

1 **Title**

2 It happened again: convergent evolution of acylglucose specialized metabolism in black
3 nightshade and wild tomato

4
5 **Authors**

6 Yann-Ru Lou¹, Thilani M. Anthony¹, Paul D. Fiesel¹, Rachel E. Arking³, Elizabeth M.
7 Christensen², A. Daniel Jones¹, and Robert L. Last^{1,2*}

8
9 **Affiliations**

10 ¹Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA.

11 ²Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.

12 ³Department of Neuroscience, East Lansing, MI 48824, USA.

13 *Corresponding author. Email: lastr@msu.edu

14
15 **Abstract**

16 Plants synthesize myriad phylogenetically-restricted specialized (aka ‘secondary’)
17 metabolites with diverse structures. Metabolism of acylated sugar esters in epidermal glandular
18 secreting trichomes across the Solanaceae (nightshade) family are ideal for investigating the
19 mechanisms of evolutionary metabolic diversification. We developed methods to structurally
20 analyze acylhexose mixtures by 2D NMR, which led to the insight that the Old World species
21 black nightshade (*Solanum nigrum*) accumulates acylglucoses and acylinositols in the same tissue.
22 Detailed *in vitro* biochemistry – cross validated by *in vivo* virus induced gene silencing – revealed
23 two unique features of the four-step acylglucose biosynthetic pathway: a trichome-expressed,
24 neofunctionalized invertase-like enzyme, SnASFF1, converts BAHD-produced acylsucroses to
25 acylglucoses, which in turn are substrates for the first-reported acylglucose acyltransferase,
26 SnAGAT1. This biosynthetic pathway evolved independently from that recently described in the
27 wild tomato *S. pennellii*, reinforcing that acylsugar biosynthesis is evolutionarily dynamic with
28 independent examples of primary metabolic enzyme cooption and additional variation in BAHD
29 acyltransferases.

30
31 **Teaser**

32 Analysis of plant protective surface hair chemistry revealed evolutionary mechanisms leading to
33 metabolic innovation.

34
35
36
37
38
39
40

41 **MAIN TEXT**

42

43 **Introduction**

44 Plants synthesize myriad structurally diverse specialized (historically referred to as
45 ‘secondary’) metabolites, which provide humans with medicines, food additives and natural
46 insecticides. In recent years, an increasing number of these compounds were documented to serve
47 ecological functions. Specialized metabolites group into taxonomically-restricted classes, such as
48 glucosinolates in Brassicales and benzoxazinoid alkaloids in Poaceae (1–4). Because the full
49 diversity of specialized metabolites is distributed across the Kingdom, the vast majority of plant
50 specialized metabolites are uncharacterized. In contrast to the typically conserved enzymes and
51 pathways of primary metabolism, variation of the lineage-specific specialized metabolic pathways
52 is observed within and across species (2, 5–7). In fact, specialized metabolic pathway enzymes
53 have characteristics that promote product diversity. These include promiscuity (seen for example
54 in BAHD (BEAT, AHCT, HCBT, and DAT) acyltransferases), gene duplication (e.g. cytochromes
55 P450) and neo- and sub-functionalization of gene expression and enzyme activity.

56 Glandular-secreted trichome acylsugars are a structurally diverse group of specialized
57 metabolites produced in the nightshade (Solanaceae) family that have developed into an exemplary
58 system for investigating the emergence and diversification of novel metabolites in plants (8–16).
59 Their anti-herbivory and anti-microbial activities make them of interest in understanding their
60 ecological roles and as breeding targets for crop protection (17–20). Most characterized acylsugars
61 consist of a sucrose, glucose or inositol sugar core esterified at various positions with acyl chains
62 that are often aliphatic, differing in length – typically C2-C12 (8, 16, 21–29). A single species in
63 Solanaceae can produce dozens of distinct acylsugars consisting of these simple biosynthetic
64 building blocks (14, 28).

65 Sugar esters with a sucrose core (acylsucroses) are predominant in Solanaceae (8, 14). The
66 well-characterized biosynthetic pathway producing cultivated tomato (*Solanum lycopersicum*)
67 acylsucroses involves four trichome-expressed BAHD-family acylsugar acyltransferases
68 (ASAT1, ASAT2, ASAT3 and ASAT4) (9–11). Evolutionarily-related ASATs were characterized
69 in acylsugar biosynthetic pathways in species across the family, with all but one using sucrose or
70 acylsucroses as acceptor substrate (12–14, 16). These enzymes sequentially transfer acyl chains
71 from acyl-Coenzyme A substrates (acyl-CoAs) onto specific positions of the sugar core, with
72 individual ASATs varying in both acyl donor and acyl acceptor substrates. Substrate specificity

73 and promiscuity of these enzymes influence inter- and intra-specific diversification of acyl chain
74 length and acylation position as reported in *Petunia axillaris*, *Salpiglossis sinuata*, and wild tomato
75 species (12–14).

76 While the majority of characterized acylsugars are based on sucrose, acylhexoses have also
77 been described. Specifically, acylglucoses were reported in species within the *Solanum*, *Datura*
78 and *Nicotiana* genera (21, 27, 29–31), while acylinositols were characterized from the South
79 American native fruit crop naranjilla (*Solanum quitoense*) and the Central American native
80 orangeberry nightshade (*Solanum lanceolatum*) (16, 23). Relatively little is reported about
81 acylhexose biosynthesis, with one published example of a BAHD acylinositol acetyltransferase in
82 *S. quitoense* (16), and no reported glucose or acylglucose acyltransferases. In fact, we recently
83 showed that wild tomato *S. pennellii* acylglucoses are produced by acylations of sucrose followed
84 by conversion to acylglucoses by a trichome-expressed invertase-like enzyme, acylsucrose
85 fructofuranosidase 1 (SpASFF1) (15). This enzyme utilizes the acylsucrose substrates with all acyl
86 chains on the pyranose ring (P-type acylsucroses) found in *S. pennellii* but not in cultivated tomato
87 or other wild tomato relatives. Hence, this ASFF1-dependent pathway evolved independently in
88 the lineage of *S. pennellii*. This leaves open the question of how acylglucose biosynthesis proceeds
89 outside the tomato sub-clade of the *Solanum* genus.

90 To address this problem, we investigated trichome acylsugar specialized metabolism in the
91 predominantly Eurasian black nightshade species, *S. nigrum*. In contrast to the New World species
92 *S. pennellii*, which produces both acylglucoses and acylsucroses, our previous work suggested that
93 the trichomes on *S. nigrum* young leaves, stems and fruits exclusively accumulate acylhexoses
94 (14). Moreover, while other characterized species accumulate acylsugars with three or more acyl
95 chains, *S. nigrum* accumulates acylhexoses with as few as two acyl chains.

96 Understanding the enzymology behind acylhexose biosynthesis requires accurate
97 assessment of acylation sites on sugar cores, for which mass spectrometry yields limited
98 information, but Nuclear Magnetic Resonance (NMR) spectroscopy is well suited. Despite the
99 power of modern methodologies, obtaining NMR-resolved chemical structures requires samples
100 of sufficient abundance and purity. As a result, while we have liquid chromatography–mass
101 spectrometry (LC–MS) data for hundreds of structurally diverse acylsugars, only a subset have
102 unambiguous structure annotations. For example, of more than 38 surveyed Solanaceae species

103 that accumulate acylsugars, fewer than half have NMR structures elucidated for at least one
104 acylsugar (14, 16, 21–26, 28, 29).

105 Here we report the structural characterization of *S. nigrum* acylhexoses – consisting of
106 triacylinositols, diacylglucoses and triacylglucoses – and describe characterization of the *S. nigrum*
107 acylglucose biosynthetic pathway. Our accelerated structure elucidation pipeline provided NMR-
108 resolved structures for *S. nigrum* acylsugars from leaf-surface extracts without purifying individual
109 compounds. A combination of *in vitro* biochemical pathway reconstruction and *in vivo* virus-
110 induced gene silencing (VIGS) validated the acylglucose biosynthetic pathway. This pathway
111 consists of four trichome-expressed enzymes. The two acylsucrose acyltransferases SnASAT1 and
112 SnASAT2 produce diacylsucroses, which serve as substrates for an acylsucrose fructofuranosidase
113 1 (SnASFF1), which synthesizes diacylglucoses. Triacylglucoses are then produced by SnAGAT1.
114 While it is analogous to the *S. pennellii* acylglucose pathway, the *S. nigrum* pathway evolved
115 independently, with co-option of the neofunctionalized SnASFF1 from a distinct lineage of the
116 invertase gene phylogeny, and evolution of a BAHD acyltransferase that acts on acylglucoses.
117 This work demonstrates the value of leveraging analytical chemistry, phylogenetics, enzymology
118 and knowledge gained from model organisms to facilitate chemical structure assignment and
119 biochemical pathway elucidation in a non-model organism.

120

121 **Results**

122 ***S. nigrum* accumulates di- and triacylated acylglucoses and triacylated acylinositols**

123 Understanding the atomic connections of products and metabolic intermediates is critical
124 for biochemical pathway elucidation. NMR-resolved structures of purified compounds are the gold
125 standard for small molecule metabolites, but obtaining these data can be time-consuming. We
126 sought to speed up the process by subjecting total extract or partially purified mixtures to 2D NMR
127 analyses, taking advantage of previously developed acylsugar analysis methods (28, 29, 32) (Fig.
128 1C).

129 As a first step to characterize the number and diversity of acylsugars, we profiled
130 metabolites in *S. nigrum* leaf-dip extracts using LC-MS (Fig. 1A). We annotated 99 metabolite
131 features – among them 45 with high confidence based on criteria discussed in the Supplementary
132 Method – with characteristics similar to known acylsugars based on molecular ion adduct masses,
133 mass spectra at elevated collision energies and fractional hydrogen content as quantified by relative

134 mass defect (32, 33) (Fig. 1A, Fig. S1 and S2, Table S1 and Supplementary Text). In contrast to
135 the majority of Solanaceous plants, we only detected acylhexoses in *S. nigrum* extracts (14). LC-
136 MS fragmentation patterns revealed that the acylsugars are typically composed of hexose cores
137 and two or three short- (C2, C4, C5) or medium-chain length (C8, C9, and C10) acyl esters (Table
138 S1 and Supplementary Text). Thirty-nine of the 45 high-confidence putative acylsugars showed
139 fragmentation patterns similar to previously analyzed *S. pennellii* acylglucoses (29). However, we
140 observed six triacylhexoses with unusually low fragment ion signals in negative mode (Fig. S2);
141 this is a characteristic reminiscent of *S. quitoense* acylinositols (16, 34).

142 The acylglucose- and acylinositol-like peaks in *S. nigrum* appear as groups of two or more
143 chromatographic peaks sharing identical molecular masses and indistinguishable mass spectra at
144 high collision energy as consistent with multiple isomeric forms (Fig. S3). Extracted ion
145 chromatograms for fragment ion masses characteristic of acyl groups (carboxylate anions)
146 revealed that many individual chromatographic peaks result from multiple co-eluting metabolite
147 features, some of which present indistinguishable mass spectra. For example, the second-eluting
148 peak annotated as H3:16 (Fig. S3) exhibits asymmetry suggestive of four isomers, two of which
149 overlap in chromatographic retention. All four components share the same acyl chain complements
150 – C2, C4 and C10 acyl chains on a hexose core as evidenced from the extracted fragment ion
151 chromatograms that confirm C4 and C10 acyl groups (Fig. S3). We expect each acylglucose to
152 resolve as two peaks corresponding to α and β anomers that interconvert, accounting for the
153 presence of two peaks for each (29). The observation of additional peaks led to the hypothesis that
154 four isomeric forms reflected different acyl chain branching patterns rather than differences in acyl
155 group lengths or positions of esterification. This is supported by the ethyl ester profile revealed by
156 gas chromatography-MS (GC-MS) after saponification and transesterification (Fig. 1B). The ethyl
157 ester profile demonstrated both abundant straight (nC8, nC9, nC10) and branched (iC4, iC5, aiC5,
158 branched C8, branched C9, branched C10) aliphatic chains (Fig. 1B), supporting the conclusion
159 that acyl chain branching variation contributes to the number of structural isomers.

160 The combined information gained from LC-MS and GC-MS led us to conclude that *S.*
161 *nigrum* accumulates at least 45 different types of di- and triacylhexoses including structural
162 isomers with acyl chains differing in branching patterns. While unequivocal structural information
163 of these acylhexoses is essential for a complete understanding of their biosynthesis, the
164 chromatography-based analytical techniques do not provide needed resolution or selectivity. For

165 example, MS has not yet developed technologies that identify acylation positions or differentiate
166 isomeric hexose cores (e.g. glucose or inositol). Such information is central to understanding
167 functions of key metabolic enzymes that shape acylsugar chemical diversity. We employed NMR
168 spectroscopy, which provides important information about molecular topology, to obtain these
169 structural details.

