A malonylhex

The chemical formula of this ion was $C_{22}H_{31}O_{11}$ (m/z 471.18690). No MS^2 spectrum was obtained, but in-source decarboxylation yielded the ion at m/z 427.19708 ($C_{21}H_{31}O_9$, Δppm = -0.65) for which a MS^2 spectrum was recorded. The first product ion at m/z 385 due to 42 Da loss together with the in-source decarboxylation is typical for malonate moieties (Pollier et al., 2011). The base peak at m/z 367 resulted from a further water loss and, when subjected to MS^3 fragmentation, rendered the characteristic fragment ions for blumenol A. Therefore, this compound was characterized as blumenol A malonyl hexoside.

142. blumenol A acetylmalonylhex

The ion had $C_{24}H_{33}O_{12}$ (m/z 513.19789) as chemical formula. No MS^2 spectrum was recorded, but in-source fragmentation produced the ion at m/z 469.20771 ($C_{23}H_{33}O_{10}$, Δppm = -0.45) that

dissociated into the first product ions at m/z 427 and 385 due to two subsequent 42 Da losses. Together with the in-source decarboxylation, these ions refer to the presence of a malonyl and an acetyl moiety. Water losses from both first product ions rendered the peaks at m/z 409 and 367. MS^3 fragmentation of the m/z 409 ion showed a very similar spectrum as the MS^3 spectrum of the m/z 223 ion of corchoionoside C **140**. Therefore, this compound was characterized as blumenol A acetyl malonyl hexoside.

143. blumenol A acetylmalonylhex

The ion had $C_{24}H_{33}O_{12}$ (m/z 513.19761) as chemical formula. No MS^2 spectrum was recorded, but the compound was connected to blumenol A malonylhex **141** via an "acetylation" conversion in the CSPP network. The levels of both compounds were highly correlated across biological replicates. Therefore, this compound was annotated as another isomer of blumenol A acetyl malonyl hexoside.

Others

144. glutathione

This ion had $C_{10}H_{16}O_6N_3S$ (m/z 306.07695) as chemical formula. Matching its MS^2 spectra with those in the MassBank database, yielded an almost perfect match to the spectrum of glutathione.

145. butyl acetylhex

The chemical formula of the ion was $C_{12}H_{21}O_7$ (m/z 277.12977). No MS^2 spectrum was recorded, but an in-source fragment at m/z 235.11946 ($C_{10}H_{19}O_6$, Δppm = 3.18) indicated the loss of an acetyl group. A MS^2 spectrum for the latter fragment ion was obtained that was very similar to that of a hexose (March and Stadey, 2005), except for the presence of the first product ion at m/z 191. The base peak at m/z 161 indicated that the charge mainly remained with the hexose moiety. Therefore, this in-source fragment is a hexose to which an apolar butyl moiety is attached. Taking into account that an acetyl group dissociated in-source, this compound is acetyl-butyl hexoside. With MetFrag, 203 structural isomers of the in-source fragment were downloaded from the PubChem database. After *in silico* fragmentation, several butyl glucosides were found among the 20 best hits. For all of them, all 10 product ions were predicted.

Interpretation of CSPPs Associated with Moderate to High Correlation Coefficients

In this study, CSPPs in which the abundances of the "substrate" and "product" peaks were moderately or highly correlated, could be classified into four groups. The first group contained CSPPs representing true enzymatic reactions (group 1). Examples were the oxygenations in aliphatic glucosinolate biosynthesis (Figures 4B and 5) and the glycosylations in flavonoid metabolism (Figure 5 and Supplemental Figure 5A).

CSPPs from group 2 represented organic reactions that occur during the biosynthesis of the compound class, but more upstream in the biosynthetic pathway than suggested by the CSPP. For example, the sequence of methylations and oxygenations in the sub-network of the aliphatic glucosinolates did not reflect the biochemical reaction sequence. Methylene insertions (represented by methylation CSPPs) take place earlier in glucosinolate biosynthesis than the oxygenations (Figures 4A and 5). This side-chain elongation occurs before the glucosinolate is synthesized, i.e., at the level of the amino acid precursor of the glucosinolate. However, using CSPPs to characterize unknown compounds still yielded valid structures. In flavonoid metabolism (Figure 5 and Supplemental Figure 5A), kaempherol glycosides and their quercetin analogues, e.g., kaempherol 3-*O*-rhamnosyl-7-*O*-rhamnoside **42**, kaempherol 3-*O*-glucosyl-7-*O*-rhamnoside **37** and kaempherol 3-*O*-glucosyl(1→2)rhamnosyl-7-*O*-rhamnoside **31**, and their respective quercetin analogues **35**, **32** and **29** (see Supplemental Data Set 1 Online, Supplemental Figure 4), are associated with CSPPs representing oxygenations occurring more upstream in the pathway, i.e., those performed by flavonoid 3´-hydroxylase. Again, although the position of the reaction in the biosynthetic pathway cannot be unambiguously inferred from the CSPP network, the network still yields valid structural information about unknown metabolites.

