nature plants

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41477-018-0350-3

In the format provided by the authors and unedited.

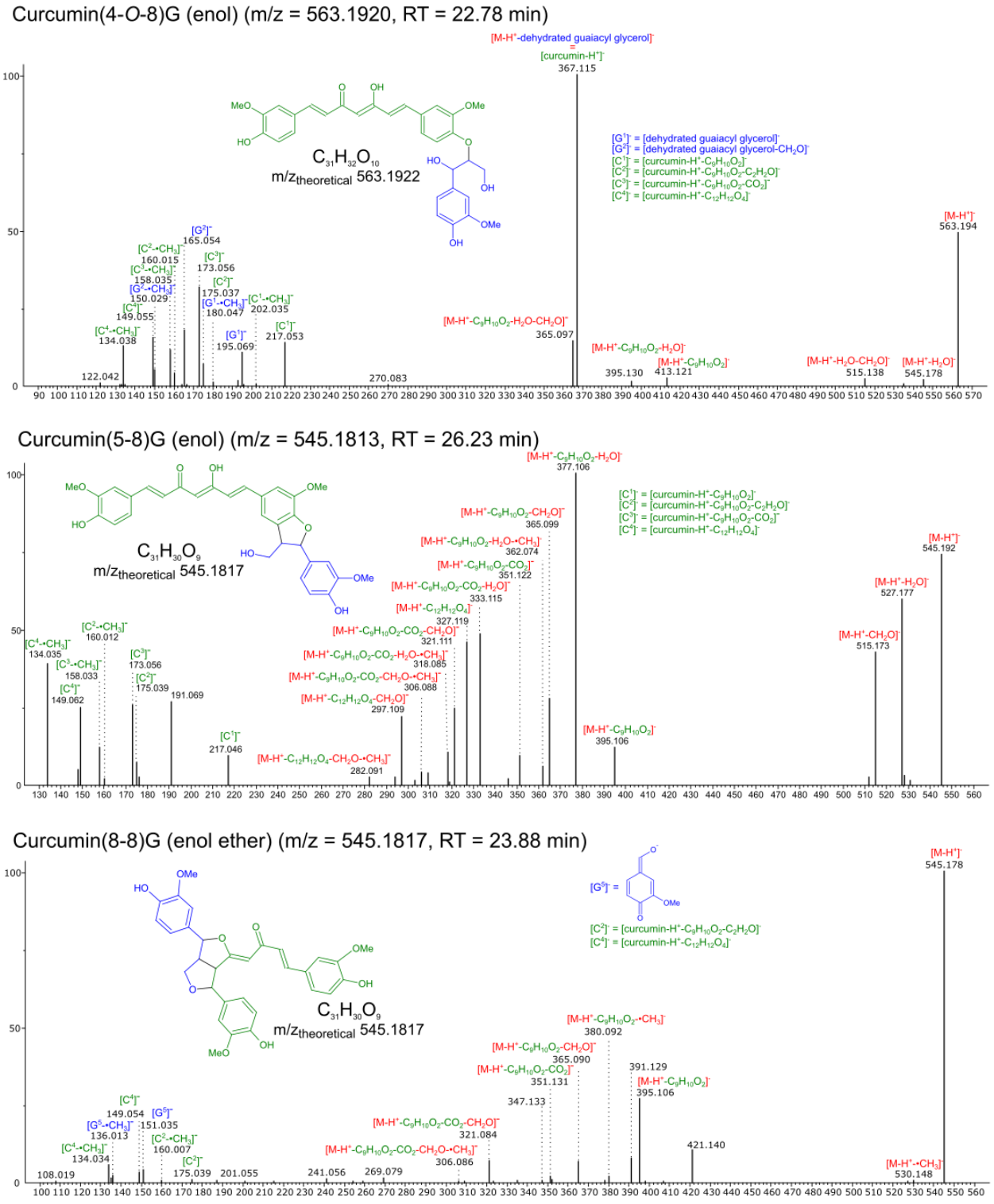
# Introducing curcumin biosynthesis in *Arabidopsis* enhances lignocellulosic biomass processing

**Paula Oyarce[1,2,8], Barbara De Meester[1,2,8], Fernando Fonseca[1,2], Lisanne de Vries[1,2], Geert Goeminne[1,2,3], Andreas Pallidis[1,2], Riet De Rycke[1,2,3,4], Yukiko Tsuji[5,6], Yanding Li[5,6], Sander Van den Bosch[7], Bert Sels[7], John Ralph[5,6], Ruben Vanholme[1,2,3,9] and Wout Boerjan[1,2,3,9]***

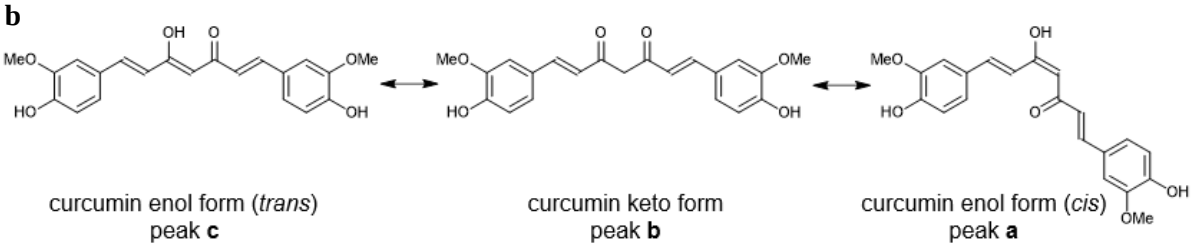
[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. [2]VIB Center for Plant Systems Biology, Ghent, Belgium. [3]VIB Metabolomics Core, Ghent, Belgium. [4]Ghent University Expertise Centre for Transmission Electron Microscopy and VIB BioImaging Core, Ghent, Belgium. [5]Department of Biochemistry, University of Wisconsin, Madison, WI, USA. [6]US Department of Energy, Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, Madison, WI, USA. [7]Center for Surface Chemistry and Catalysis, KU Leuven, Heverlee, Belgium. [8]These authors contributed equally:
Paula Oyarce, Barbara De Meester. [9]These authors jointly supervised this work: Ruben Vanholme, Wout Boerjan. *e-mail: woboe@psb.vib-ugent.be

**Supplementary Figure 1. MS/MS spectra and structural characterization of curcumin coupling products in oxidative coupling assays.** (**a**) MS/MS spectra of curcumin-coniferyl alcohol coupling products observed in in vitro oxidative coupling assays. A representative MS/MS spectrum from seven, ten, and five recorded spectra for curcumin(8-*O*-4)G, curcumin(8-5)G and curcumin(8-8)G, respectively, is shown. Fragments and neutral losses that originate from the coniferyl alcohol moiety are given in blue, neutral losses that originate from the linkage type[28,29] are given in red. The annotation of the curcumin fragments (in green) is explained in panels (**b**) to (**f**). Curcumin(8-*O*-4)G and curcumin(8-5)G were assigned as enol tautomers, based on the relative signal intensities of m/z 158.0, 160.0, 173.1 and 175.0 in the respective MS/MS spectra. For curcumin(8-8)G, a keto tautomer is impossible because the compound has an enol ether functionality. (**b**) Fragmentation of curcumin isomers and tautomers observed in *Curcuma longa*. The differences in MS/MS spectra between keto and enol tautomers is explained by the fragmentation pathways in panel (**e**) and (**f**). The *cis* and *trans* annotation is based on the observation that peak a and peak c have similar MS/MS spectra, but different retention times. In the *trans* conformation, the alcohol and ketone can interact with an intramolecular hydrogen bridge, making the molecule less polar and thus more retentive in the reversed-phase separation. In the *cis* conformation, the alcohol and ketone can not interact, making the molecule more polar as compared with the *trans* conformation, and thus less retentive in the reversed-phase separation. Upon ionization, the *cis* and *trans* equilibrate instantly, explaining the similar MS/MS spectra. (**c**) Fragmentation of demethoxycurcumin isomers and tautomers observed in *C. longa*. The differences in MS/MS spectra between keto and enol tautomers is explained by the fragmentation pathways similar to those of curcumin given in panel (**e**) and (**f**). (**d**) Fragmentation of bisdemethoxycurcumin isomers and tautomers observed in *C. longa*. The differences in MS/MS spectra between keto and enol tautomers is explained by the fragmentation pathways similar to those of curcumin given in panel (**e**) and (**f**). For (**b**), (**c**) and (**d**) one *Curcuma longa* sample was analyzed and one MS/MS spectrum was generated for each compound. (**e**) Fragmentation pathway of curcumin keto form. (**f**) Fragmentation pathway of curcumin enol form.

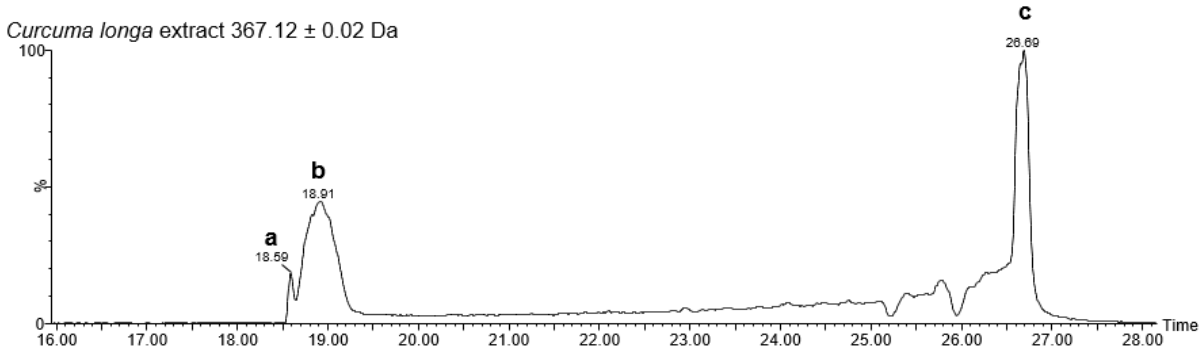
**a**


Curcumin(4-O-8)G (enol) (m/z = 563.1920, RT = 22.78 min)
C31H32O10
m/ztheoretical 563.1922
[M-H+-dehydrated guaiacyl glycerol]-
=
[curcumin-H+]-
367.115
[M-H+]-
563.194
[M-H+-C9H10O2-H2O-CH2O]-
365.097
[M-H+-C9H10O2-H2O]-
[M-H+-C9H10O2]-
395.130
413.121
[M-H+-H2O-CH2O]-
515.138
[M-H+-H2O]-
545.178
[G2]- 165.054
[C2-•CH3]- 160.015
[C3]- 173.056
[C3-•CH3]- 158.035
[C2]- 175.037
[G5-•CH3]- 150.029
[G1-•CH3]- 180.047
[C1-•CH3]- 202.035
[C4]- 149.055
[C1]- 217.053
[C4-•CH3]- 134.038
[G1]- 195.069
122.042
270.083
[G1]- = [dehydrated guaiacyl glycerol]-
[G2]- = [dehydrated guaiacyl glycerol-CH2O]-
[C1]- = [curcumin-H+-C9H10O2]-
[C2]- = [curcumin-H+-C9H10O2-C2H2O]-
[C3]- = [curcumin-H+-C9H10O2-CO2]-
[C4]- = [curcumin-H+-C12H12O4]-
Curcumin(5-8)G (enol) (m/z = 545.1813, RT = 26.23 min)
C31H30O9
m/ztheoretical 545.1817
[M-H+-C9H10O2-H2O]-
377.106
[M-H+-C9H10O2-CH2O]-
365.099
[M-H+-C9H10O2-H2O-•CH3]-
362.074
[M-H+-C9H10O2-CO2]-
351.122
[M-H+-C9H10O2-CO2-H2O]-
333.115
[M-H+-C12H12O4]-
327.119
[M-H+-C9H10O2-CO2-CH2O]-
321.111
[M-H+-C9H10O2-CO2-H2O-•CH3]-
318.085
[M-H+-C9H10O2-CO2-CH2O-•CH3]-
306.088
[M-H+-C12H12O4-CH2O]-
297.109
[M-H+-C12H12O4-CH2O-•CH3]-
282.091
[M-H+]-
545.192
[M-H+-H2O]-
527.177
[M-H+-CH2O]-
515.173
[M-H+-C9H10O2]-
395.106
[C2-•CH3]- 160.012
[C4-•CH3]- 134.035
[C3-•CH3]- 158.033
[C5]- 173.056
[C2]- 175.039
[C4]- 149.062
191.069
[C1]- 217.046
[C1]- = [curcumin-H+-C9H10O2]-
[C2]- = [curcumin-H+-C9H10O2-C2H2O]-
[C3]- = [curcumin-H+-C9H10O2-CO2]-
[C4]- = [curcumin-H+-C12H12O4]-
Curcumin(8-8)G (enol ether) (m/z = 545.1817, RT = 23.88 min)
C31H30O9
m/ztheoretical 545.1817
[G5]- =
[C2]- = [curcumin-H+-C9H10O2-C2H2O]-
[C4]- = [curcumin-H+-C12H12O4]-
[M-H+]-
545.178
[M-H+-C9H10O2-•CH3]-
380.092
[M-H+-C9H10O2-CH2O]-
365.090
[M-H+-C9H10O2-CO2]-
351.131
391.129
[M-H+-C9H10O2]-
395.106
347.133
[M-H+-C9H10O2-CO2-CH2O]-
321.084
[M-H+-C9H10O2-CO2-CH2O-•CH3]-
306.086
421.140
[M-H+-•CH3]-
530.148
[C4]- 149.054
[G5]- 151.035
[G5-•CH3]- 136.013
[C2-•CH3]- 160.007
[C4-•CH3]- 134.034
[C2]- 175.039
108.019
201.055
241.056
269.079

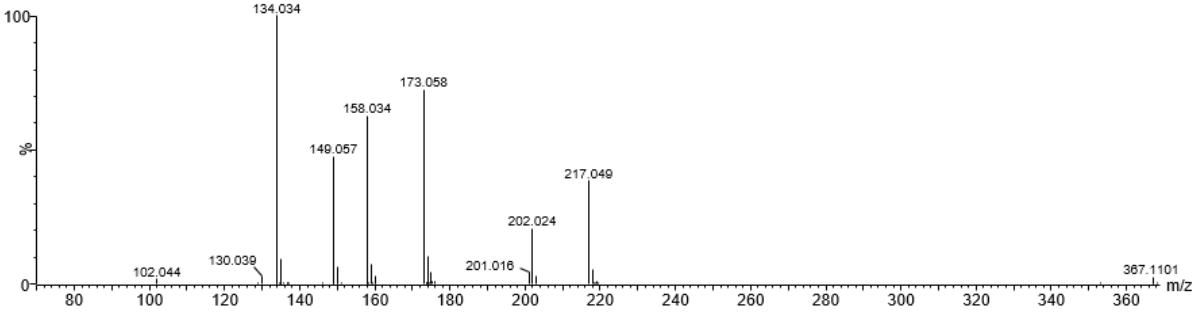
**b**



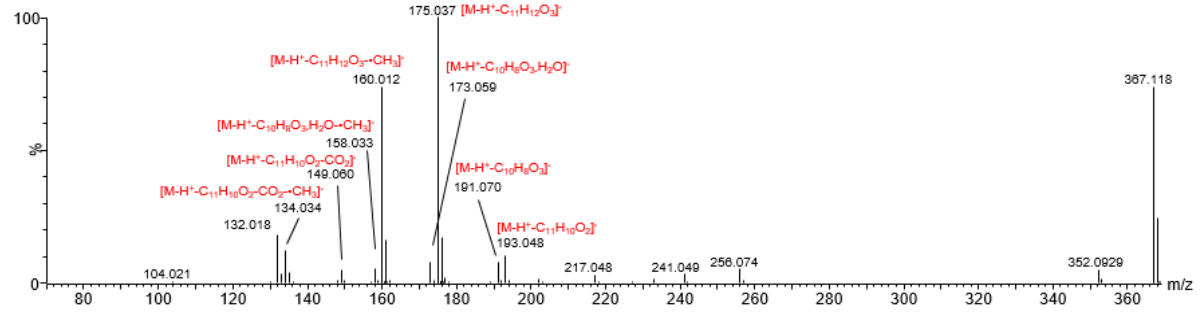

curcumin enol form (*trans*) peak **c** | curcumin keto form peak **b** | curcumin enol form (*cis*) peak **a**



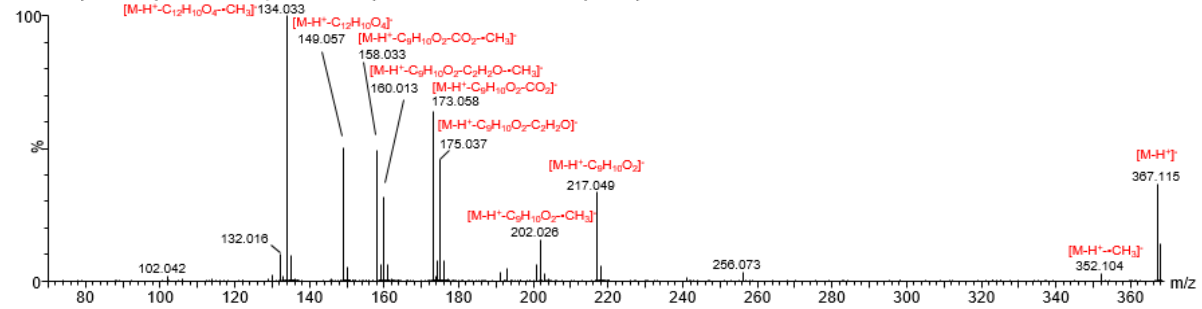

MS/MS peak **a** (m/z 367.12, 18.59 min): curcumin enol form (*cis*)



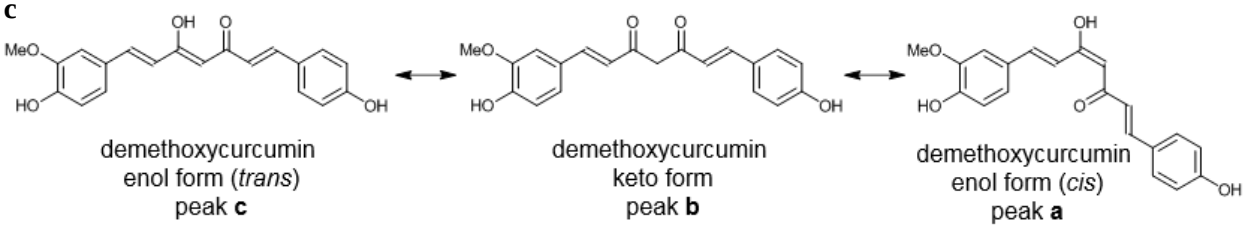

MS/MS peak **b** (m/z 367.12, 18.91 min): curcumin keto form



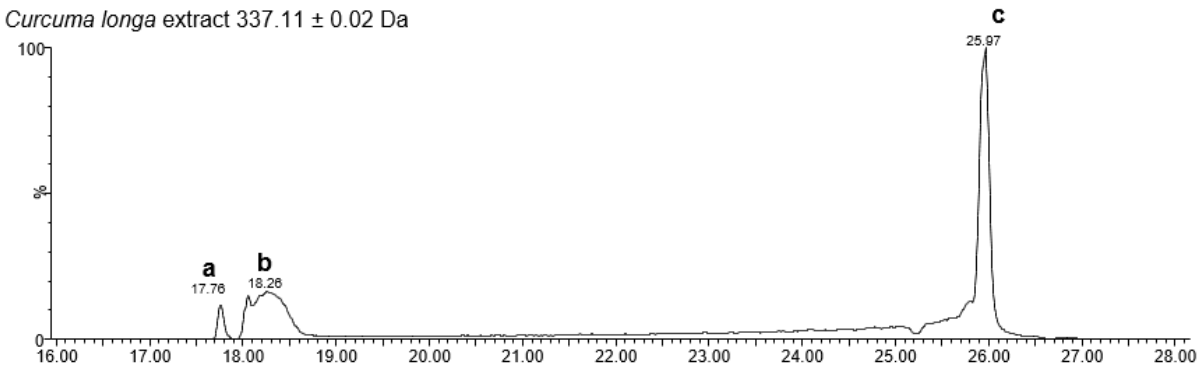

MS/MS peak **c** (m/z 367.12, 26.69 min): curcumin enol form (*trans*)



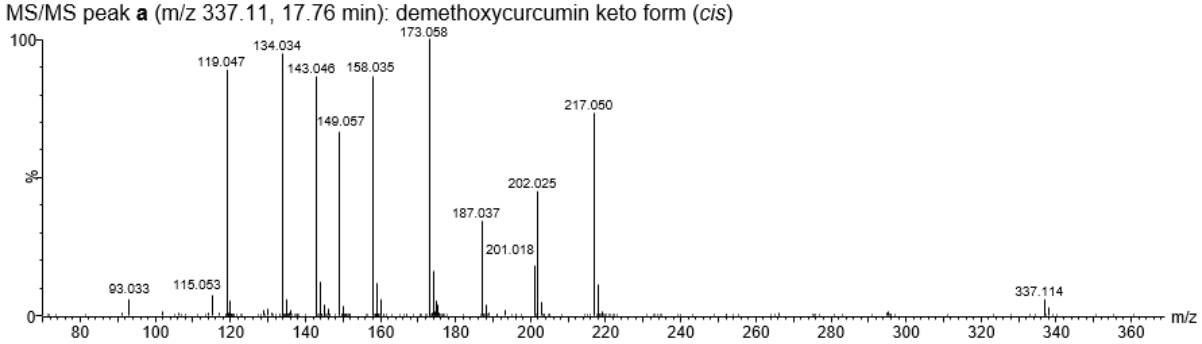
**c**



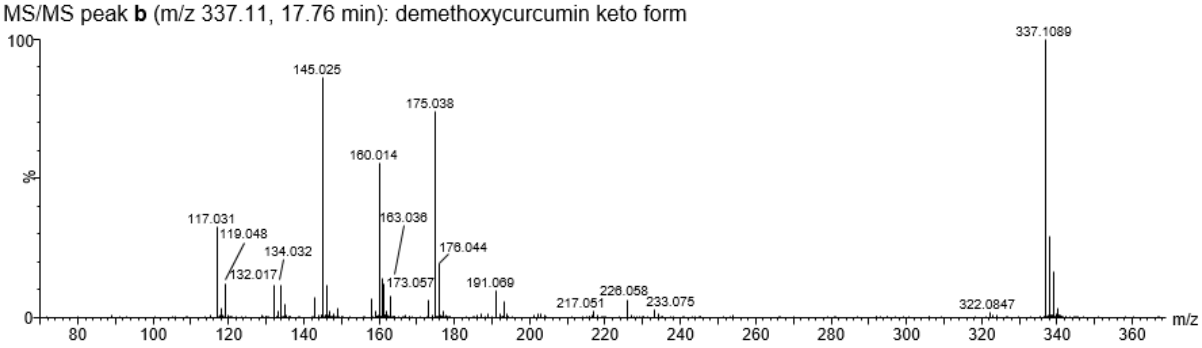

*Curcuma longa* extract 337.11 ± 0.02 Da



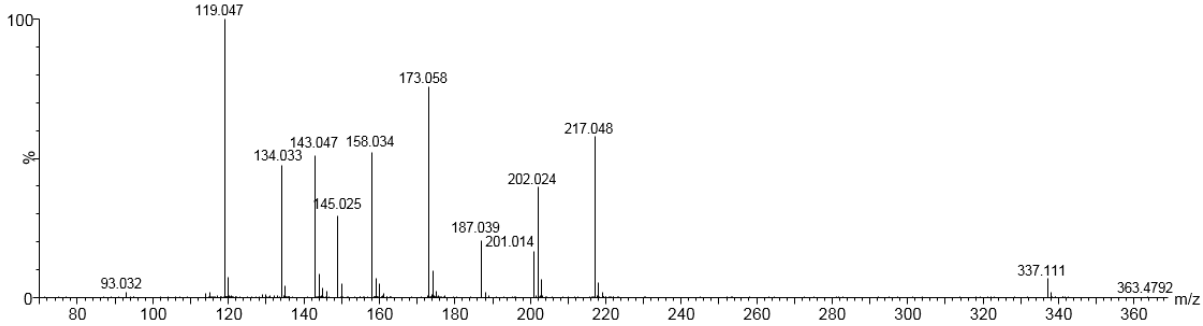

MS/MS peak **a** (m/z 337.11, 17.76 min): demethoxycurcumin keto form (*cis*)



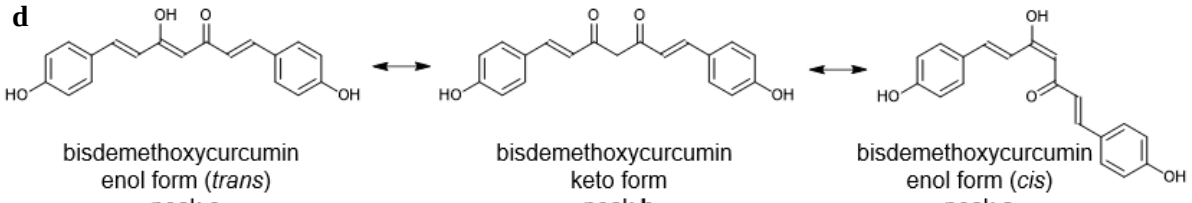

MS/MS peak **b** (m/z 337.11, 17.76 min): demethoxycurcumin keto form




MS/MS peak **c** (m/z 337.11, 25.97 min): demethoxycurcumin enol form (*trans*)



**d**

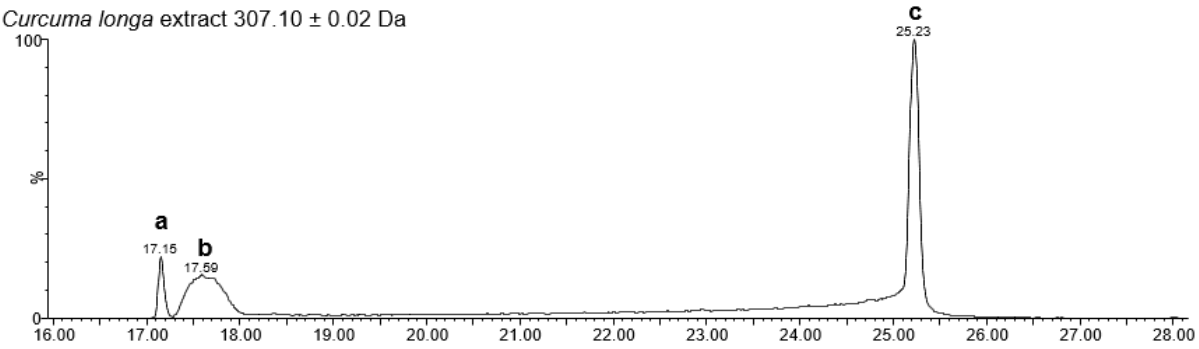



bisdemethoxycurcumin enol form (*trans*) peak **c**

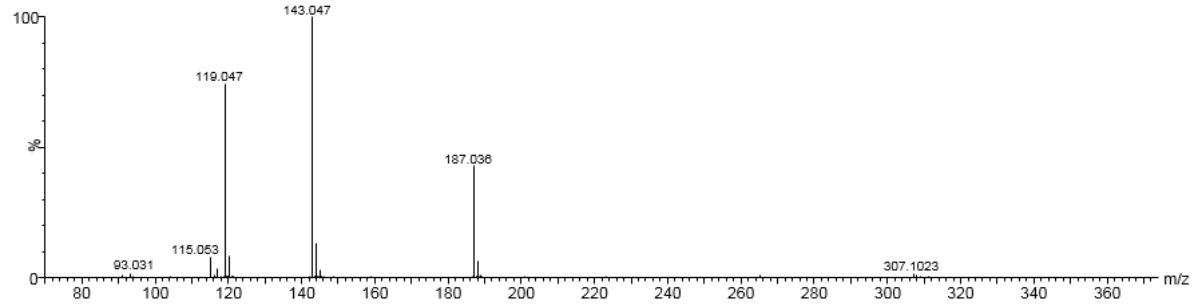
bisdemethoxycurcumin keto form peak **b**

bisdemethoxycurcumin enol form (*cis*) peak **a**

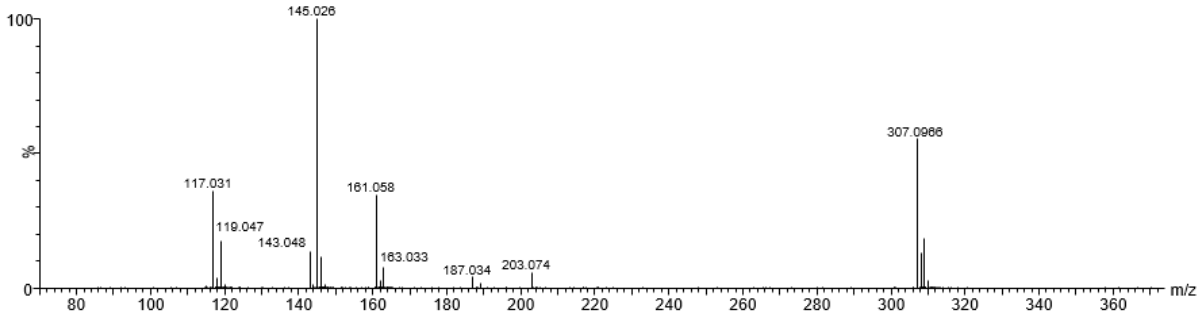
*Curcuma longa* extract 307.10 ± 0.02 Da



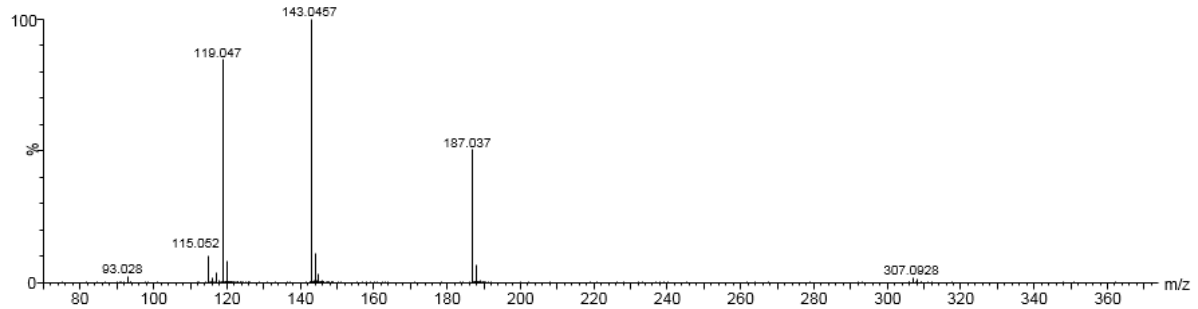

MS/MS peak **a** (m/z 307.10, 17.15 min) bisdemethoxycurcumin enol form (*cis*)