170 We sought to avoid the time-consuming process of purifying metabolites to homogeneity
171 by analyzing unfractionated and partially purified fractions (Fig. 1C). This took advantage of the
172 2D NMR spectroscopy methods Heteronuclear Single Quantum Coherence (HSQC), which
173 provides information about hydrogen-carbon attachments, and Heteronuclear Multiple Bond
174 Correlation (HMBC), which yields information about atoms separated by 2-3 bonds. Analysis of
175 the total, unfractionated acylsugar extracts from young plants revealed the dominant presence of a
176 glucose backbone with acyl chains at the 2-, 3- and 4-positions (Fig. 1D). However, abundant
177 triacylglucoses obscured the signals of diacylglucoses and acylinositols in the total extract: we
178 employed partial purification to obtain similar information about these less abundant acylsugars.

179 Silica gel chromatography was used to separate triacylglucoses (fraction 1) from
180 diacylglucoses and triacylinositols (fraction 2) based largely on the number of hydroxyl groups:
181 there are two hydroxyls on triacylglucoses; three on diacylglucoses and triacylinositols (Fig. 2A-
182 C). Analysis of the NMR spectra from fraction 1 confirmed that it predominantly contains
183 triacylglucoses with one medium-length (8-10 carbon) acyl chain esterified to the 4-position, a
184 short (4-5 carbon) acyl chain at the 3-position, and an acetyl (C2) group at the 2-position (Fig. 2D
185 and Table S2). In agreement with LC-MS data, NMR spectra of fraction 2 exhibited signals of two
186 distinct groups of acylhexoses in the sugar region ($\delta_H = 3-6$ ppm; $\delta_C = 60-100$ ppm) (Fig. S4). The
187 anomeric positions in hexoses stand out in having larger (downfield) ^{13}C chemical shifts ($\delta_C \sim 90-$
188 100 ppm) than other acylsugar carbon atoms (Fig. S4). We employed 2D HSQC-TOtal Correlation
189 Spectroscopy (HSQC-TOCSY), which generates spectra at long mixing times to provide
190 connectivity information between coupled nuclei in a molecule (35). Such information aids in
191 associating sugar core hydrogen and carbon signals within a molecule and distinguishing signals
192 from other molecules in a mixture. This approach provided a useful strategy for distinguishing
193 acylation positions on acylglucoses and acylinositols in fraction 2 where signals often overlap. Our
194 analysis revealed acylations at the 3- and 4-positions on a sugar core with an anomeric center
195 (glucose) and acylations at 2-, 3-, and 4- positions on a hexose core with no anomeric center

196 (inositol) (Fig. 2E and Fig. S4). HMBC data analysis confirmed the presence of abundant
197 diacylglucoses with a medium-length acyl chain as R₄ and a short acyl chain as R₃ (Fig. 2F and
198 Table S3). This approach also revealed the presence of triacylinositols with two short- and one
199 medium-length acyl chain as R₂, R₃ and R₄ (Fig. 2G and Table S4).

200 Analysis of the combined LC-MS, GC-MS and NMR results led to the conclusion that *S.*
201 *nigrum* accumulates diacylglucoses esterified at the 3- and 4-positions, with triacylglucoses and
202 triacylinositols esterified at the 2-, 3-, and 4- positions. For the remainder of this report we follow
203 the nomenclature established in Schillmiller et al. (2010) to describe these compounds:
204 SX:Y(A^a,B^b,C^c), in which **S** indicates the type of sugar core (**G** for glucose, **I** for inositol), **X** is the
205 total number of substituting acyl chains, **Y** signifies the sum of acyl chain carbons and (A^a,B^b,C^c)
206 represents the number of carbons of each chains with the superscript documenting the site of
207 substitution. Based on this nomenclature, G2:14(4^{R3},10^{R4}), G3:16(2^{R2},4^{R3},10^{R4}), and
208 I3:17(4^{R4},5^{R2},8^{R3}) represent some of the most abundant acylsugars in each class (Fig 2 D, F and
209 G). In cases where the identity of the hexose core remains ambiguous, the letter **H** for hexose is
210 employed.

211

212 **Identification of a trichome-expressed β-fructofuranosidase involved in *S. nigrum* acylsugar** 213 **biosynthesis**

214 The observation that acylglucoses are predominant in *S. nigrum* trichome extracts led us to
215 posit that, as previously shown for *S. pennellii* LA0716 acylglucose biosynthesis, acylsucroses are
216 intermediates in acylglucose production. To test the hypothesis, we identified trichome-expressed
217 transcripts in *S. nigrum* predicted to encode proteins of the glycoside hydrolase family 32, which
218 includes SpASFF1. Twenty invertase-like homologs were identified by BLAST from *S. nigrum*
219 RNAseq data (Fig. S5A). We reasoned that – as acylsugars appear to accumulate on *S. nigrum*
220 trichome tip cells (Fig. S5B) – the acylglucose biosynthetic genes in *S. nigrum* would be highly
221 expressed and enriched in trichomes. Among the twenty homologs, six demonstrate trichome-
222 enriched expression (Fig. S5C). Five of the six have sequence lengths comparable to functional
223 invertases, and have WXNDPNG, RDP and EC sequences characteristic of β-fructofuranosidases
224 (Fig. S5D). Because we previously observed a **PSTP** non-canonical substrate binding site in
225 SpASFF1(15), we sought invertase-like sequences also lacking the canonical **DXXXK** in *S. nigrum*
226 trichome expressed transcripts. Indeed, assembly c70979_g1 has **PLTY** in place of that conserved

227 substrate binding site (Fig. S5D). Based on the *in vivo* transient silencing and *in vitro* enzyme
228 activities results described below, we designated this protein SnASFF1.

229 We developed and deployed *S. nigrum* VIGS to test the hypothesis that SnASFF1 is
230 involved in *S. nigrum* acylglucose biosynthesis *in vivo*. Plants with silenced *SnASFF1* accumulated
231 diacylsucroses (referred to as ‘*in vivo* diacylsucroses’ below) that were not detected in the control
232 plants (Fig. 3, Fig. S5E and F and Table S5). NMR spectra of the early- and late-eluting
233 S2:14(4,10) isomers revealed that the acylation positions – on positions 3 and 4 (Fig. S5G) – mirror
234 those of *S. nigrum* acylglucoses (Fig. 2D and F), supporting their role as intermediates in the
235 acylglucose biosynthetic pathway. As characterized ASATs show acylation position specificity,
236 this result is consistent with the hypothesis that there are two or more BAHD acyltransferases that
237 produce diacylsucrose substrates for this SnASFF1 invertase-like enzyme.

238

239 **Identification of trichome-specific BAHD acyltransferases**

240 Based on published results with acylsucrose- and acylinositol-producing plants in
241 Solanaceae, we hypothesized that acyltransferase enzymes of *S. nigrum* acylsugar biosynthesis
242 would be trichome-enriched and evolutionarily related to characterized ASATs (14, 16). The
243 combined tissue-specific expression and homology-guided approach led us to consider six highly
244 trichome-enriched BAHD family members in the *S. nigrum* transcriptome (Fig. 4A and Fig. S6).
245 These candidates are predicted to encode proteins with characteristics of active BAHD family
246 enzymes (36). First, they have the conserved HXXXD catalytic motif and the DFGWG-like
247 structural motif (Fig. S7). Second, the predicted proteins all have lengths comparable to
248 characterized functional BAHDs (8–16) (Fig. S6B).

249 We tested the hypothesis that these genes encode proteins of acylsugar biosynthesis by
250 examining the activities of *Escherichia coli*-expressed and His-tag purified enzymes. C10, C8,
251 iC5, iC4 and C2-CoAs were chosen as acyl chain donors in our *in vitro* assays, based on the
252 abundance of these ester groups in our structural annotations (Fig. 1B). Using LC-MS, activity was
253 detected with sucrose and the closest *S. nigrum* homolog of tomato ASAT1 (SlASAT1), which is
254 encoded by the transcript *c63608_g1* (Fig. 4B and Fig. S8). This enzyme catalyzes the formation
255 of monoacylsucroses – S1:8 and S1:10 – from sucrose and nC8, iC10 and nC10 acyl-CoAs (Fig.
256 4B and Fig. S8). The apparent K_m values for SnASAT1 and nC8- and nC10-CoAs are 114 ± 21
257 and 11.7 ± 2.1 (μM), respectively (Fig. S9A). An apparent K_m of 5 mM was measured for SnASAT1

258 and the acceptor substrate sucrose (Fig. S9A). Based on its *in vitro* activity and the *in planta* results
259 presented below, we named this enzyme SnASAT1. SnASAT1-produced S1:10(nC10) co-eluted
260 with SlASAT1-produced S1:10(nC10^{R4}) but not with SsASAT1-produced S1:10(nC10^{R2}) (Fig.
261 S9B) (10, 11, 13, 14). Positive-mode MS fragmentation supports localization of the C10 chain on
262 the six-member pyranose ring (Fig. S9C). This result was extended by NMR analysis, revealing
263 S1:10(nC10^{R4}) as the major product of SnASAT1 *in vitro* enzyme assays (Fig. 4E and Fig. S9D).

264 A second, chromatographically separable S1:10 isomer, S1:10(10^{R6}), accumulated in
265 SnASAT1 enzyme assays (Fig. S10A and S9D). This 6-position acylated isomer appeared as a
266 minor product in assays carried out at pH 6.0 with short incubation periods (5-30 min) (Fig. S10A).
267 The concentration of this second isomer increases after extended incubation time, especially under
268 neutral-to-alkaline pH conditions (Fig. S10B-D). For example, comparable amounts of
269 S1:10(10^{R4}) and S1:10(10^{R6}) accumulated after 30 min at 30°C under pH 8.0 (Fig. S10C). We also
270 observed non-enzymatic conversion of S1:10(10^{R4}) to S1:10(10^{R6}) after brief exposure to elevated
271 temperature (65°C) (Fig. S10B and D) or extended incubation of purified S1:10(10^{R4}) in
272 unbuffered distilled water at room temperature. Non-enzymatic chain migration was previously
273 documented for a R₄ monoacylsucrose produced by SlASAT1, where the short acyl chain moved
274 from R₄ to R₆ (11).

275 We characterized the activity of trichome-expressed *c65670_g1*-encoded SnASAT2, the
276 closest homolog of tomato SlASAT2 and SpASAT2 (Fig. 4A). SnASAT2 utilized iC4-CoA as
277 acyl-chain donor to decorate a second ring position on the pyranose ring of both S1:10(nC10^{R4})
278 and S1:10(iC10) (Fig. 4CD and Fig. S11A). MS fragmentation patterns and retention times of the
279 produced S2:14(iC4,iC10) and S2:14(iC4,nC10) (referred to below as ‘*in vitro* diacylsucroses’)
280 were indistinguishable from the early- and late-eluting *in vivo* S2:14(4,10) metabolites,
281 respectively (Fig. 4D and Fig. S11). We observed broad acyl donor and acceptor substrate
282 specificity with iC4-, iC5- and nC8- serving as substrates with both S1:10(10^{R4}) and S1:8(8)
283 acceptors, producing diacylsucroses chromatographically indistinguishable from those detected in
284 invertase-like deficient VIGS-*SnASFF1* lines (Fig. 4E and Fig. S12A). In contrast, no products of
285 SnASAT2 were detected with unacylated sucrose, glucose, the SsASAT1 product S1:10(10^{R2}), or
286 the S1:10(10^{R6}) SnASAT1 rearrangement product as acceptor (Fig. S12B-D). As seen with
287 SnASAT1-generated monoacylsucroses, a second diacylsucrose isomer accumulated in a pH- and
288 time-dependent manner after SnASAT1 + SnASAT2 reaction (Fig. S13). Together with the

289 identification of SnASFF1, these results are consistent with the hypothesis that S2:14(4,10) from
290 SnASAT1 + SnASAT2 sequential reactions are precursors of the *S. nigrum* di- and triacylglycerols
291 (e.g. G2:14(4^{R3},10^{R4}) and G3:16(2^{R2},4^{R3},10^{R4})).

292

293 **SnASFF1 is an acylsucrose hydrolase that evolved independently from SpASFF1**

294 We tested SnASFF1 protein activity on *in vitro*- and *in vivo*-produced diacylsucroses, using
295 recombinant His-tagged SnASFF1 protein expressed and purified from the *Nicotiana benthamiana*
296 transient expression system(15). Indeed, SnASFF1 hydrolyzed all three tested diacylsucroses: the
297 two *in vivo* pathway intermediates and the unrearranged *in vitro* diacylsucrose product generated
298 by SnASAT1 and SnASAT2 sequential reactions (Fig. 5A and Fig. S14A). The reactions with *in*
299 *vitro* and *in vivo* S2:14(4,10) isomers each yielded two major products that are the anomers of *S.*
300 *nigrum* G2:14(4^{R3},10^{R4}), as suggested by their identical retention times and mass spectra (Fig. 5A
301 and Fig. S14B). In fact, the G2:14(4,10) isomers produced from *in vitro* S2:14(iC4,nC10) and *in*
302 *vivo* late-eluting S2:14(4,10) are indistinguishable from the late-eluting anomers of *S. nigrum*
303 G2:14(4^{R3},10^{R4}). Additionally, SnASFF1 converted the *in vivo* early-eluting S2:14(4,10) into
304 anomers co-eluting with those of the early-eluting *S. nigrum* G2:14(4^{R3},10^{R4}) (Fig. 5A and Fig.
305 S14B). We hypothesize that these two sets of diacylglycerols, differing in an nC10 or iC10 acyl
306 chain as R₄, collectively explain the overlapping signals of *S. nigrum* G2:14(4^{R3},10^{R4}) (Fig. 5A).
307 This result is consistent with the presence of abundant nC10 and iC10 in the acylsugar-derived
308 ethyl ester profile (Fig. 1B). Taken together, these results provide strong evidence that SnASAT1,
309 SnASAT2 and SnASFF1 are sufficient to reconstitute production of diacylglycerols produced in *S.*
310 *nigrum* from sucrose and acyl-CoAs.