Confounding the CSPP-based structural elucidation were those CSPPs in which the "substrate" was connected with a structurally similar isomer of the expected "product" (group 3). This can readily happen in, e.g., phenylpropanoid/(neo)lignan metabolism. In phenylpropanoid metabolism (Figure 5), the side-chain of feruloyl-CoA might be reduced or oxidized, leading to coniferyl alcohol (a hydroxycinnamyl alcohol) or ferulic acid (a hydroxycinnamic acid), respectively. By the addition of a methoxy group to the benzene ring, coniferyl alcohol and ferulic acid can be further converted to sinapyl alcohol (a hydroxycinnamyl alcohol) and sinapic acid (a hydroxycinnamic acid). During radical cross-coupling reactions, coniferyl and sinapyl alcohol provide the guaiacyl (G) and syringyl (S) units in lignins and (neo)lignans, whereas units derived from ferulic and sinapic acids are denoted as FA and SA. Noticeably, units derived from

both hydroxycinnamic acids (FA and SA) as well as from both hydroxycinnamyl alcohols (G and S) differ only by a methoxyl group. This has to be taken into account when characterizing the structures of their coupling products. For example, the guaiacyl (Gun) CSPP between sinapoyl glucose **55**, and **76** (Supplemental Figure 5B), suggested G(8–O–4)SA hexose, which is the β-aryl ether crossed dimer derived from a G monomer (coniferyl alcohol) coupling to sinapoyl hexose, as the structure for the latter compound. However, when the MS^2 spectrum of compound **76** was recorded, it was deemed to be S(8–O–4)FA hexose, which is a structural isomer – a β-aryl ether crossed dimer derived from an S monomer (sinapyl alcohol) coupling to feruloyl hexose. In other words, a methoxy substituent had to be swapped to obtain an agreement between the CSPP-based and the MS^2-based structural elucidation.

In a fourth group of CSPPs, often no biochemical support for the presumed conversion type was obtained based on the molecular structures as the structural moieties of the "substrate" were shuffled in the "product" structure. For example, kaempherol 3-arabinosyl-rhamnosyl-7-rhamnoside **36** and kaempherol 3-rhamnosyl-7-malonylglucoside **39** are both derived from the same aglycone precursor, yet follow different glycosylation paths. The sequence of glycosylation reactions for each resulted in a final mass difference corresponding with a glycerol moiety, i.e., the mass difference between the malonyl-glucosyl and arabinosyl-rhamnosyl moieties. Obviously, structural annotations based on the CSPP network in which "low-correlated" edges are removed, will be erroneous in this case, necessitating the inclusion of an MS^2 spectral similarity algorithm.

(Bio)chemical Validity of the "High CSPP" Group Conversions

Because of the higher fraction of true (bio)chemical CSPPs in the sub-network of the "high CSPP" group versus that of the network containing all CSPPs, the former sub-network was, as expected, more similar to the topology of a metabolic network (Figures 2C and 2D). Furthermore, the levels of the "substrate" and "product" of each CSPP of the "high CSPP" group were on average more highly correlated than those of the "low CSPP" group. Although reductions (Red, Table 1) and the addition of syringyl (S) units to oligolignols/(neo)lignans (Sun, Table 1) are true conversions in *Arabidopsis*, they belonged to the "low CSPP" group, yet the correlations found for their CSPPs were still higher than those for the other CSPPs of the "low CSPP" group. Furthermore, when selecting the node file for "reduction" peak pairs based solely

on mass differences and checking subsequently their retention time differences (Figure 3E, Supplemental Figure 3), the "substrate" and "product" peaks often eluted too closely to each other to be annotated as a CSPP and, hence, the number of CSPPs was underestimated. The classification of the "Sun" conversion as a "low CSPP" group conversion was a borderline case as it had an intermediate number of CSPPs (245, Table 1). Conversely, glycerol addition is not a well-known reaction in *Arabidopsis* metabolism, yet the conversion belonged to the "high CSPP number" conversions. However, MS^2 data of some *Arabidopsis* (neo)lignans/oligolignols showed neutral losses corresponding to the expected mass of a dissociated glycerol moiety (e.g., feruloyl glycerol **52**, Supplemental Data Set 1), indicating that this conversion was correctly classified as a "high CSPP number" conversion. Finally, the use of correlations as support for true metabolism-based CSPPs emanates also from the observation that the sub-networks of highly correlated peaks in the CSPP network corresponded with biochemically related compounds, such as those of the glucosinolate, flavonoid, sinapate and oligolignol/(neo)lignan metabolism.