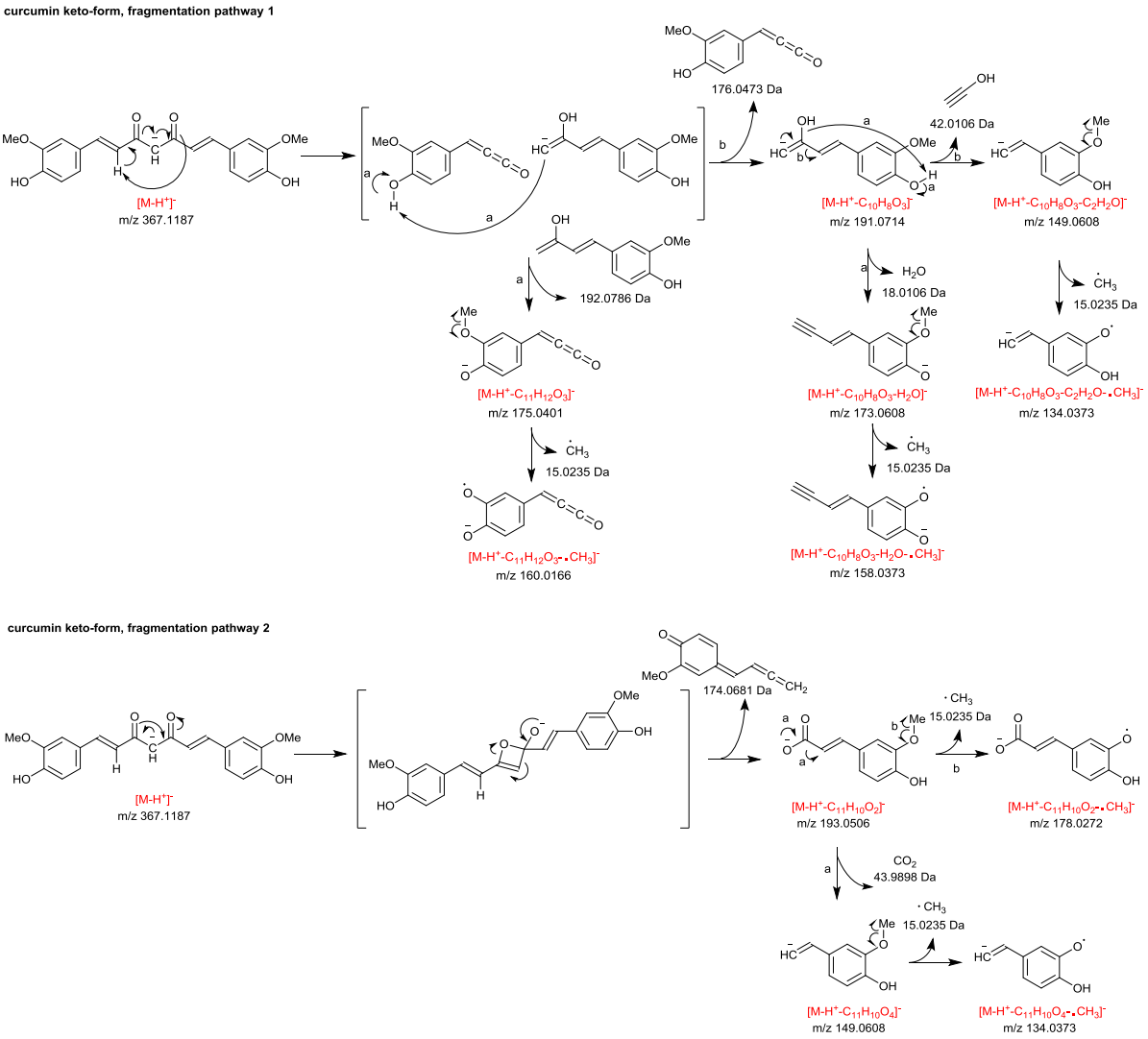

MS/MS peak **b** (m/z 307.10, 17.59 min): bisdemethoxycurcumin keto form



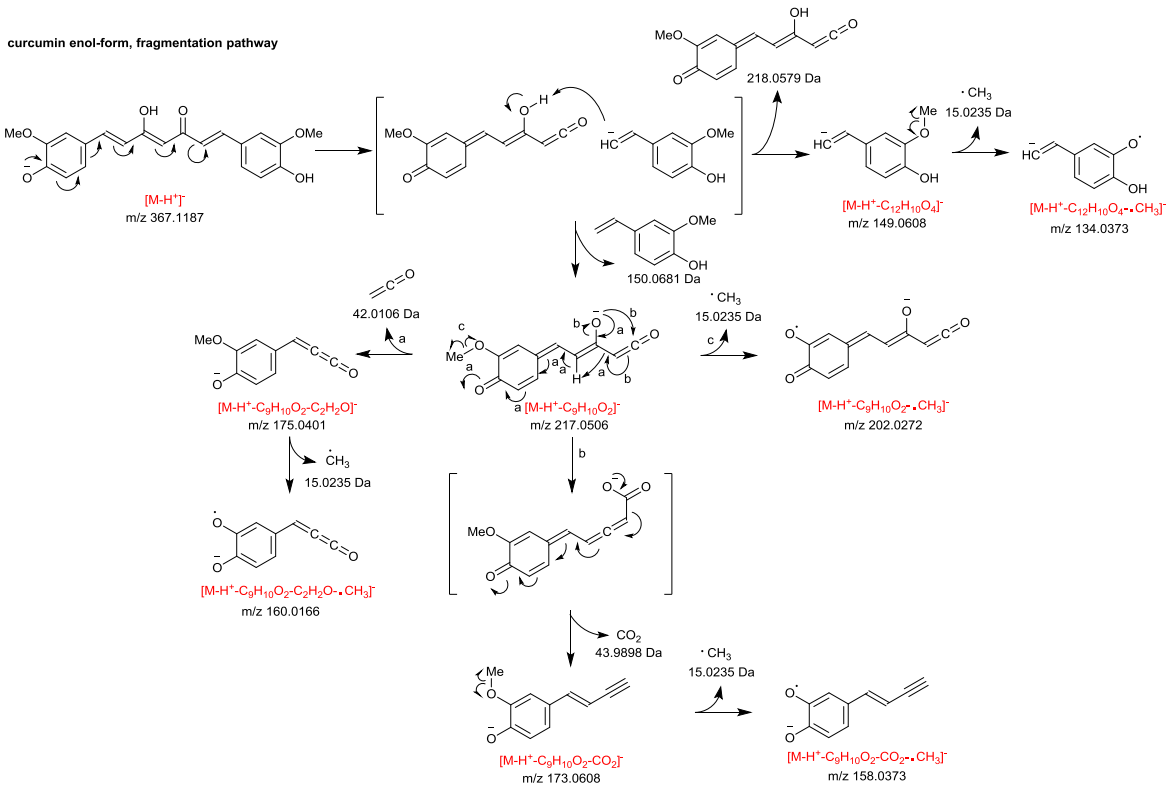

MS/MS peak **c** (m/z 307.10, 25.23 min): bisdemethoxycurcumin enol form (*trans*)



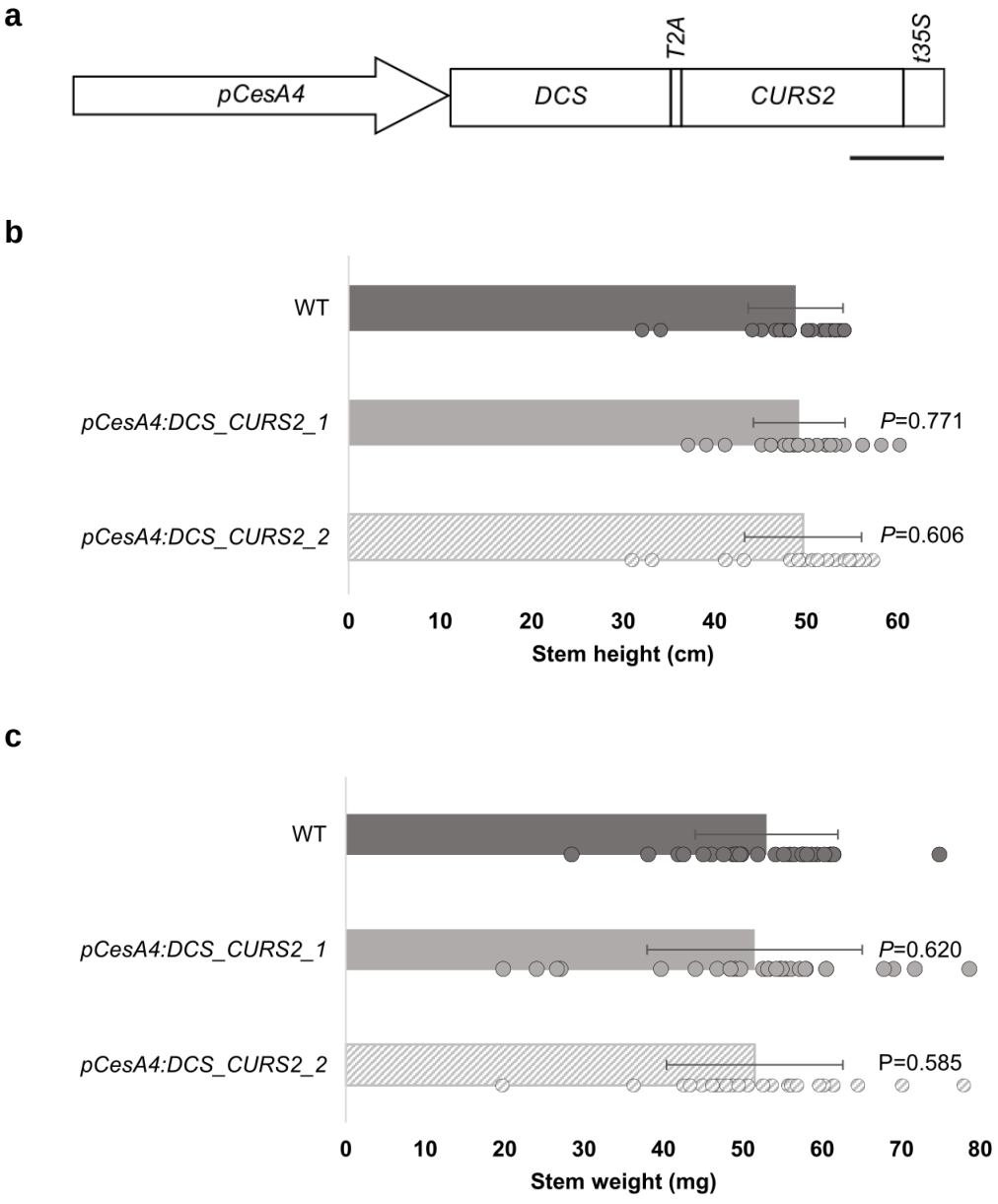
**e**


curcumin keto-form, fragmentation pathway 1
[M-H+]-
m/z 367.1187
176.0473 Da
42.0106 Da
192.0786 Da
[M-H+-C10H8O3]-
m/z 191.0714
[M-H+-C10H8O3-C2H2O]-
m/z 149.0608
H2O
18.0106 Da
15.0235 Da
[M-H+-C11H12O3]-
m/z 175.0401
[M-H+-C10H8O3-H2O]-
m/z 173.0608
[M-H+-C10H8O3-C2H2O-•CH3]-
m/z 134.0373
[M-H+-C11H12O3-•CH3]-
m/z 160.0166
[M-H+-C10H8O3-H2O-•CH3]-
m/z 158.0373
curcumin keto-form, fragmentation pathway 2
[M-H+]-
m/z 367.1187
174.0681 Da
· CH3
15.0235 Da
[M-H+-C11H10O2]-
m/z 193.0506
[M-H+-C11H10O2-•CH3]-
m/z 178.0272
CO2
43.9898 Da
[M-H+-C11H10O4]-
m/z 149.0608
[M-H+-C11H10O4-•CH3]-
m/z 134.0373

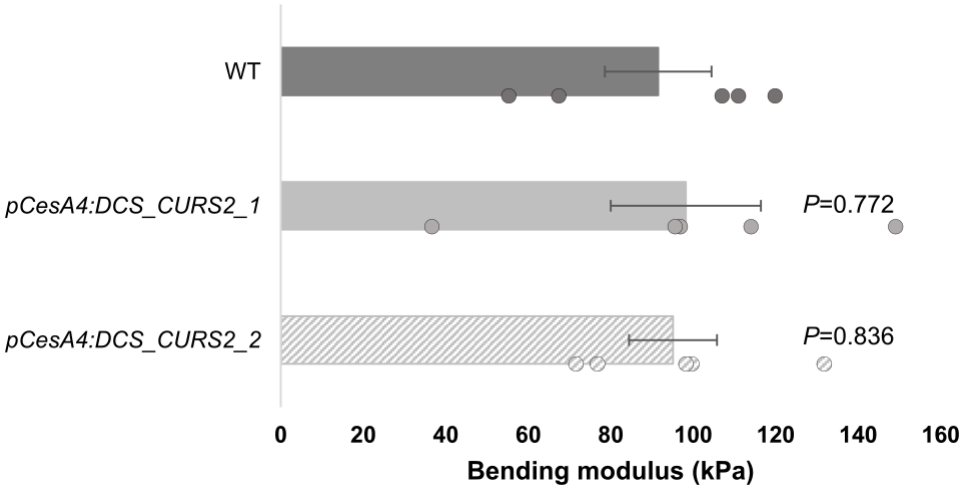
**f**


curcumin enol-form, fragmentation pathway
[M-H+]-
m/z 367.1187
218.0579 Da
150.0681 Da
[M-H+-C12H10O4]-
m/z 149.0608
· CH3
15.0235 Da
[M-H+-C12H10O4-·CH3]-
m/z 134.0373
42.0106 Da
[M-H+-C9H10O2-C2H2O]-
m/z 175.0401
[M-H+-C9H10O2]-
m/z 217.0506
[M-H+-C9H10O2-·CH3]-
m/z 202.0272
[M-H+-C9H10O2-C2H2O-·CH3]-
m/z 160.0166
CO2
43.9898 Da
[M-H+-C9H10O2-CO2]-
m/z 173.0608
[M-H+-C9H10O2-CO2-·CH3]-
m/z 158.0373

**Supplementary Figure 2. Gene construct and biomass measurements of *pCesA4:DCS_CURS2* transgenic Arabidopsis lines.** (**a**) Graphical representation of the *pCesA4:DCS_CURS2* construct. *DCS* and *CURS2* genes were linked together by the *T2A* sequence, controlled by the *CesA4* promoter and terminated with a *35S* terminator sequence (t35S). The different elements are drawn to scale. Scale bar: 500 bp. (**b**) Height measurements of the *pCesA4:DCS_CURS2* lines upon senescence. (**c**) Dry weight of the senesced main inflorescence stem of which siliques, cauline leaves and side branches were removed. Individual values (dots) and means (bars) of 29, 30 and 26 independent biological replicates for WT, *pCesA4:DCS_CURS2_1* and *pCesA4:DCS_CURS2_2*, respectively, are shown. No significant differences were observed for stem height and weight between the lines ($P > 0.05$; two-tailed Student's *t*-test). Error bars indicate standard deviation.



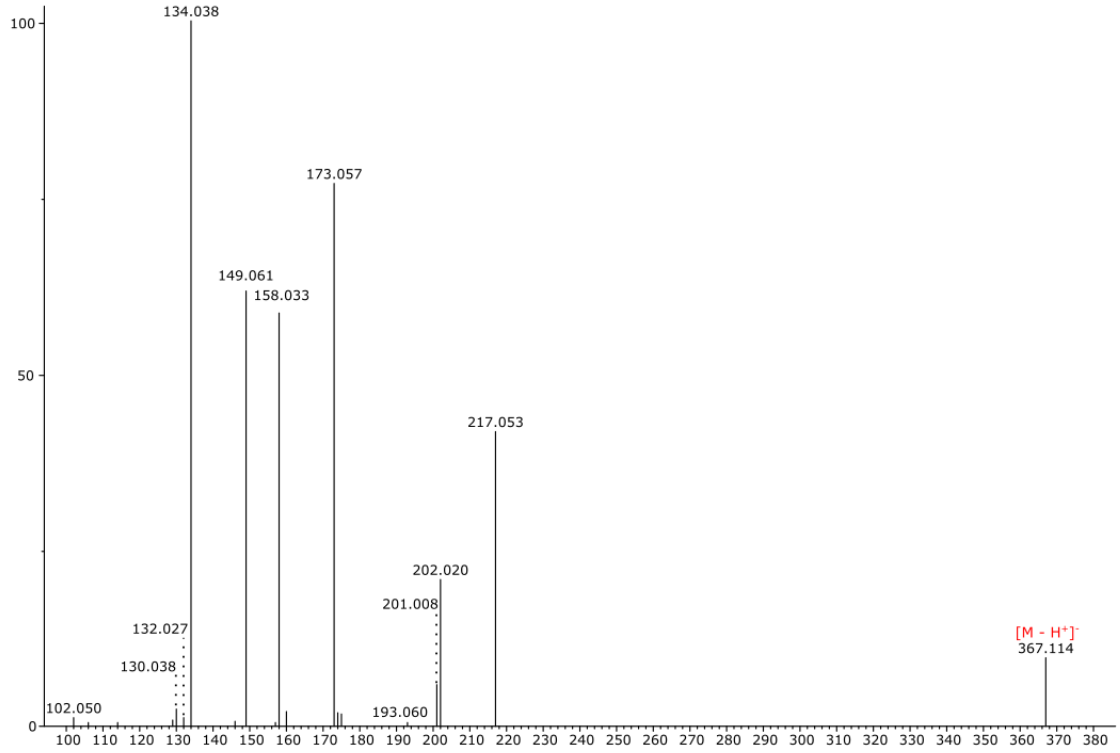
**Supplementary Figure 3. Bending modulus determined by two-point bending tests.** The bending modulus was determined to estimate the stiffness of the stems (*n*=5). Bars and error bars indicate means and standard errors, respectively. No significant differences were observed between the lines (*P* > 0.05; two-tailed Student's *t*-test).



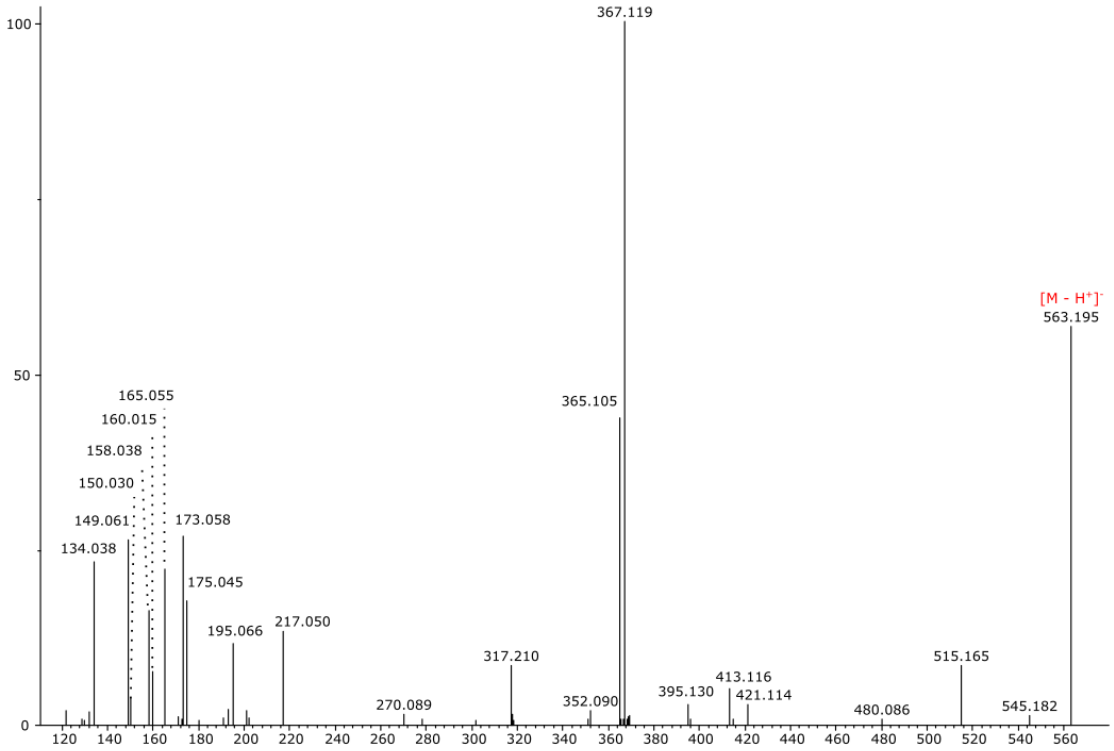
**Supplementary Figure 4. Phenolic profiling of *pCesA4:DCS_CURS2* plants.** MS/MS fragmentation spectra for compounds detected in extracts of *pCesA4:DCS_CURS2* plants. For each compound, a representative MS/MS spectrum is shown from at least four independent MS/MS spectra. The interpretation of the MS/MS spectra of compound **1**-**5** is given in Supplementary Figure 1. The fragments indicated with a red triangle in the MS/MS spectra of compound **6**-**21**, **26**, **27**, **32**-**38** and **40** refer to fragments from dihydroferuloyl-β-keto acid, as explained by the fragmentation routes shown for compound **6** (dihydroferuloyl-β-keto acid). In the same spectra, neutral losses of $C_9H_{10}O_2$ and $C_{10}H_{10}O_3$ are also explained by the fragmentation routes shown for compound **6** (dihydroferuloyl-β-keto acid). For compound **7**, the glycerol is thought to be on the 4-*O* position of dihydroferuloyl-β-keto acid, because the observed $[M - H^+ - C_2H_4O_2 - H_2O]^-$ and $[M - H^+ - CO_2 – C_2H_4O_2]^-$ fragments can only be formed if the 9CHOH and 11COOH positions are free. The fragments indicated with a red circle in the MS/MS spectra of compound **17**, **22**-**25** and **41** refer to fragments from tetrahydroferuloyl-β-keto acid, as explained by the fragmentation routes shown for compound **22** (tetrahydroferuloyl-β-keto acid). In the same spectra, the neutral loss corresponding to $C_2H_4O_2$ and $C_{10}H_{12}O_3$ is also explained by the fragmentation routes shown for compound **22** (tetrahydroferuloyl-β-keto acid). For compound **28**, fragmentation of the 8-5 (phenylcoumaran) structure resulting into $H_2O$, $CH_2O$ and $^{1,2}A$ neutral losses and $^{1,2}B^-$ fragments, is as described in Morreel et al., 2004[35]. In the same MS/MS spectrum, $CO_2$, $C_2H_4O$, $C_2H_4O_2$ and $C_4H_6O_3$ neutral losses are derived from the dihydroferuloyl-β-keto acid, as shown for compound **6** (dihydroferuloyl-β-keto acid). For compound **30** and **31**, fragments indicated with a red star are explained in the fragmentation route for compound **30** (dihydroferuloyl-β-keto acid(8-5)G + glutamate). For compound **32** [dihydroferuloyl-β-keto acid(δ-lactone β-*O*-4)dihydroferuloyl-β-keto acid], the peaks indicated with the red diamond in the MS/MS spectrum are explained in the fragmentation route. For compound **33**, fragmentation resulting into the $^{2,5}X$ fragment, is as described in Morreel et al., 2004[35]. In the same spectrum, $CO_2$ and $C_2H_4O_2$ neutral losses are derived from the dihydroferuloyl-β-keto acid, as shown for compound **6**. The $C_3H_4O_3$ neutral loss is also derived from the dihydroferuloyl-β-keto acid moiety, as explained in the fragmentation route. Compound **39** is characterized as dihydroferuloyl-β-keto acid sinapic acid cyclobutane dimer, because the peak with m/z 223.0614 co-elutes with a peak with m/z 461.1476 (see insert). The latter indicates that m/z 223.0614 is an in-source fragment of m/z 461.1476. There is no MS/MS spectrum available for m/z 461.1476, but the MS/MS of m/z 223.0614 is highly similar to that of the sinapic acid standard. In addition, the experimental m/z 461.1476 closely matches the theoretical m/z of dihydroferuloyl-β-keto acid sinapic acid cyclobutane dimer (m/z 461.1476, Δppm 5.0). The observed neutral losses of $C_2H_4O_2$, $C_3H_6O_3$, $C_4H_8O_4$ in the spectra of compound **13**, **14**, **29** are derived from the hexose cross-ring cleavages. For compound **41**, fragmentation of the 4-*O*-8 (β-aryl ether) structure resulting into $H_2O$ and $CH_2O$ neutral losses and $[G^1]^-$ and $[G^2]^-$ fragments, is as described in Morreel et al., 2004[35]. In the same MS/MS spectrum, the $C_2H_4O_2$ neutral loss is derived from the tetrahydroferuloyl-β-keto acid, as shown for compound **22** (tetrahydroferuloyl-β-

keto acid). Compound **42**-**49** matched retention time, m/z and MS/MS spectra as described in De Meester et al., 2018[36].

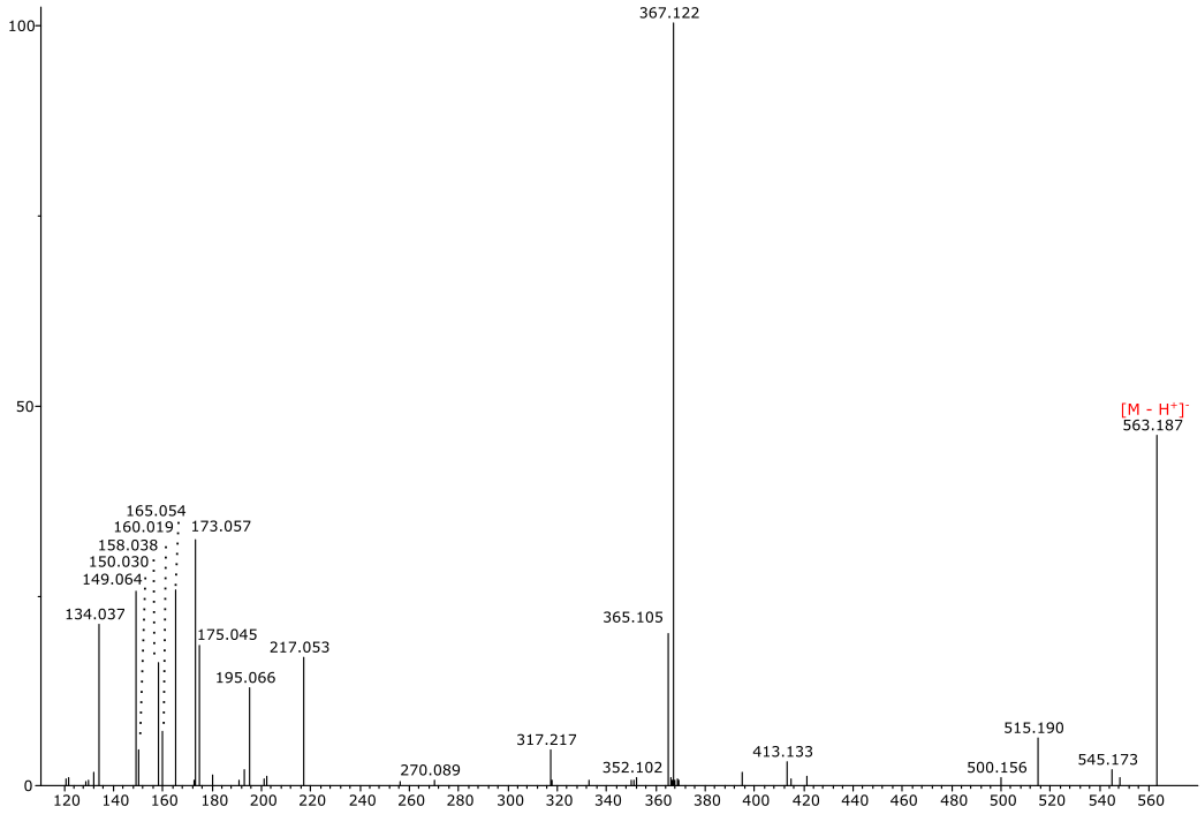
1. MS/MS curcumin (enol) (26.46 min, m/z 367.1202)



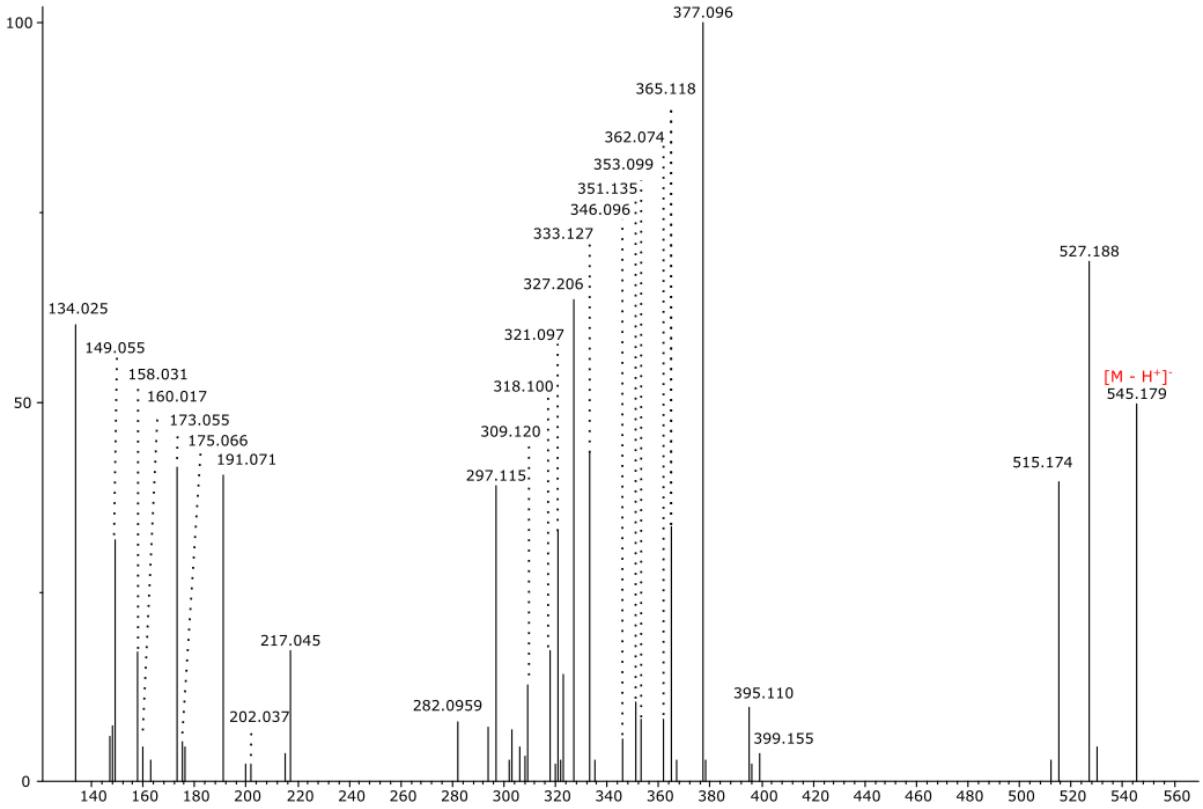
2. MS/MS curcumin(4-*O*-8)G (enol) 1 (23.83 min, m/z 563.1925)



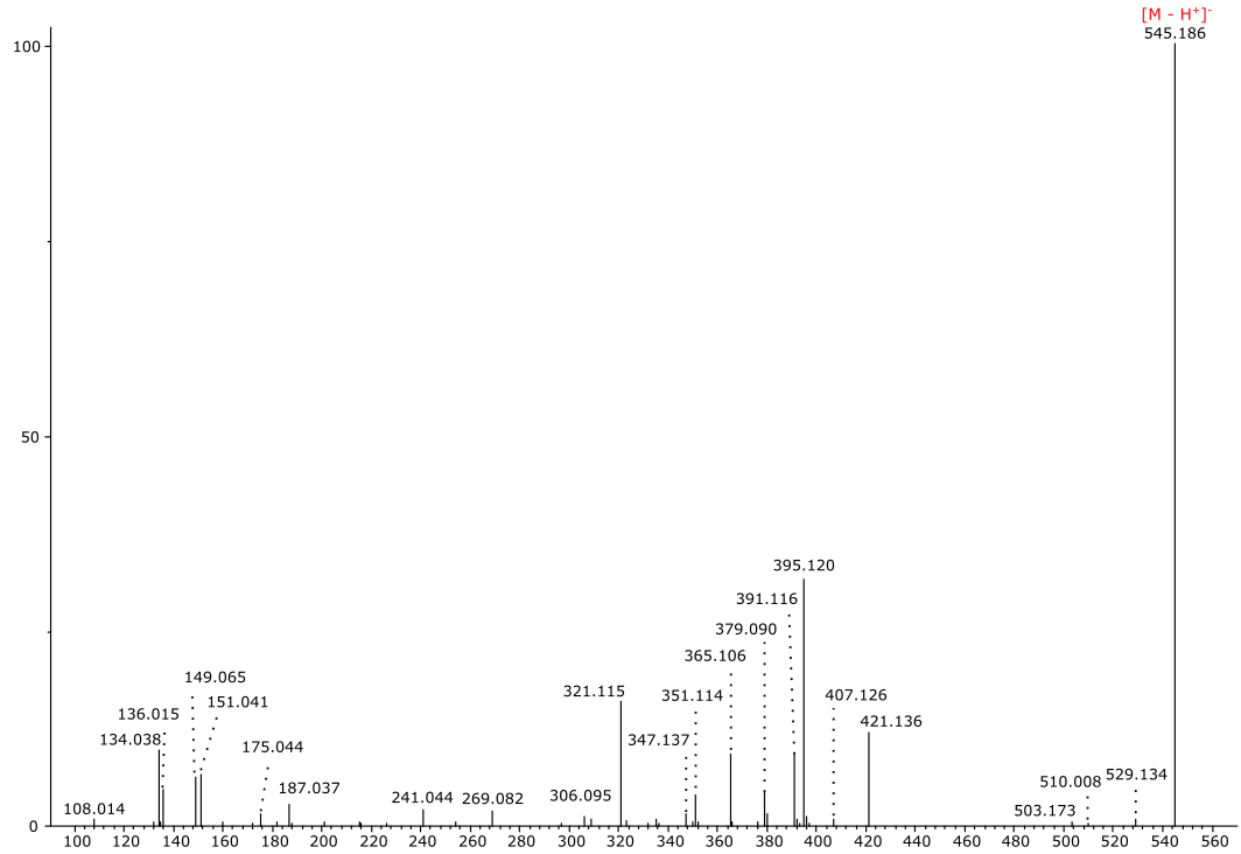

3. MS/MS curcumin(4-*O*-8)G (enol) 2 (23.93 min, m/z 563.1928)



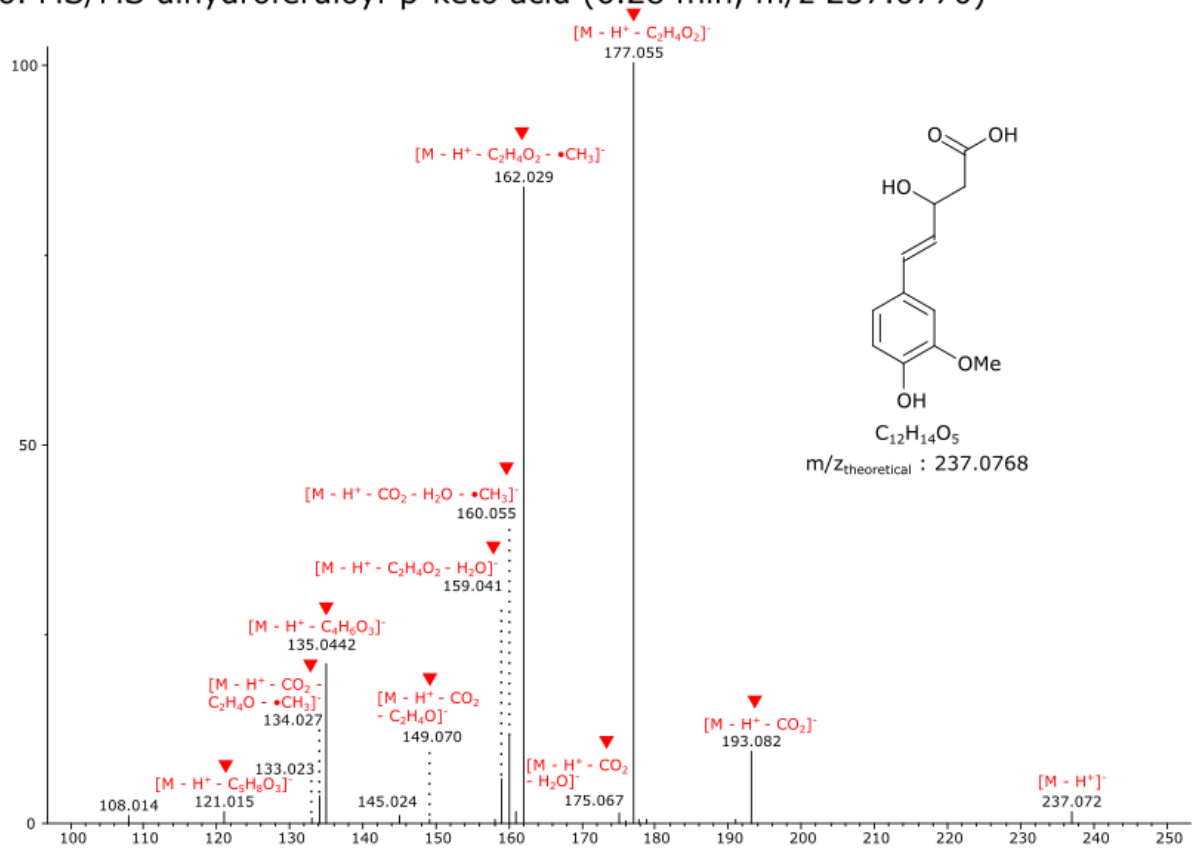
4. MS/MS curcumin(5-8)G (enol) (27.55 min, m/z 545.1822)