311 SnASFF1 hydrolytic activity was not detected using sucrose, in contrast to the canonical
312 yeast invertase activity (Fig. S15A). Conversely, yeast invertase failed to use any tested
313 S2:14(4,10) as substrate (Fig. S15B). These data are consistent with the hypothesis that SnASFF1
314 has acylsucrose-specific activity. We also observed SnASFF1 hydrolytic activity with *in vitro*-
315 produced S1:10(10^{R4}) and the *S. pennellii* triacylsucrose S3:18(4^{R2},4^{R4},10^{R3}) as substrate (Fig.
316 S15B). However, the monoacylglycerol product, G1:10(10), does not appear to be an acceptor
317 substrate for SnASAT2 with nC10-, nC8-, iC5-, iC4- and acetyl-CoA (Fig. S12A). Taken together,
318 the results support the hypothesis that SnASFF1 is an acylsucrose hydrolase that contributes to *S.*
319 *nigrum* acylglucose biosynthesis by converting diacylsucroses to diacylated glycerols.

320 To explore the evolution of acylglucoses in *S. nigrum* and wild tomato, we generated a
321 gene tree with invertase-like genes and transcripts from *S. lycopersicum*, *S. pennellii* and *S. nigrum*
322 (Fig. 6). Phylogenetic analysis showed that the *S. nigrum ASFF1* is a close homolog of the tomato
323 *LIN5* (Solyc09g010080) and *LIN7* (Solyc09g010090) paralogs – tandem duplicates of sucrose-
324 specific invertases. In contrast, the *S. pennellii ASFF1* is in a clade distinct from *SnASFF1*, with
325 greater similarity to tomato Solyc03g121680 and Solyc06g064620 and the two *S. nigrum*
326 transcripts, *c65240_g1* and *c62944_g1*. We observed no acylsugar phenotype upon *in planta*
327 silencing *c65240_g1*, the closest BLAST hit of SpASFF1 in *S. nigrum* transcriptome (Fig. S15B).
328 This result is consistent with the low *S. nigrum* trichome expression of *c65240_g1*. Together, these
329 relationships indicate that the acylsugar hydrolyzing activity of SnASFF1 and SpASFF1 evolved
330 independently in *S. nigrum* and the wild tomato, *S. pennellii*.

331

332 **Identification of a trichome-specific BAHD acylglucose acetyltransferase**

333 The *in vitro* production of diacylglucoses by sequential reaction of SnASAT1, SnASAT2
334 and SnASFF1 is consistent with the hypothesis that the acetylation step leading to *S. nigrum*
335 triacylglucoses is carried out on diacylglucose substrates. Indeed, the search among the remaining
336 trichome-enriched ASAT candidates led to identification of acetylation by SnAGAT1 (*S. nigrum*
337 **acylglucose acyltransferase 1**; encoded by *c65306_01*), as the last step of triacylglucose
338 biosynthesis in *S. nigrum*.

339 As hypothesized, SnAGAT1 acetylated the diacylglucoses produced by SnASFF1
340 activities with *in vitro* and *in vivo* diacylsucrose S2:14(4,10) substrates (Fig. 5B). The
341 G3:16(2,4,10) anomers obtained from both *in vitro* S2:14(iC4,nC10) and *in vivo* late-eluting
342 S2:14(4,10) have retention times and mass spectra identical to the late-eluting peaks of *S. nigrum*
343 G3:16(2^{R2},4^{R3},10^{R4}) (Fig. 5B and Fig. S16A). SnASFF1 and SnAGAT1 sequential activities with
344 *in vivo* early-eluting S2:14(4,10) – which shares identical retention times and mass spectra with *in*
345 *vitro* generated S2:14(iC4,iC10) – generated early-eluting triacylglucose anomers (Fig. 5B and
346 Fig. S16A). We hypothesize that these two sets of triacylglucoses, with an nC10 or iC10 chain,
347 collectively explain the multiple peak signals of *S. nigrum* G3:16(2^{R2},4^{R3},10^{R4}) (Fig. 5B and Fig.
348 S3). This result is consistent with the abundance of nC10 and iC10 acyl chains on *S. nigrum*
349 acylsugars (Fig. 1B). We also observed SnAGAT1 activity with the diacylglucose-rich *S. nigrum*
350 extract fraction 2 as acceptor substrates and acetyl-CoA as donor, leading to accumulation of

351 triacylglycerols that are indistinguishable to *S. nigrum* triacylglycerols (Fig. S16C). These assays
352 provide strong evidence that SnAGAT1 catalyzes the last step in *S. nigrum* triacylglycerol
353 production (Fig. 5C).

354

355 ***In vitro* validation of the *S. nigrum* acylglucose biosynthetic pathway**

356 We independently assessed the functions of the three BAHD enzymes on the reconstructed
357 pathway by taking advantage of BAHD acyltransferase reversibility in the presence of free
358 coenzyme A. Using plant-derived acylglucosyls as substrates for reverse enzyme assays, we
359 verified that SnAGAT1 and free coenzyme A converts *S. nigrum* extract fraction 1 triacylglycerols
360 into diacylglycerols (Fig. 7A). The ion masses and mass spectra of these diacylglycerols are
361 consistent with loss of an acetyl group (Fig. S16B), supporting the designation of SnAGAT1 as an
362 acyltransferase (Fig. 5C and 7B).

363 SnASAT2 was tested for reverse activity with both early- and late-eluting *in vivo*
364 S2:14(4,10) from VIGS-*SnASFF1* lines and the major and minor SnASAT1 + SnASAT2 *in vitro*-
365 generated S2:14(iC4,nC10) isomers (Fig. 7C). SnASAT2 converted the *in vivo*-derived and the
366 major *in vitro*-produced diacylsucroses to monoacylsucroses S1:10(10) (Fig. 7C). The
367 monoacylsucroses produced from both *in vitro* S2:14(iC4,nC10) and *in vivo* late-eluting
368 S2:14(4,10) are chromatographically indistinguishable from the SnASAT1-forward reaction
369 product S1:10(nC10^{R4}) (Fig. 7C). Addition of SnASAT1 to either of the SnASAT2 reverse
370 reaction assays reduced the accumulation of S1:10(10) (Fig. 7C). In contrast, no reverse activity
371 was detected using the minor *in vitro* diacylsucrose rearrangement product as substrate (Fig. 7C).
372 Similarly, reverse assays using *in vitro*-generated monoacylsucroses as substrates revealed
373 SnASAT1 enzymatic activity with S1:10(10^{R4}), but not with the S1:10(10^{R6}) rearrangement
374 product (Fig. S17). Together, these data provide independent evidence for the *S. nigrum*
375 acylglucose biosynthetic pathway, with SnASAT1, SnASAT2, SnASFF1 producing
376 diacylglycerols, and SnAGAT1 catalyzing the final step in triacylglycerol biosynthesis (Fig. 7B).

377

378 **VIGS validation of three acyltransferases in *S. nigrum* acylglucose biosynthesis**

379 We deployed VIGS to test the hypothesis that SnASAT1, SnASAT2 and SnAGAT1
380 activities are necessary for *in vivo* acylglucose biosynthesis (Fig. 8). As expected for early steps in
381 the core acylsugar biosynthetic pathway, silencing *SnASAT1* and *SnASAT2* each led to reduction

382 of the four most abundant di- and triacylglucoses (Fig. 8A and B, Fig. S18A and B and Table S6
383 and S7). Silencing SnAGAT1 caused statistically significant increases in the ratio of
384 diacylglucoses to corresponding triacylglucoses relative to the controls (Fig. 8C, Fig. S18C and
385 Table S8). These results provide strong support that di- and triacylglucoses are produced *in vivo*
386 via the three- and four-step biosynthetic pathway demonstrated through *in vitro* biochemistry,
387 respectively (Fig. 7B).

388

389 **Discussion**

390 While evolutionary diversification occurs by relatively simple mechanisms of gene
391 duplication, changes in gene expression, and modification of gene product activities, the resulting
392 phenotypic diversity is remarkable. Plant specialized metabolism is a poster child for how a matrix
393 of a relatively small number of chemical feedstocks – generally from core metabolism – and varied
394 enzyme classes, lead to hundreds of thousands of metabolites of diverse structure and function.
395 The sheer number and structural variety of these compounds presents challenges for the structural
396 elucidation essential for rigorous metabolic pathway dissection. Conversely, the ability to use *in*
397 *vitro* pathway reconstruction to test hypotheses regarding pathway evolution is a strength of evo-
398 metabolism.

399 Acylsugar metabolism has become an exemplary system for understanding both the general
400 principals and specific mechanisms by which enzymes are repurposed and combined to generate
401 phenotypic diversity. Here we describe three characteristics of *S. nigrum* acylsugar diversity and
402 metabolism not previously reported for other species. First, although only acylhexoses were
403 detected, they include a mixture of both inositol and glucose esters (Fig. 1 and 2). Second, an
404 **acylglucose** acyltransferase is involved in making triacylglucoses (Fig. 5). Finally, the acylglucose
405 pathway intermediates are produced from a neofunctionalized invertase, which evolved
406 independently from the recently reported wild tomato SpASFF1 triacylsucrose β -fructofuranosidase
407 (Fig. 5 and 6). This work was facilitated by integration of LC- and GC-MS combined with 2-D
408 NMR methods, bypassing a requirement for metabolite purification (Fig. 1).

409

410 **Integrative approaches to elucidate the *S. nigrum* acylglucose biosynthetic pathway**

411 We present four lines of evidence that *S. nigrum* di- and triacylglucoses are synthesized by
412 SnASFF1 and SnAGAT1 from the diacylsucrose intermediates produced by sequential reactions
413 of SnASAT1 and SnASAT2. First, we reconstructed the biosynthetic pathway *in vitro* using

414 purified enzymes, sucrose and acyl-CoAs (Fig. 4 and 5). The acylglucoses generated are
415 chromatographically indistinguishable from those observed in *S. nigrum*, while intermediates of
416 the reconstructed pathway have acylation positions consistent with the *in vivo* final products.
417 Second, genetic evidence supports the involvement of these enzymes in *S. nigrum* acylglucose
418 biosynthesis (Fig. 3 and 8): VIGS silencing of the BAHD acyltransferases *SnASAT1* and *SnASAT2*
419 reduces total acylglucoses, *SnAGAT1* reduces total triacylglucoses, while *SnASFF1* silencing
420 causes accumulation of diacylsucrose pathway intermediates. Third, we cross-validated that the
421 major *in vitro* diacylsucrose intermediates produced by the reconstructed pathway behave
422 chromatographically and biochemically identically to *in vivo* diacylsucrose intermediates purified
423 from VIGS-*SnASFF1* lines (Fig. 4D, 5 and 7C). Notably, SnASFF1 converted both *in vivo* and *in*
424 *vitro* synthesized diacylsucroses to diacylglucose products that can be converted to triacylglucoses
425 by SnAGAT1. In addition, both the di- and triacylglucose anomers produced are indistinguishable
426 from those extracted from *S. nigrum*. While this biochemical and genetic evidence provided strong
427 support for the reconstructed pathway, the observation of mono- and diacylated *in vitro*
428 rearrangement products under pH >6 and elevated temperature (Fig. S10 and S13) led us to seek
429 a fourth line of evidence for pathway validation using reverse BAHD assays.

430 We compared reverse enzymatic products of *in vitro* and *in vivo* intermediates to verify the
431 relevance of all the *in vitro*-produced intermediates. SnASAT2-reverse activity deacylates the *in*
432 *vivo* – and major *in vitro* – diacylsucroses to monoacylsucroses that are indistinguishable from
433 those obtained from SnASAT1 forward assays. All of these monoacylsucrose products are
434 substrates for SnASAT1 reverse activities, supporting the hypothesis that SnASAT1 acylates at
435 the 4-position. The lack of SnASAT2 reverse enzyme activity with *in vitro* minor diacylsucrose is
436 consistent with the hypothesis that the rearranged product is not a biosynthetic intermediate. This
437 result is consistent with NMR analysis suggesting that the 6-position acylation of the medium-
438 length chain on the minor monoacylsucrose products is not observed on *S. nigrum* acylglucoses.
439 Hence, we conclude that the minor *in vitro* artifacts are not intermediates of the *S. nigrum*
440 acylglucose biosynthetic pathway, despite their similar – but not identical – retention times and
441 mass spectra.