Searching Compound Classes and New Enzymatic Reactions with the CSPP Algorithm

The information obtained from the CSPP algorithm not only aids structural characterization of compounds but provides a new metabolomics tool to analyze profiled data. Firstly, by using one or a few representatives of a compound class as "bait", the CSPP network strategy illustrates the ease of teasing out all compounds with similar structures from the "haystack" of chromatographic peaks at once. Comparing the MS^n ion trees of all similar compounds aids in interpretation of the gas-phase fragmentation pathways and, thus, the spectral interpretation. The power of this approach is illustrated with, e.g., the trisinapoylgentiobiose isomers **66** and **68** (Supplemental Data Set 1, Supplemental Figure 4) for which the structure would likely never have been proposed using solely the complex MS^2 spectrum. Perhaps most importantly, the CSPP-based annotation procedure allows (tentative) assignments of structures for unknowns that are only present in minute amounts that compromise their purification and, thus, their identification by NMR. Examples of such structurally characterized low-abundance compounds are G(8–O–4)SA malate **106** and G(8–O–4)FA malate hexoside **69** (Supplemental Data Set 1, Supplemental Figure 4). The metabolic class of some sub-networks could not be elucidated as

none of their members could be structurally annotated. Purification followed by NMR-based identification of the more abundant compounds would allow further characterization of these sub-networks of unknown compound classes.

Secondly, the use of CSPPs enabled the pinpointing of as yet unknown enzymatic reactions. For example, the levels of kaempherol 3-*O*-arabinosyl-rhamnosyl-7-*O*-rhamnoside **36** and kaempherol 3-*O*-arabinosyl-7-*O*-rhamnoside **38** were mutually strongly correlated (Pearson r^2 = 0.92; Supplemental Data Set 1), and the CSPP conversion between them suggests a rhamnosyltransferase-catalyzed reaction. Furthermore, the presence of a sinapic acid derivative of a kaempherol dirhamnoside (compound **43**) indicates that hydroxycinnamate derivatization within flavonoid metabolism is not restricted to the anthocyanins (Nakabayashi et al., 2009).

Overall, the CSPP algorithm allowed the structural annotation of 145 compounds in *Arabidopsis* leaves from a total of 229 profiled compounds. Consequently, 60% of all profiled compounds were annotated/characterized, a percentage that has never been obtained before in a metabolomics experiment. Searching the Scifinder database revealed that only 40 compounds had been previously detected in *Arabidopsis* leaves (Supplemental Data Set 1), although another 14 compounds had been described in other *Arabidopsis* tissues. Ten compounds were found before in one or more *Brassicaceae* species but were observed here for the first time in *Arabidopsis*. From the remainder of the annotated compounds, 20 have been described in other plant families and 61 structures were not found in the database at all. Compound classes that are already well-known in *Arabidopsis* leaves were the phenylpropanoids, flavonoids and glucosinolates (Figure 5). Nevertheless, the subclass of the hydroxy-(methylsulfinyl)-alkyl glucosinolates has not yet been described in *Arabidopsis* and only a few members have been found in the plant kingdom (Fahey et al., 2001). Because no hydroxy-(methylthio)-alkyl glucosinolates were detected, this strongly suggests that the hydroxy-(methylsulfinyl)-alkyl glucosinolates arise by hydroxylation of the methylsulfinylalkyl glucosinolate which is itself formed from the oxidation of the corresponding methylthioalkyl glucosinolate (Figure 4A). Thus, CSPP networks also provide information on biosynthetic routes. In addition to the indolic glucosinolates, their breakdown products were detected as well (Supplemental Data Set 1). These breakdown products are formed whenever indolic glucosinolates are released from the vacuole and encounter myrosinase. For the first time dihydroascorbigen-based glucosinolate catabolites were observed in *Arabidopsis*. Their existence was supported by the high MS^2 similarity of 5´-

Glc-dihydroascorbigen **137** with a previously recorded MS fragmentation spectrum for this compound (Montaut and Bleeker, 2010). Strikingly, a high number of (neo)lignans/oligolignols (Morreel et al., 2004; 2010a; 2010b) were present in the leaves. These structures displayed the various units and combinatorial linkages that are found in lignin, yet they were all adorned with hexose, malate and/or glutamate which classifies them as (neo)lignans. Some of these compounds have recently been described in flax stem tissues (Huis et al., 2012) and were also found in *Arabidopsis* stems (Vanholme et al., 2012). Finally, apocarotenoids implicated here were unknown to exist in *Arabidopsis* to date.

SUPPLEMENTAL REFERENCES

Al-Shehbaz, I.A., and Al-Shammary, K.I. (1987). Distribution and chemotaxonomic significance of glucosinolates in certain Middle-Eastern cruciferae. Biochem. Syst. Ecol. **15:** 559-569.

Attygalle, A., Ruzicka, J., Varughese, D., and Sayed, J. (2006). An unprecedented ortho effect in mass spectrometric fragmentation of even-electron negative ions from hydroxyphenyl carbaldehydes and ketones. Tetrahedron Lett. **47:** 4601-4603.

Bandu, M.L., Grubbs, T., Kater, M., and Desaire, H. (2006). Collision induced dissociation of alpha hydroxy acids: Evidence of an ion-neutral complex intermediate. Int. J. Mass Spectrom. **251:** 40-46.