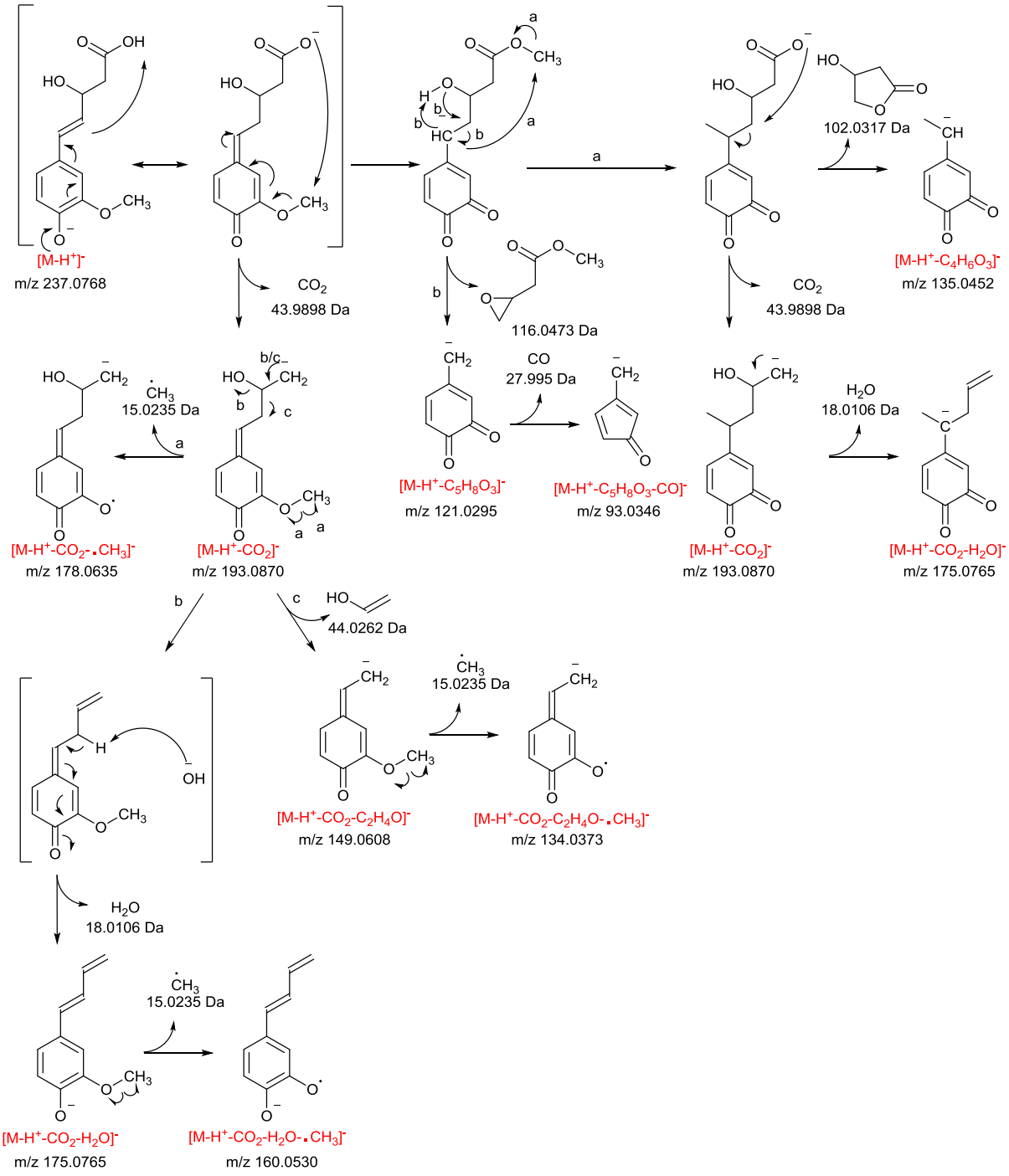

5. MS/MS curcumin(8-8)G (enol ether) (25.09 min, m/z 545.1828)



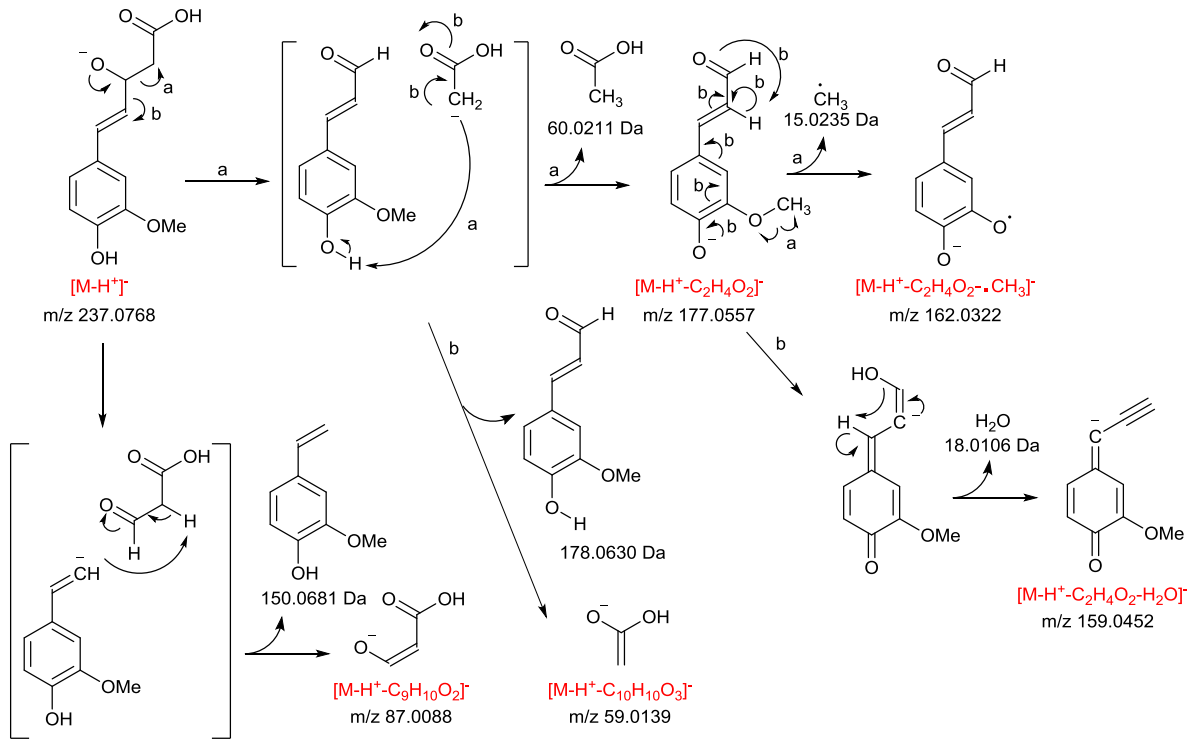
6. MS/MS dihydroferuloyl-β-keto acid (6.28 min, m/z 237.0770)



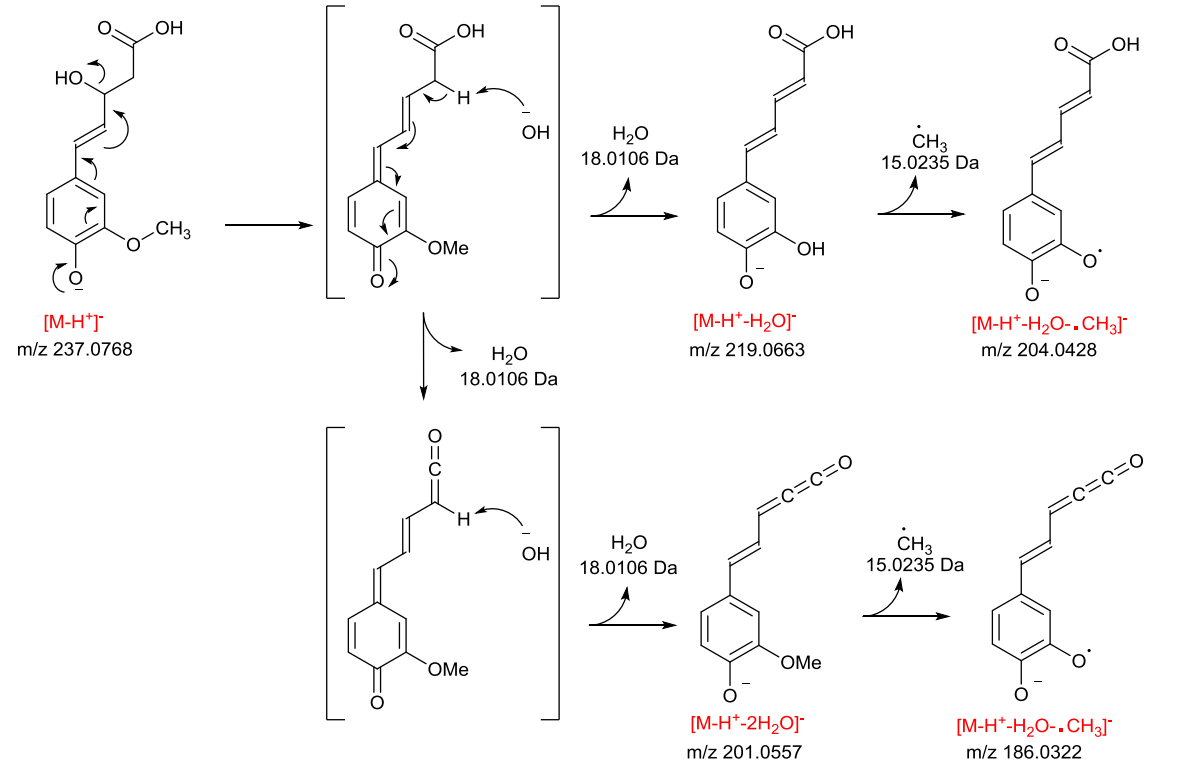
dihydroferuloyl-β-keto acid fragmentation pathway 1



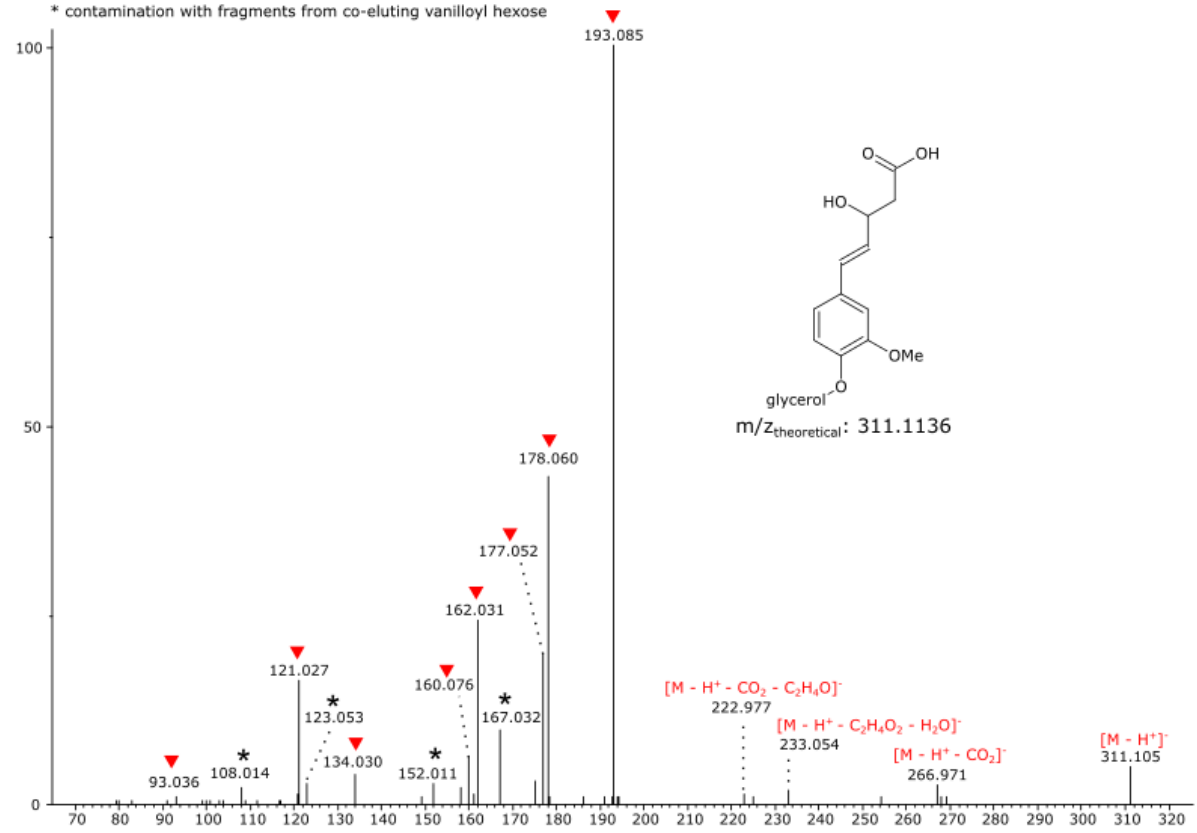
dihydroferuloyl-β-keto acid fragmentation pathway 2



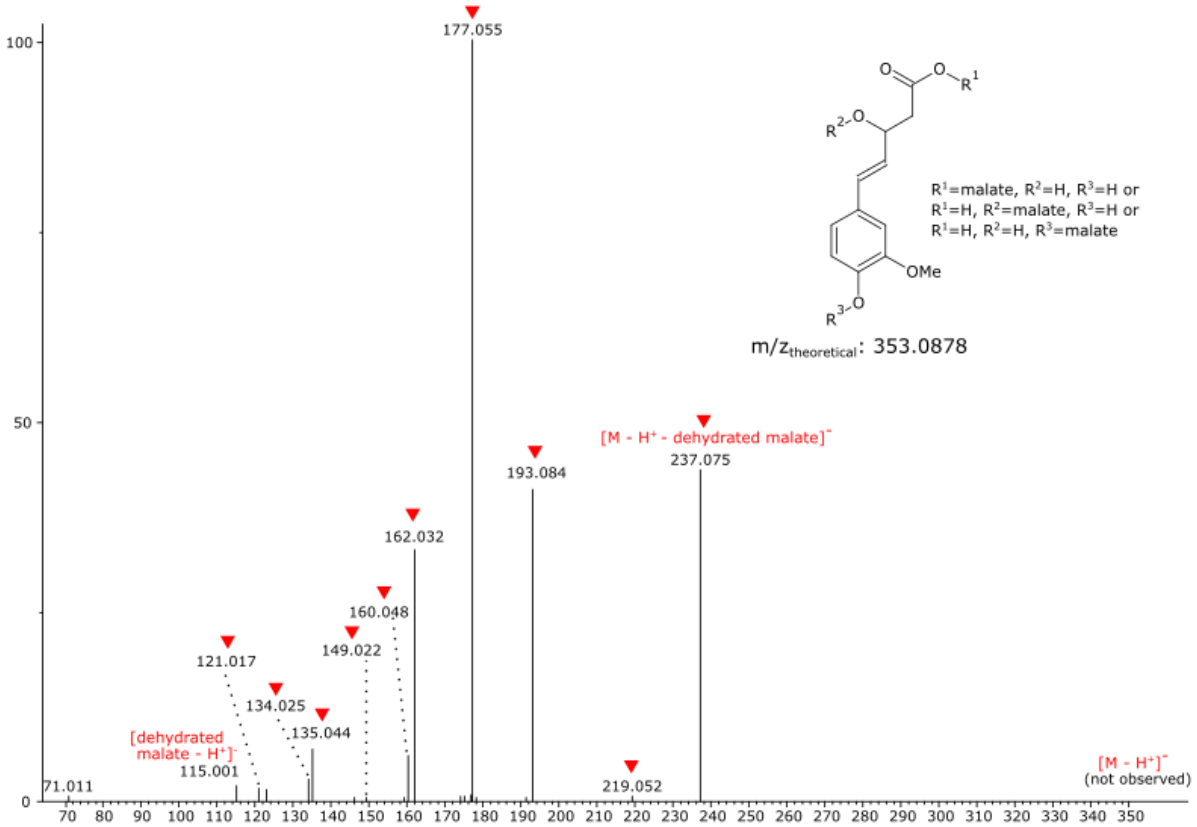

dihydroferuloyl-β-keto acid fragmentation pathway 3



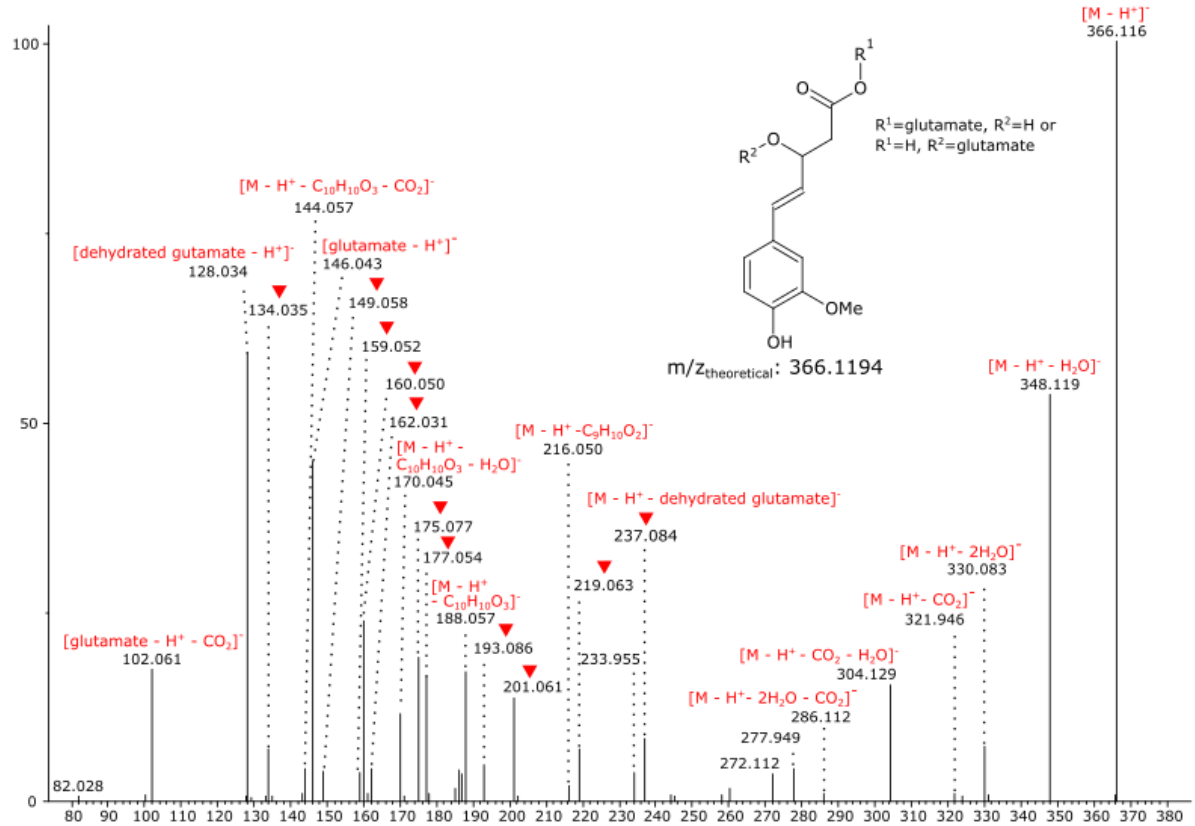
7. MS/MS dihydroferuloyl-β-keto acid-4-*O*-glycerol (4.90 min, m/z 311.1148)



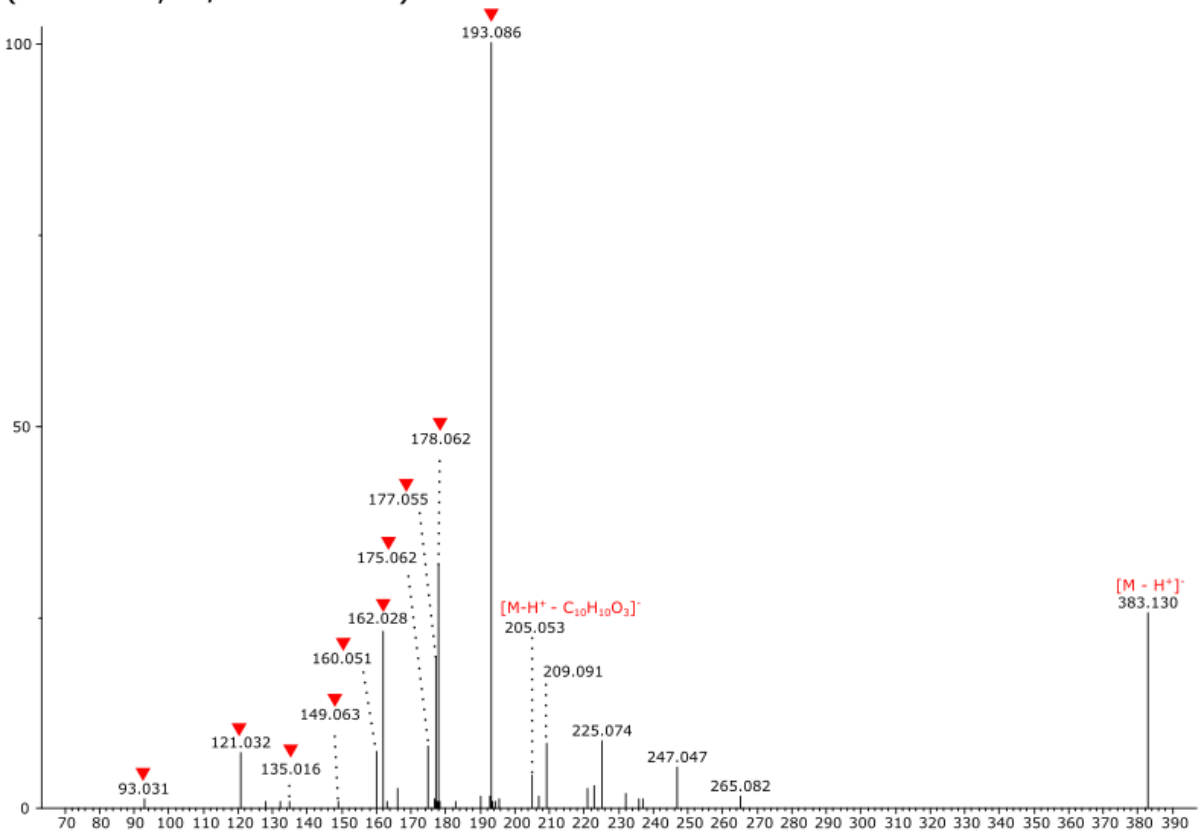

8. MS/MS dihydroferuloyl-β-keto acid + malate (7.00 min, m/z 353.0894)



9. MS/MS dihydroferuloyl-β-keto acid + glutamate (4.80 min, m/z 366.1210)

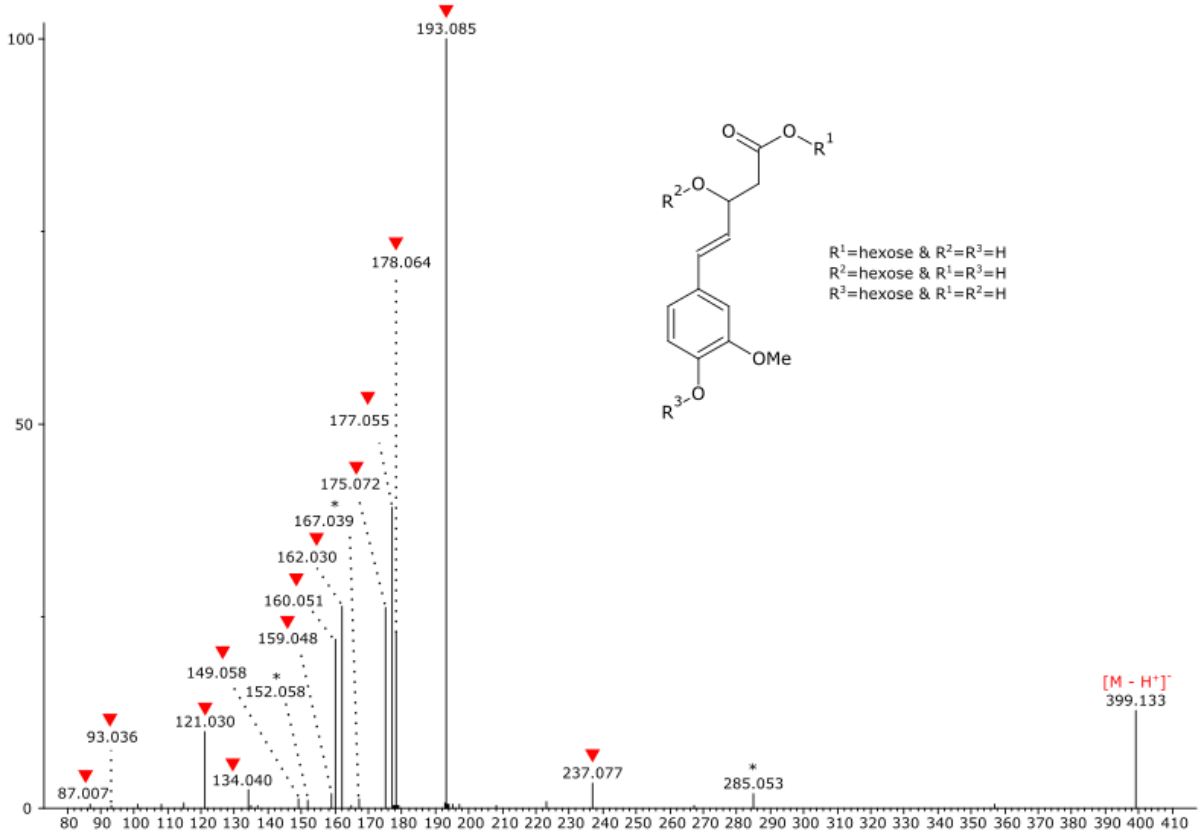



10. MS/MS dihydroferuloyl-β-keto acid + deoxyhexose (5.39 min, m/z 383.1365)

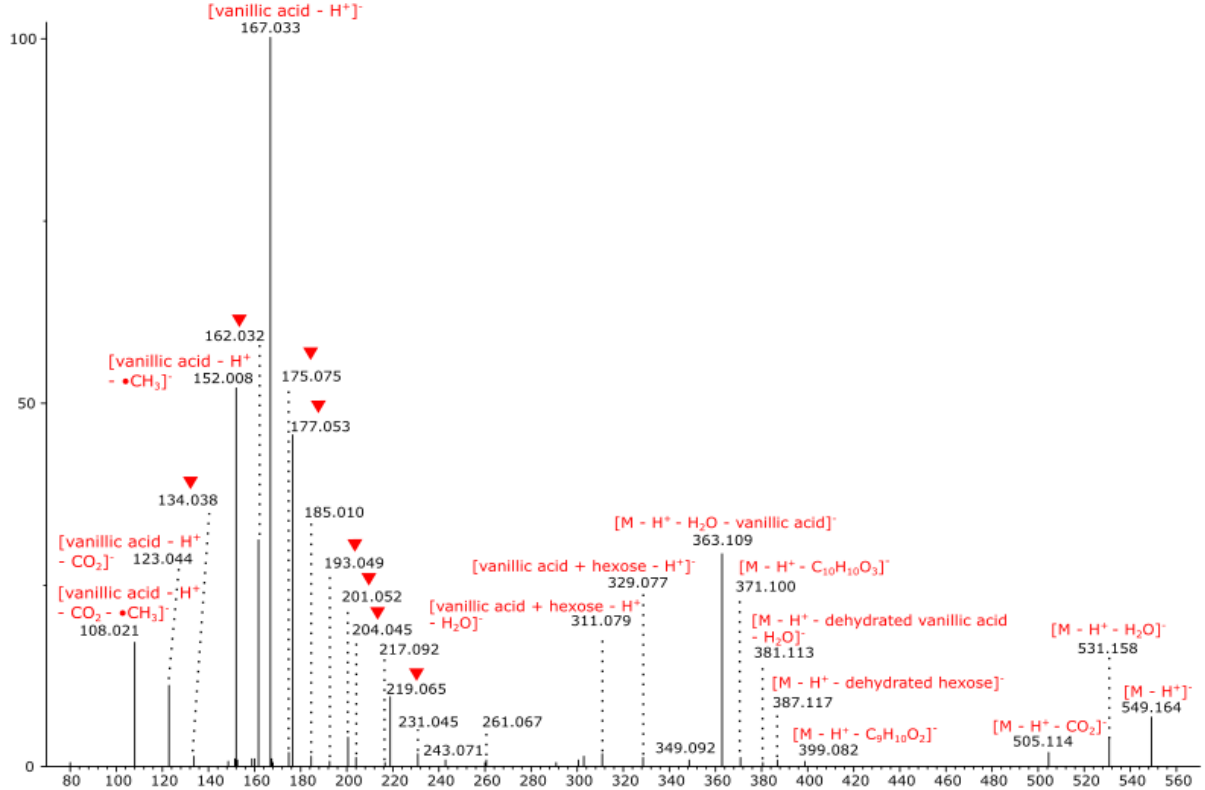


## 11. MS/MS dihydroferuloyl-β-keto acid + hexose (3.15 min, m/z 399.1306)

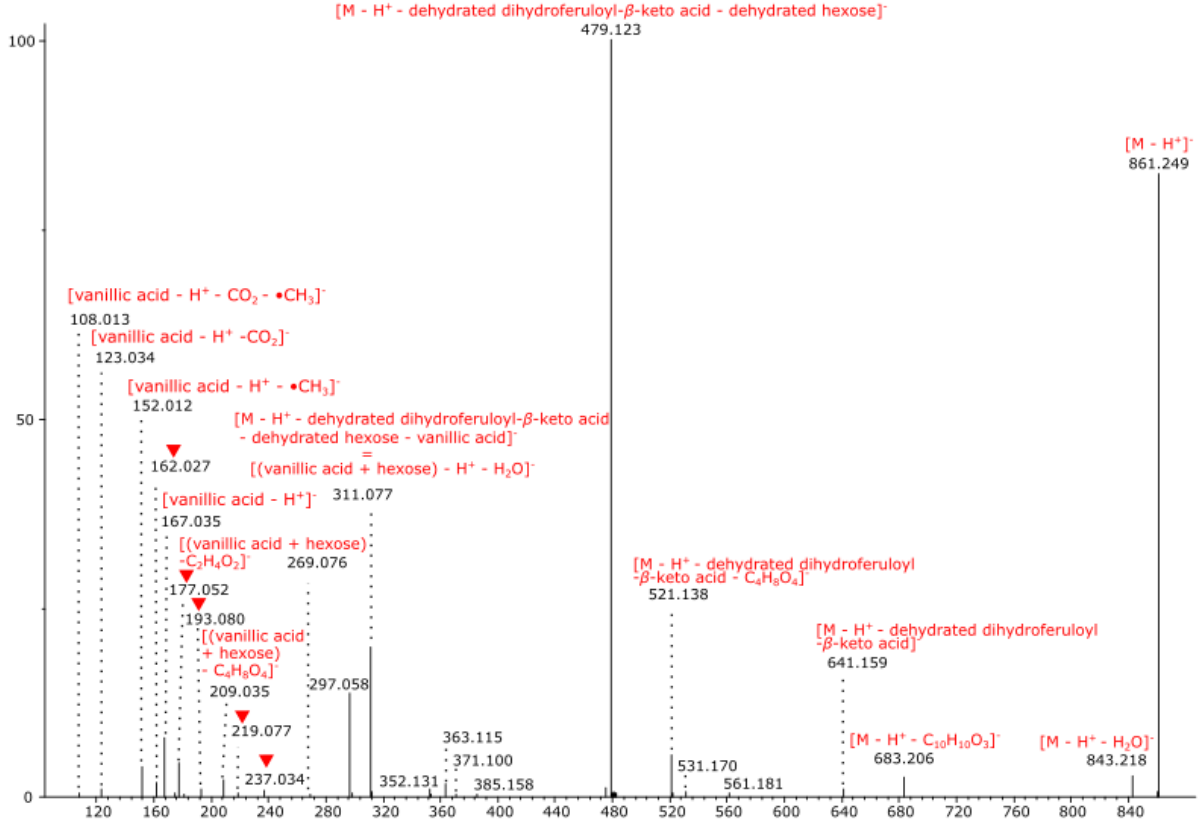
* contamination with fragments from co-eluting dihydroxybenzoic acid pentoside