442 The reverse activity also enabled us to facilitate pathway discovery. Using abundant plant-
443 derived products, we discovered the first reported acylglucose acyltransferase, SnAGAT1, acting
444 as the last step in *S. nigrum* triacylglucose biosynthesis (Fig. 7A). This acetyltransferase activity

445 was later verified by forward enzyme assays with acetyl-CoA and *in vitro* diacylglucoses produced
446 by the multistep – SnASAT1 + SnASAT2 + SnASFF1 – reactions (Fig. 5B). SnAGAT1 is the third
447 phylogenetically related acetyltransferase in acylsugar biosynthesis, falling in the same BAHD
448 cluster with tomato SlASAT4, *S. quitoense* TAIAT and *S. sinuata* SsASAT5 (Fig 3A). Deploying
449 BAHD reverse enzyme assays accelerated SnAGAT1 identification and allowed *in vitro* pathway
450 reconstruction in the non-model species. In fact, the *S. nigrum* acylglucose biosynthetic pathway
451 is the first completely *in vitro* reconstructed acylhexose pathway in a non-model organism without
452 the powerful genetic resources that enabled the discovery of *S. pennellii* acylglucose biosynthetic
453 pathway.

454

455 **Convergent evolution of acylsucrose hydrolyzing enzymes**

456 Phylogenetic analysis of SnASFF1 and restricted accumulation of acylhexoses across the
457 Solanaceae is consistent with the hypothesis that acylglucose production arose independently in
458 the Old World black nightshade *S. nigrum* and New World wild tomato *S. pennellii* (Fig. 6). One
459 strong line of evidence is that the two ASFF enzymes are members of different clades within the
460 β -fructofuranosidase gene tree (Fig. 6). Second, SpASFF1 – and synthesis of its P-type
461 triacylsucrose substrates (15) – are restricted within a subclade of the tomato Solanum group (12),
462 consistent with this being a derived trait within the New World Solanum. The independent
463 evolution of acylhexose biosynthesis raises the intriguing question of whether there are phenotypic
464 advantages of producing acylsugars with a glucose core. Similarly, the observation of mixed
465 acylsucroses and acylglucoses in *S. pennellii*, and mixed acylglucose and acylinositols in *S. nigrum*
466 begs the question of whether there are synergistic effects of acylglucoses with other acylsugars.

467 Despite co-option from distinct lineages, the two independently evolved ASFFs are both
468 trichome-enriched, neofunctionalized GH32 β -fructofuranosidases, which lack the canonical
469 **DDTK** sucrose-binding pocket of Arabidopsis Cell-Wall Invertase 1 (ACWINV1; AT3G13790)
470 (37–39) (Fig. S4). This is consistent with the observation that neither of the ASFFs utilize un-
471 acylated sucrose as substrate (15) (Fig. S12C). In addition, both ASFFs are expressed in the same
472 tip cells as their acylsucrose substrates (15, 40). It will be interesting to learn whether invertase-
473 like enzymes evolved to produce the acylglucoses reported in *Datura* and *Nicotiana* species (21,
474 22). The diversity of apparent *in vivo* substrates for both ASFFs – diacylsucroses for SnASFF1

475 and triacylsucroses for SpASFF1 – also present opportunities for enzyme structure and function
476 analysis.

477

478 **Streamlined analytical approaches for metabolite annotation**

479 It is remarkable that we continue to discover metabolic variations as new Solanaceae
480 species are analyzed. Because biosynthetic pathway dissection requires knowledge of product and
481 intermediate structures, this process can be a bottleneck. We developed a structural analysis
482 pipeline that reduces the cost and time involved in sample purification. The presence of three
483 distinct acylhexose classes – di- and triacylglucoses and triacylinositols – in *S. nigrum* presented
484 a challenging case study for this integrative approach. We designed the pipeline to leverage the
485 substantial chemical shift library assembled from acylsugar studies in other Solanaceous plants
486 (10, 11, 13, 29, 34, 41).

487 While the pipeline performed as expected for the most abundant triacylsugars in *S. nigrum*,
488 the many overlapping NMR signals in total trichome acylsugar extracts led us to perform partial
489 purification using silica gel chromatography. We obtained clear signals of *S. nigrum*
490 triacylglucoses from the less polar fraction, while 2D HSQC-TOCSY successfully separated
491 proton correlation signals from diacylglucoses and triacylinositols in the high polarity fraction.
492 With the simple partial purification, our NMR approach provided sufficient information to address
493 key characteristics of the enzymes involved in acylsugar biosynthesis without the need to purify
494 individual compounds: (1) recognition of different acylated sugar cores in mixtures and (2)
495 assignments of ester linkage positions on sugar cores. This approach extends the use of 2D ^1H - ^{13}C
496 HSQC NMR spectra beyond previously published characterization of lignin structures (42).
497 Limiting the need for purification is particularly beneficial for investigating anomers and isomers
498 differing in acyl chain branching patterns, where overlapping retention times can frustrate attempts
499 at large-scale purification to homogeneity. As acylsugars are sporadically found in Martyniaceae,
500 Rosaceae, Geraniaceae, Caryophyllaceae and Brassicaceae plants, our approach can further assist
501 with large scale screening for sugar esters beyond Solanaceae (43–47). The vast complexity of
502 specialized metabolism in the plant kingdom will be more efficiently addressed by exploiting the
503 variety of multidimensional NMR strategies to aid in metabolite annotations from spectra of
504 mixtures.

505 In addition to illustrating principles by which the evolution of form occurs, our deepening
506 understanding of specialized metabolic innovation has practical benefits. Metabolic engineering
507 continues to be a trial and error enterprise (48). Understanding the successful evolutionary
508 outcomes leading to synthesis of biologically active metabolites provides new enzymes,
509 transporters and transcriptional regulators for the synthetic biology toolkit. Documenting recurring
510 themes, such as the independent recruitment of invertases in multiple acylglucose-producing
511 *Solanum* lineages, indicates robust strategies for engineering synthesis of these protective
512 compounds and other structurally similar compounds of economic or pharmaceutical importance.

513

514 **Materials and Methods**

515 **Plant Material**

516 Seeds of *S. nigrum* were obtained from New York Botanical Garden (14). For germination, seeds were treated
517 with half-strength bleach for 5 min and rinsed six times in deionized water before sowing on moist filter paper in petri
518 dishes at 28°C. Seedlings were transferred to Jiffy-7® peat pellets (Jiffy Products of America, OH, USA) upon
519 germination. Plants used for analysis were grown in a growth chamber at 22°C under a 16-hour photoperiod (70 μmol
520 $\text{m}^{-2} \text{s}^{-1}$ photosynthetic photon flux density) with relative humidity set to 50%.

521

522 **Acylsugar metabolite annotation**

523 The acylsugar extraction protocol for LC-MS analysis is available in Protocols.io at
524 <https://dx.doi.org/10.17504/protocols.io.xj2fkqe>. As described previously in Leong et al. (2019), leaf surface
525 acylsugar extraction were carried out by gently agitating a single leaflet in 1 mL acetonitrile:isopropanol:water (3:3:2
526 v/v with 0.1% formic acid and 1 μM telmisartan; isopropanol obtained from J.T.Baker, Phillipsburg, NJ, USA, all
527 others obtained from Sigma-Aldrich, St. Louis, MO, USA) for 2 min. Telmisartan acts as an internal standard for
528 high-performance liquid chromatography (HPLC). The extraction solvent was collected and stored in 2-mL LC-MS
529 vials at -20°C. All extracts were analyzed on LC-MS (Waters Corporation, MA, USA) using 7-min, 30-min or 110-
530 min LC gradients on an Ascentis Express C18 HPLC column (10 cm x 2.1 mm, 2.7 μm) (Sigma-Aldrich, St. Louis,
531 MO, USA), which was maintained at 40°C. The 110-min method minimized chromatographic overlap in support of
532 metabolite annotation. The HPLC-MS methods are described in protocols.io and Table S10.

533 Acylsugars structures were inferred by positive and negative mode MS collision-induced dissociation as
534 described previously (11, 29, 32, 41). In short, co-eluting fragments generated by collision-induced dissociation in
535 negative ion mode were compared among three energy potentials to confirm acylsugar metabolites. Annotation
536 strategy, thresholds and confidence levels were described in detail in Supplementary Text.

537 Acylsugar acyl chain composition was determined by ethyl ester derivatization and subsequent GC-MS
538 analysis. To create fatty acid ethyl esters, acylsugar samples were saponified and transesterified as previously
539 described in Ning et al. (2015) with some modification. In short, a leaflet was immersed in 1 mL of
540 acetonitrile/isopropanol (1:1 v/v) for 2 min with gentle agitation. The extracts were transferred, evaporated to dryness
541 under flowing air and redissolved in ethanol with 300 μL of 21% (v/v) sodium ethoxide (Sigma-Aldrich, St. Louis,
542 MO, USA). The reaction was gently vortexed every 5 min for 30 min in the fume hood. Four hundred μL hexane with
543 55 $\mu\text{g}/\text{mL}$ of tetradecane (internal standard; Sigma-Aldrich, St. Louis, MO, USA) was added to the reaction mixture
544 for phase separation. The hexane layer was transferred and extracted three times using 500 μL saturated aqueous
545 sodium chloride each time. The final hexane phase (~100 μL) was transferred to autosampler vials with glass inserts
546 and analyzed by capillary GC-MS on an Agilent J&W DB-5 column (10-m, 0.1-mm (i.d.) fused silica column with a
547 0.34- μm -thick stationary phase; Agilent). One microliter of each hexane extract was injected using splitless mode.
548 The gas chromatography program is described in Table S10. All compounds were analyzed using an Agilent 6890N

549 gas chromatograph/Agilent 5975B single quadrupole mass spectrometer using 70 eV electron ionization. Fatty acid
550 ethyl esters were identified by library search against Agilent RTL library and compared with commercially available
551 standards.

552 For total extract NMR profiling, acylsugars were extracted from fifteen 3-4 weeks old *S. nigrum* plants by
553 dipping aerial tissue into 500 mL of ethanol with 0.1% (v/v) formic acid with gentle agitation in a 1-L beaker. The
554 ethanol was evaporated under reduced pressure using a rotary evaporator with a warm water bath (40°C). The dried
555 residue (~5 mg) of *S. nigrum* surface plant extracts were dissolved in CDCl₃ (99.8 atom % D-, Sigma Aldrich, St.
556 Louis, MO, USA) and transferred to solvent-matched (5 mm) NMR Shigemi tube (Shigemi Co., LTD., Tokyo, Japan)
557 for analysis. ¹H, HSQC, HMBC, and TOCSY spectra were recorded using the Avance 900 MHz spectrometer (Bruker,
558 Billerica, MA, USA) equipped with a TCI triple resonance probe at the Michigan State University Max T. Rogers
559 NMR facility (See Table S2-4 for more details). All spectra were referenced to non-deuterated CDCl₃ solvent signals
560 (δ H = 7.26 (s) and δ C = 77.2 (t) ppm).

561 For silica gel column chromatography, *S. nigrum* surface plant ethanol extract was concentrated *in vacuo* (~8
562 mg), dissolved in ethyl acetate/ hexane (1 mL, 1:1 v/v) and loaded to the silica gel column (100 g, 200-425 mesh, 60
563 A, Jade Scientific Inc, Westland, MI, USA) packed with ethyl acetate/hexane (200 mL, 1:1 v/v) slurry. The compounds
564 were eluted using mobile phase ethyl acetate/hexane (400 mL, 2:1 v/v in 0.02% acetic acid) with compressed air flash
565 column chromatography. After TLC analysis (4:1 ethyl acetate/hexane, p-Anisaldehyde stain), fractions *F*₁₅₋₁₇ and *F*₂₄₋
566 ₂₈ were combined, concentrated *in vacuo* to give fraction 1 (~2 mg, R_f = 0.7) and fraction 2 (~2 mg, R_f =
567 0.3). Combined samples were analyzed on NMR as described above.

568

569 **Gene Identification and Phylogenetic Analysis**

570 All transcript assemblies and expression data are from Moghe et al. (2017) and were analyzed using Geneious
571 R9.1.8 and R, respectively. To identify BAHD candidates, BLAST and TBLASTN searches were performed using
572 ASAT sequences from tomato and *S. sinuata*. The TBLASTN hits were parsed to include only those with HXXXD
573 motifs and DFGWG-like motif (one mismatch). The trichome-stem expression data of the remaining sequences were
574 obtained from expression data available in the study by Moghe et al. (2017), whereas a length of 400 to 500 amino
575 acids and the relative positions of the two motifs on these sequences were confirmed manually(36). To identify ASFF
576 candidates, 23 invertase-like sequences from tomato were used to BLAST and TBLASTN search *S. nigrum*
577 transcriptome. The obtained sequences were checked for trichome-enrichment and WXNDPNG, RDP, EC and DXKK
578 motifs as described above.