Bandu, M.L., Watkins, K.R., Bretthauer, M.L., Moore, C.A., and Desaire, H. (2004). Prediction of MS/MS Data. 1. A focus on pharmaceuticals containing carboxylic acids. Anal. Chem. **76:** 1746-1753.

Bialecki, J.B., Ruzicka, J., Weisbecker, C.S., Haribal, M., and Attygalle, A.B. (2010). Collision-induced dissociation mass spectra of glucosinolate anions. J. Mass Spectrom. **45:** 272-283.

Binkley, R.W., Binkley, E.R., Duan, S.M., Tevesz, M.J.S., and Winnik, W. (1996). Negative-ion mass spectrometry of carbohydrates. A mechanistic study of the fragmentation reactions of dideoxy sugars. Journal of Carbohydrate Chemistry **15:** 879-895.

Born, M., Ingemann, S., and Nibbering, N.M.M. (1997). Formation and chemistry of radical anions in the gas phase. Mass Spectrom. Rev. **16:** 181-200.

Bowie, J.H. (1990). The fragmentations of even-electron organic negative ions. Mass Spectrom. Rev. **9:** 349-379.

Carroll, J.A., Willard, D., and Lebrilla, C.B. (1995). Energetics of cross-ring cleavages and their relevance to the linkage determination of oligosaccharides. Anal. Chim. Acta **307:** 431-447.

Cataldi, T.R.I., Lelario, F., Orlando, D., and Bufo, S.A. (2010). Collision-induced dissociation of the A+2 isotope ion facilitates glucosinolates structure elucidation by electrospray ionization-tandem mass spectrometry with a linear quadrupole ion trap. Anal. Chem. **82:** 5686-5696.

Chapple, C.C.S., Vogt, T., Ellis, B.E., and Somerville, C.R. (1992). An Arabidopsis mutant defective in the general phenylpropanoid pathway. Plant Cell **4:** 1413-1424.

Cheng, C., and Gross, M.L. (2000). Applications and mechanisms of charge-remote fragmentation. Mass Spectrom. Rev. **19:** 398-420.

Cutillo, F., Dellagreca, M., Previtera, L., and Zarrelli, A. (2005). C_{13} norisoprenoids from *Brassica fruticulosa*. Nat. Prod. Res. **19:** 99-103.

Cuyckens, F., and Claeys, M. (2004). Mass spectrometry in the structural analysis of flavonoids. J. Mass Spectrom. **39:** 1-15.

Dauwe, R., Morreel, K., Goeminne, G., Gielen, B., Rohde, A., Van Beeumen, J., Ralph, J., Boudet, A.-M., Kopka, J., Rochange, S.F., Halpin, C., Messens, E., and Boerjan, W. (2007). Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration. Plant J. **52:** 263-285.

Debrauwer, L., Paris, A., Rao, D., Fournier, F., and Tabet, J.-C. (1992). Mass spectrometric studies on 17β-estradiol-17-fatty acid esters: Evidence for the formation of anion-dipole intermediates. Organic Mass Spectrometry **27:** 709-719.

De Kok, L.J., and Graham, M. (1989). Levels of pigments, soluble proteins, amino acids and sulfhydryl compounds in foliar tissue of *Arabidopsis thaliana* during dark-induced and natural senescence. Plant Physiol. Biochem. **27:** 203-209.

DePuy, C.H. (2000). An introduction to the gas phase chemistry of anions. Int. J. Mass Spectrom. **200:** 79-96.

Deyama, T., Ikawa, T., Kitagawa, S., and Nishibe, S. (1986). The constituents of *Eucommia ulmoides* Oliv. IV. Isolation of a new sesquilignan glycoside and iridoids. Chem. Pharm. Bul. **34:** 4933-4938.

Deyama, T., Ikawa, T., and Nishibe, S. (1985). The constituents of *Eucommia ulmoides* Oliv. II. Isolation and structures of three new lignan glycosides. Chem. Pharm. Bull. **33:** 3651-3657.

Eichinger, P.C.H., Dua, S., and Bowie, J.H. (1994). A comparison of skeletal rearrangement reactions of even-electron anions in solution and in the gas phase. Int. J. Mass Spectrom. Ion Processes **133:** 1-12.

Eklund, P.C., Backman, M.J., Kronberg, L.Å., Smeds, A.I., and Sjöholm, R.E. (2008). Identification of lignans by liquid chromatography-electrospray ionization ion-trap mass spectrometry. J. Mass Spectrom. **43:** 97-107.

Fabre, N., Rustan, I., de Hoffmann, E., and Quetin-Leclercq, J. (2001). Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry. J. Am. Soc. Mass Spectrom. **12:** 707-715.

Fabre, N., Poinsot, V., Debrauwer, L., Vigor, C., Tulliez, J., Fourasté, I., and Moulis, C. (2007). Characterisation of glucosinolates using electrospray ion trap and electrospray quadrupole time-of-flight mass spectrometry. Phytochem. Anal. **18:** 306-319.

Fahey, J.W., Zalcmann, A.T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochem. **56:** 5-51.