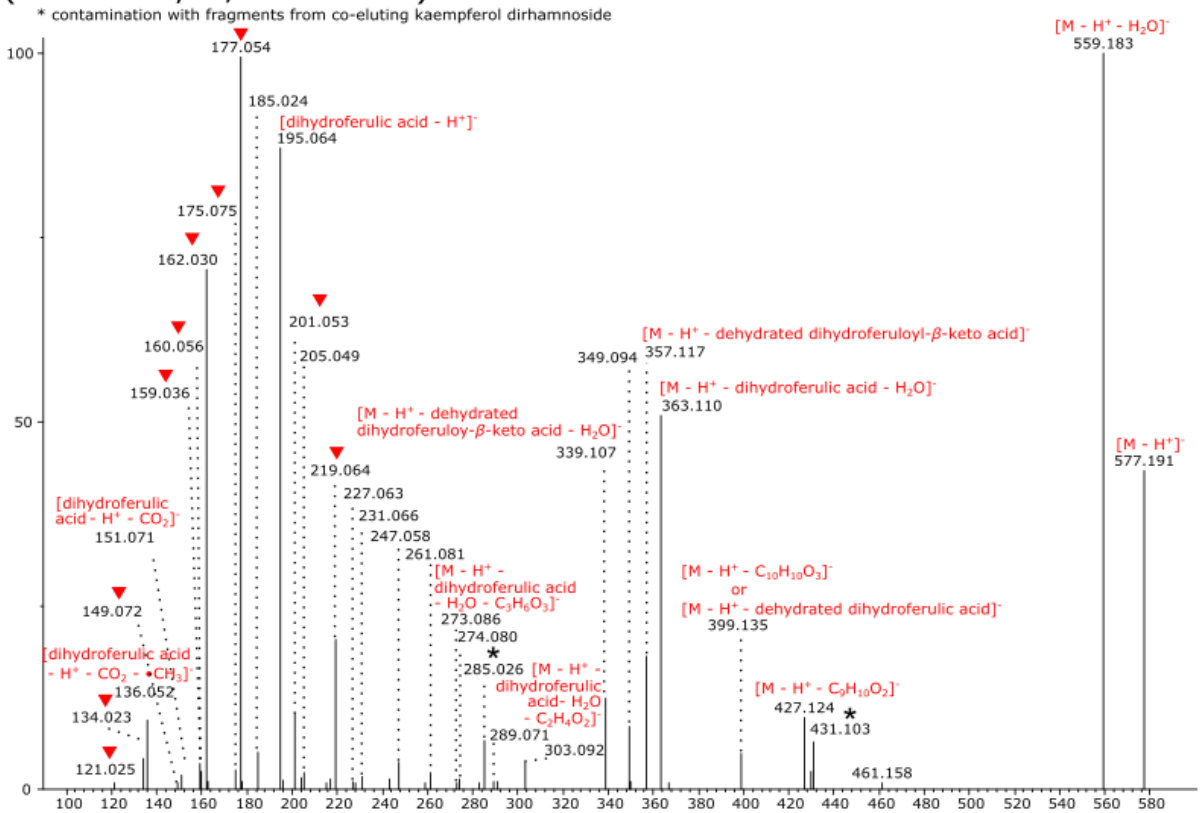

## 12. MS/MS dihydroferuloyl-β-keto acid + vanillic acid + hexose (9.72 min, m/z 549.1615)



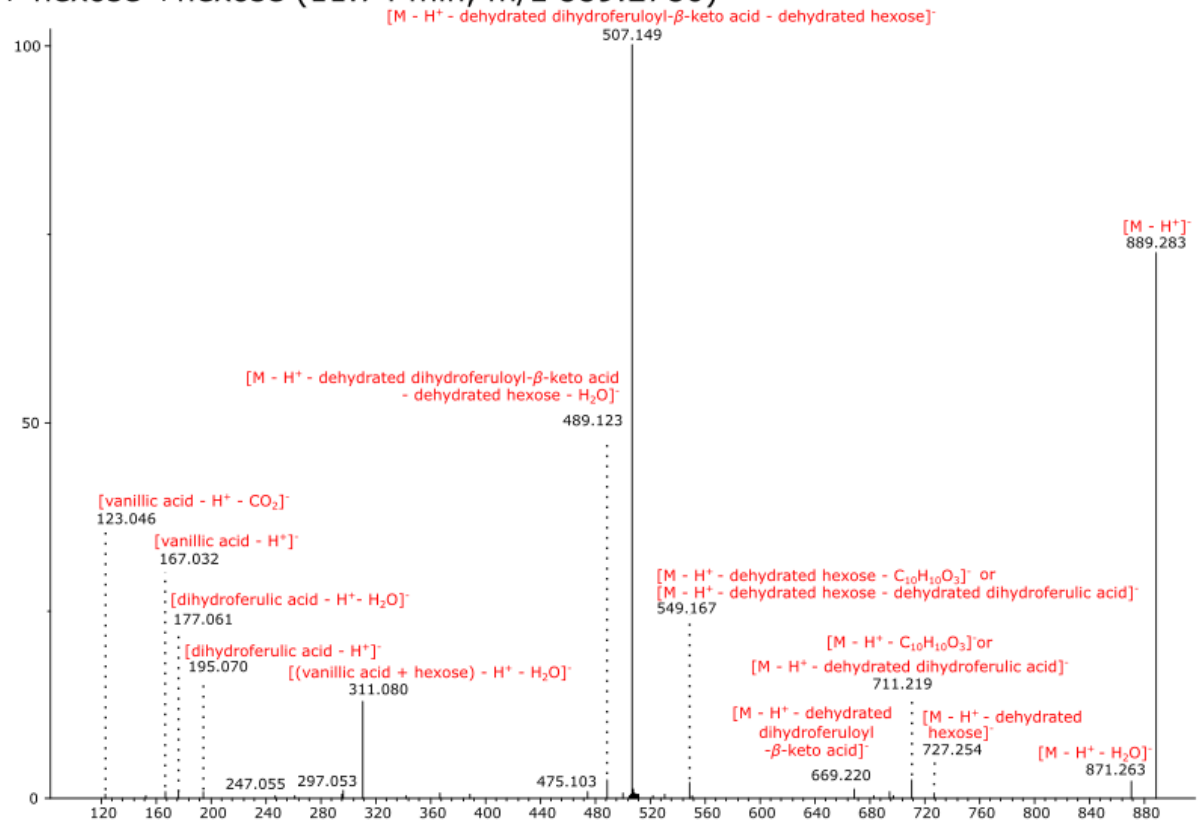
13. MS/MS dihydroferuloyl-β-keto acid + vanillic acid + vanillic acid + hexose + hexose (11.13 min, m/z 861.2464)



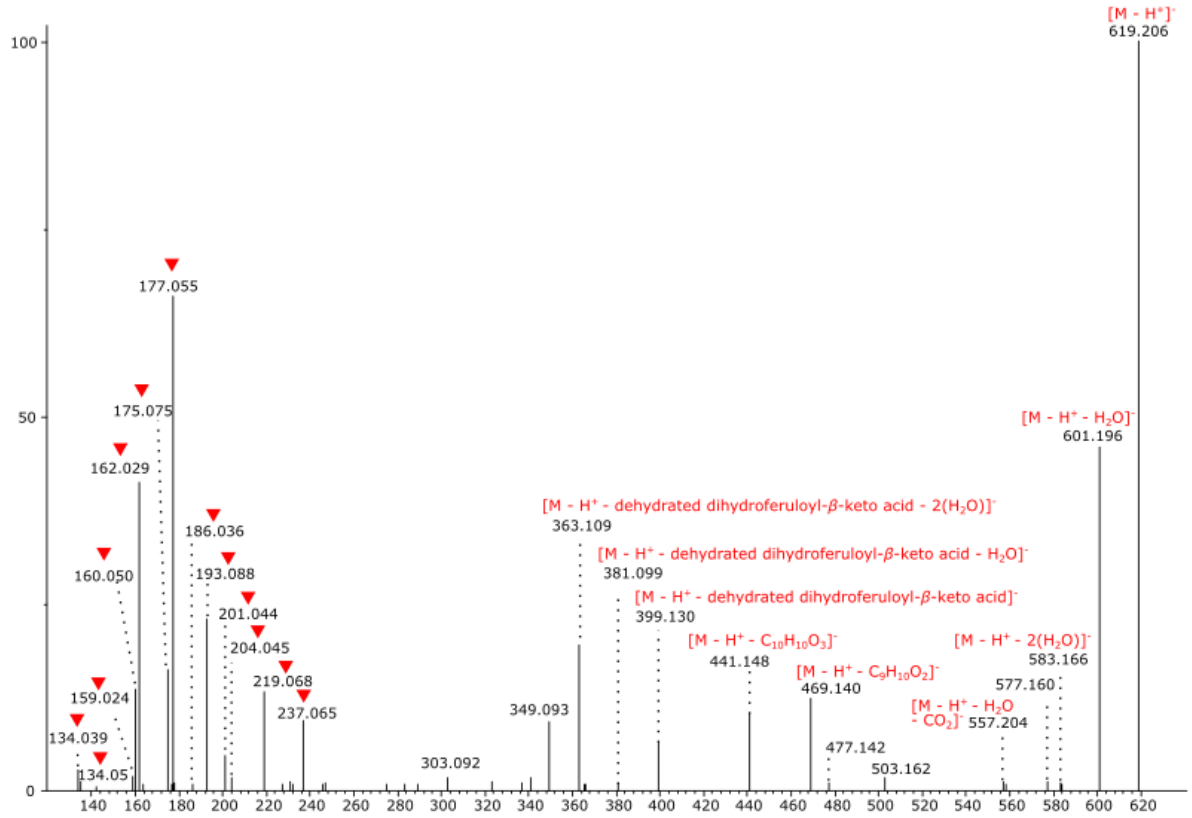

14. MS/MS dihydroferuloyl-β-keto acid + dihydroferulic acid + hexose (11.87 min, m/z 577.1909)



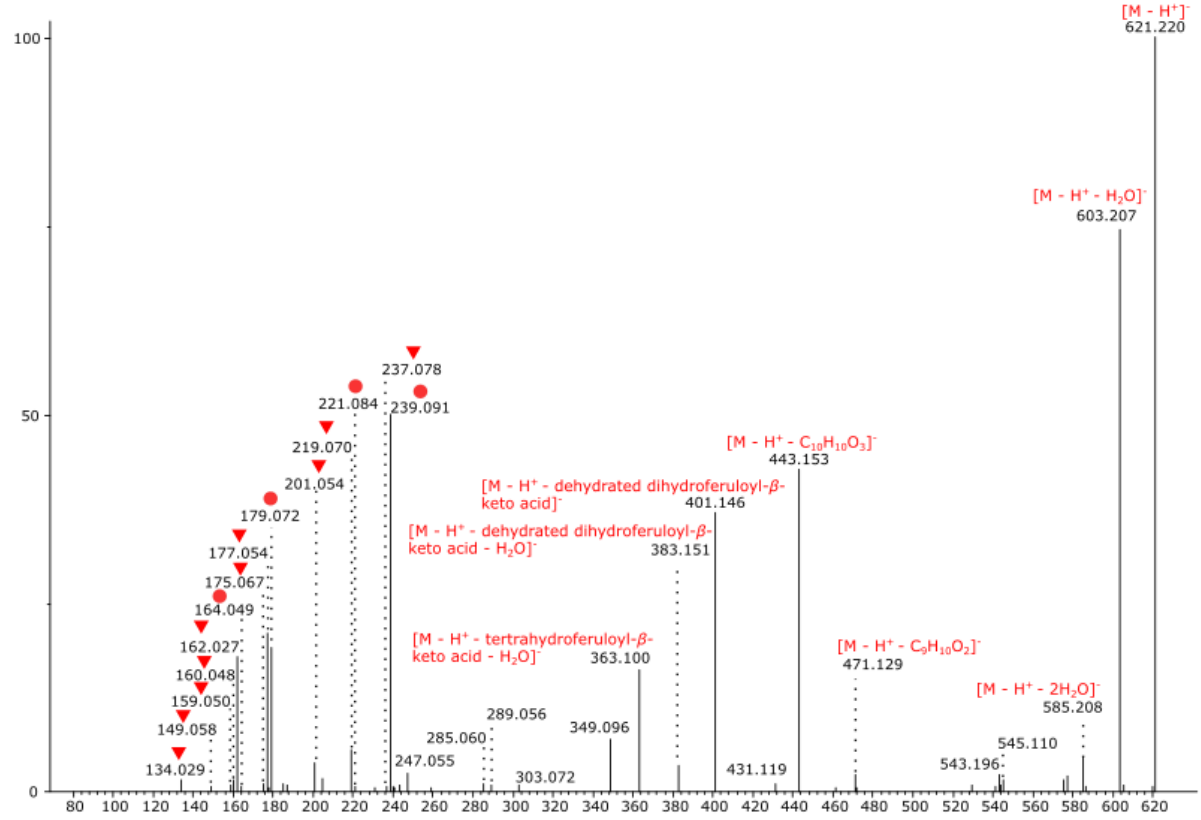
15. MS/MS dihydroferuloyl-β-keto acid + dihydroferulic acid + vanillic acid + hexose +hexose (11.74 min, m/z 889.2780)



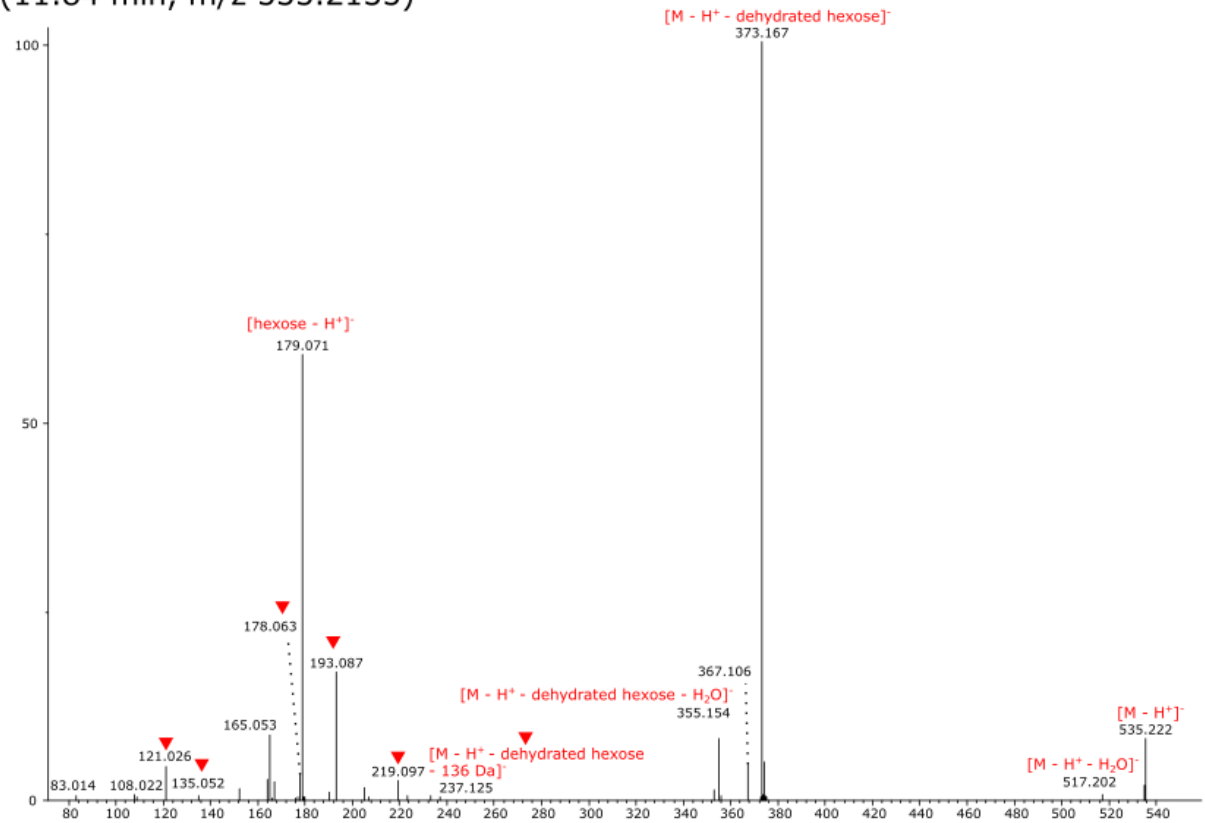

16. MS/MS dihydroferuloyl-β-keto acid + dihydroferuloyl-β-keto acid + hexose (9.96 min, m/z 619.2030)



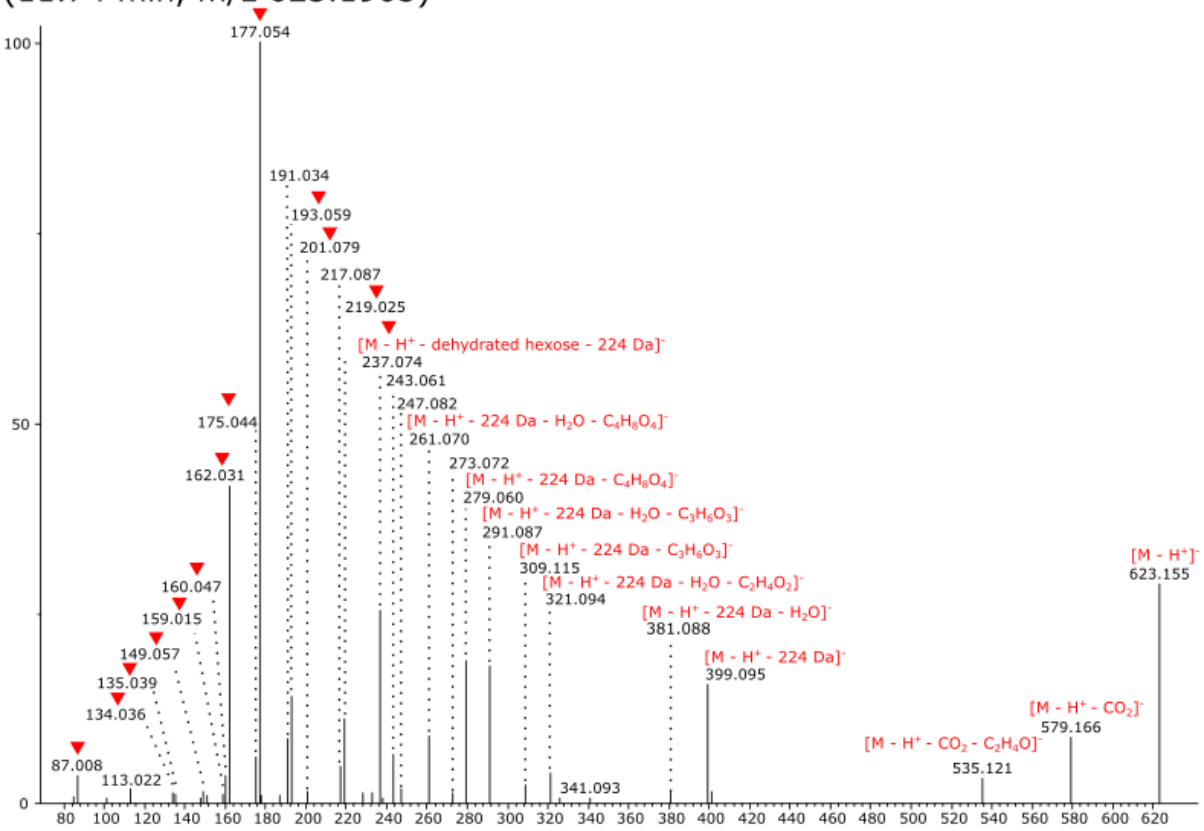
17. MS/MS dihydroferuloyl-β-keto acid + tetrahydroferuloyl-β-keto acid + hexose (10.76 min, m/z 621.2186)



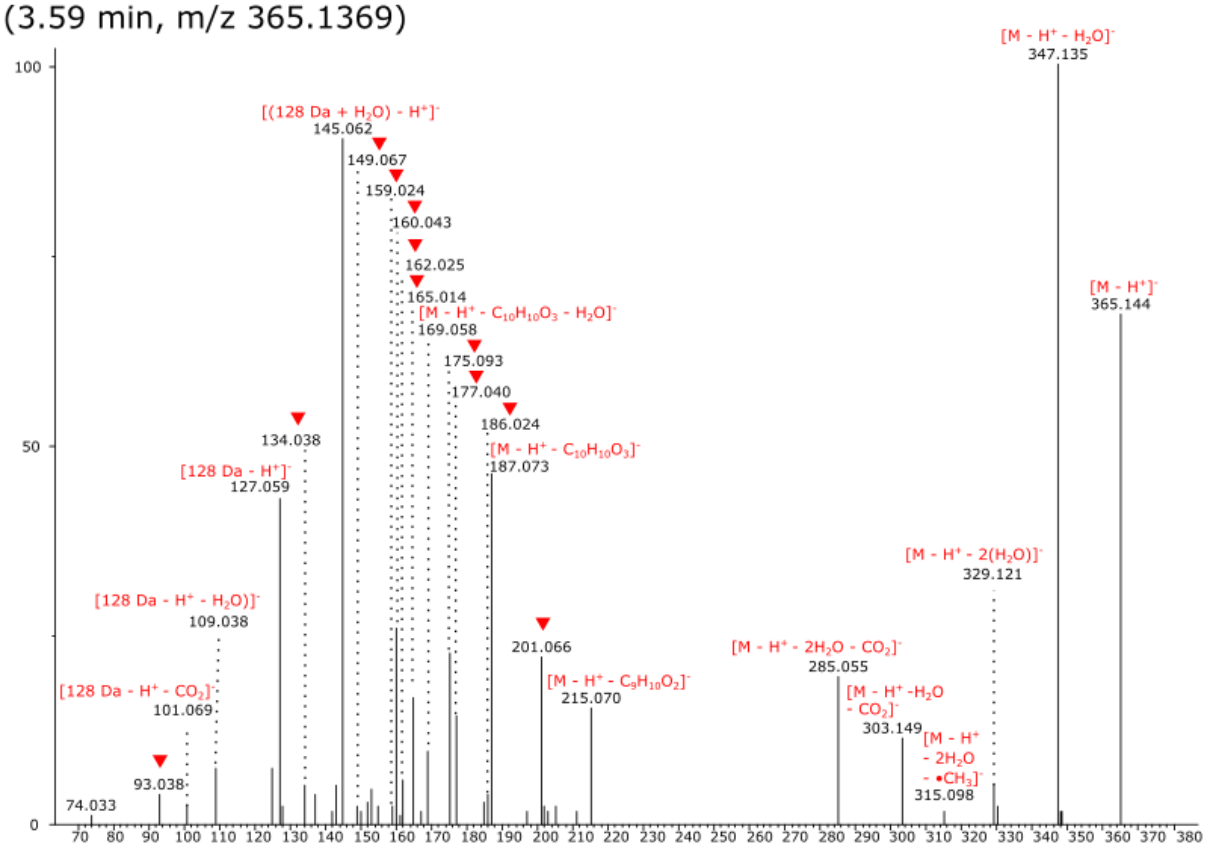

18. MS/MS dihydroferuloyl-β-keto acid + hexose + 136 Da (11.84 min, m/z 535.2135)



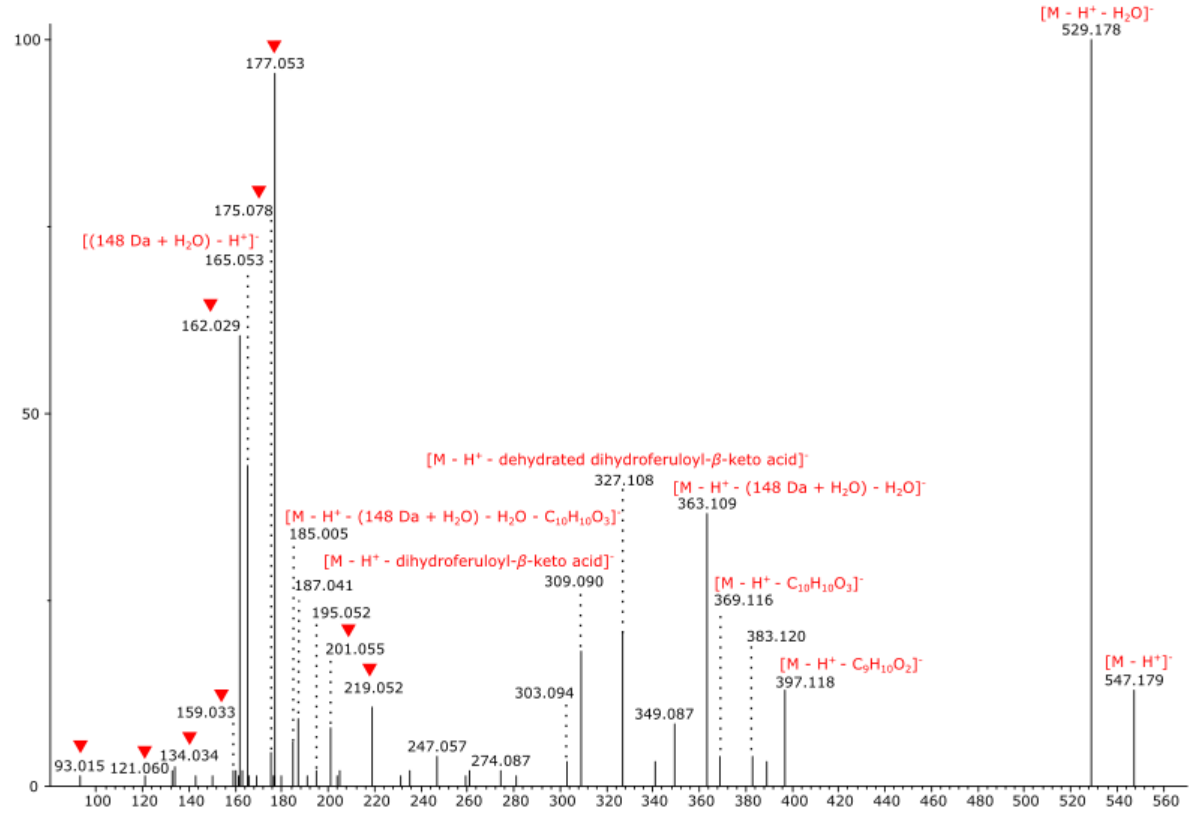
19. MS/MS dihydroferuloyl-β-keto acid + hexose + 224 Da (11.74 min, m/z 623.1903)



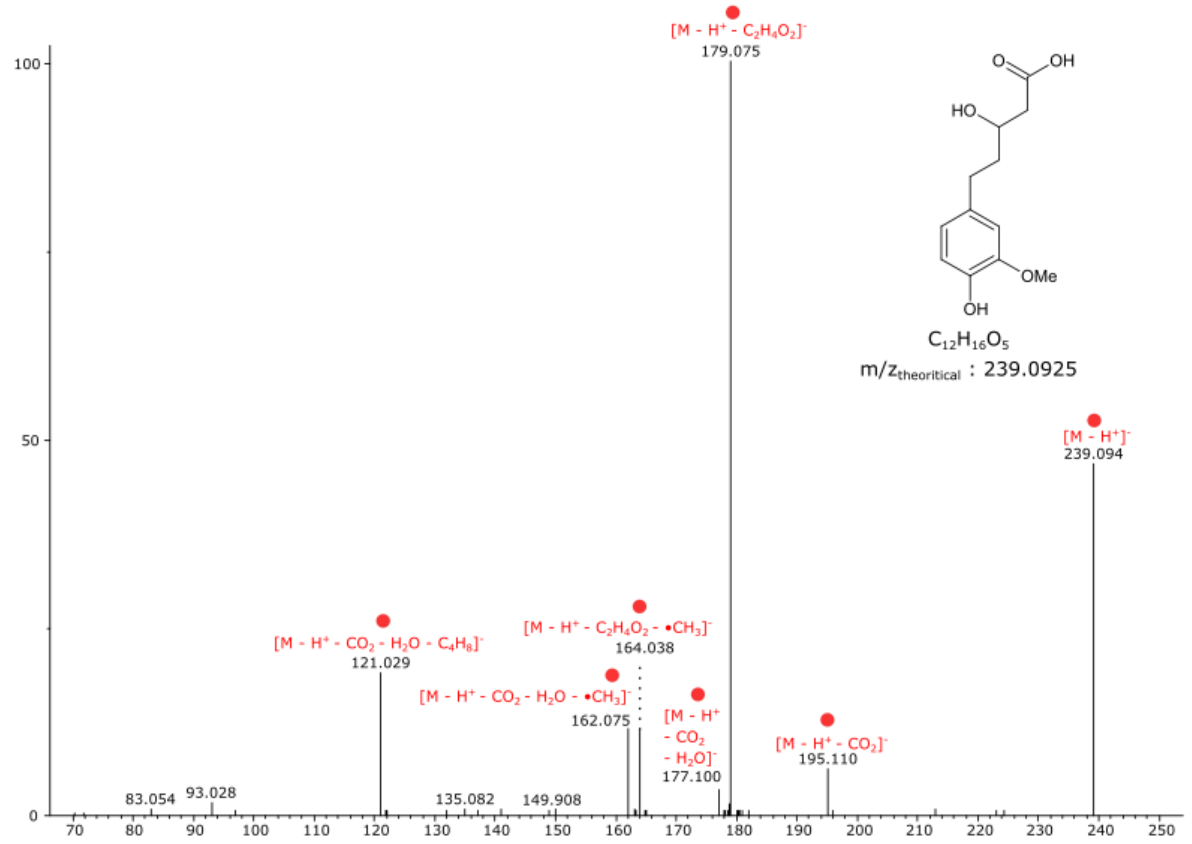

20. MS/MS dihydroferuloyl-β-keto acid + 128 Da (3.59 min, m/z 365.1369)



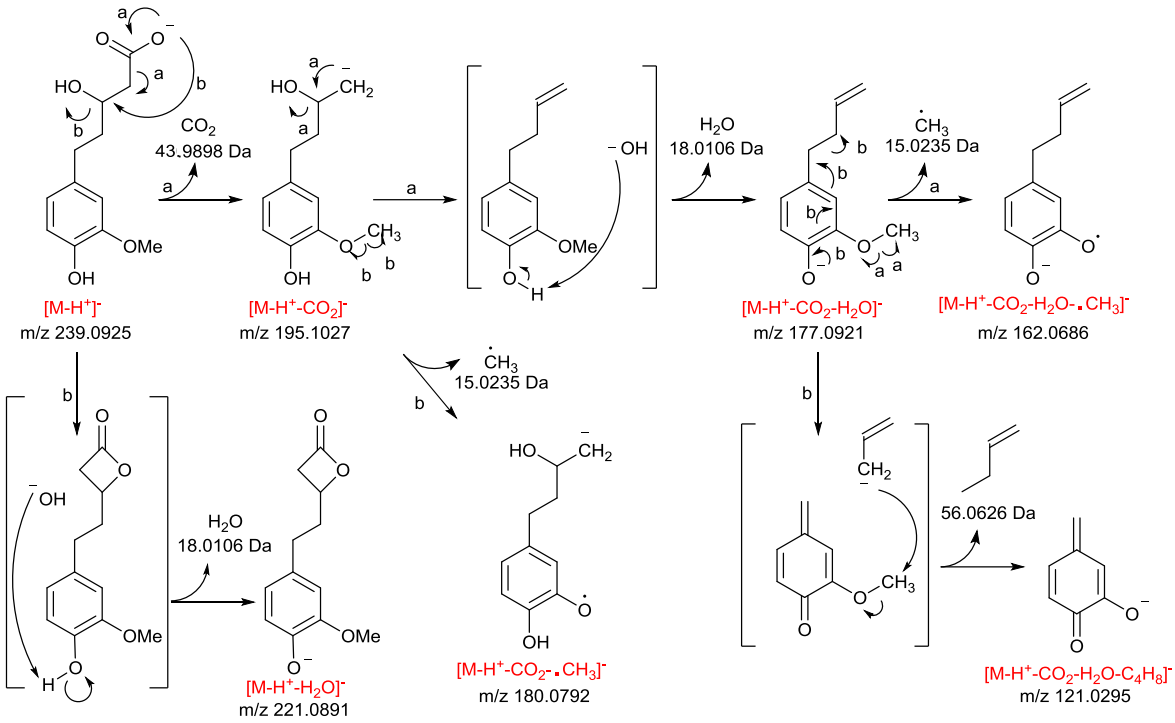
21. MS/MS dihydroferuloyl-β-keto acid + hexose + 148 Da (11.76 min, m/z 547.1825)