579 Phylogenetic reconstructions were performed using MEGA X(49). To obtain the BAHD tree, BAHD
580 candidate sequences were aligned against several characterized ASATs(9-11, 14) and several other BAHD sequences
581 from D'Auria (2006) using the MUSCLE algorithm under default parameters. To obtain the invertase tree, invertase-
582 like proteins from *S. lycopersicum*, *S. pennellii* and *S. nigrum* were aligned with the same algorithm and settings.
583 Maximum likelihood estimations were performed with Jones-Taylor-Thornton (JTT)+G+I with 5 rate categories as
584 the substitution model. One thousand bootstrap replicates were performed using partial deletion (30% gaps) for tree
585 reconstruction.

586

587 **Transient expression and purification of BAHD protein**

588 All Sanger DNA sequencing confirmations in this study were performed with the indicated sequencing
589 primers at the Research Technology Support Facility Genomics Core, Michigan State University, East Lansing, MI.
590 All primer sequences are listed in Table S9.

591 Recombinant BAHD proteins were generated using *E. coli* as the host for enzyme assays. The full-length
592 open reading frames of *c63608_g1*, *c65670_g1*, *c71009_g1*, *c53868_g2*, *c60145_g1*, *c65306_g1* were amplified from
593 *S. nigrum* young leaf and peduncle cDNA and cloned into pET28b(+) (EMD Millipore, MA, USA) using BamHI and
594 XhoI restriction sites and 2× Gibson Assembly master mix (NEB, Ipswich, MA, USA) according to the manufacturer's
595 instructions. The assembled constructs were transformed into BL21 Rosetta (DE3) cells (EMD Millipore, MA, USA)
596 and submitted for Sanger sequencing.

597 Protein expression was carried out as described before (11, 16). In short, 1 L Luria-Bertani (LB) media with
598 kanamycin (50 $\mu\text{g mL}^{-1}$) and chloramphenicol (33 $\mu\text{g mL}^{-1}$) were inoculated 500:1 with an overnight culture obtained
599 from the bacterial strain with the desired construct. After incubating at 37°C, 225 rpm and reaching OD600 between
600 0.5 and 0.6, large scale cultures were chilled on ice for 20 min before a final concentration of 50 μM isopropylthio- β -
601 galactoside was added. Cultures were incubated at 16°C and 180 rpm for 16 h before cell pellets were harvested by
602 centrifuging at 4,000g for 10 min under 4°C. The following protein extraction steps were also performed on ice or at
603 4°C. The cell pellets were resuspended in 25 mL of extraction buffer (50 mM NaPO₄, 300 mM NaCl, 20 mM
604 imidazole, and 5 mM 2-mercaptoethanol, pH 8) by vortexing and submitted to eight cycles of sonication (30 s on ice
605 with 30 s intervals for cooling). The obtained cellular extracts were centrifuged twice at 30,000g for 10 min to obtain
606 clear supernatant. Ni-NTA resin (Qiagen, Venlo, The Netherlands) was washed 3 times and resuspended in 1 mL of
607 extraction buffer before incubated with the centrifuged extracts at 4°C for 1 h with nutation. After removing the
608 supernatant by centrifuging the slurry at 3,200g for 5 min, the resins were transferred to a gravity flow column (Bio-
609 Rad Laboratories, Hercules, CA, USA) and washed with 3 column volumes of wash buffer (50 mM NaPO₄, 300 mM
610 NaCl, 40 mM imidazole, and 5 mM 2-mercaptoethanol, pH 8). Two milliliters of elution buffer (50 mM NaPO₄, 300
611 mM NaCl, 3 M imidazole, and 5 mM 2-mercaptoethanol, pH 8) was added to the column and allowed to incubate
612 with the resin for 1 min. The elutes were then diluted into 15 mL of storage buffer (extraction buffer without
613 imidazole), concentrated using 10-kD centrifugal filter units (EMD Millipore, MA, USA) till 1,000-fold dilution. A
614 final volume of 40% (v/v) glycerol-elution solution was prepared and stored at -20°C. Immunoblot with the anti-His-
615 antibody conjugated to peroxidase (BMG-His-1 monoclonal antibody; Roche, Basel, Switzerland) was used to
616 confirm the presence of enzymes.

617

618 **Transient expression and purification of SnASFF1 protein**

619 ASFF1 protein expression and purification was carried out exactly as described before(15). SnASFF1 full-
620 length open reading frame was amplified from *S. nigrum* young leaf and peduncle cDNA using c70979_Fw and
621 c70979_Rv primer and cloned into the pEAQ-HT vector(50) using NruI-HF and SmaI restriction sites and 2× Gibson
622 Assembly master mix (NEB, Ipswich, MA) according to the manufacturer's instructions. The completed vector was
623 subsequently transformed into *Agrobacterium tumefaciens* LBA4404 cells. Single colonies from LB agar plates with
624 rifampicin (50 $\mu\text{g mL}^{-1}$) and kanamycin (50 $\mu\text{g mL}^{-1}$) were used to inoculate 50 mL of YEP medium with the same
625 antibiotics in the same concentration. After overnight incubation at 28°C, 300 rpm, culture was harvested and washed
626 in 50 ml of buffer A [10 mM 2-ethanesulfonic acid (MES; Sigma-Aldrich, St. Louis, MO, USA) at pH 5.6 and 10 mM
627 MgCl₂] before resuspended to a final OD600 = 1.0 with buffer A with 200 μM acetosyringone (Sigma-Aldrich, St.
628 Louis, MO, USA). The suspension was incubated at room temperature with gentle rocking for 4 hours and then
629 infiltrated into fully expanded leaves of 6-week-old *Nicotiana benthamiana* plants using a needleless 1-mL tuberculin
630 syringe. Infiltrated leaves were harvested, deveined, and flash-frozen in liquid nitrogen after 7-8 days. Tissue was
631 powdered under liquid nitrogen and added to 140 mL of ice-cold buffer B [25 mM 3-[4-(2-hydroxyethyl)piperazin-1-
632 yl]propane-1-sulfonic acid (EPPS) at pH 8.0, 1.5 M NaCl, 1 mM EDTA with 2 mM dithiothreitol (DTT), 1 mM
633 benzamidine, 0.1 mM phenylmethanesulfonyl fluoride, 10 μM trans-epoxysuccinyl-l-leucylamido(4-
634 guanidino)butane (E-64), and 5% (w/v) polyvinylpyrrolidone (PVPP); all reagents were obtained from Sigma-
635 Aldrich (St. Louis, MO, USA) except DTT, which was obtained from Roche Diagnostics (Risch-Rotkreuz,
636 Switzerland)]. The mixture was stirred for 4 hours at 4°C before filtered through six layers of Miracloth and
637 centrifuged at 27,000g at 4°C for 30 min. After passed through a 0.22- μm polyethersulfone filter (EMD Millipore,
638 Billerica, MA), the supernatant was loaded onto a HisTrap HP 1-mL affinity column and eluted using a gradient of
639 10 to 500 mM imidazole in buffer B using an ÄKTA start FPLC module (GE Healthcare, Uppsala, Sweden). Fractions
640 were analyzed by SDS-polyacrylamide gel electrophoresis, and the presence of SnASFF1-HT was confirmed by
641 immunoblot using the BMG-His-1 monoclonal antibody (Roche, Basel, Switzerland). Purified SnASFF1 proteins
642 were stored in 100 mM sodium acetate (pH 4.5) with 40% (v/v) glycerol at -20°C.

643

644 **Enzyme assays**

645 Unless otherwise specified, all BAHD forward enzyme assays were performed by incubating purified
646 recombinant proteins in 60 μ L of 100 mM ammonium acetate (pH 6.0) buffer with 100 μ M acyl-CoA and an acyl
647 chain acceptor (unmodified sugar or purified or fractionated acylsugar acceptors in an ethanol:water mixture (1:1
648 v/v)) at 30°C for 30 min. After the incubation, 2 volumes of stop solution [acetonitrile:isopropanol (1:1) with 0.1%
649 (v/v) formic acid and 1 μ M telmisartan as internal standard] was added to the assays and mixed. Reactions were
650 centrifuged at 17,000g for 5 min, and the supernatants were stored at -20°C after being transferred to LC-MS vials.
651 SnASFF1 enzyme assays were carried out in a similar manner in the absence of acyl-CoAs. For BAHD reverse enzyme
652 assays, 100 μ M free Coenzyme A were used in place of acyl-CoAs. For negative controls, enzymes boiled at 95°C for
653 10 min were used in place of active enzymes. Methods used to determine the apparent K_m value for different substrates
654 were performed as previously described(11). Tested sugars, free coenzyme A and nC10-, nC8-, iC5-, iC4-CoAs were
655 obtained from Sigma-Aldrich, St. Louis, MO, USA, whereas iC10-CoA and aiC5-CoA were produced following
656 method described in Kawaguchi et al., 1981(51).

657

658 **Purification of S1:10 and S2:14**

659 Purifications were performed using a Waters 2795 Separations Module (Waters Corporation) and an Acclaim
660 120 C18 HPLC column (4.6 mm \times 150 mm, 5 μ m; Thermo Fisher Scientific, Waltham, MA, USA) with a column
661 oven temperature of 30°C and flow rate of 1 mL/min. For S1:10 purification, the mobile phase consisted of water with
662 0.1% (solvent A) and acetonitrile (solvent B). For *in vivo* S2:14 purification, two purification methods where solvent
663 B consist of acetonitrile and methanol, respectively, were used in a sequential manner to enhance purity of compounds.
664 The HPLC methods used for separating S1:10 and S2:14 are described in Table S10. Fractions were collected using a
665 2211 Superrac fraction collector (LKB Bromma, Stockholm, Sweden).

666

667 **VIGS Analysis**

668 VIGS target regions were selected to have a low chance of altering expression of trichome expressed non-
669 target genes(14). These fragments were amplified from *S. nigrum* young leaf and peduncle cDNA using primers listed
670 in Table S9 and were cloned into pTRV2-LIC as previously described (16). In short, the PstI-HF-linearized pTRV2-
671 LIC vector and all PCR fragments were separately incubated in 5- μ L T4 DNA polymerase reactions with 5 mM dATP
672 or dTTP, respectively. The reactions were incubated at 22°C for 30 min, then 70°C for 20 min, and then stored on ice.
673 Two microliters of desired PCR reaction were mixed with 1 μ L vector reaction and incubated at 65°C for 2 min,
674 followed by 22°C for 10 min. Sequenced constructs and pTRV1 were transformed into *A. tumefaciens* strain GV3101
675 using the protocol described previously(15).

676 The vacuum infiltration protocol was adapted from Hartl et al. (2008) (52). Twenty-five milliliters of LB
677 medium with kanamycin (50 μ g mL⁻¹), rifampicin (50 μ g mL⁻¹), and gentamicin (10 μ g mL⁻¹) were inoculated with a
678 single colony of the respective Agrobacterium strain from plate. After overnight incubation at 28°C and 225 rpm, cells
679 were harvested by centrifugation at 3,200g for 10 min and resuspended to a final OD600 = 1 with the induction media
680 (10 mM MES, pH 5.6, and 10 mM MgCl₂). An equal volume of pTRV1 suspension was mixed with different pTRV2-
681 LIC constructs suspension. 200 μ M acetosyringone was added to each mixture before the suspension was incubated
682 in the dark at room temperature with gentle rocking for 3 hours. Seven to ten days-old *S. nigrum* seedlings were
683 carefully transferred from the petri dishes to the bacteria solution. Vacuum was applied in a desiccator for 2 min,
684 followed by slow release of the vacuum. Infiltrated seedlings were kept for 3 d under indirect light, then planted in
685 Jiffy-7® peat pellets and grown in a climate chamber as described above. Metabolite and tissue samples were
686 harvested approximately 3 weeks post infiltration and phytoene desaturase silencing efficiency was monitored.

687 The QuanLynx function in MassLynx v4.1 (Waters Corporation, MA, USA) was used to integrate extracted
688 ion chromatograms from untargeted LC-MS data as described previously (29). In short, all quantifications were
689 performed using extracted ion chromatograms of the m/z value for the relevant $[M+NH_4]^+$ or $[M+HCOO]^-$ adduct ions
690 using a mass window of m/z 0.05. Isomeric forms of acylsugars (including anomers) were quantified in corresponding
691 groups as shown in Table S1. The retention time window was adjusted for each compound to include all isomers. Peak
692 area of telmisartan was quantified and used as an internal reference. Then, acylsugar quantities (per mg) were

693 calculated by normalizing peak areas to the internal standard peak area and dry leaf weight, whereas acylsugar ratios
694 were calculated by comparing peak areas over internal standard peak area of acylsugars of interest.