Fernández-Arroyo, S., Barrajón-Catalán, E., Micol, V., Segura-Carretero, A., and Fernández-Gutiérrez, A. (2010). High-performance liquid chromatography with diode array detection coupled to electrospray time-of-flight and ion-trap mass spectrometry to identify phenolic compounds from a *Cistus ladanifer* aqueous extract. Phytochem. Anal. **21:** 307-313.

Ferreres, F., Sousa, C., Valentão, P., Seabra, R.M., Pereira, J.A., and Andrade, P.B. (2007). Tronchuda cabbage (*Brassica oleracea* L. var. *costata* DC) seeds: phytochemical characterization and antioxidant potential. Food Chem. **101:** 549-558.

Ferreres, F., Llorach, R., and Gil-Izquierdo, A. (2004). Characterization of the interglycosidic linkage in di-, tri-, tetra- and pentaglycosylated flavonoids and differentiation of

positional isomers by liquid chromatography/electrospray ionization tandem mass spectrometry. J. Mass Spectrom. **39:** 312-321.

Fournier, F., Perlat, M.-C., and Tabet, J.-C. (1995). Control of internal proton transfers on ion-dipole complexes from $[M-H]^-$ ions of diphenol esters. Rapid Commun. Mass Spectrom. **9:** 13-17.

Fournier, F., Remaud, B., Blasco, T., and Tabet, J.C. (1993). Ion-dipole complex formation from deprotonated phenol fatty acid esters evidenced by using gas-phase labeling combined with tandem mass spectrometry. J. Am. Soc. Mass Spectrom. **4:** 343-351.

Fraser, C.M., Thompson, M.G., Shirley, A.M., Ralph, J., Schoenherr, J.A., Sinlapadech, T., Hall, M.C., and Chapple, C. (2007). Related Arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. Plant Physiol. **144:** 1986-1999.

Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., Lu, F., Ralph, J., Mila, I., Barrière, Y., Lapierre, C., and Jouanin, L. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of *O*-methyltransferase impacts lignins and sinapoyl esters. Plant Mol. Biol. **51:** 973-989.

Gronert, S. (2001). Mass spectrometric studies of organic ion/molecule reactions. Chem. Rev. **101:** 329-360.

Gronert, S. (2005). Ouadrupole ion trap studies of fundamental organic reactions. Mass Spectrom. Rev. **24:** 100-120.

Grossert, J.S., Cook, M.C., and White, R.L. (2006). The influence of structural features on facile McLafferty-type, even-electron rearrangements in tandem mass spectra of carboxylate anions. Rapid Commun. Mass Spectrom. **20:** 1511-1516.

Guo, H., Liu, A.-H., Ye, M., Yang, M., and Guo, D.-A. (2007). Characterization of phenolic compounds in the fruits of *Forsythia suspensa* by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry. Rapid Commun. Mass Spectrom. **21:** 715-729.

Hagemeier, J., Schneider, B., Oldham, N.J., and Hahlbrock, K. (2001). Accumulation of soluble and wall-bound indolic metabolites in *Arabidopsis thaliana* leaves infected with virulent or avirulent *Pseudomonas syringae* pathovar tomato strains. Proc. Natl. Acad. Sci. U.S.A. **98:** 753-758.

Hanhineva, K., Rogachev, I., Aura, A.-M., Aharoni, A., Poutanen, K., and Mykkänen, H. (2012). Identification of novel lignans in the whole grain rye bran by non-targeted LC-MS metabolite profiling. Metabolomics **8:** 399-409.

Hansen, C.H., Wittstock, U., Olsen, C.E., Hick, A.J., Pickett, J.A., and Halkier, B.A. (2001). Cytochrome P450 CYP79F1 from Arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. J. Biol. Chem. **276:** 11078-11085.

Harrison, A.G. (1992). Chemical ionization mass spectrometry. (Boca Raton, Florida: CRC Press).

Haughn, G.W., Davin, L., Giblin, M., and Underhill, E.W. (1991). Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana*. Plant Physiol. **97:** 217-226.

Heinonen, M., Rantanen, A., Mielikäinen, T., Kokkonen, J., Kiuru, J., Ketola, R.A., and Rousu, J. (2008). FiD: a software for *ab initio* structural identification of product ions from tandem mass spectrometric data. Rapid Commun. Mass Spectrom. **22:** 3043-3052.

Ho, J.-C., Chen, C.-M., and Row, L.-C. (2003). Neolignans from the parasitic plants. Part 1. *Aeginetia indica*. J. Chin. Chem. Soc. **50:** 1271-1274.

Hou, S., Zhu, J., Ding, M., and Lv, G. (2008). Simultaneous determination of gibberellic acid, indole-3-acetic acid and abscisic acid in wheat extracts by solid-phase extraction and liquid chromatography-electrospray tandem mass spectrometry. Talanta **76:** 798-802.

Hughes, R.J., Croley, T.R., Metcalfe, C.D., and March, R.E. (2001). A tandem mass spectrometric study of selected characteristic flavonoids. Int. J. Mass Spectrom. **210-211:** 371-385.