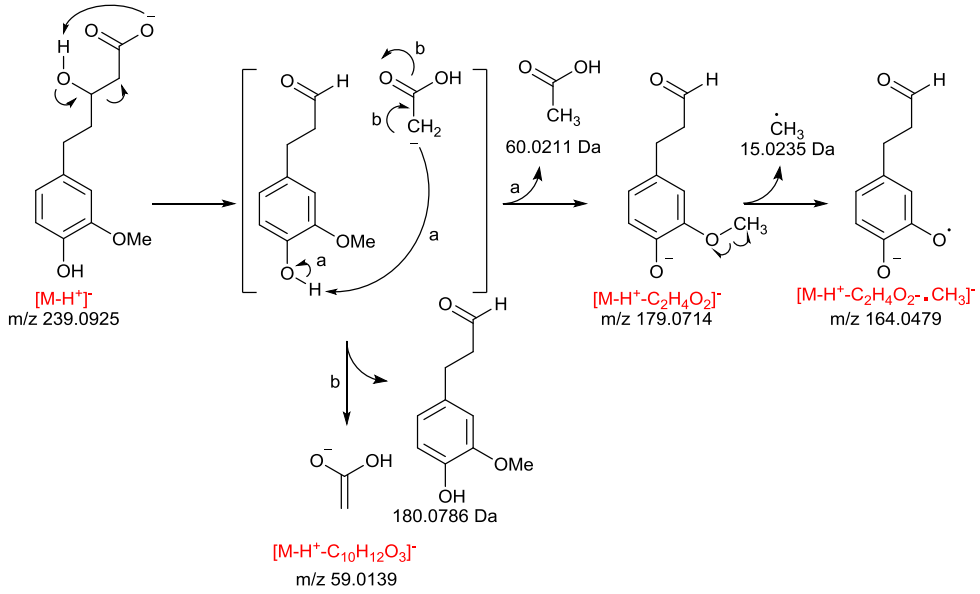

22. MS/MS tetrahydroferuloyl-β-keto acid (7.00 min, m/z 239.0930)



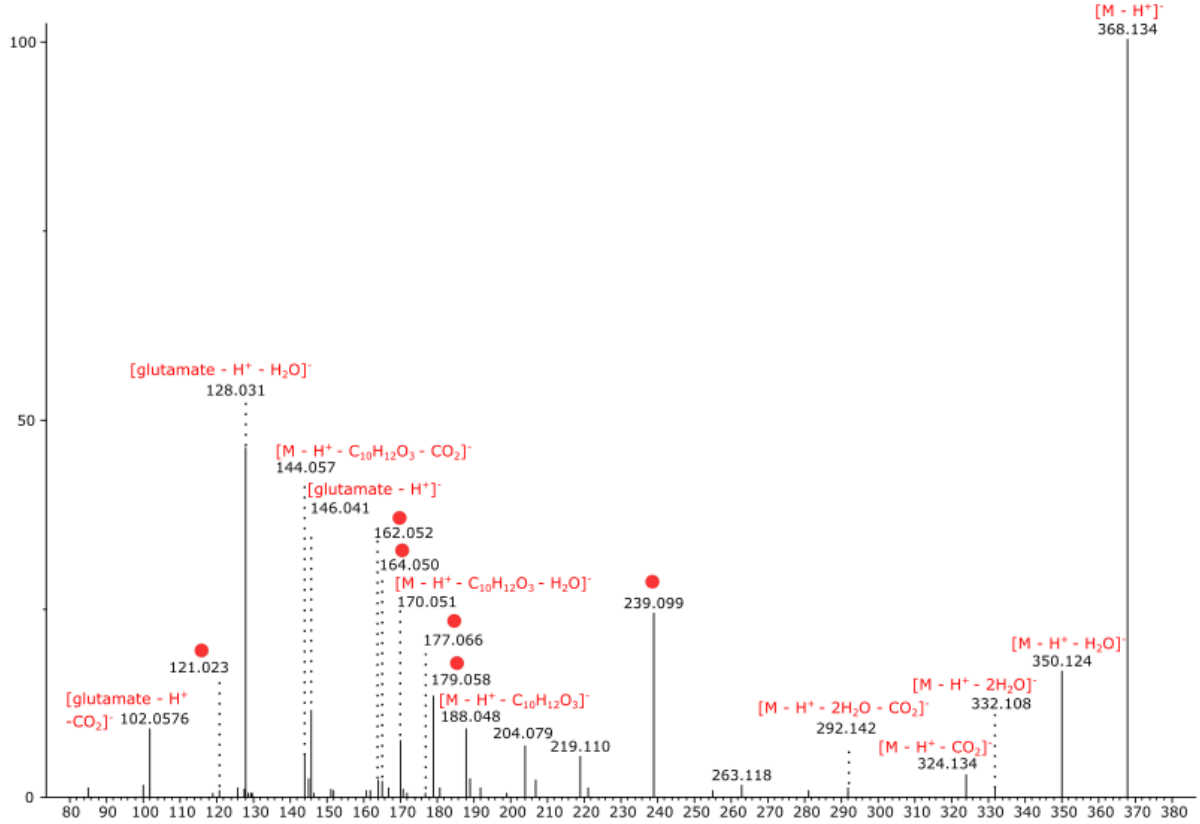
tetrahydroferuloyl-β-keto acid fragmentation pathway 1



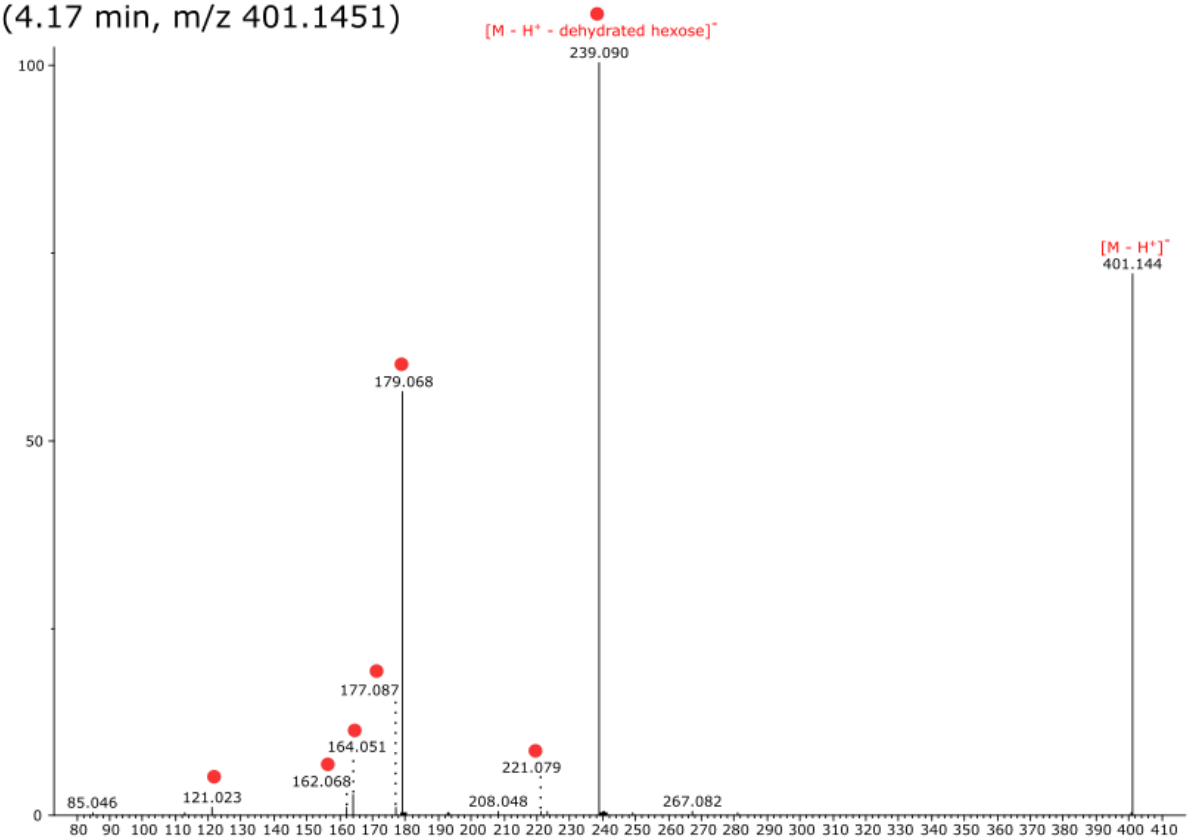

tetrahydroferuloyl-β-keto acid fragmentation pathway 2



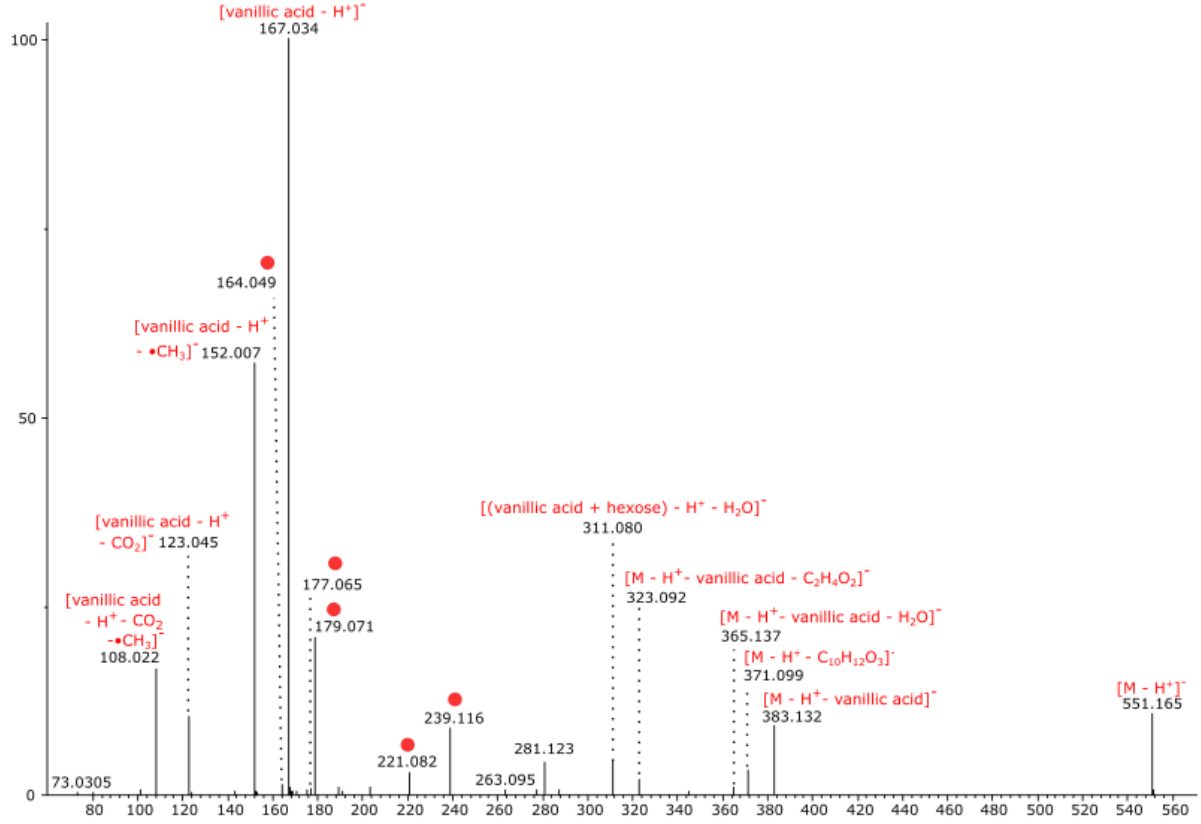
23. MS/MS tetrahydroferuloyl-β-keto acid + glutamate
(6.07 min, m/z 368.1367)



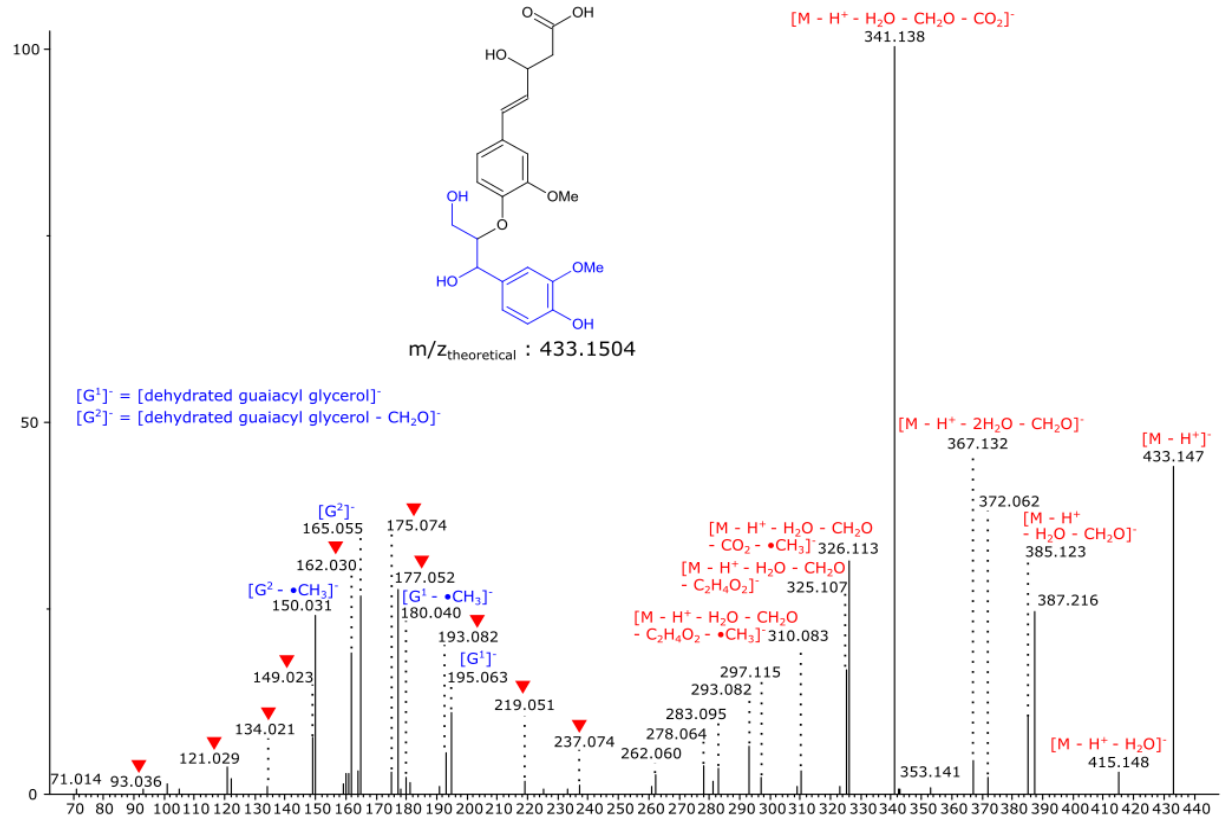

24. MS/MS tetrahydroferuloyl-β-keto acid + hexose
(4.17 min, m/z 401.1451)



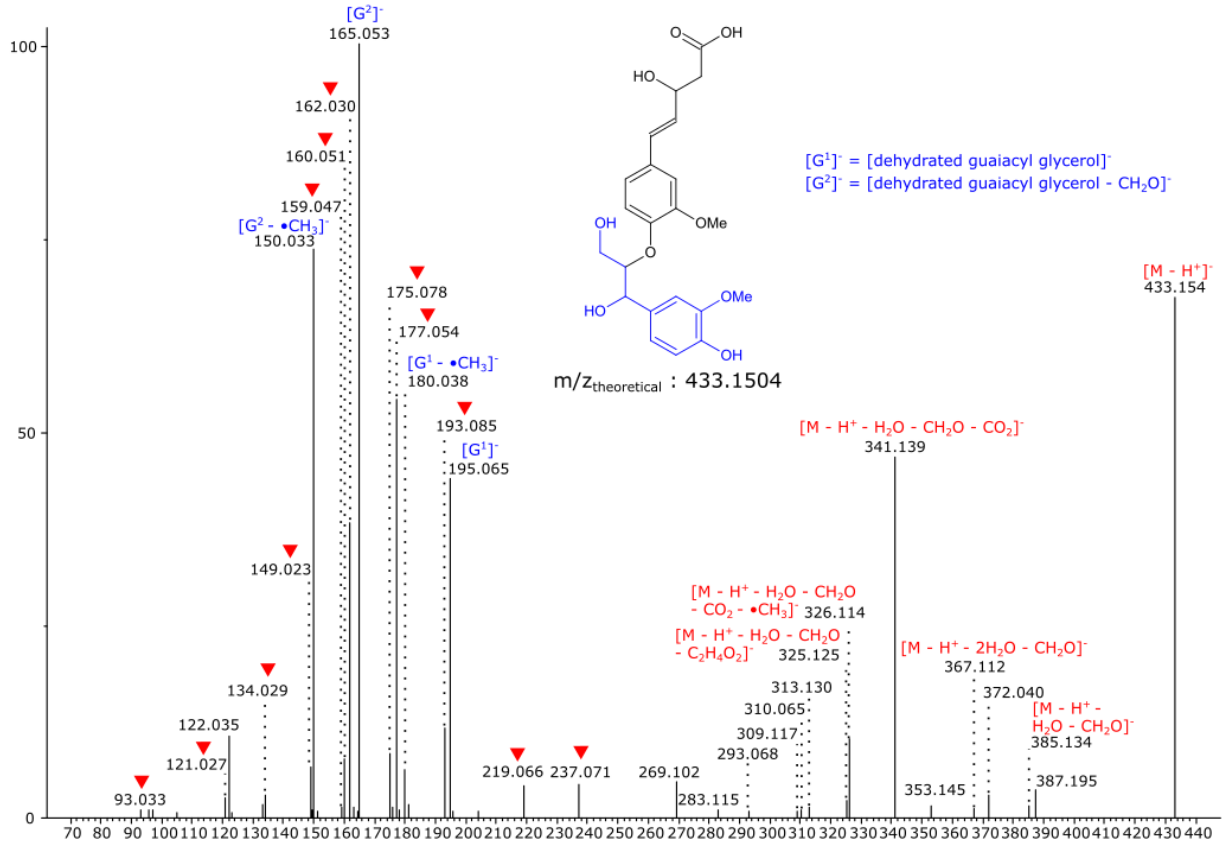
25. MS/MS tetrahydroferuloyl-β-keto acid + vanillic acid + hexose (10.22 min, m/z 551.1775)



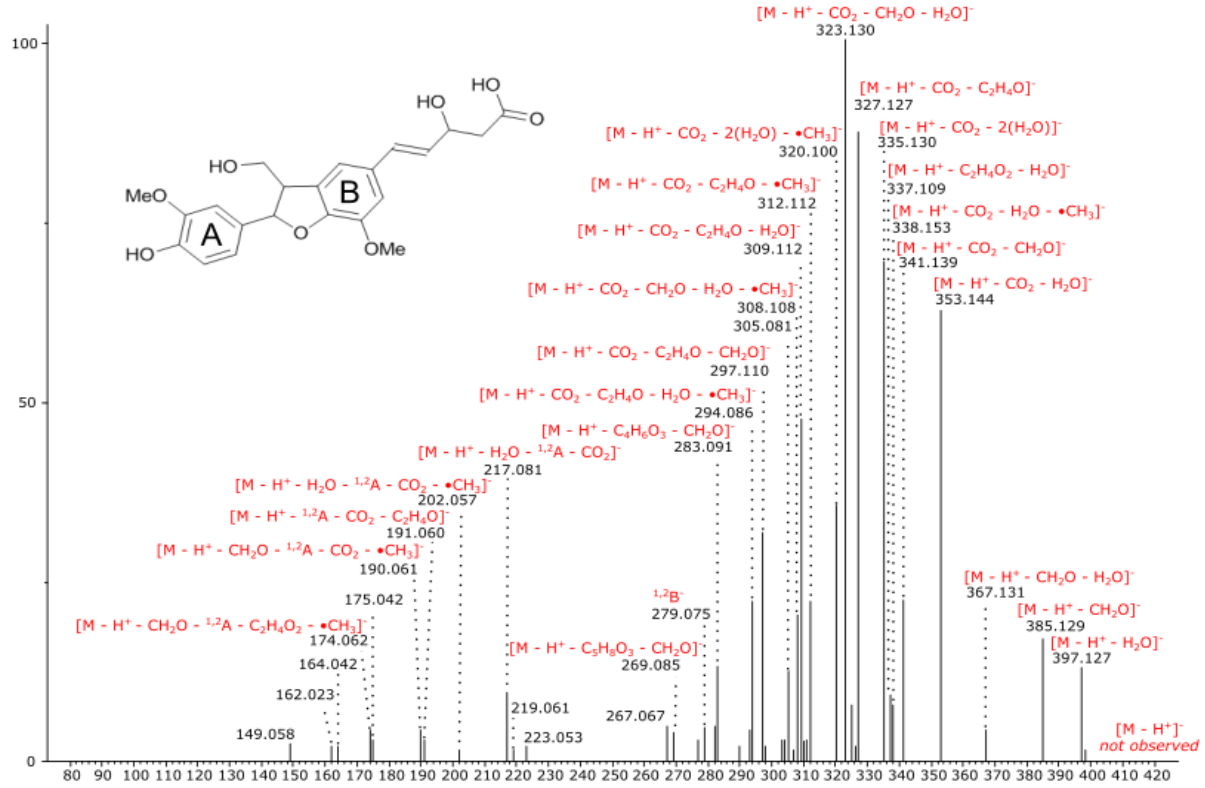

26. MS/MS dihydroferuloyl-β-keto acid(4-*O*-8)G 1 (7.80 min, m/z 433.1521)



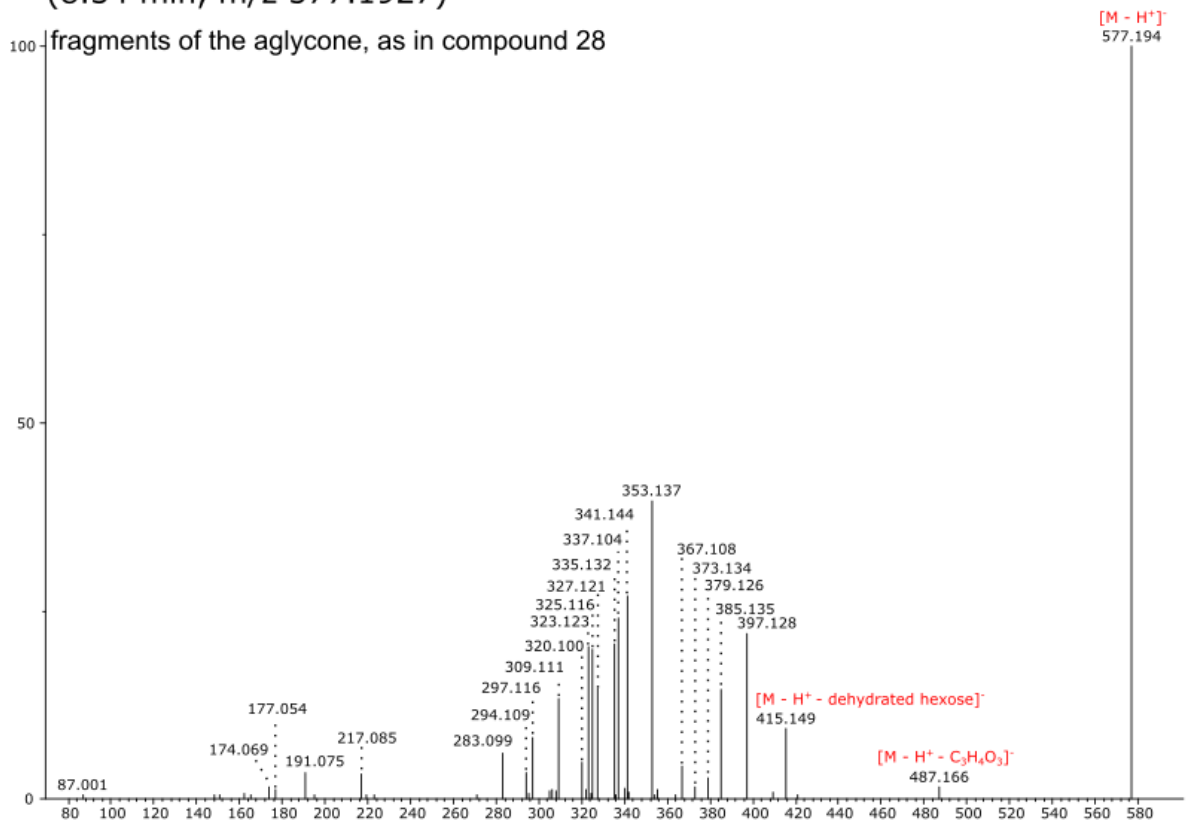
27. MS/MS dihydroferuloyl-β-keto acid(4-*O*-8)G 2 (8.13 min, m/z 433.1544)