695

696 **qPCR Analysis**

697 RNA was extracted with the RNeasy Plant Mini Kit including on-column DNase digestion (Qiagen, Venlo,
698 The Netherlands), according to the manufacturer's instructions. RNA was quantified with a Nanodrop 2000c
699 instrument (Thermo Fisher Scientific, Waltham, MA, USA). cDNA was synthesized using 1mg of the isolated RNA
700 and SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA). The cDNA samples were diluted 40-fold
701 (10-fold initial dilution and 4-fold dilution into qPCRs). qPCRs (10 μ L) were created with SYBR Green PCR Master
702 Mix (Thermo Fisher Scientific, Waltham, MA, USA), and primers were used at a final concentration of 200 nM.
703 RT_SnASAT1_F/R, RT_SnASAT2_F/R, RT_SnASFF1_F/R, RT_SnAGAT1_F/R, RT_Actin_1_F/R, and
704 RT_Actin_3_F/R primers were used to detect *SnASAT1*, *SnASAT2*, *SnASFF1*, *SnAGAT1*, *ACTIN1*, and *ACTIN3*
705 transcripts, respectively (Table S9). Reactions were carried out with a QuantStudio 7 Flex Real-Time PCR System
706 (Applied Bio-systems) by the Michigan State University RTSF Genomics Core. The following temperature cycling
707 conditions were applied: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. Relative
708 expression of *SnASAT1*, *SnASAT2*, *SnASFF1*, and *SnAGAT1* was calculated with the $\Delta\Delta$ Ct method (53) and
709 normalized to the geometric mean of *ACTIN1* and *ACTIN3* transcript levels. The mean expression values of the
710 transcripts in the control plants were used for normalization. Three to four technical replicates were used for all the
711 qPCRs.

712

713 **Statistical Analysis**

714 All statistical analyses were performed using the “stats” R package (R Core Team, 2017). Welch two-sample
715 *t* tests were executed on metabolites and transcript abundance data using the “t.test” command. The power of these
716 analyses was determined using the “power.t.test” function.

717

718 **References**

719

- 720 1. N. Agerbirk, C. E. Olsen, Glucosinolate structures in evolution. *Phytochemistry*. **77**, 16–45 (2012).
- 721 2. B. Barco, N. K. Clay, Evolution of glucosinolate diversity via whole-genome duplications, gene
722 rearrangements, and substrate promiscuity. *Annu. Rev. Plant Biol.* **70**, 585–604 (2019).
- 723 3. H. M. Niemeyer, Hydroxamic acids derived from 2-hydroxy-2 H-1, 4-benzoxazin-3 (4 H)-one: key defense
724 chemicals of cereals. *J. Agric. Food Chem.* **57**, 1677–1696 (2009).
- 725 4. S. Zhou, A. Richter, G. Jander, Beyond Defense: Multiple Functions of Benzoxazinoids in Maize
726 Metabolism. *Plant Cell Physiol.* **59**, 1528–1537 (2018).
- 727 5. J.-K. Weng, R. N. Philippe, J. P. Noel, The rise of chemodiversity in plants. *Science*. **336**, 1667–1670
728 (2012).
- 729 6. L. Chae, T. Kim, R. Nilo-poyanco, S. Y. Rhee, Genomic Signatures of Specialized Metabolism in Plants.
730 *Science*. **344**, 510–513 (2014).
- 731 7. G. D. Moghe, R. L. Last, Something Old, Something New: Conserved Enzymes and the Evolution of
732 Novelty in Plant Specialized Metabolism. *Plant Physiol.* **169**, 1512–23 (2015).
- 733 8. P. Fan, B. J. Leong, R. L. Last, Tip of the trichome : evolution of acylsugar metabolic diversity in
734 Solanaceae. *Curr. Opin. Plant Biol.* **49**, 8–16 (2019).
- 735 9. A. L. Schillmiller, A. L. Charbonneau, R. L. Last, Identification of a BAHD acetyltransferase that produces
736 protective acyl sugars in tomato trichomes. *Proc. Natl. Acad. Sci.* **109**, 16377–16382 (2012).
- 737 10. A. L. Schillmiller, G. D. Moghe, P. Fan, B. Ghosh, J. Ning, A. D. Jones, R. L. Last, Functionally divergent
738 alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in
739 *Solanum* trichomes. *Plant Cell.* **27**, 1002–17 (2015).
- 740 11. P. Fan, A. M. Miller, A. L. Schillmiller, X. Liu, I. Ofner, A. D. Jones, D. Zamir, R. L. Last, *In vitro*
741 reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. *Proc.*
742 *Natl. Acad. Sci.* **113**, E239–E248 (2016).
- 743 12. P. Fan, A. M. Miller, X. Liu, A. D. Jones, R. L. Last, Evolution of a flipped pathway creates metabolic
744 innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* **8**, 1–13 (2017).
- 745 13. S. S. Nadakuduti, J. B. Uebler, X. Liu, A. D. Jones, C. S. Barry, Characterization of trichome-expressed
746 BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly mechanisms within the

- 747 Solanaceae. *Plant Physiol.* **175**, 36–50 (2017).
- 748 14. G. D. Moghe, B. J. Leong, S. Hurney, A. D. Jones, R. L. Last, Evolutionary routes to biochemical
749 innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife*
750 (2017).
- 751 15. B. J. Leong, D. B. Lybrand, Y. Lou, P. Fan, A. L. Schillmiller, R. L. Last, Evolution of metabolic novelty : A
752 trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci. Adv.*, 1–14 (2019).
- 753 16. B. J. Leong, S. M. Hurney, P. D. Fiesel, G. D. Moghe, A. D. Jones, R. L. Last, Specialized metabolism in a
754 nonmodel nightshade: trichome acylinositol biosynthesis. *Plant Physiol.* **183**, 915–924 (2020).
- 755 17. B. M. Leckie, D. A. D’Ambrosio, T. M. Chappell, R. Halitschke, D. M. De Jong, A. Kessler, G. G.
756 Kennedy, M. A. Mutschler, Differential and synergistic functionality of acylsugars in suppressing
757 oviposition by insect herbivores. *PLoS One.* **11**, 1–19 (2016).
- 758 18. J. R. Smeda, A. L. Schillmiller, R. L. Last, M. A. Mutschler, Introgression of acylsugar chemistry QTL
759 modifies the composition and structure of acylsugars produced by high-accumulating tomato lines. *Mol.*
760 *Breed.* **36**, 160 (2016).
- 761 19. A. Weinhold, I. T. Baldwin, Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them
762 for predation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7855–7859 (2011).
- 763 20. V. T. Luu, A. Weinhold, C. Ullah, S. Dressel, M. Schoettner, K. Gase, E. Gaquerel, S. Xu, I. T. Baldwin, O-
764 acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. *Plant*
765 *Physiol.* **174.1**, 370-386 (2017).
- 766 21. R. R. King, L. A. Calhoun, R. P. Singh, 3,4-Di-O- and 2,3,4-tri-O-acylated glucose esters from the glandular
767 trichomes of nontuberous *Solanum* species. *Phytochemistry.* **27**, 3765–3768 (1988).
- 768 22. T. Matsuzaki, Y. Shinozaki, S. Suhara, T. Tobita, Leaf Surface Glycolipids from *Nicotiana acuminata* and
769 *Nicotiana pauciflora*. *Agric. Biol. Chem.* **55**, 1417–1419 (1991).
- 770 23. Y. Herrera-Salgado, M. L. Garduño-Ramírez, L. Vázquez, M. Y. Rios, L. Alvarez, Myo-inositol-derived
771 glycolipids with anti-inflammatory activity from *Solanum lanceolatum*. *J. Nat. Prod.* **68**, 1031–1036 (2005).
- 772 24. E. Maldonado, F. R. Torres, M. Martinez, A. L. Pérez-Castorena, Sucrose esters from the fruits of *Physalis*
773 *nicandroides* var. *attenuata*. *J. Nat. Prod.* **69**, 1511–1513 (2006).
- 774 25. C. Dutra, M. V. Cesio, P. Moyna, H. Heinzen, Acyl sucroses from *Salpichroa origanifolia*. *Nat. Prod.*
775 *Commun.* **3**, 539–542 (2008).
- 776 26. J. M. Alba, M. Montserrat, R. Fernández-Muñoz, Resistance to the two-spotted spider mite (*Tetranychus*
777 *urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant
778 inbred line population. *Exp. Appl. Acarol.* **47**, 35–47 (2009).
- 779 27. J. Kim, K. Kang, E. Gonzales-Vigil, F. Shi, A. D. Jones, C. S. Barry, R. L. Last, Striking natural diversity in
780 glandular trichome acylsugar composition is shaped by variation at the Acyltransferase2 locus in the wild
781 tomato *Solanum habrochaites*. *Plant Physiol.* **160**, 1854–1870 (2012).
- 782 28. X. Liu, M. Enright, C. S. Barry, A. D. Jones, Profiling, isolation and structure elucidation of specialized
783 acylsucrose metabolites accumulating in trichomes of *Petunia* species. *Metabolomics.* **13**, 1–10 (2017).
- 784 29. D. B. Lybrand, T. M. Anthony, A. D. Jones, An integrated analytical approach reveals trichome acylsugar
785 metabolite diversity in the wild tomato *Solanum pennellii*. *Metabolites.* **10**, 401 (2020).
- 786 30. J. C. Goffreda, M. A. Mutschler, D. A. Avé, W. M. Tingey, J. C. Steffens, Aphid deterrence by glucose
787 esters in glandular trichome exudate of the wild tomato, *Lycopersicon pennellii*. *J. Chem. Ecol.* **15**, 2135–
788 2147 (1989).
- 789 31. T. Matsuzaki, Y. Shinozaki, S. Suhara, H. Shigematsu, A. Koiwai, Isolation and characterization of tetra-
790 and triacylglycerol from the surface lipids of *Nicotiana glauca*. **1369**, 2–5 (2014).
- 791 32. B. Ghosh, A. D. Jones, Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar
792 esters). *Anal. Methods.* **9**, 892–905 (2017).
- 793 33. E. A. P. Ekanayaka, M. D. Celiz, A. D. Jones, Relative mass defect filtering of mass spectra: a path to
794 discovery of plant specialized metabolites. *Plant Physiol.* **167**, 1221–1232 (2015).
- 795 34. S. M. Hurney, "Strategies for profiling and discovery of acylsugar specialized metabolites," thesis, Michigan
796 State University, East Lansing, MI (2018).
- 797 35. K. Bingol, L. Bruschweiler-Li, D.-W. Li, R. Bruschweiler, Customized metabolomics database for the
798 analysis of NMR 1H--1H TOCSY and 13C--1H HSQC-TOCSY spectra of complex mixtures. *Anal. Chem.*
799 **86**, 5494–5501 (2014).
- 800 36. J. C. D’Auria, Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340
801 (2006).
- 802 37. K. Le Roy, W. Lammens, M. Verhaest, B. De Coninck, A. Rabijns, A. Van Laere, W. den Ende, Unraveling

- 803 the difference between invertases and fructan exohydrolases: a single amino acid (Asp-239) substitution
804 transforms Arabidopsis cell wall invertase1 into a fructan 1-exohydrolase. *Plant Physiol.* **145**, 616–625
805 (2007).
- 806 38. W. Lammens, K. Le Roy, A. Van Laere, A. Rabijns, W. den Ende, Crystal structures of *Arabidopsis*
807 *thaliana* cell-wall invertase mutants in complex with sucrose. *J. Mol. Biol.* **377**, 378–385 (2008).
- 808 39. V. A. Reddy, F. Maley, Identification of an active-site residue in yeast invertase by affinity labeling and site-
809 directed mutagenesis. *J. Biol. Chem.* **265**, 10817–10820 (1990).
- 810 40. E. Fridman, D. Zamir, Functional divergence of a syntenic invertase gene family in tomato, potato, and
811 arabidopsis 1. **131**, 603–609 (2003).
- 812 41. B. Ghosh, T. C. Westbrook, A. D. Jones, Comparative structural profiling of trichome specialized
813 metabolites in tomato (*Solanum lycopersicum*) and *S. habrochaites* : acylsugar profiles revealed by UHPLC
814 / MS and NMR. *Metabolomics*, 496–507 (2014).
- 815 42. J. C. Del Rio, J. Rencoret, P. Prinsen, A. T. Martinez, J. Ralph, A. Gutiérrez, Structural characterization of
816 wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. *J. Agric.*
817 *Food Chem.* **60**, 5922–5935 (2012).
- 818 43. T. Asai, N. Hara, Y. Fujimoto, Phytochemistry Fatty acid derivatives and dammarane triterpenes from the
819 glandular trichome exudates of *Ibicella lutea* and *Proboscidea louisiana*. *Phytochemistry.* **71**, 877–894
820 (2010).
- 821 44. T. A. Sai, T. S. Akai, K. O. Hyama, Y. F. Ujimoto, n-Octyl α -L-Rhamnopyranosyl-(1→2)- β -D-
822 glucopyranoside Derivatives from the Glandular Trichome Exudate of *Geranium carolinianum*. *Chem*
823 *Pharm Bull.* **59**, 747–752 (2011).
- 824 45. T. Asai, Y. Nakamura, Y. Hirayama, K. Ohyama, Y. Fujimoto, Cyclic glycolipids from glandular trichome
825 exudates of *Cerastium glomeratum*. *Phytochemistry.* **82**, 149–157 (2012).
- 826 46. Q. Wu, J. G. Cho, D. S. Lee, D. Y. Lee, N. Y. Song, Y. C. Kim, K. T. Lee, H. G. Chung, M. S. Choi, T. S.
827 Jeong, E. M. Ahn, G. S. Kim, N. I. Baek, Carbohydrate derivatives from the roots of *Brassica rapa* ssp.
828 *campestris* and their effects on ROS production and glutamate-induced cell death in HT-22 cells.
829 *Carbohydr. Res.* **372**, 9–14 (2013).
- 830 47. Y. Liu, S. Jing, S. Luo, S. Li, Non-volatile natural products in plant glandular trichomes: chemistry,
831 biological activities and biosynthesis. *Natural Product Reports* **36.4**, 626–665 (2019).
- 832 48. L. J. Sweetlove, R. L. Last, A. R. Fernie, Predictive metabolic engineering: a goal for systems biology. *Plant*
833 *Physiol.* **132**, 420–425 (2003).
- 834 49. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis
835 across computing platforms. *Mol. Biol. Evol.* **35**, 1547 (2018).
- 836 50. H. Peyret, G. P. Lomonosoff, The pEAQ vector series: the easy and quick way to produce recombinant
837 proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).
- 838 51. A. Kawaguchi, T. YOSHIMURA, S. Okuda, A new method for the preparation of acyl-CoA thioesters. *J.*
839 *Biochem.* **89**, 337–339 (1981).
- 840 52. M. Hartl, H. Merker, D. D. Schmidt, I. T. Baldwin, Optimized virus-induced gene silencing in *Solanum*
841 *nigrum* reveals the defensive function of leucine aminopeptidase against herbivores and the shortcomings of
842 empty vector controls. *New Phytol.* **179**, 356–365 (2008).
- 843 53. M. W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids*
844 *Res.* **29**, e45–e45 (2001).