Huis, R., Morreel, K., Fliniaux, O., Lucau-Danila, A., Fenart, S., Grec, S., Neutelings, G., Chabbert, B., Mesnard, F., Boerjan, W., and Hawkins, S. (2012). Natural hypolignification is associated with extensive oligolignol accumulation in flax stems. Plant Physiol. **158:** 1893-1915.

Hvattum, E., and Ekeberg, D. (2003). Study of the collision-induced radical cleavage of flavonoid glycosides using negative electrospray ionization tandem quadrupole mass spectrometry. J. Mass Spectrom. **38:** 43-49.

Justesen, U. (2000). Negative atmospheric pressure chemical ionisation low-energy collision activation mass spectrometry for the characterisation of flavonoids in extracts of fresh herbs. J. Chromatogr. A **902:** 369-379.

Kanawati, B., and Schmitt-Kopplin, P. (2010). Exploring rearrangements along the fragmentation of glutaric acid negative ion: a combined experimental and theoretical study. Rapid Commun. Mass Spectrom. **24:** 1198-1206.

Kanawati, B., Joniec, S., Winterhalter, R., and Moortgat, G.K. (2007). Mass spectrometric characterization of small oxocarboxylic acids and gas phase ion fragmentation mechanisms studied by electrospray triple quadrupole-MS/MS-TOF system and DFT theory. Int. J. Mass Spectrom. **266:** 97-113.

Kanawati, B., Herrmann, F., Joniec, S., Winterhalter, R., and Moortgat, G.K. (2008). Mass spectrometric characterization of *β*-caryophyllene ozonolysis products in the aerosol studied using an electrospray triple quadrupole and time-of-flight analyzer hybrid system and density functional theory. Rapid Commun. Mass Spectrom. **22:** 165-186.

Kitamura, S., Matsuda, F., Tohge, T., Yonekura-Sakakibara, K., Yamazaki, M., Saito, K., and Narumi, I. (2010). Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. Plant J. **62:** 549-559.

Kjær, A., and Schuster, A. (1970). Glucosinolates in *Erysimum hieracifolium* L.; three new, naturally occurring glucosinolates. Acta Chem. Scand. **24:** 1631-1638.

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

Kuang, H., Sun, S., Yang, B., Xia, Y., and Feng, W. (2009). New megastigmane sesquiterpene and indole alkaloid glucosides from the aerial parts of *Bupleurum chinense* DC. Fitoterapia **80:** 35-38.

Le Gall, G., Metzdorff, S.B., Pedersen, J., Bennett, R.N., and Colquhoun, I.J. (2005). Metabolite profiling of *Arabidopsis thaliana* (L.) plants transformed with an antisense chalcone synthase gene. Metabolomics **1:** 181-198.

Leplé, J.-C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Pollet, B., Naumann, A., Kang, K.-Y., Kim, H., Ruel, K., Lefèbvre, A., Joseleau, J.-P., Grima-Pettenati, J., De

Rycke, R., Andersson-Gunnerås, S., Erban, A., Fehrle, I., Petit-Conil, M., Kopka, J., Polle, A., Messens, E., Sundberg, B., Mansfield, S.D., Ralph, J., Pilate, G., and Boerjan, W. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. Plant Cell **19:** 3669-3691.

López-Carbonell, M., and Jáuregui, O. (2005). A rapid method for analysis of abscisic acid (ABA) in crude extracts of water stressed *Arabidopsis thaliana* plants by liquid chromatography-mass spectrometry in tandem mode. Plant Physiol. Biochem. **43:** 407-411.

Levsel, K., Schiebel, H.-M., Terlouw, J.K., Jobst, K.J., Elend, M., Preiß, A., Thiele, H., and Ingendoh, A. (2007). Even-electron ions: a systematic study of the neutral species lost in the dissociation of quasi-molecular ions. J. Mass Spectrom. **42:** 1024-1044.

Mancel, V., Sellier, N., Lesage, D., Fournier, F., and Tabet, J.-C. (2004). Gas phase enantiomeric distinction of (*R*)- and (*S*)-aromatic hydroxy esters by negative ion chemical ionization mass spectrometry using a chiral reagent gas. Int. J. Mass Spectrom. **237:** 185-195.

March, R.E., and Stadey, C.J. (2005). A tandem mass spectrometric study of saccharides at high mass resolution. Rapid Commun. Mass Spectrom. **19:** 805-812.

Marzouk, M.M., Al-Nowaihi, A.-S.M., Kawashty, S.A., and Saleh, N.A.M. (2010). Chemosystematic studies on certain species of the family Brassicaceae (Cruciferae) in Egypt. Biochem. Syst. Ecol. **38:** 680-685.

Matsubara, Y., Kumamoto, H., Sawabe, A., Iizuka, Y., and Okamoto, K. (1985). Structure and physiological activity of phenylpropanoid glycosides in lemon, unshui and orange peelings. Kinki Daigaku Igaku Zasshi **10:** 51-58.