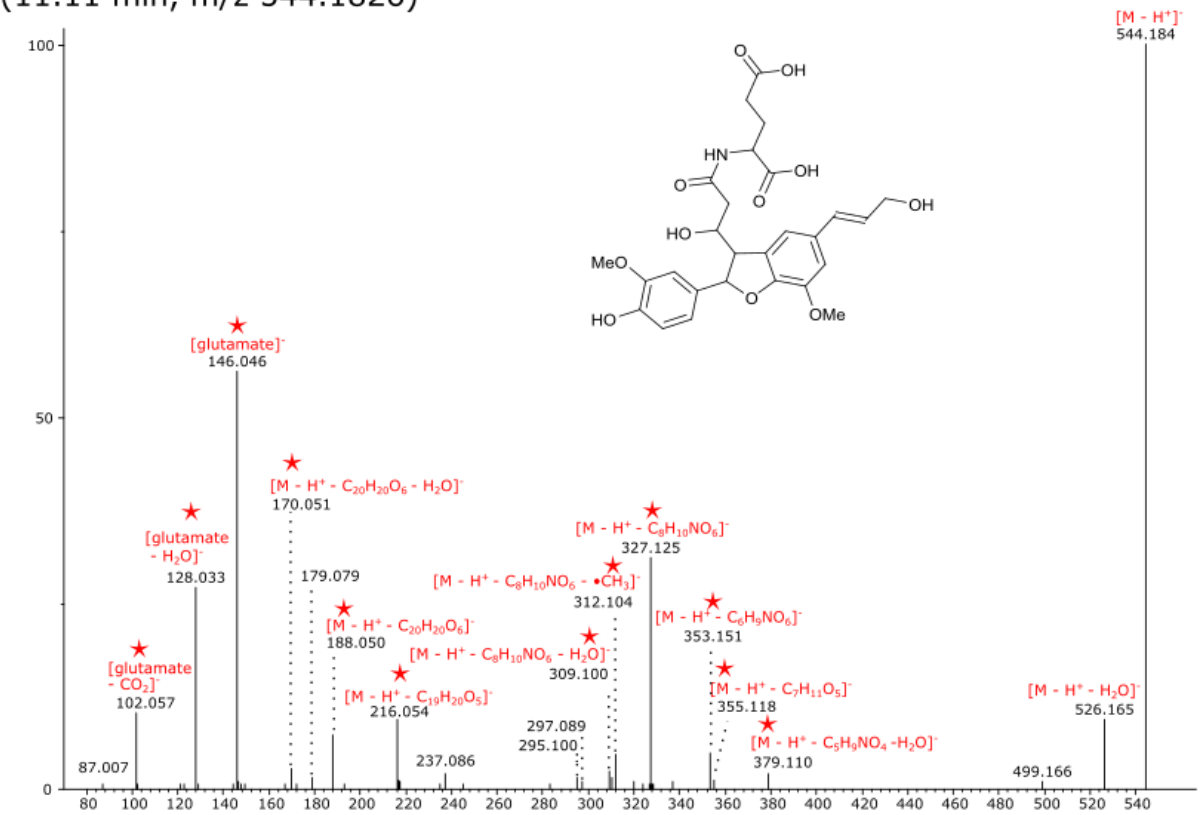

28. MS/MS dihydroferuloyl-β-keto acid(5-8)G (12.0 min, m/z 415.1418)

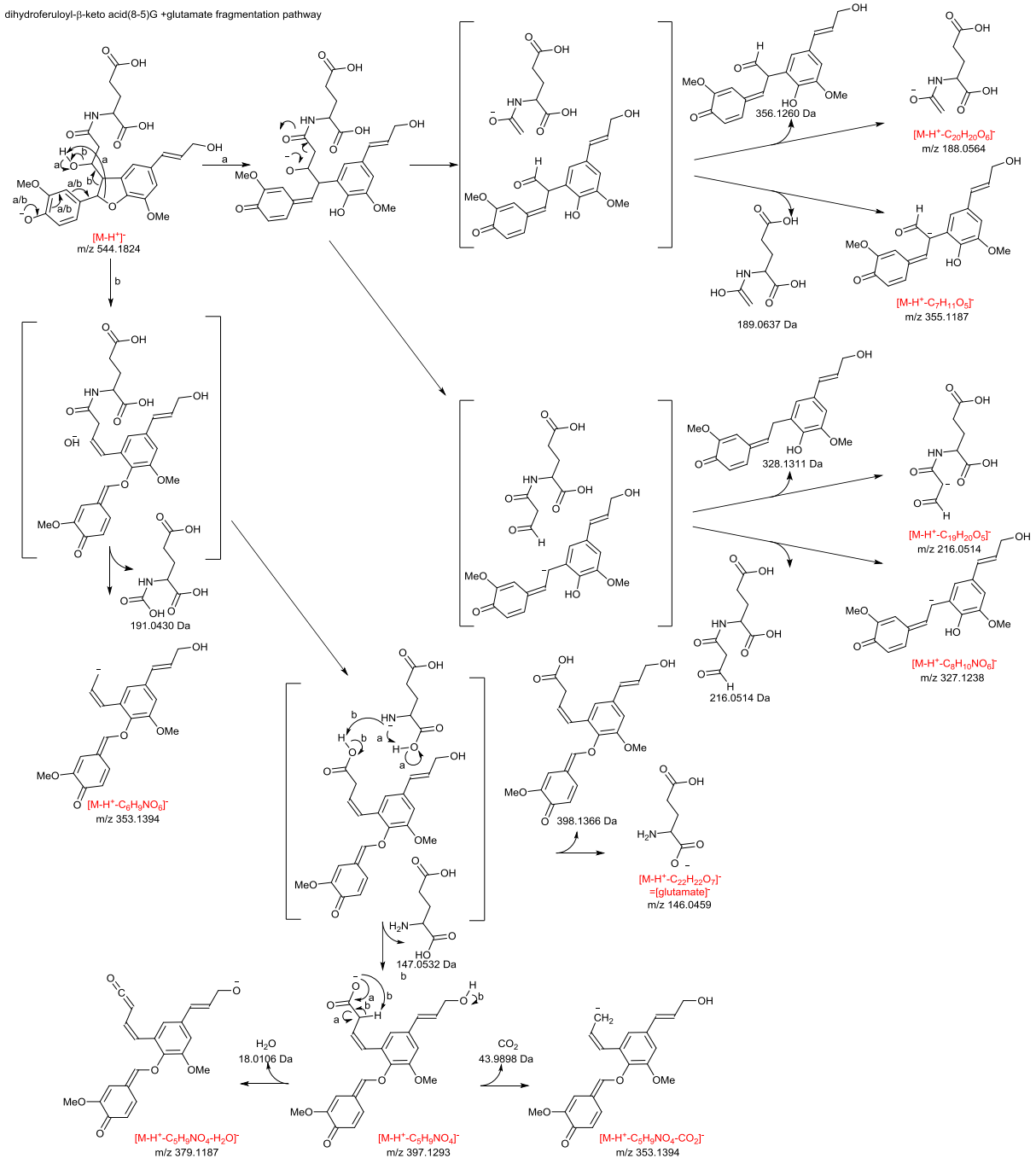


29. MS/MS dihydroferuloyl-β-keto acid(5-8)G + hexose (8.34 min, m/z 577.1927)

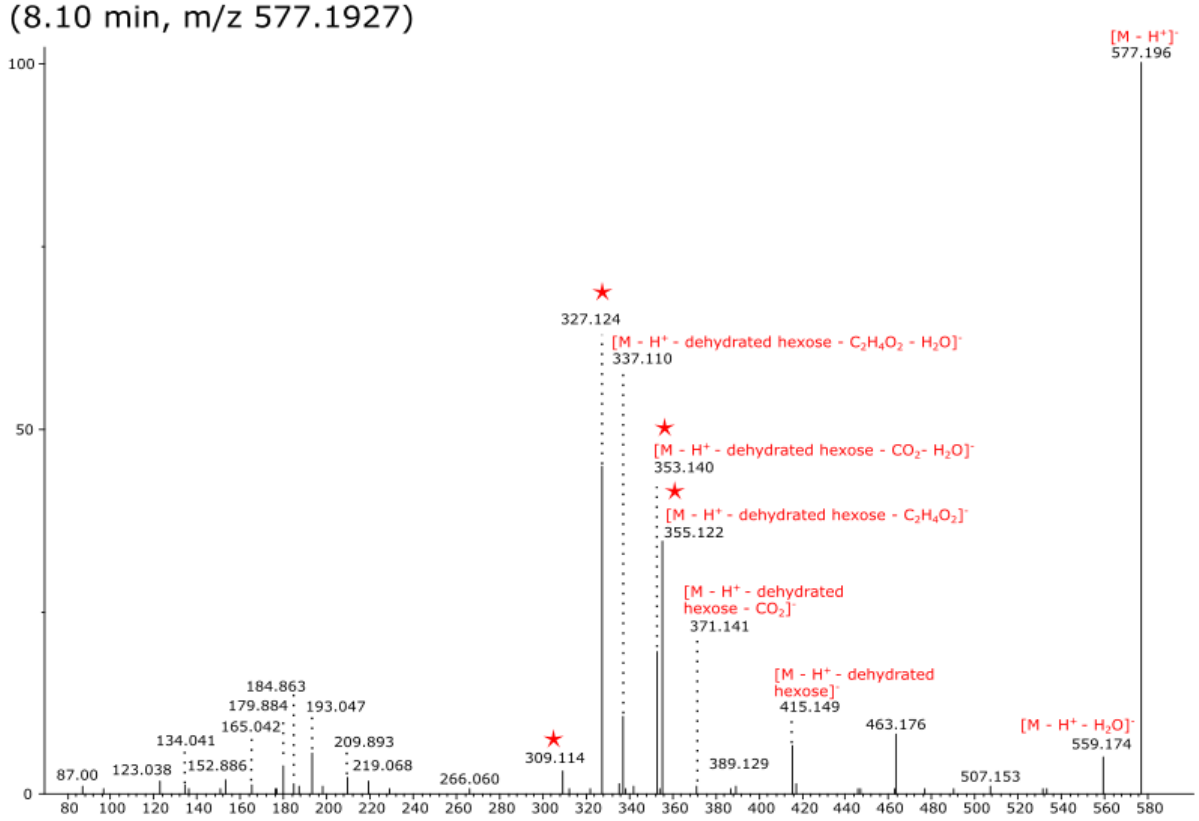



30. MS/MS dihydroferuloyl-β-keto acid(8-5)G + glutamate (11.11 min, m/z 544.1826)

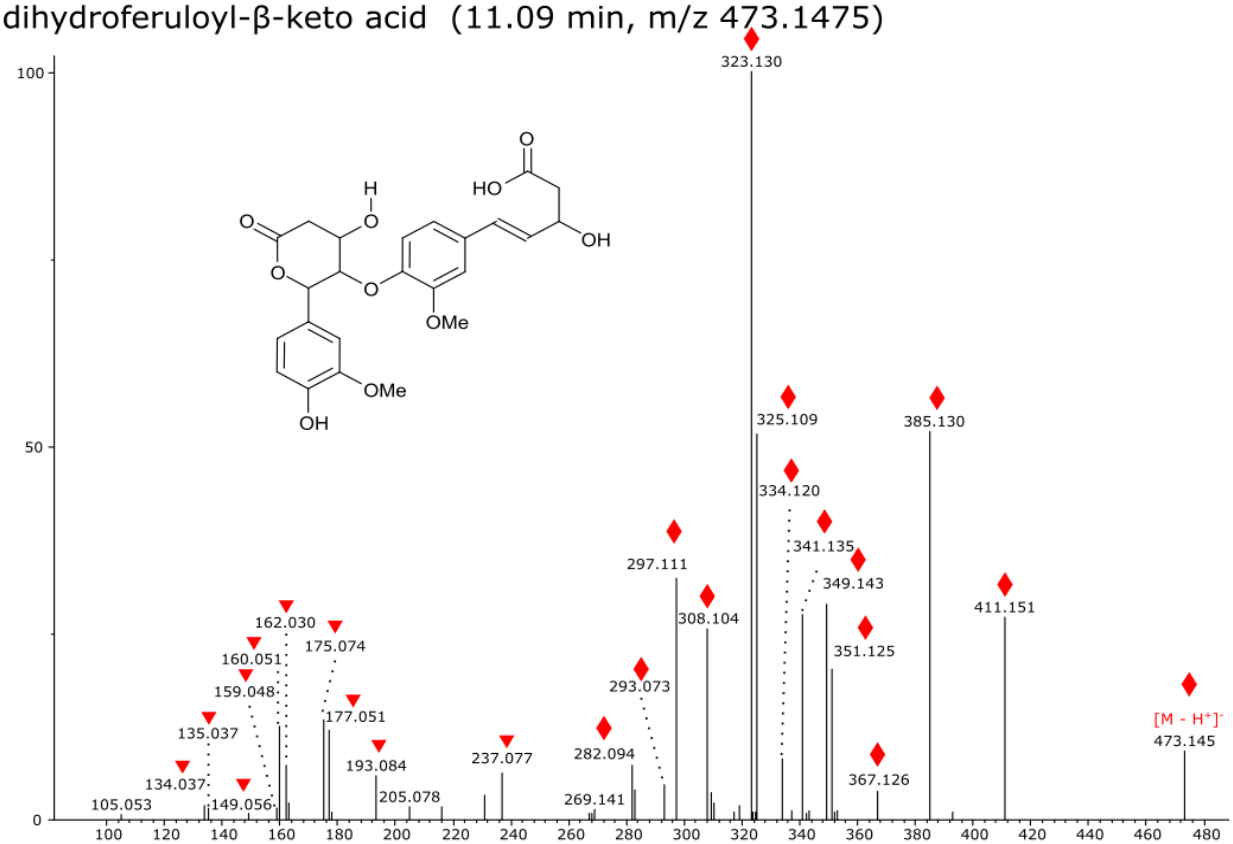



dihydroferuloyl-β-keto acid(8-5)G +glutamate fragmentation pathway
[M-H+]-
m/z 544.1824
356.1260 Da
[M-H+-C20H20O6]-
m/z 188.0564
189.0637 Da
[M-H+-C7H11O5]-
m/z 355.1187
191.0430 Da
328.1311 Da
[M-H+-C19H20O5]-
m/z 216.0514
216.0514 Da
[M-H+-C8H10NO6]-
m/z 327.1238
[M-H+-C8H9NO6]-
m/z 353.1394
398.1366 Da
[M-H+-C22H22O7]-
=[glutamate]-
m/z 146.0459
147.0532 Da
H2O
18.0106 Da
CO2
43.9898 Da
[M-H+-C5H9NO4-H2O]-
m/z 379.1187
[M-H+-C5H9NO4]-
m/z 397.1293
[M-H+-C5H9NO4-CO2]-
m/z 353.1394

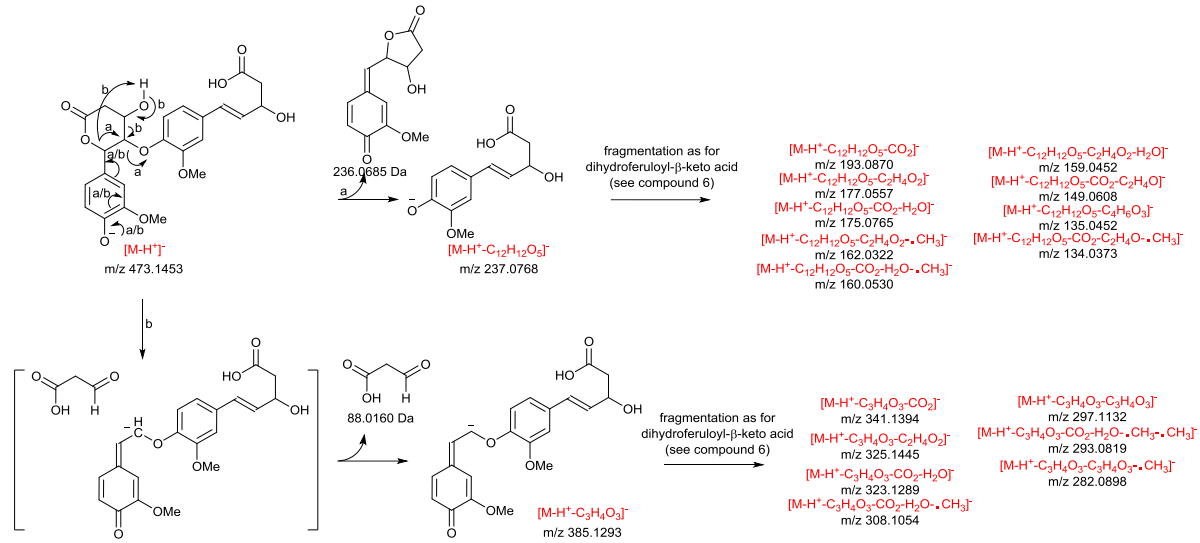
31. MS/MS dihydroferuloyl-β-keto acid(8-5)G + hexose (8.10 min, m/z 577.1927)



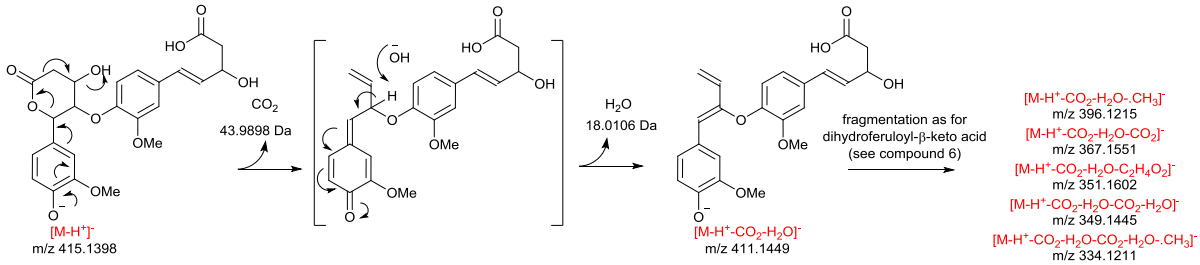

32. MS/MS dihydroferuloyl-β-keto acid(δ-lactone 8-*O*-4) dihydroferuloyl-β-keto acid (11.09 min, m/z 473.1475)



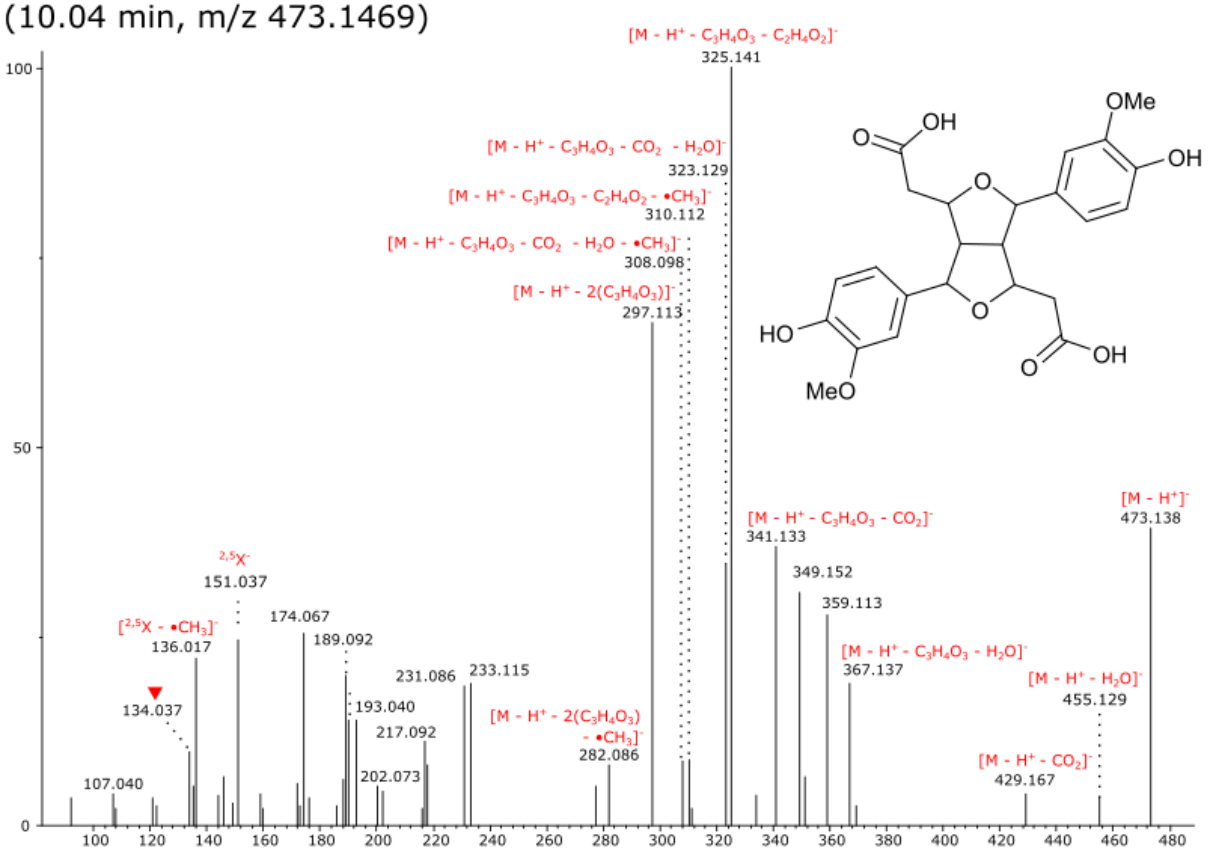
dihydroferuloyl-β-keto acid(δ-lactone 8-O-4)dihydroferuloyl-β-keto acid fragmentation pathway 1



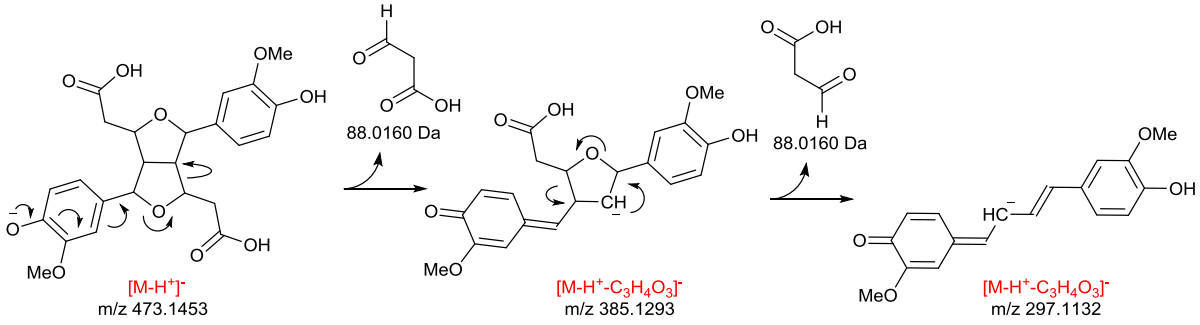

dihydroferuloyl-β-keto acid(δ-lactone 8-O-4)dihydroferuloyl-β-keto acid fragmentation pathway 2



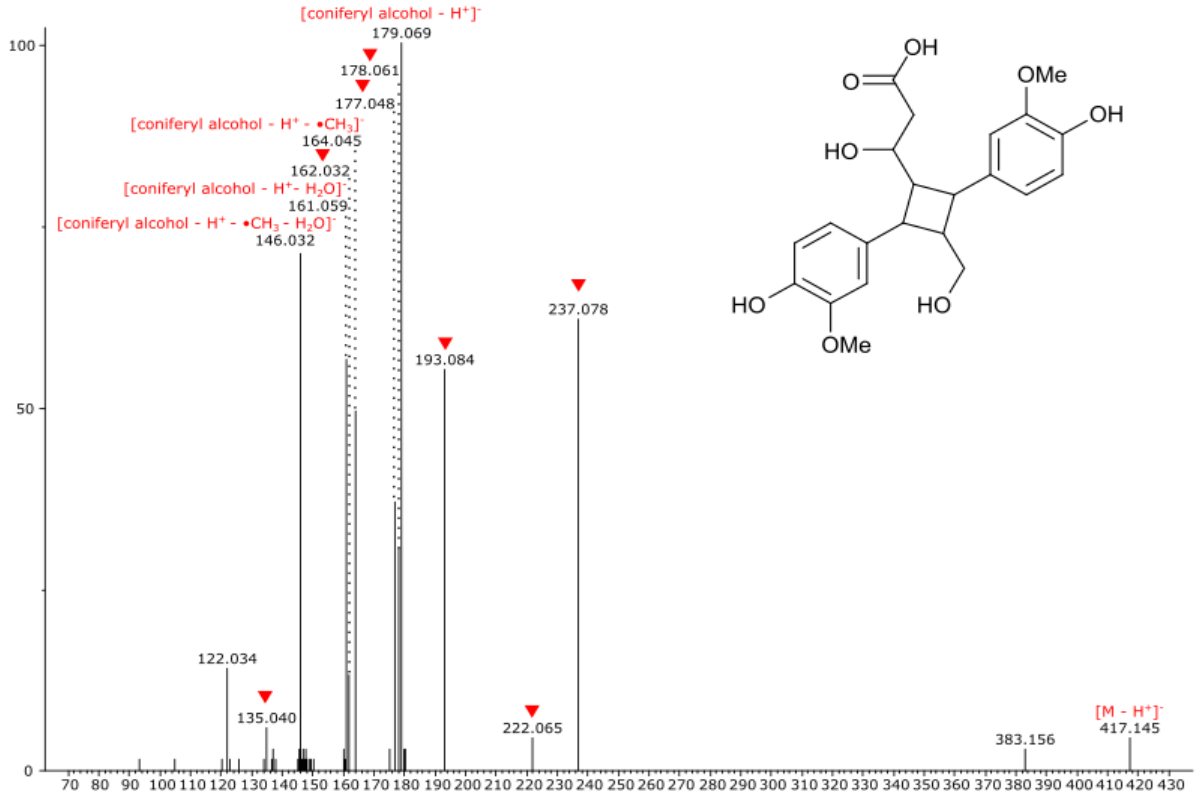
## 33. MS/MS dihydroferuloyl-β-keto acid(8-8)dihydroferuloyl-β-keto acid (10.04 min, m/z 473.1469)



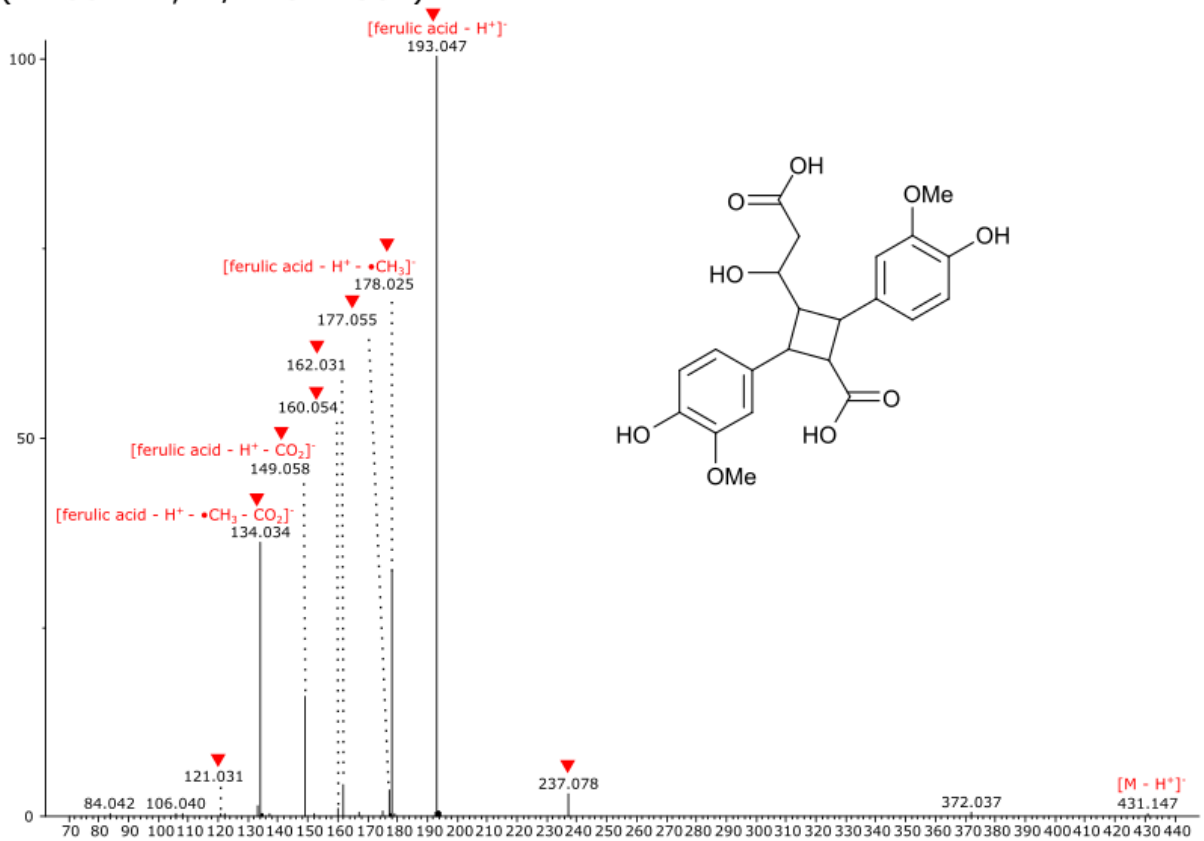

dihydroferuloyl-β-keto acid(8-8)dihydroferuloyl-β-keto acid fragmentation pathway



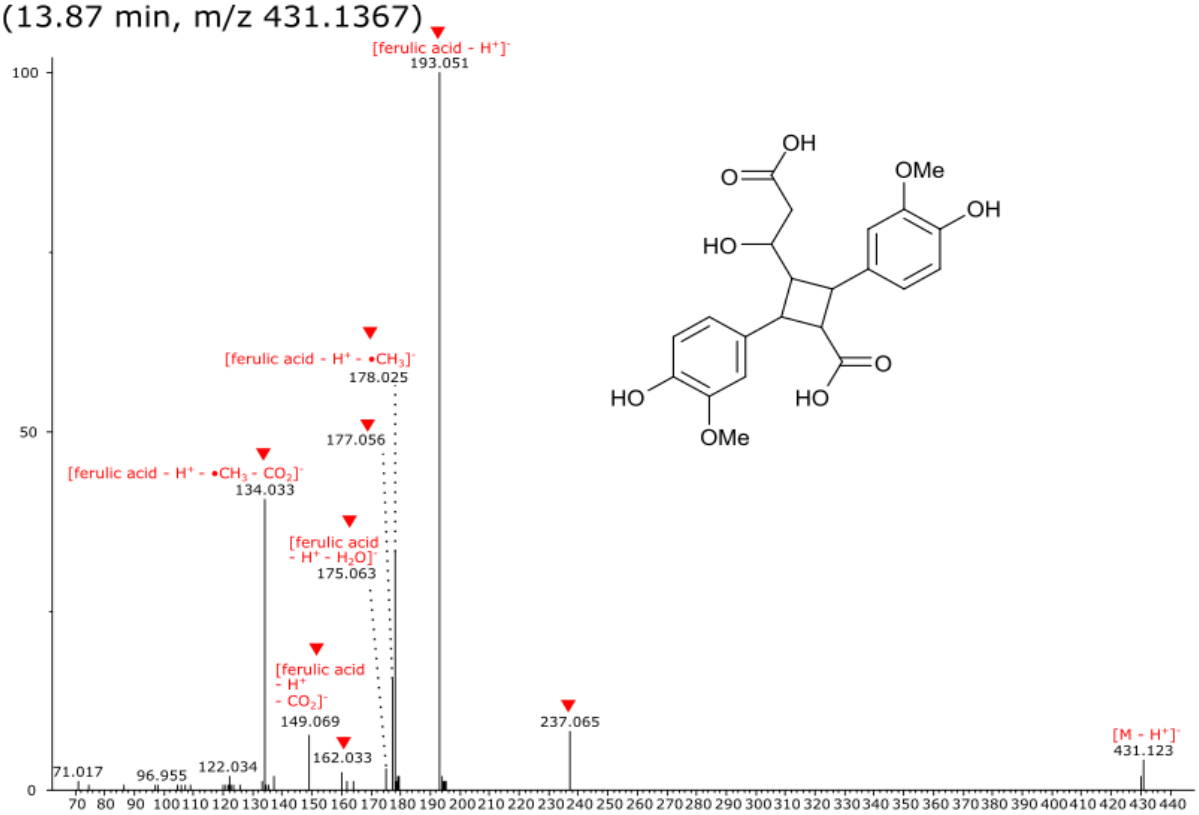
34. MS/MS dihydroferuloyl-β-keto acid + coniferyl alcohol cyclobutane dimer (12.47 min, m/z 417.1570)



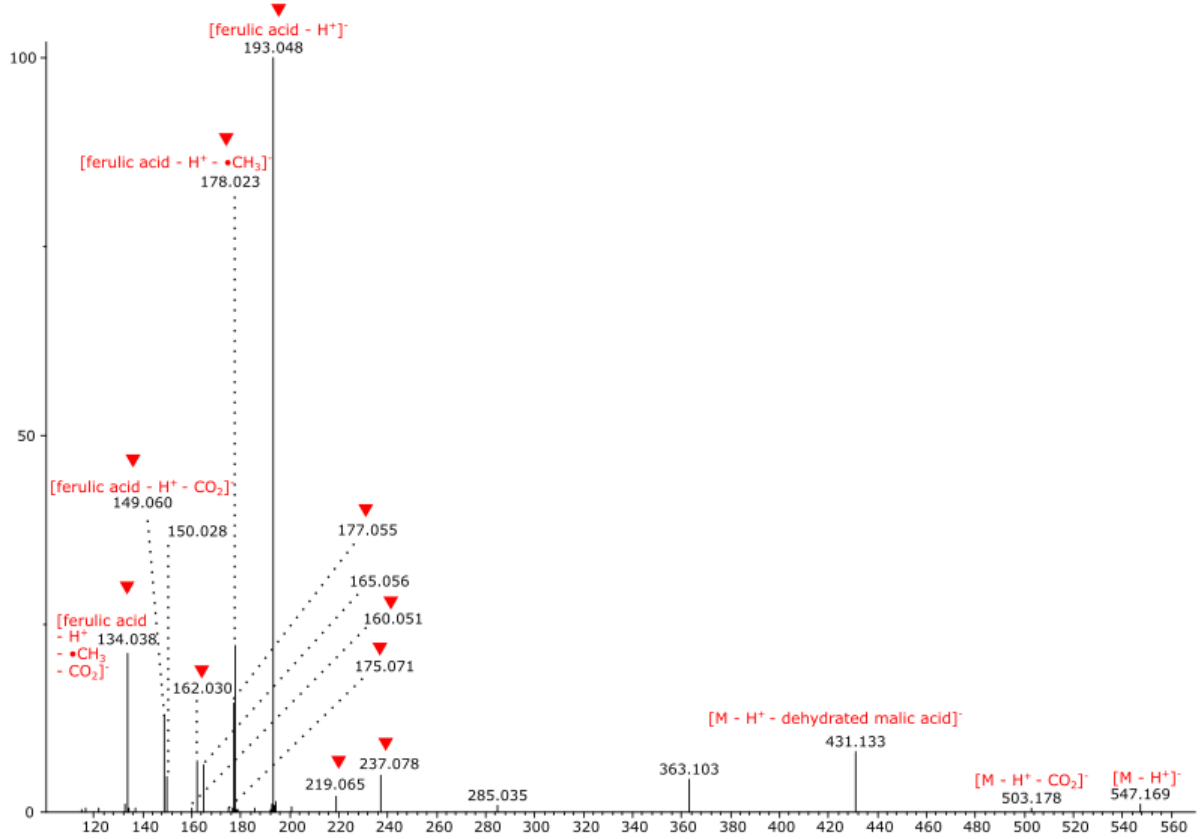

35. MS/MS dihydroferuloyl-β-keto acid ferulic acid cyclobutane dimer 1 (12.55 min, m/z 431.1362)



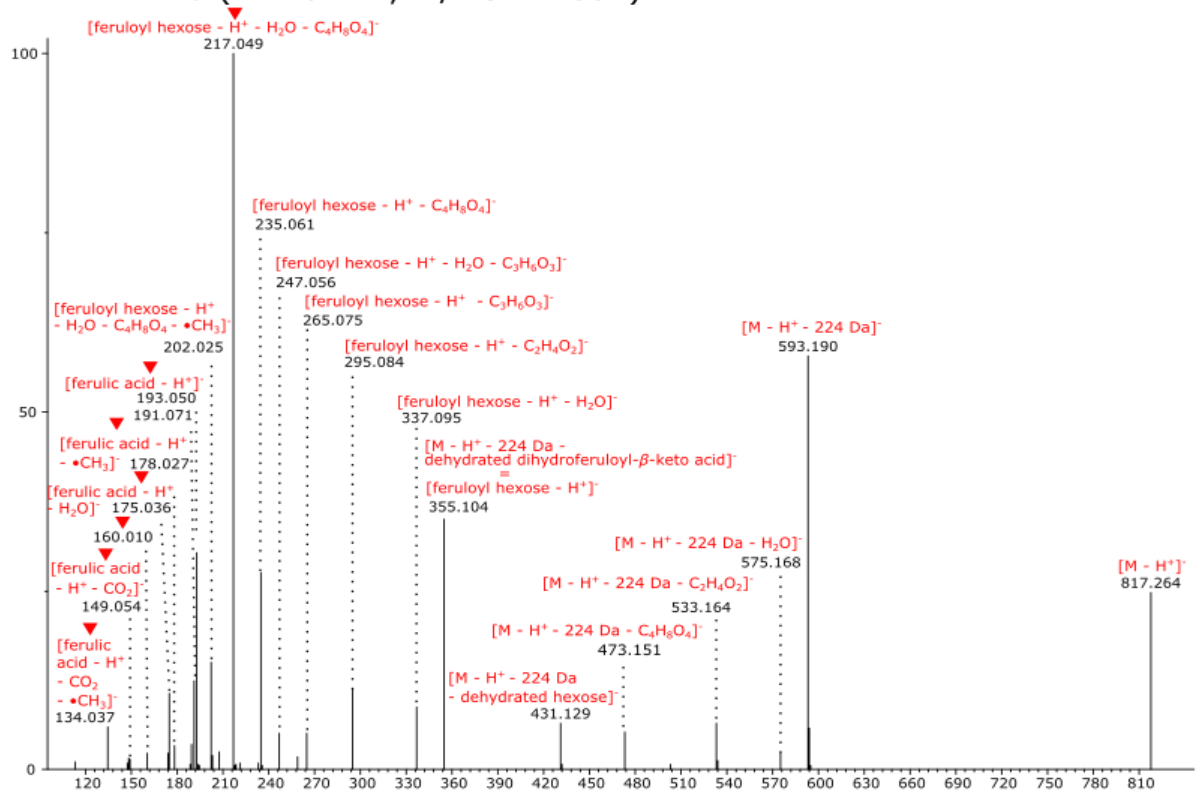
36. MS/MS dihydroferuloyl-β-keto acid ferulic acid cyclobutane dimer 2 (13.87 min, m/z 431.1367)