845
846

847 Acknowledgments

848 We acknowledge the MSU RTSF Mass Spectrometry and Metabolomics Core Facilities
849 for their support with LC-MS analysis. We thank Dr. D. Holms and L. Xie at the Michigan
850 State University Max T. Rogers NMR Facility for technical assistance. We thank Dr. Dan
851 Lybrand for helpful advice and generously providing the pET28b-SnASAT1 construct. We
852 also acknowledge the valuable feedback received from members of the Last lab.

853

854 **Funding:** This work was supported by the US National Science Foundation Plant
855 Genome Research Program grant IOS-1546617 to R.L.L. and A.D.J. and the National

856 Institute of General Medical Sciences of the National Institutes of Health graduate
857 training grant no. T32–GM110523 to P.D.F.

858

859 **Author contributions:**

860 Design: YRL, TMA, ADJ, RLL

861 NMR experiments and data analysis: TMA, PDF, ADJ

862 Biochemical and biological experiments and data analysis: YRL, PDF, REA and EMC

863 Interpretation and preparation of figures and table: YRL, TMA, PDF

864 Writing – original draft: YRL, RLL

865 Writing – review and editing: YRL, TMA, PDF, REA, EMC, ADJ and RLL

866

867 **Competing interests:** The authors declare that they have no competing interests.

868

869 **Data and materials availability:** All data needed to evaluate the conclusions in the paper
870 are present in the paper and/or the Supplementary Materials. Additional data related to this
871 paper may be requested from the authors. The constructs pET28b-SnASAT1, pET28b-
872 SnASAT2, pET28b-SnAGAT1, and pEAQ-HT-SnASFF1 can be provided by Y.-R.L.
873 pending a completed material transfer agreement. Requests for biological materials or data
874 should be submitted to R.L.L. at lastr@msu.edu.

875

876

877 **Figures and Tables**

878

879 **Figure 1. Integrated analysis of *S. nigrum* acylhexose structures.** (A) LC-MS profile of *S.*
880 *nigrum* leaf dip extracts reveals acylsugar-like compounds. Metabolites a, b and c represent the
881 most abundant acylinositol, diacylglucose and triacylglucose, respectively. See Fig. S1, S2, S3,
882 Table S1 and Supplementary Methods for details. (B) Fatty acid ethyl ester profile of saponified
883 acylsugars analyzed using GC-MS reveal abundant iC4, iC5, aiC5, as well as straight and branched
884 C8, C9 and C10 acyl chains. The abbreviations “i” and “ai” refer to *iso*- and *anteiso*- branched
885 isomers (terminally and sub-terminally branched, respectively). (C) Schematic of pipeline used
886 for acylsugar structure determination. (D) NMR analysis of total acylsugars revealed that the most
887 abundant acylsugar contains a glucose core and acyl chains on C2, C3 and C4 ring protons. BPI,
888 base peak intensity; ESI-, negative electrospray ionization. All ESI- mode acylsugars are identified
889 as formate adducts. The corresponding *m/z* of these compounds are listed in Table S1 and
890 Supplementary Methods.

891

892 **Figure 2. *S. nigrum* accumulates triacylinositols and di- and triacylated acylglucoses.** (A)

893 Silica flash column separation of leaf dip extracts into less polar fraction 1 and more polar fraction

894 2, as analyzed by thin layer chromatography. (B) LC-MS profile of fraction 1 reveals mostly

895 triacylglucoses. The peaks of metabolite c (cyan dashed rectangle) with identical m/z and mass
896 spectra represent the two anomers each of two structural isomers of triacylglucose G3:16(2,4,10).
897 (C) LC-MS profile of fraction 2 reveals abundant diacylglucoses and triacylinositols. The two
898 peaks with m/z and mass spectra consistent with metabolite a (orange dashed rectangle) represent
899 two structural isomers of I3(4,4,8), whereas metabolite b (magenta dashed rectangle) resolves into
900 multiple peaks representing the two anomers of the two structural isomers of G2:14(4,10). (D)
901 NMR-resolved structure of metabolite c, the most abundant triacylglucose in fraction 1. See Table
902 S2 for details. (E) HSQC-TOCSY separated spin systems between diacylglucoses and
903 triacylinositols in fraction 2. See Fig. S4 for details. (F) NMR-resolved structure of metabolite b,
904 the most abundant diacylglucose, and (G) metabolite a, the most abundant triacylinositol in
905 fraction 2. See Table S3 and S4 for details. BPI, base peak intensity; ESI-, negative electrospray
906 ionization. All chromatograms showing telmisartan as internal standard (I.S.). All ESI- mode
907 acylsugars are identified as formate adducts. The corresponding m/z 's of these compounds are
908 listed in Table S1.

909

910 **Figure 3. Diacylsucroses accumulate in *SnASFF1* virus induced gene silencing (VIGS) lines.**

911 LC-MS analysis of acylsugars extracted from *SnASFF1*-targeted (purple trace) and empty vector
912 VIGS plants (black trace) show that *SnASFF1*-silenced lines accumulate diacylsucroses that are
913 undetectable in control plants. The acylation at positions 3 and 4 on S2:14(4,10) was verified by
914 NMR as shown in Fig. S5. Extracted ion chromatogram (EIC) values indicate telmisartan as
915 internal standard (I.S.) [m/z 513.23] and the formic adducts of S2:12 (m/z 583.26), S2:13 (m/z
916 597.28), S2:14 (m/z 611.29), and S2:15 (m/z 625.31). ESI-, negative electrospray ionization.

917

918 **Figure 4. SnASAT1 and SnASAT2 sequential reaction produces diacylsucroses. (A)**

919 Maximum likelihood phylogeny analysis of amino acid sequences of ASAT candidates showed
920 that SnASAT1 and SnASAT2 cluster with characterized ASAT1 and ASAT2 in tomato species,
921 while SnAGAT1 clusters with the Solanaceae acylsugar acetyltransferases, SlASAT4, SsASAT5
922 and SqTAIAT. (B) LC-MS analysis of *in vitro* assay products showing that SnASAT1 produces
923 monoacylsucrose S1:10(10^{R4}) from sucrose and nC10-CoA substrates. The acylation at 4-position
924 was verified by NMR as shown in Fig. S9. Blue trace represents full enzyme assays, whereas the
925 black trace has heat-denatured SnASAT1. (C) LC-MS analysis of *in vitro* assay products showing

926 that SnASAT2 produces diacylsucrose S2:14(4,10) from monoacylsucrose S1:10(10^{R4}) and iC4-
927 CoA substrates. Green trace represents full enzyme assays, whereas the black trace has heat-
928 denatured SnASAT2. **(D)** *In vitro*-generated S2:14(iC4, nC10) and S2:14(iC4, iC10) co-elute (top
929 and middle, respectively) with the *in vivo* S2:14(4,10) isomers from *SnASFF1*-silenced lines
930 (bottom). S2:14(iC4, nC10) products in C and D are from two independent technical replicates.
931 **(E)** Summary of the reactions catalyzed by SnASAT1 and SnASAT2 with sucrose and
932 monoacylsucroses, respectively, and acyl-CoA substrates of different chain lengths. For more
933 details, see Fig. S12. *Acylation at 3-position verified from SnASAT2-products co-elution with
934 NMR characterized S2:14(4,10) with acyl chains as R₃ and R₄ from *S. nigrum*. All chromatographs
935 are combined extracted ion chromatogram (ESI-) with telmisartan internal standard (I.S.) (*m/z*
936 513.23) and the formic adducts of S1:10(10) (*m/z* 541.25) or S2:12(4,10) (*m/z* 611.29).

937

938 **Figure 5. SnASFF1 and SnAGAT1 sequential activities with both *in vitro* and *in vivo***
939 **diacylsucrose substrates produces triacylglycerols.** **(A)** SnASFF1 activity (purple traces) with
940 the major *in vitro* S2:14(iC4, nC10) from consecutive SnASAT1 and SnASAT2 reactions (top);
941 and the early- and late-eluting *in vivo* diacylsucroses S2:14(4,10) from VIGS-*SnASFF1* lines
942 (bottom and middle, respectively) produced two distinct sets of diacylglycerols that collectively
943 explains the multiple peaks of *S. nigrum* G2:14(4^{R3},10^{R4}) (magenta trace). The diacylglycerol
944 products from *in vitro* S2:14(iC4, nC10) and the late-eluting *in vivo* S2:14(4,10) are
945 chromatographically indistinguishable from the late-eluting *S. nigrum* G2:14(4^{R3},10^{R4}). Black
946 traces represent reactions with heat-denatured enzymes. **(B)** SnAGAT1 activities (orange traces)
947 with the diacylglycerols produced by SnASFF1 activities with *in vitro*- and *in vivo* diacylsucrose
948 S2:14(4,10) substrates produce two distinct set of triacylglycerol products that collectively explains
949 the triplet-like peaks of G3:14(2^{R2},4^{R3},10^{R4}) in *S. nigrum* leaf extracts (cyan trace). Combined
950 extracted ion chromatogram (EIC) under negative electrospray ionization (ESI-) showing formic
951 adducts of S2:14(4,10) (*m/z* 611.29), G2:14(4,10) (*m/z* 449.24) and G3:16(2,4,10) (*m/z* 491.25).
952 **(C)** Summary of results of SnASFF1 and SnAGAT1 activities in the reconstructed acylglycerol
953 biosynthetic pathway.

954

955 **Figure 6. Phylogenetic analysis of SnASFF1 and SpASFF1 reveals independent evolution.**
956 Invertase-like proteins from *S. lycopersicum*, *S. pennellii* and *S. nigrum* were aligned using

957 MUSCLE with default parameters in MEGA X. SnASFF1 is marked bold in purple, whereas
958 SpASFF1 is marked bold in black. Maximum likelihood tree constructed using the Jones-Taylor-
959 Thornton algorithm with 1000 bootstrap support using the topology-only tree in MEGA X.

960

961 **Figure 7. Independent validation of the reconstructed acylglucose biosynthetic pathway with**
962 ***in vitro* reverse activities. (A)** LC-MS analysis of reverse enzyme assay products from SnAGAT1
963 activity (orange trace) shows SnAGAT1 deacetylating *S. nigrum* fraction 1 tri-acylglucoses in the
964 presence of free coenzyme A. Combined extracted ion chromatogram (EIC) under negative
965 electrospray ionization (ESI-) showing telmisartan as internal standard (I.S.) (m/z 513.23) and
966 formic adducts of G2:12, G2:13, G2:14 and G2:15. The corresponding m/z 's of these compounds
967 are listed in Table S1. **(B)** Schematic representation of the reconstructed *S. nigrum* triacylglucose
968 biosynthetic pathway. Blue arrows indicate *in vitro* pathway reconstruction starting from sucrose,
969 green arrows indicate enzyme activities with plant-derived products. **(C)** SnASAT2 activities with
970 the major (top) – but not minor (second from the top) – *in vitro* S2:14(iC4, nC10) products and the
971 early- and late-eluting *in vivo* diacylsucrose S2:14(4,10) substrates from VIGS-*SnASFF1* lines
972 (second from the bottom and bottom, respectively) produce monoacylsucrose products that can
973 further be deacylated by SnASAT1 activity (bottom right). The monoacylsucrose intermediates
974 from the late-eluting *in vivo* S2:14(4,10) and *in vitro* S2:14(iC4, nC10) (top and third from top)
975 are chromatographically indistinguishable from S1:10(nC10^{R4}). Combined extracted ion
976 chromatogram (EIC) under negative electrospray ionization (ESI-) showing formic adducts of
977 S2:14(4,10) (m/z 611.29) and S1:10(10) (m/z 541.25).