Matsuda, F., Hirai, M.Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N.J., Sakurai, T., Shimada, Y., and Saito, K. (2010). AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol. **152:** 566-578.

Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., Herdewijn, P., Devreese, B., Van Beeumen, J., Marita, J.M., Ralph, J., Chen, C., Burggraeve, B., Van Montagu, M., Messens, E., and Boerjan, W. (2000). Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of

caffeoyl-coenzyme A *O*-methyltransferase, an enzyme involved in lignin biosynthesis. J. Biol. Chem. **275:** 36899-36909.

Ming, D.-S., Jiang, R.-W., But, P.P.-H., Towers, G.H.N., and Yu, D.-Q. (2002). A new compound from *Geum rivale* L. J. Asian Nat. Prod. Res. **4:** 217-220.

Miyaichi, Y., and Tomimori., T. (1998). Studies on constituents of *Scutellaria* species. XIX. Lignan glycosides of roots of *Scutellaria baicalensis* GEORGI. Nat. Med. **52:** 82-86.

Montaut, S., and Bleeker, R.S. (2010). Isolation and structure elucidation of 5′-*O*-β-d-glucopyranosyl-dihydroascorbigen from *Cardamine diphylla* rhizome. Carbohydr. Res. **345:** 1968-1970.

Morreel, K., Ralph, J., Kim, H., Lu, F., Goeminne, G., Ralph, S., Messens, E., and Boerjan, W. (2004). Profiling of oligolignols reveals monolignol coupling conditions in lignifying poplar xylem. Plant Physiol. **136:** 3537-3549.

Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., Breyne, P., Steenackers, M., Georges, M., Messens, E., and Boerjan, W. (2006). Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. Plant J. **47:** 224-237.

Morreel, K., Kim, H., Lu, F., Dima, O., Akiyama, T., Vanholme, R., Niculaes, C., Goeminne, G., Inzé, D., Messens, E., Ralph, J., and Boerjan, W. (2010a). Mass spectrometry-based fragmentation as an identification tool in lignomics. Anal. Chem. **82:** 8095-8105.

Morreel, K., Dima, O., Kim, H., Lu, F., Niculaes, C., Vanholme, R., Dauwe, R., Goeminne, G., Inzé, D., Messens, E., Ralph, J., and Boerjan, W. (2010b). Mass spectrometry-based sequencing of lignin oligomers. Plant Physiol. **153:** 1464-1478.

Mulroney, B., Peel, J.B., and Traeger, J.C. (1999). Theoretical study of deprotonated glucopyranosyl disaccharide fragmentation. J. Mass Spectrom. **34:** 856-871.

Nagatani, Y., Warashina, T., and Noro, T. (2002). Studies on the constituents from the aerial part of *Baccharis dracunculifolia* DC. II. Chem. Pharm. Bull. **50:** 583-589.

Nakabayashi, R., Kusano, M., Kobayashi, M., Tohge, T., Yonekura-Sakakibara, K., Kogure, N., Yamazaki, M., Kitajima, M., Saito, K., and Takayama, H. (2009). Metabolomics-oriented isolation and structure elucidation of 37 compounds including two anthocyanins from *Arabidopsis thaliana*. Phytochem. **70:** 1017-1029.

Pedras, M.S.C., and Zheng, Q.-A. (2010). Metabolic responses of *Thellungiella halophila/salsuginea* to biotic and abiotic stresses: metabolite profiles and quantitative analyses. Phytochemistry **71:** 581-589.

Petersen, B.L., Chen, S., Hansen, C.H., Olsen, C.E., and Halkier, B.A. (2002). Composition and content of glucosinolates in developing *Arabidopsis thaliana*. Planta **214:** 562-571.

Plumb, G.W., Price, K.R., Rhodes, M.J.C., and Williamson, G. (1997). Antioxidant properties of the major polyphenolic compounds in broccoli. Free Radical Res. **27:** 429-435.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J. Nat. Prod. **74:** 1462-1476.

Rasoanaivo, P., Ratsimamanga-Urverg, S., Messana, I., De Vicente, Y., and Galeffi, C. (1990). Cassinopin, a kaempferol trirhamnoside from *Cassinopsis madagascariensis*. Phytochemistry **29:** 2040-2043.

Reeks, L.B., Eichinger, P.C.H., Bowie, J.H. (1993). *Ortho*-rearrangements of O-alkylphenoxide anions. Rapid Commun. Mass Spectrom. **7:** 286-287.

Reichelt, M., Brown, P.D., Schneider, B., Oldham, N.J., Stauber, E., Tokuhisa, J., Kliebenstein, D.J., Mitchell-Olds, T., and Gershenzon, J. (2002). Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. Phytochemistry **59:** 663-671.

Ricci, A., Fiorentino, A., Piccolella, S., D'Abrosca, B., Pacifico, S., and Monaco, P. (2010). Structural discrimination of isomeric tetrahydrofuran lignan glucosides by tandem mass spectrometry. Rapid Commun. Mass Spectrom. **24:** 979-985.