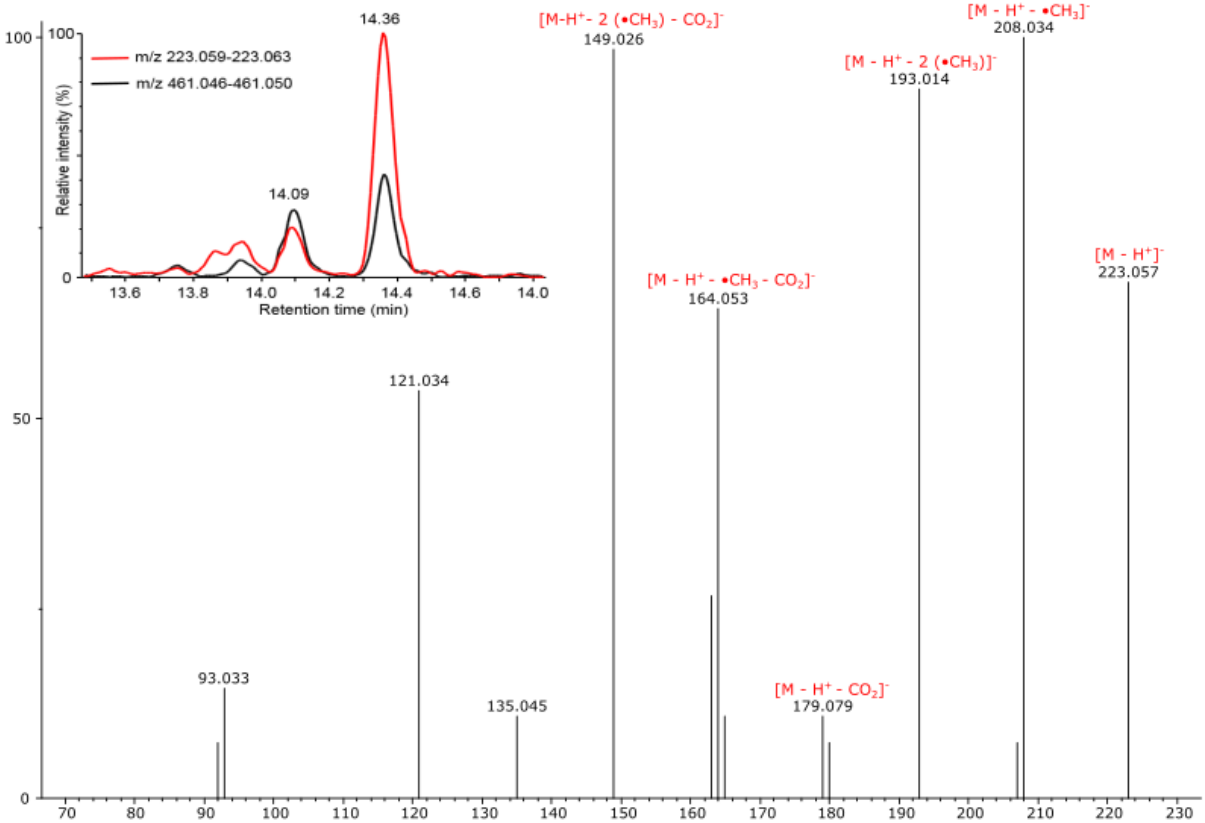

37. dihydroferuloyl-β-keto acid ferulic acid cyclobutane dimer + malate (12.52 min, m/z 547.1517)



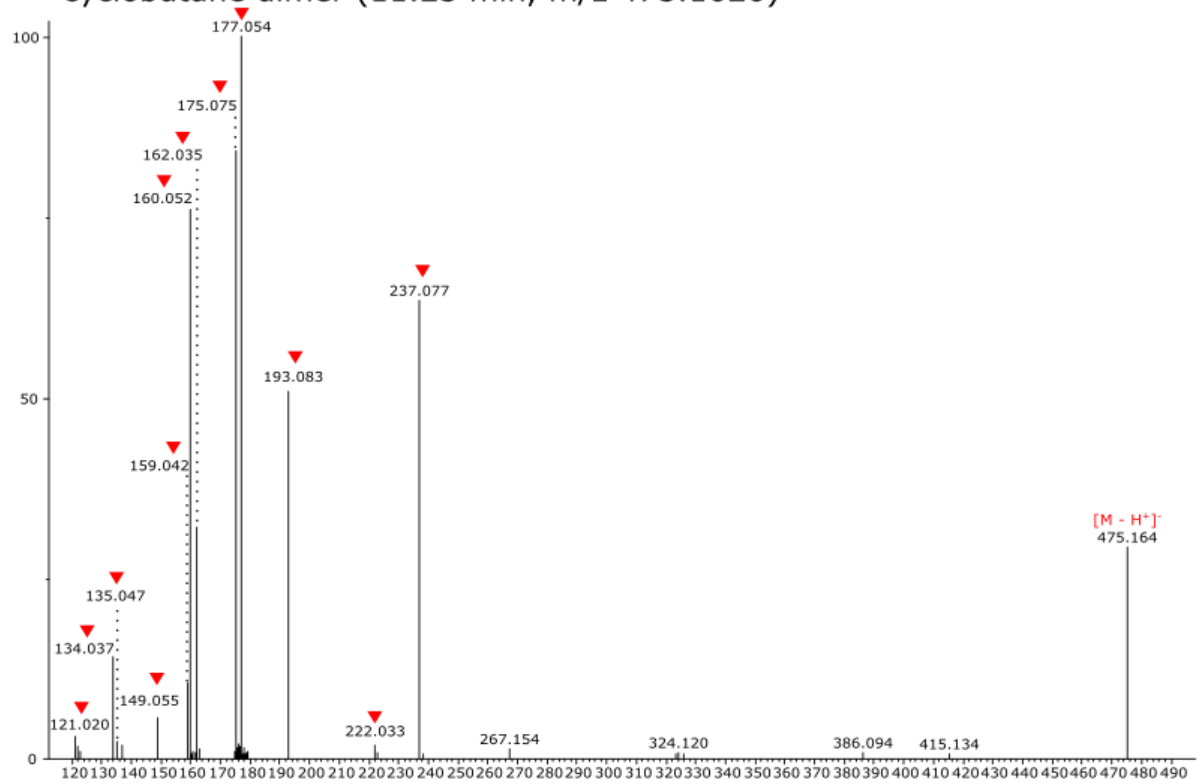
38. MS/MS dihydroferuloyl-β-keto acid feruloyl hexose cyclobutane dimer + 224 Da (14.16 min, m/z 817.2557)



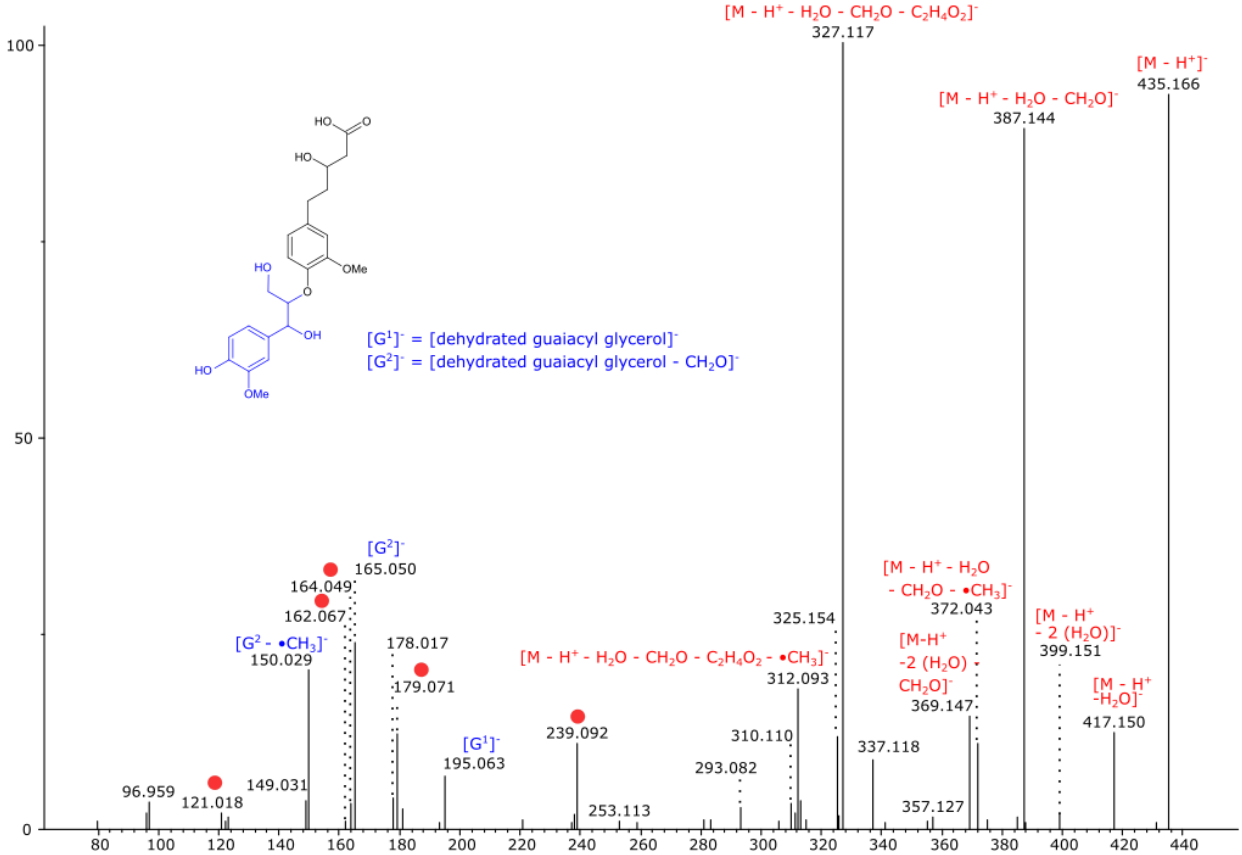

39. MS/MS dihydroferuloyl-β-keto acid sinapic acid cyclobutane dimer (14.36 min, m/z 223.0614, fragment of m/z 461.1476)



40. MS/MS dihydroferuloyl-β-keto acid dihydroferuloyl-β-keto acid cyclobutane dimer (11.23 min, m/z 475.1626)




41. MS/MS tetrahydroferuloyl-β-keto acid(4-*O*-8)G (8.82 min, m/z 435.1676)

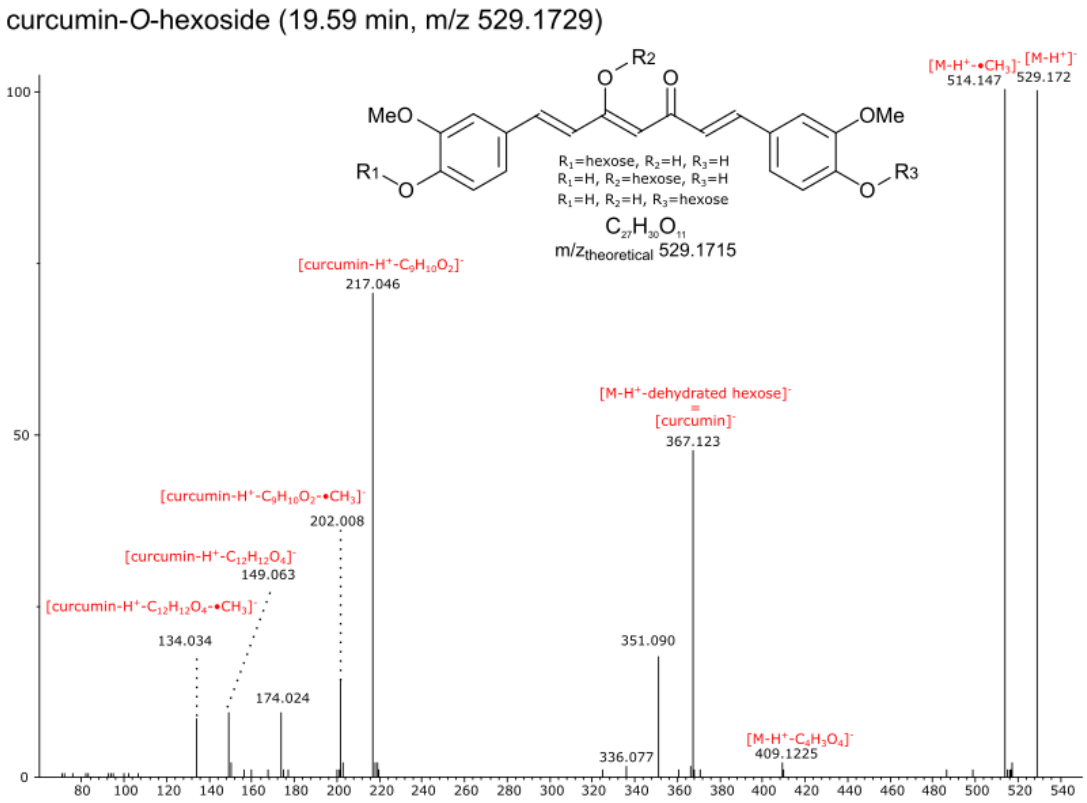


**Supplementary Figure 5. Targeted search for curcuminoid metabolites in the *pCesA4:DCS_CURS2* lines.** (**a**) Peak area of curcumin-*O*-hexoside and a curcumin-containing trimer in which coniferyl alcohol is linked to curcumin(8-8)G. Because their respective peak areas were below 1000, the latter two compounds did not end up as one of the 49 compounds in **Table 1**. *n* = 8 independent biological replicates per line. stdev: standard deviation; R.T.: retention time; inf: infinity; ratio: average *pCesA4:DCS_CURS2* / average WT. (**b**) MS/MS-based structural elucidation of curcumin-*O*-hexoside. (**c**) MS/MS-based structural elucidation of the trimer in which coniferyl alcohol is linked to curcumin(8-8)G. The compound is either curcumin(8-8)G(4-*O*-8)G, G(8-*O*-4)curcumin(8-8)G, or G(8-*O*-4')curcumin(8-8)G. For (**b**) and (**c**), a representative MS/MS spectrum is shown from at least four independent MS/MS spectra.

**a**

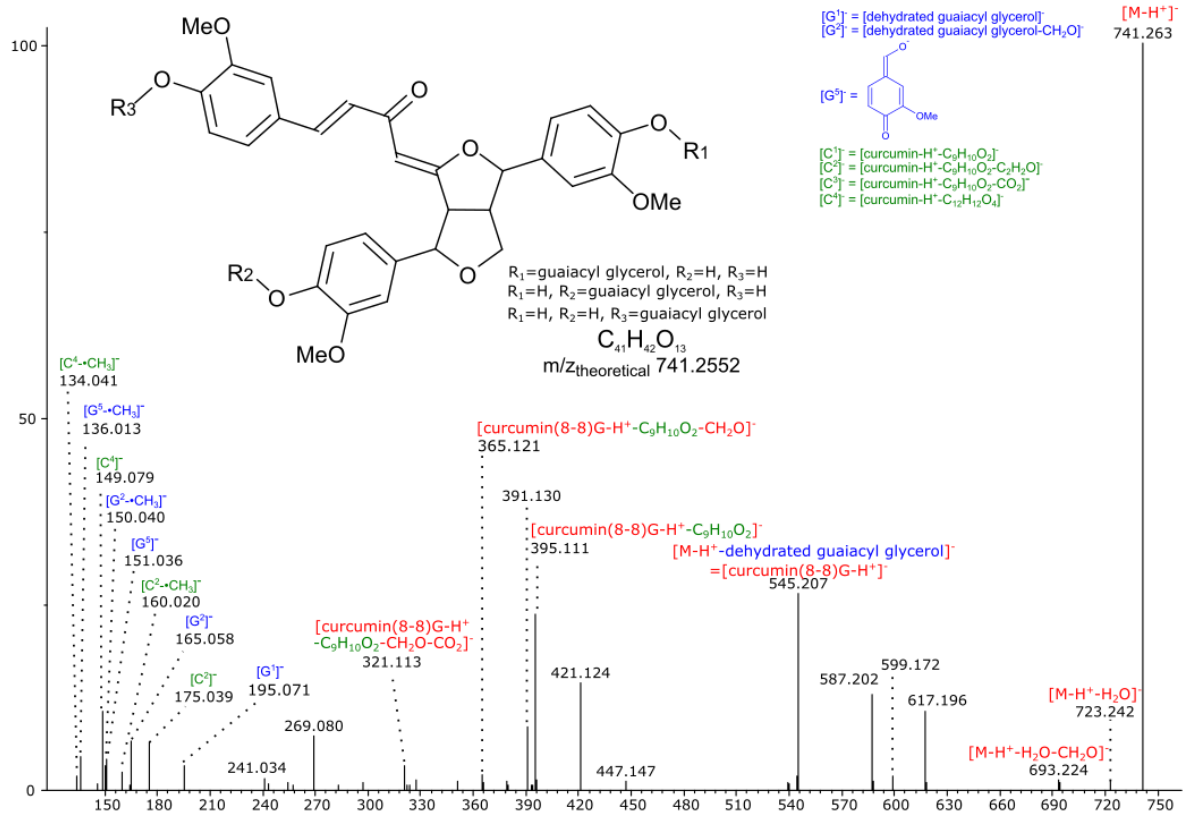
| | | R.T. | Wild type | *pCesA4:DCS_CURS2_1* | | *pCesA4:DCS_CURS2_2* | |
|---|---|---|---|---|---|---|---|
| **compound name** | ***m/z*** | **(min)** | **mean ± stdev** | **mean ± stdev** | **ratio** | **mean ± stdev** | **ratio** |
| curcumin-*O*-hexoside | 529.173 | 19.59 | n.d. | 168 ± 130 | inf | 214 ± 140 | inf |
| curcumin(8-8)G(8-*O*-4)G | 741.255 | 22.78 | n.d. | 367 ± 368 | inf | 941 ± 733 | inf |

**b**

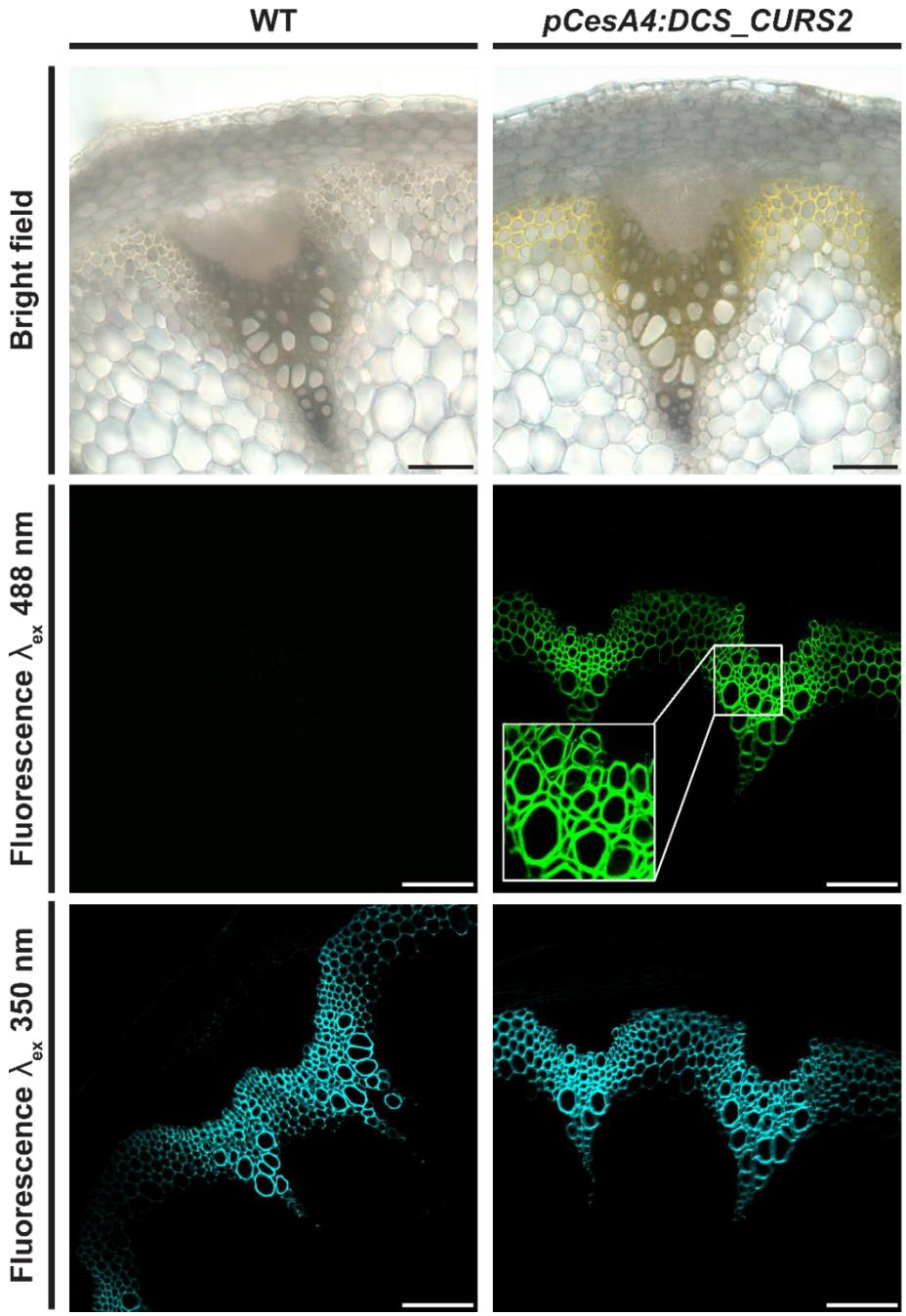


**c**

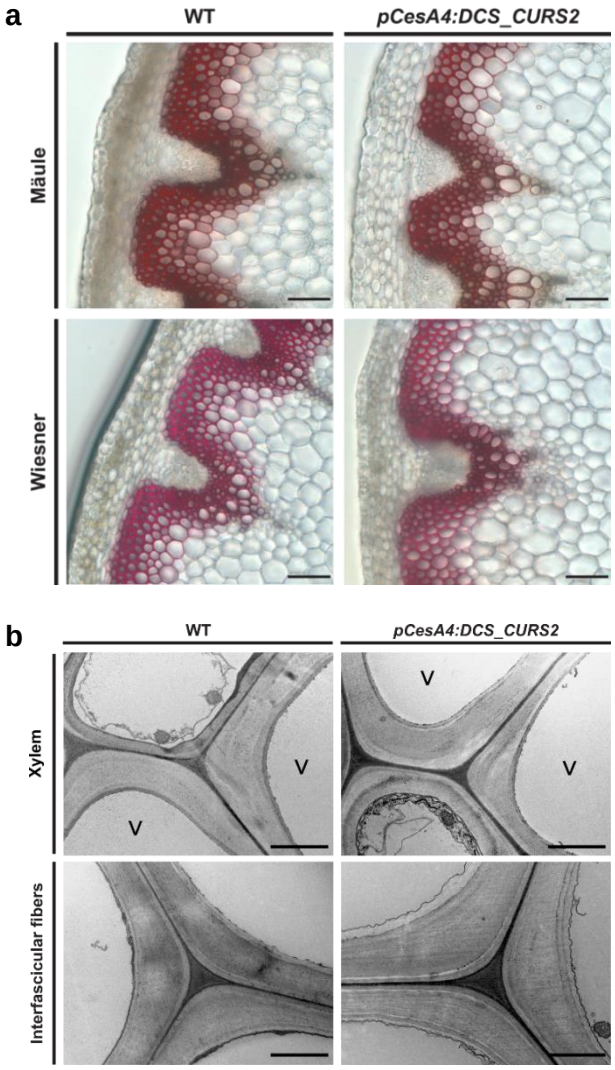
curcumin(8-8)G(4-*O*-8)G or G(8-*O*-4/4')curcumin(8-8)G (22.78 min, m/z 741.2545)



**Supplementary Figure 6. Bright field and fluorescence microscopy of WT and *pCesA4:DCS_CURS2* cross sections.** The yellow colour and fluorescence at an excitation wavelength of 488 nm remained in the secondary cell walls of cross sections of *pCesA4:DCS_CURS2* inflorescence stems after a 2 h incubation in acetone. Representative images are shown from six independent biological replicates per line, eight sections for each biological replicate and three regions for each section. Scale bar, 100 µm.

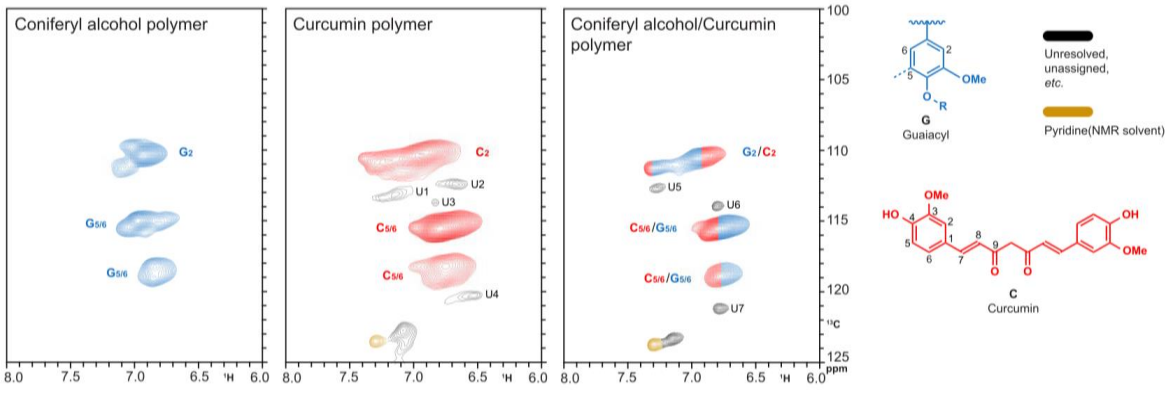


**Supplementary Figure 7. Lignin staining and TEM of WT and *pCesA4:DCS_CURS2* stem cross sections.** (a) Mäule stains S units red and G units light brown and is generally used to visualize the presence of S units in lignin, whereas the Wiesner reagent colors cinnamaldehyde end-groups in lignin and is used to visualize the presence of lignin in a concentration-dependent manner. Representative images are shown of three independent biological replicates per line, eight sections from each biological replicate and three regions for each section. With both Mäule and Wiesner staining, no differences were observed between *pCesA4:DCS_CURS2* and WT. Scale bar, 100 µm. (b) Transmission electron microscopy (TEM) demonstrates the ultrastructure of the interfascicular fibers and xylem regions. Representative images are shown from three independent biological replicates per line, one section for each biological replicate and a visual screen of the complete xylem and interfascicular fiber region for each section. No differences between *pCesA4:DCS_CURS2* and WT were observed. v, xylary vessel. Scale bar, 2 µm.

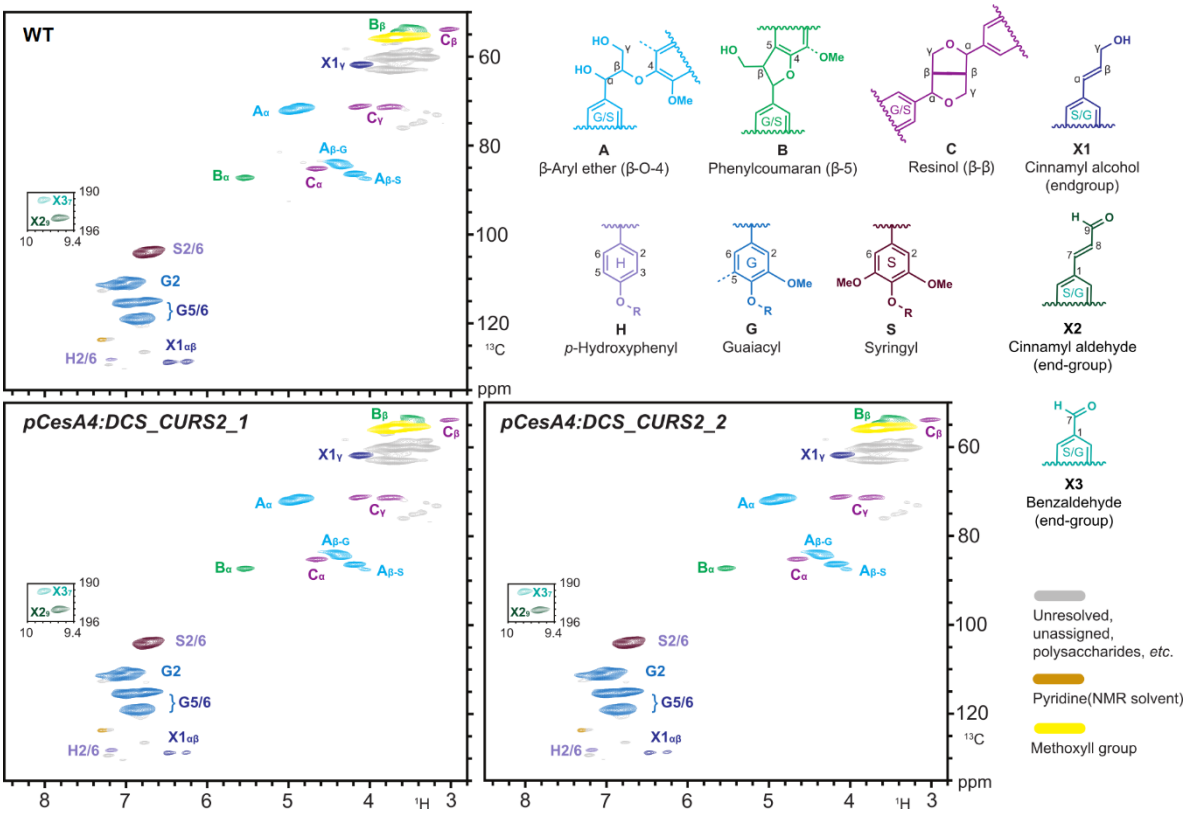


**Supplementary Figure 8. Structural characterization of dehydrogenation polymers (DHPs) and lignins by 2D HSQC NMR.** It was investigated which 2D HSQC NMR signals were specific for curcumin, anticipating that curcumin might provide signatures of its incorporation into lignins. (**a**) Pure coniferyl alcohol DHP; pure curcumin DHP; 50% coniferyl alcohol and 50% curcumin DHP (*n*=1). In the DHPs made from pure curcumin and coniferyl alcohol/curcumin, the main signals originated from the aromatic rings, i.e. the $C_2$, $C_5$ and $C_6$ signals and the $G_2/C_2$, $G_5/C_5$ and $G_6/C_6$ signals, respectively. However, similar signals were also present in DHPs made from pure coniferyl alcohol ($G_2$, $G_5$ and $G_6$) and can therefore not be used as specific markers for curcumin. In the pure curcumin DHP, signals (U1-U4) with relatively low intensity were detected that were not present in the pure coniferyl alcohol DHP. However, these signals remain unassigned. Also in the coniferyl alcohol/curcumin DHP, different signals (U5-U7) with relatively low intensity were detected that were not present in the pure coniferyl alcohol DHP. These signals also remained unassigned. (**b**) HSQC spectra of the aromatic regions and the oxygenated aliphatic regions of whole cell walls from stems of WT and the *pCesA4:DCS_CURS2* lines. A representative spectrum from three independent biological repeats is shown; each biological repeat is a pool of four or five inflorescence stems. Integrated values for each monomeric unit H, G and S and the α-C/H correlation peaks from the major lignin interunit structures A-C are provided in Table 2. The colors of the substructures shown match those of the corresponding signals in the HSQC spectra (where they are resolved). None of the U1-U7 signals were present in 2D HSQC NMR spectra of lignin samples from the *pCesA4:DCS_CURS2* lines. No other signals appeared in the *pCesA4:DCS_CURS2* spectra that were different to those of WT.