Ricci, A., Fiorentino, A., Piccolella, S., Golino, A., Pepi, F., D'Abrosca, B., Letizia, M., and Monaco, P. (2008). Furofuranic glycosylated lignans: a gas-phase ion chemistry investigation by tandem mass spectrometry. Rapid Commun. Mass Spectrom. **22:** 3382-3392.

Rochfort, S.J., Trenerry, V.C., Imsic, M., Panozzo, J., and Jones, R. (2008). Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MS^n fragmentation. Phytochemistry **69:** 1671-1679.

Rohde, A., Morreel, K., Ralph, J., Goeminne, G., Hostyn, V., De Rycke, R., Kushnir, S., Van Doorsselaere, J., Joseleau, J.-P., Vuylsteke, M., Van Driessche, G., Van Beeumen, J., Messens, E., and Boerjan, W. (2004). Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. Plant Cell **16:** 2749-2771.

Sakushima, A., Coskun, M., Tanker, M., and Tanker, N. (1994) A sinapic acid ester from *Boreava orientalis*. Phytochemistry **35:** 1481-1484.

Schmidt, T.J., Alfermann, A.W., and Fuss, E. (2008). High-performance liquid chromatography/mass spectrometric identification of dibenzylbutyrolactone-type lignans: insights into electrospray ionization tandem mass spectrometric fragmentation of lign-7-eno-9,9'-lactones and application to the lignans of *Linum usitatissimum* L. (Common Flax). Rapid Commun. Mass Spectrom. **22:** 3642-3650.

Shahat, A.A., Cuyckens, F., Wang, W., Abdel-Shafeek, K.A., Husseiny, H.A., Apers, S., Van Miert, S., Pieters, L., Vlietinck, A.J., and Claeys, M. (2005). Structural characterization of flavonol di-*O*-glycosides from *Farsetia aegyptia* by electrospray ionization and collision-induced dissociation mass spectrometry. Rapid Commun. Mass Spectrom. **19:** 2172-2178.

Shimomura, H., Sashida, Y., and Mimaki, Y. (1987) Phenolic glycerides from *Lilium auratum*. Phytochemistry **26:** 844-845.

Stroobant, V., Rozenberg, R., Bouabsa, e.M., Deffense, E., and de Hoffmann, E. (1995). Fragmentation of conjugate bases of esters derived from multifunctional alcohols including triacylglycerols. J. Am. Soc. Mass Spectrom. **6:** 498-506.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., and Viant, M.R. (2007). Proposed minimum reporting standards for chemical analysis. Metabolomics **3:** 211-221.

Syrjänen, K., and Brunow, G. (1998). Oxidative cross coupling of *p*-hydroxycinnamic alcohols with dimeric arylglycerol β-aryl ether lignin model compounds. The effect of oxidation potentials. J. Chem. Soc., Perkin Trans. 1**:** 3425-3429.

Thevis, M., Schänzer, W., and Schmickler, H. (2003). Effect of the location of hydrogen abstraction on the fragmentation of diuretics in negative electrospray ionization mass spectrometry. J. Am. Soc. Mass Spectrom. **14:** 658-670.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., and Saito, K. (2005), Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J**. 42:** 218-235.

Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., and Saito, K. (2007). Phytochemical genomics in *Arabidopsis thaliana*: a case study for functional identification of flavonoid biosynthesis genes. Pure Appl. Chem. **79:** 811-823.

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J.H., Goeminne, G., Halpin, C., Rohde, A., Morreel, K., and Boerjan, W. (2012). A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. Plant Cell **24:** 3506-3529.

Veit, M., and Pauli, G.F. (1999). Major flavonoids from *Arabidopsis thaliana* leaves. J. Nat. Prod. **62:** 1301-1303.

Wang, H., Leach, D.N., Thomas, M.C., Blanksby, S.J., Forster, P.I., and Waterman, P.G. (2008). Bisresorcinols and arbutin derivatives from Grevillea banksii R. Br. Nat. Prod. Commun. **3:** 57-64.

Wolf, S., Schmidt, S., Müller-Hannemann, M., and Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics **11:** 148.

Woodhead, S., and Cooper-Driver, G. (1979). Phenolic acids and resistance to insect attack in *Sorghum bicolor*. Biochem. Syst. Ecol. **7:** 309-310.

Yan, C., Liu, S., Zhou, Y., Song, F., Cui, M., and Liu, Z. (2007). A study of isomeric diglycosyl flavonoids by SORI CID of Fourier transform ion cyclotron mass spectrometry in negative ion mode. J. Am. Soc. Mass Spectrom. **18:** 2127-2136.

Ye, M., Yan, Y.N., and Guo, D.-a. (2005). Characterization of phenolic compounds in the Chinese herbal drug Tu-Si-Zi by liquid chromatography coupled to electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. **19:** 1469-1484.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., and Saito, K. (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. Plant Cell **20:** 2160-2176.