**a**

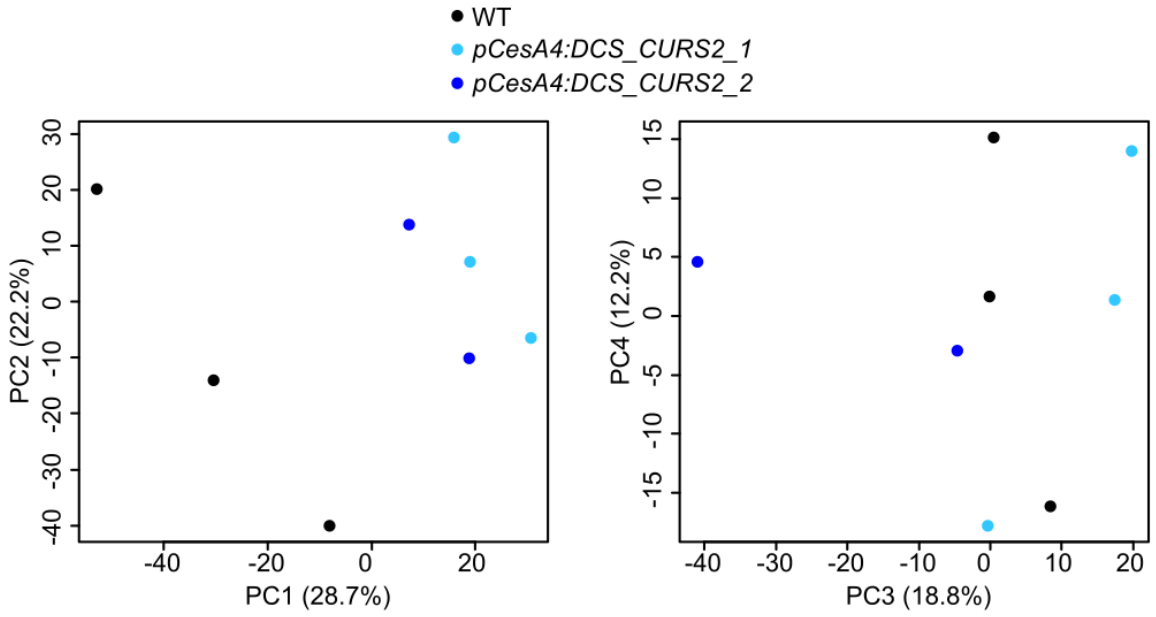


**b**


WT
pCesA4:DCS_CURS2_1
pCesA4:DCS_CURS2_2
Bβ
Cβ
X1γ
Aα
Cγ
Aβ-G
Bα
Cα
Aβ-S
X3₇
X2₉
S2/6
G2
G5/6
X1αβ
H2/6
60
80
100
120
13C
ppm
8
7
6
5
4
1H
3
A
β-Aryl ether (β-O-4)
B
Phenylcoumaran (β-5)
C
Resinol (β-β)
X1
Cinnamyl alcohol (endgroup)
H
p-Hydroxyphenyl
G
Guaiacyl
S
Syringyl
X2
Cinnamyl aldehyde (end-group)
X3
Benzaldehyde (end-group)
Unresolved, unassigned, polysaccharides, etc.
Pyridine(NMR solvent)
Methoxyll group

**Supplementary Figure 9. Characterization of catalytic hydrogenolysis lignin oil from *pCesA4:DCS_CURS2* inflorescence stems.** (**a**) PCA showing the separation between UHPLC-MS profiles of *pCesA4:DCS_CURS2* and WT lignin oils based on PC1. For WT, *pCesA4:DCS_CURS2_1* and *pCesA4:DCS_CURS2*_2, *n*=3, 3 and 2, respectively. (**b**) Peak list of 21 compounds that accumulated significantly 10-fold in lignin oils of *pCesA4:DCS_CURS2* (*n*=5) as compared with lignin oils from WT (*n*=3). Names given are based on high-resolution m/z and on MS/MS spectra given in panel (**c**). Ratio: average *pCesA4:DCS_CURS2* / average WT. SD, standard deviation. (**c**) MS/MS spectra of compounds given in (**b**). Because of the lack of appropriate authentic chemical standards for these compounds, the structural characterization remains tentative. Two independent MS/MS spectra were generated for each compound, of which one MS/MS spectrum is shown. The signals for which the m/z value is given were common in the two independent MS/MS spectra. The detection limit was set at a peak intensity of 200 counts. n.d., not detected.

**a**

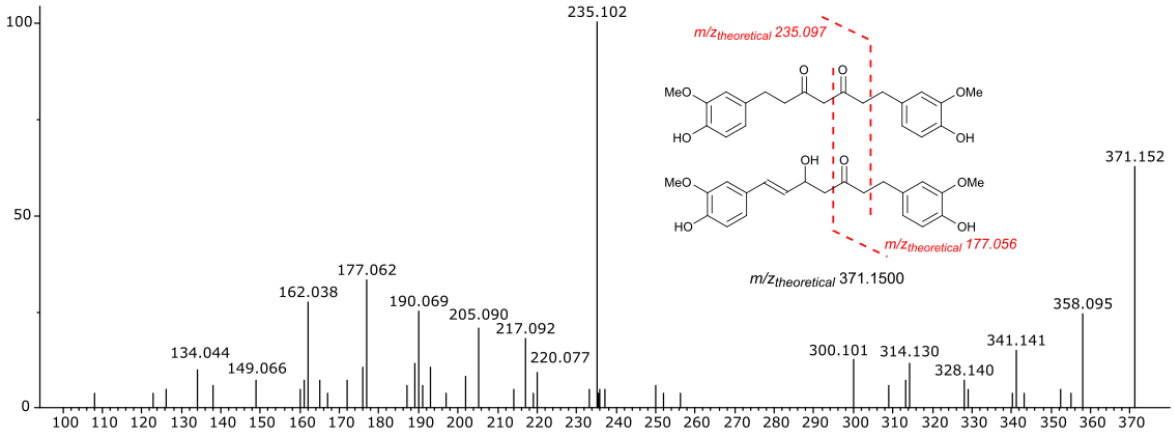


**b**

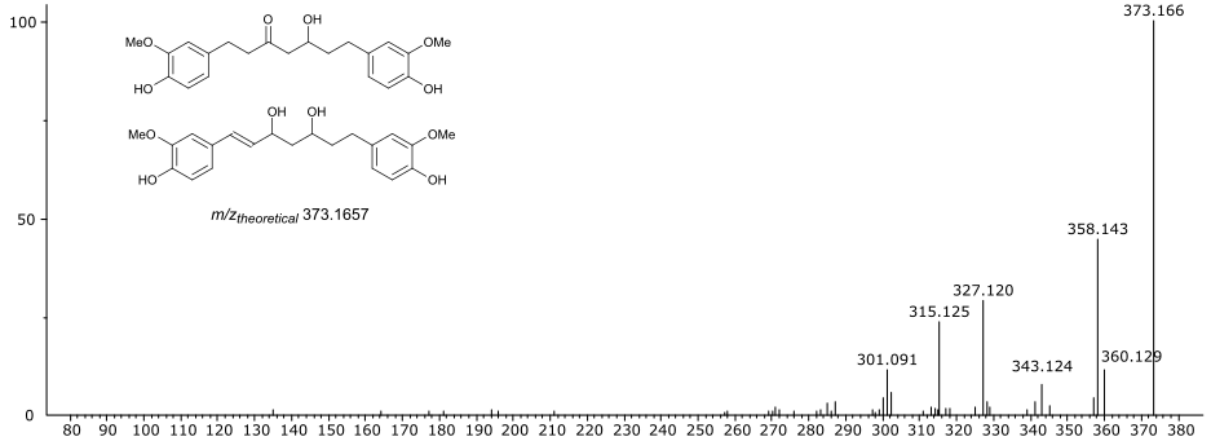
| no. | m/z | R.T. (min) | compound name | WT Average ± SD | *pCesA:DCS_CURS* Average ± SD | ratio *pCesA:DCS_CURS2*/WT |
|---|---|---|---|---|---|---|
| **1** | 249.0772 | 8.71 | - | 366 ± 187 | 3679 ± 452 | 10 |
| **2** | 263.0941 | 10.19 | - | n.d. | 1089 ± 301 | inf |
| **3** | 371.1497 | 13.27 | tetrahydrocurcumin | n.d. | 1159 ± 501 | inf |
| **4** | 373.1651 | 15.82 | hexahydrocurcumin 1 | n.d. | 8200 ± 1433 | inf |
| **5** | 389.1938 | 20.81 | dideoxyhexahydroferuloyl-β-keto acid(red5-8)G | n.d. | 4193 ± 1145 | inf |
| **6** | 187.0761 | 21.63 | - | 278 ± 77 | 2909 ± 410 | 10 |
| **7** | 417.4103 | 22.28 | - | n.d. | 1083 ± 142 | inf |
| **8** | 387.1800 | 22.56 | dideoxyhexahydroferuloyl-β-keto acid(5-8)G | n.d. | 2370 ± 558 | inf |
| **9** | 387.1524 | 23.00 | - | n.d. | 1938 ± 178 | inf |
| **10** | 537.2445 | 23.94 | deoxyoctahydrocurcumin(5-8)G | n.d. | 1053 ± 315 | inf |
| **11** | 537.2466 | 24.69 | deoxyoctahydrocurcumin(5-8)G | n.d. | 2983 ± 484 | inf |
| **12** | 201.0915 | 24.83 | - | n.d. | 1204 ± 332 | inf |
| **13** | 343.1549 | 25.14 | - | n.d. | 3207 ± 672 | inf |
| **14** | 359.1515 | 26.71 | - | n.d. | 2260 ± 305 | inf |
| **15** | 309.2075 | 26.77 | - | n.d. | 1015 ± 248 | inf |
| **16** | 359.3899 | 27.34 | - | n.d. | 1089 ± 110 | inf |
| **17** | 357.1711 | 28.18 | deoxyhexahydrocurcumin (no MSMS) | n.d. | 3809 ± 645 | inf |
| **18** | 431.2073 | 28.98 | - | n.d. | 3475 ± 549 | inf |
| **19** | 373.1666 | 29.08 | hexahydrocurcumin 2 | n.d. | 2942 ± 362 | inf |
| **20** | 509.2207 | 29.95 | - | n.d. | 3795 ± 403 | inf |
| **21** | 415.2133 | 32.47 | - | n.d. | 1479 ± 228 | inf |

**c**

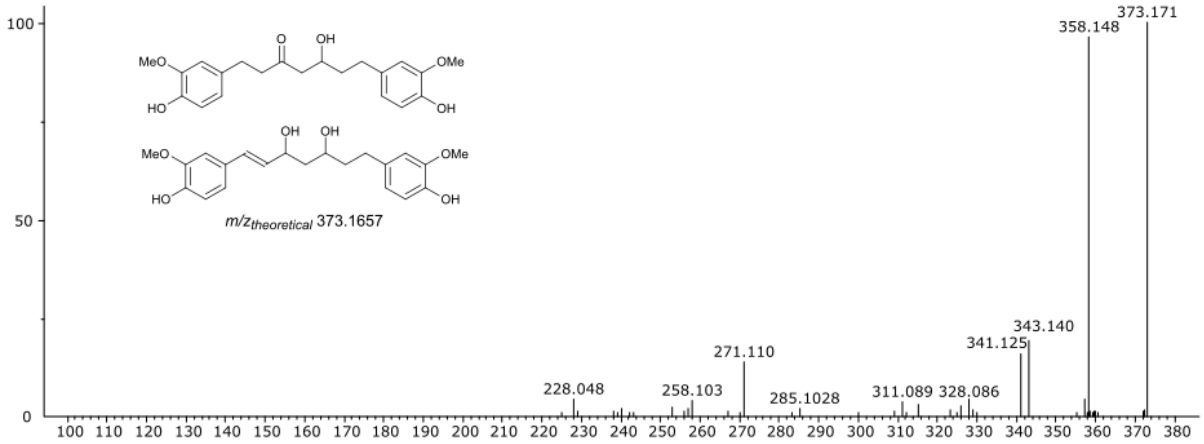
tetrahydrocurcumin
m/z 371.1497, 13.27 min



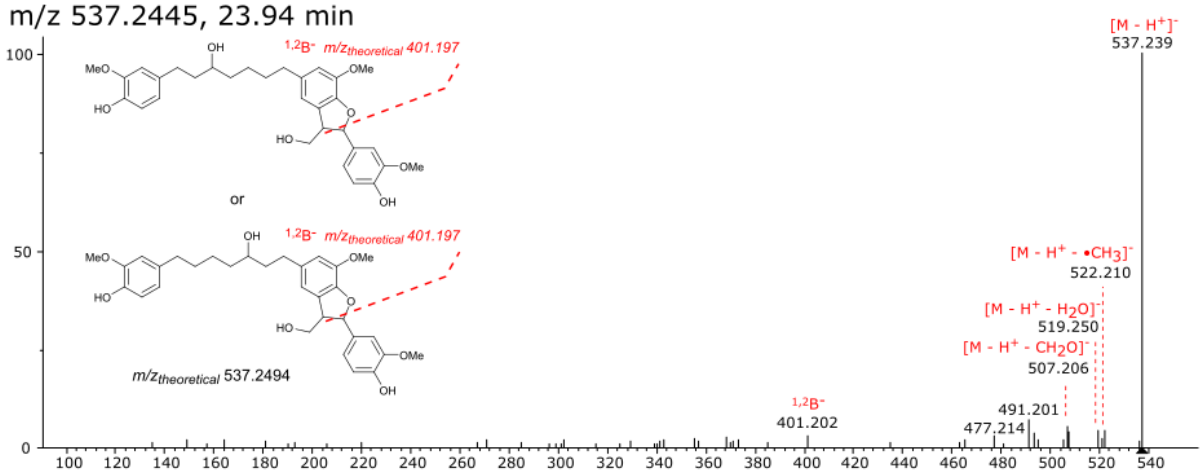

hexahydrocurcumin 1
m/z 373.1651, 15.82 min



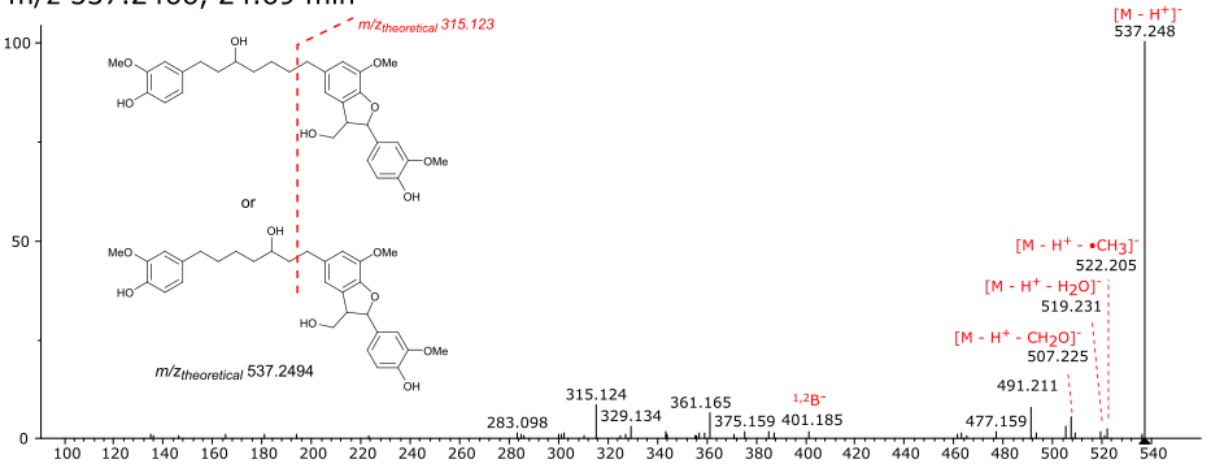

hexahydrocurcumin 2
m/z 373.1667, 29.08 min



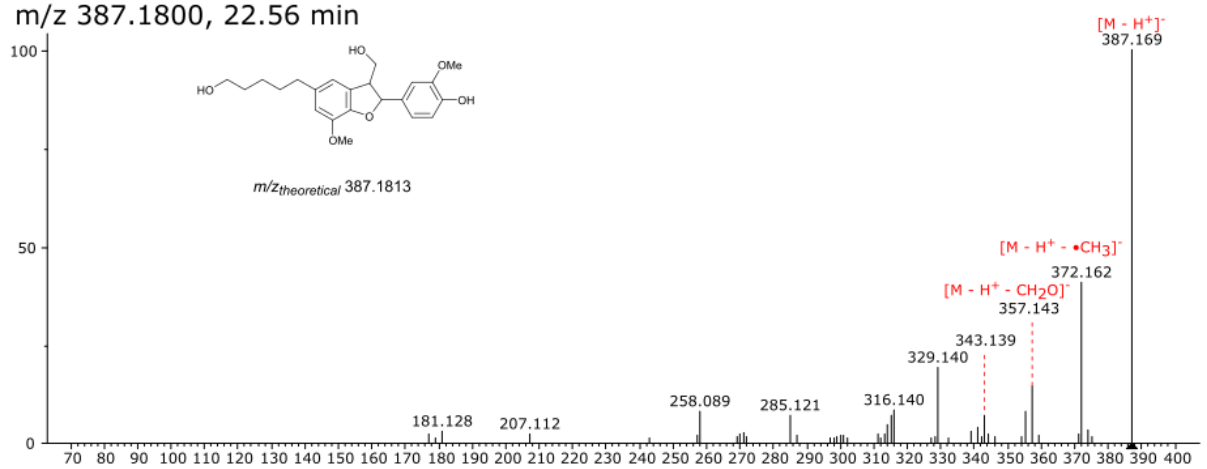
deoxyoctahydrocurcumin(5-8)G 1
m/z 537.2445, 23.94 min



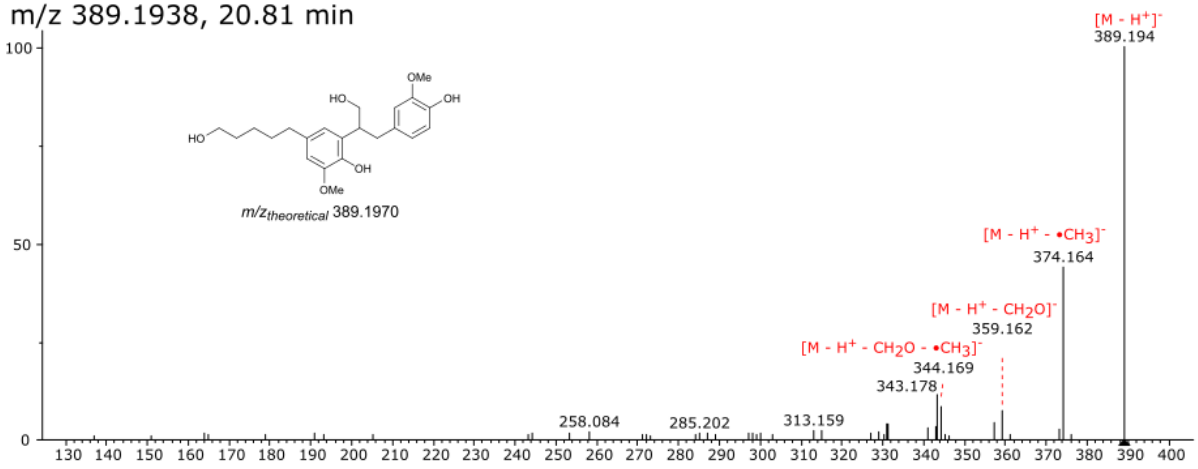

deoxyoctahydrocurcumin(5-8)G 2
m/z 537.2466, 24.69 min




dideoxyhexahydroferuloyl-β-keto acid(5-8)G
[4-hydroxy-3-methoxybenzenepentanol(5-8)G]
m/z 387.1800, 22.56 min

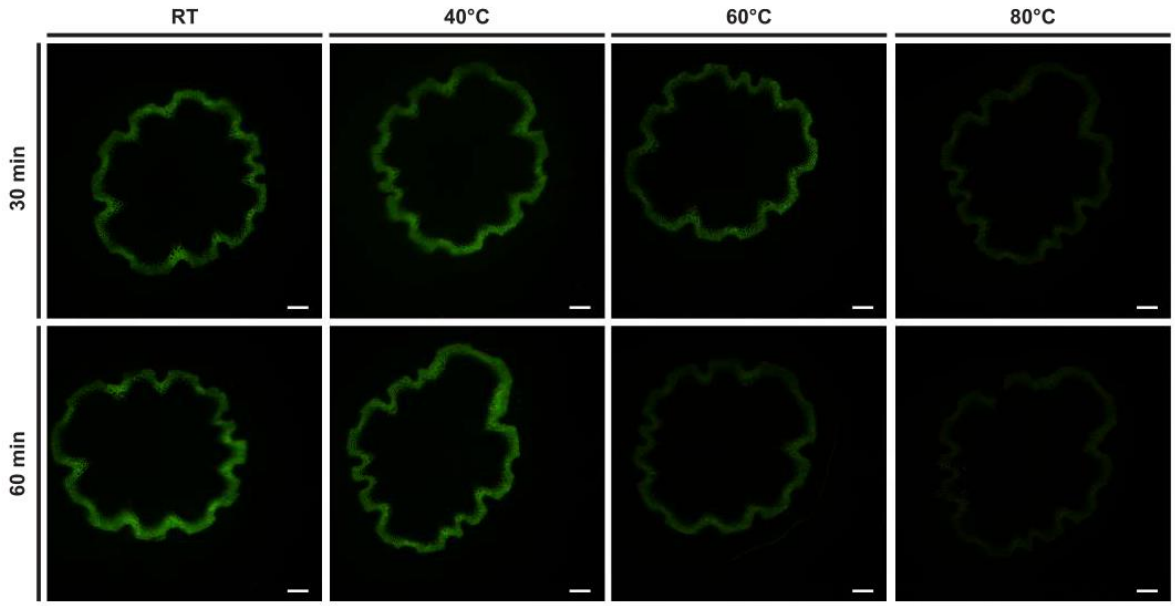


dideoxyhexahydroferuloyl-β-keto acid(red5-8)G
[4-hydroxy-3-methoxybenzenepentanol(red5-8)G]
m/z 389.1938, 20.81 min

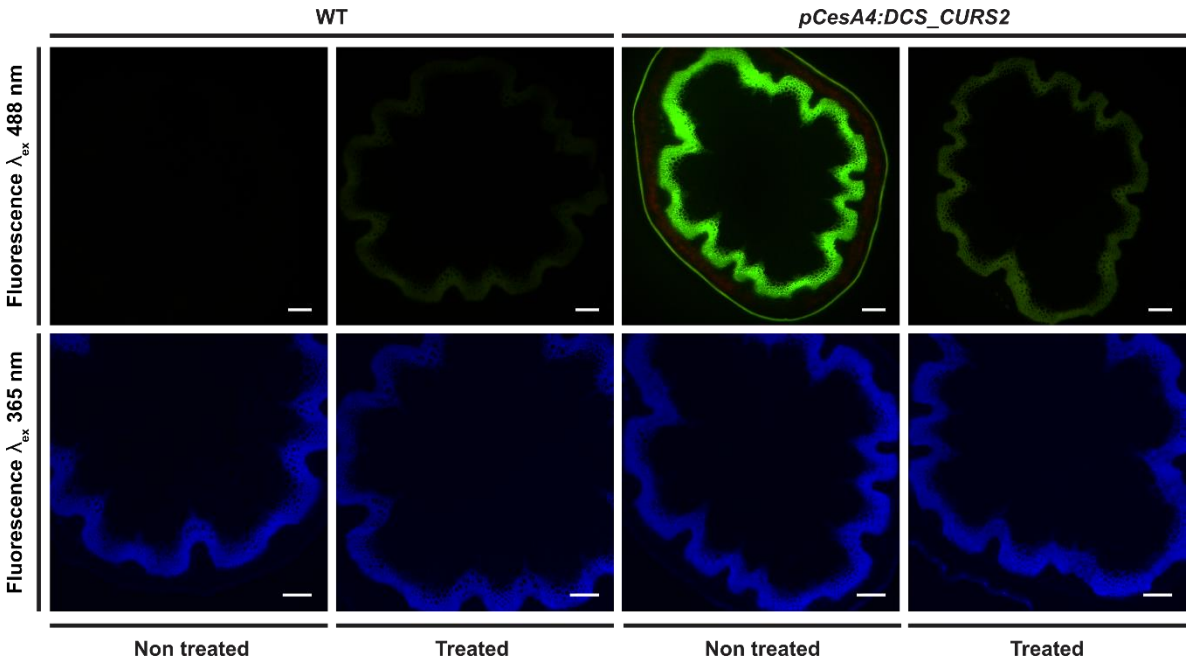


**Supplementary Figure 10. Curcumin breakdown in alkaline conditions.** (**a**) Transgenic *pCesA4:DCS_CURS2* stem cross sections were incubated in 62.5 mM NaOH (pH 12.8) for 30 and 60 min at four different temperatures: room temperature (RT), 40°C, 60°C and 80°C. Images (excitation filter of 460-500 nm) taken after extensive washing show representative pictures from 20 to 30 replicates originating from 3 independent biological replicates. Scale bars, 200 µm. (**b**) WT and *pCesA4:DCS_CURS2* stem cross sections were incubated in 62.5 mM NaOH (pH 12.8, 90ºC, 30 min). Fluorescence of *pCesA4:DCS_CURS2* at an excitation wavelength of 488 nm decreased after the treatment. WT cross sections were not fluorescent at a wavelength of 488 nm before treatment, but were slightly fluorescent after the alkaline treatment. The autofluorescence of lignin at an excitation wavelength of 365 nm was similar for WT and *pCesA4:DCS_CURS2* cross sections and appeared not to be affected by the alkaline treatment. Representative images are shown from three replicates derived from a visual screening of 24 stem sections (eight stem sections of three independent biological replicates per line) for each condition tested. Scale bars, 200 µm.

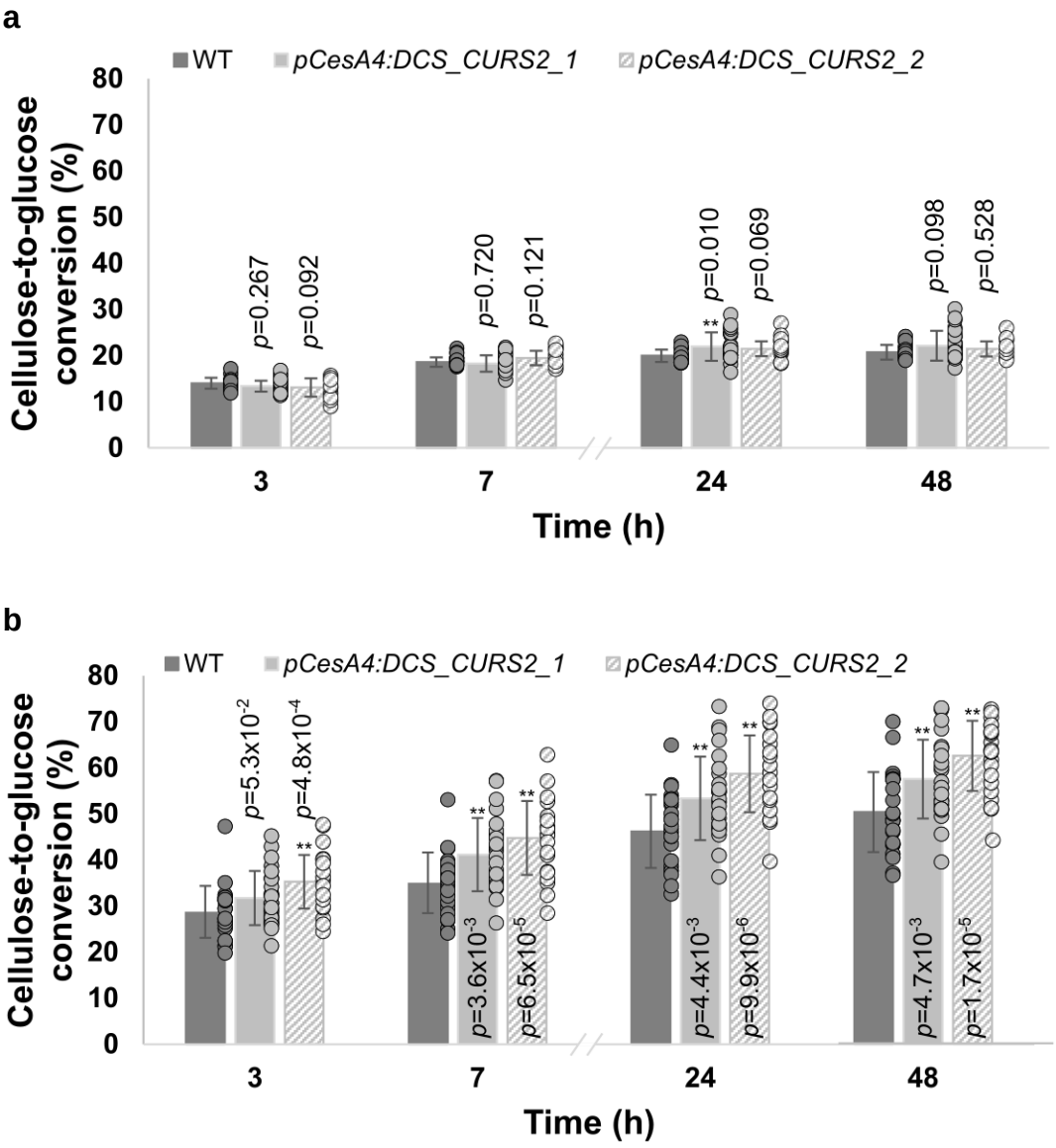
**a**



b


WT
pCesA4:DCS_CURS2
Fluorescence λex 488 nm
Fluorescence λex 365 nm
Non treated
Treated
Non treated
Treated

**Supplementary Figure 11. Cellulose-to-glucose conversion during saccharification of WT and *pCesA4:DCS_CURS2* cell wall material.** Samples were saccharified using (**a**) no or (**b**) alkaline pretreatment (62.5 mM NaOH, 90ºC, 3 h). Individual values (dots) and means (bars) of 24 biological replicates for all lines are shown. ***P* < 0.01 (two-tailed Student's *t*-test) the exact *P* value is shown. Error bars indicate standard deviation.



**Supplementary Table 1. Cell wall composition of the transgenic lines and WT plants.** CWR was determined as the fraction of material obtained after washes relative to the original dry weight. Lignin content was determined by acetyl bromide (AcBr) and Klason, and expressed as a percentage of CWR (*n* = 8 independent biological replicates for all lines). Lignin composition was determined in enzyme lignin by 2D HSQC NMR (*n* = 3 pools of each 5 independent biological replicates for all lines). Crystalline cellulose content was determined by the Updegraff method, while alpha-cellulose (i.e. the sum of crystalline and amorphous cellulose) was determined via an acetic acid/nitric acid-based protocol (see Methods) (*n* = 8 pools of each 6 independent biological replicates for all lines). Both are expressed as a percentage of CWR. No statistical differences were observed between transgenic and control lines (*P* > 0.05; two-tailed Student's *t*-test). stdev, standard deviation.

| | **WT mean ± stdev** | ***pCesA4:DCS_CURS2_1* mean ± stdev (*P* value)** | ***pCesA4:DCS_CURS2_2* mean ± stdev (*P* value)** |
|---|---|---|---|
| **CWR (% dry weight)** | 77.5 ± 3.2 | 77.0 ± 2.4 (0.747) | 76.9 ± 1.2 (0.643) |
| **Lignin** | | | |
| **AcBr lignin (% CWR)** | 11.2 ± 1.7 | 11.6 ± 1.1 (0.577) | 11.2 ± 0.3 (0.967) |
| **Klason lignin (% CWR)** | 16.1 ± 0.4 | 17.1 ± 2.0 (0.442) | 15.8 ± 0.9 (0.644) |
| **Aromatic units (%)** | | | |
| **%S** | 24.2 ± 1.2 | 24.2 ± 0.6 (0.551) | 24.8 ± 0.3 (0.911 |
| **%G** | 74.8 ± 1.1 | 74.5 ± 0.8 (0.681) | 73.9 ± 0.4 (0.534) |
| **%H** | 1.0 ± 0.2 | 1.3 ± 0.3 (0.214) | 1.3 ± 0.1 (0.132) |
| **Inter-unit linkages (%)** | | | |
| **β-aryl ether** | 83.0 ± 1.7 | 84.7 ± 1.2 (0.316) | 85.7 ± 0.5 (0.163) |
| **Phenylcoumaran** | 10.5 ± 1.4 | 9.5 ± 0.9 (0.436) | 8.4 ± 0.2 (0.131) |
| **Resinol** | 6.5 ± 0.4 | 5.8 ± 0.2 (0.076) | 5.9 ± 0.3 (0.299) |
| **Cellulose** | | | |
| **Crystalline cellulose (%CWR)** | 37.7 ± 2.1 | 39.1 ± 1.7 (0.174) | 37.8 ± 1.2 (0.885) |
| **Alpha-cellulose (%CWR)** | 45.6 ± 2.6 | 45.8 ± 2.4 (0.875) | 44.9 ± 1.8 (0.521) |

**Supplementary Table 2. GPC chromatographic comparison of acetylated lignin from *pCesA4:DCS_CURS2* and WT inflorescence stems.** Mean ± standard deviation of number-average molecular weight (Mn), weight-average molecular weight (Mw), and polydispersity index (PDI) estimated from the GPC curves (integrated between 80,000 and 600 Da). n = 3 pools of each 5 independent biological replicates for all lines. No significant differences were observed between the lines ($P > 0.05$; two-tailed Student's *t*-test).

| Line | Mw (Da) | (*P* value) | Mn (Da) | (*P* value) | PDI (Mw/Mn) | (*P* value) |
|---|---|---|---|---|---|---|
| WT | 30,019 ± 1,330 | | 10,249 ± 604 | | 2.93 ± 0.06 | |
| *pCesA4:DCS_CURS2_1* | 28,582 ± 3,340 | (0.602) | 9,636 ± 630 | (0.377) | 2.96 ± 0.15 | (0.828) |
| *pCesA4:DCS_CURS2_2* | 31,071 ± 1,292 | (0.468) | 10,430 ± 206 | (0.708) | 2.98 ± 0.07 | (0.516) |