978

979 **Figure 8. *In vivo* validation of the reconstructed acylglucose biosynthetic pathway with**
980 **VIGS. (A)** Comparison of acylglucose accumulation in *SnASAT1*-targeted and empty vector VIGS
981 plants. Acylsugars were analyzed using LC-MS in ESI+ mode. **(B)** Comparison of acylglucose
982 accumulation in *SnASAT2*-targeted and empty vector VIGS plants. Acylsugars were analyzed
983 using LC-MS in ESI- mode. **(C)** Comparison of the ratio of diacylglucose G2:14 and triacylglucose
984 G3:16 in *SnAGAT1*-targeted and empty vector VIGS plants. Diacylglucose quantities were
985 measured by integrating peak areas of G2:12, G2:13, G2:14 and G2:15, whereas triacylglucose
986 quantities were measured by integrating peak areas of G3:14, G3:15, G3:16 and G3:17. The
987 integrated peak areas were normalized to the internal standard telmisartan and dry leaf weights.

988 Peak areas of G2:14 and G3:16 were integrated under negative mode and normalized to the internal
989 standard telmisartan and dry leaf weights to measure the ratio between tri- and diacylglucoses. The
990 corresponding m/z of analyzed acylsugars are listed in Table S1 unless otherwise specified.
991 Significant levels are shown (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; Welch's two sample t-test).
992
993

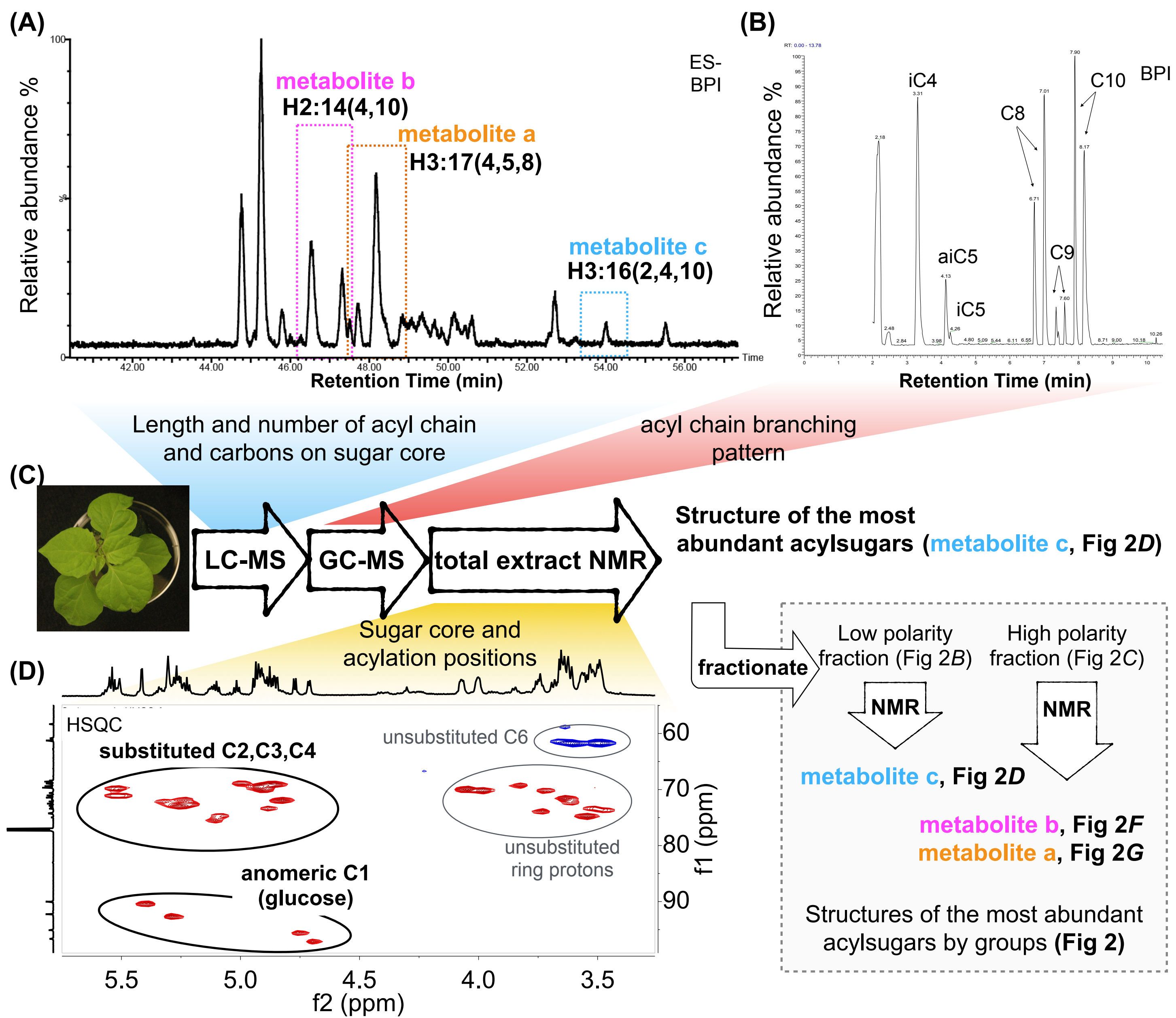


Figure 1

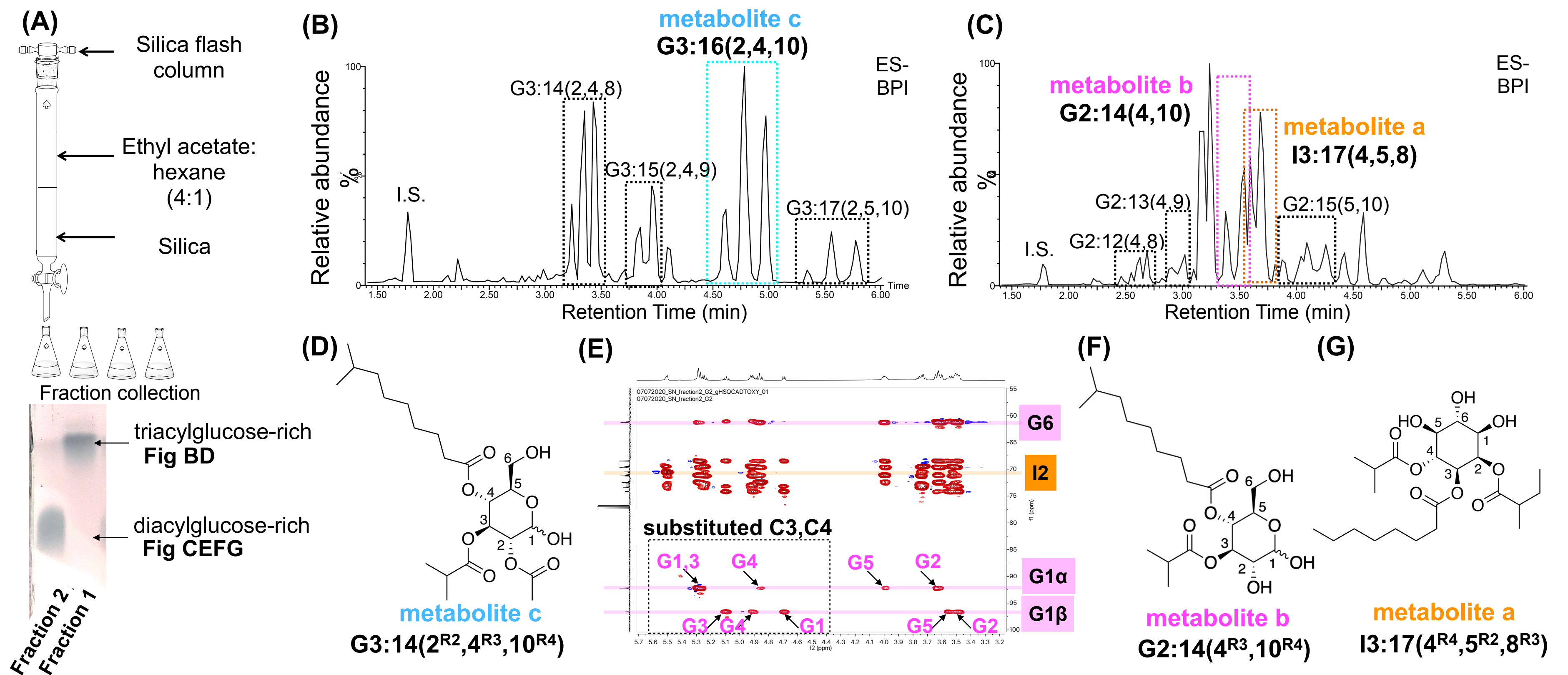


Figure 2

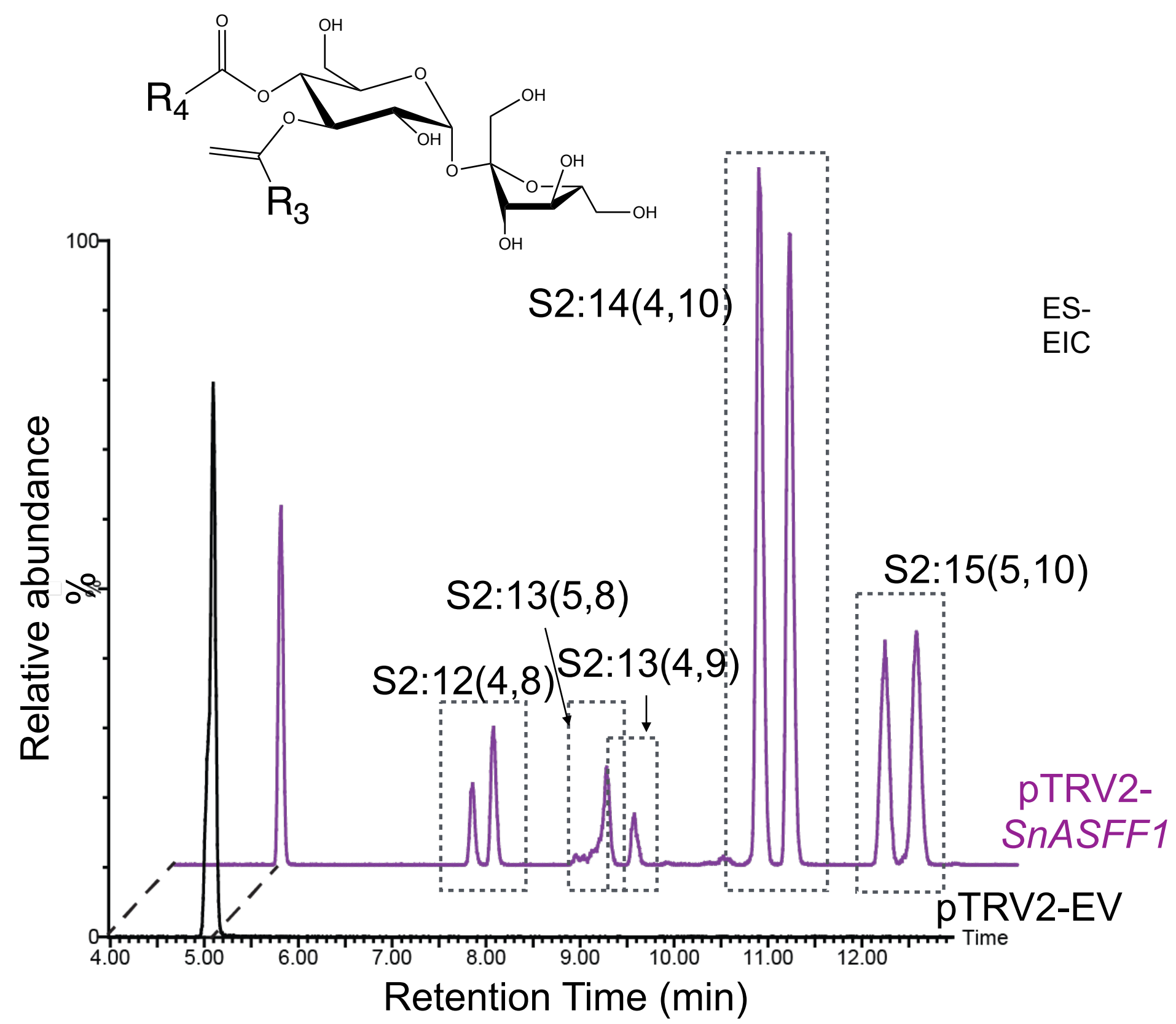


Figure 3

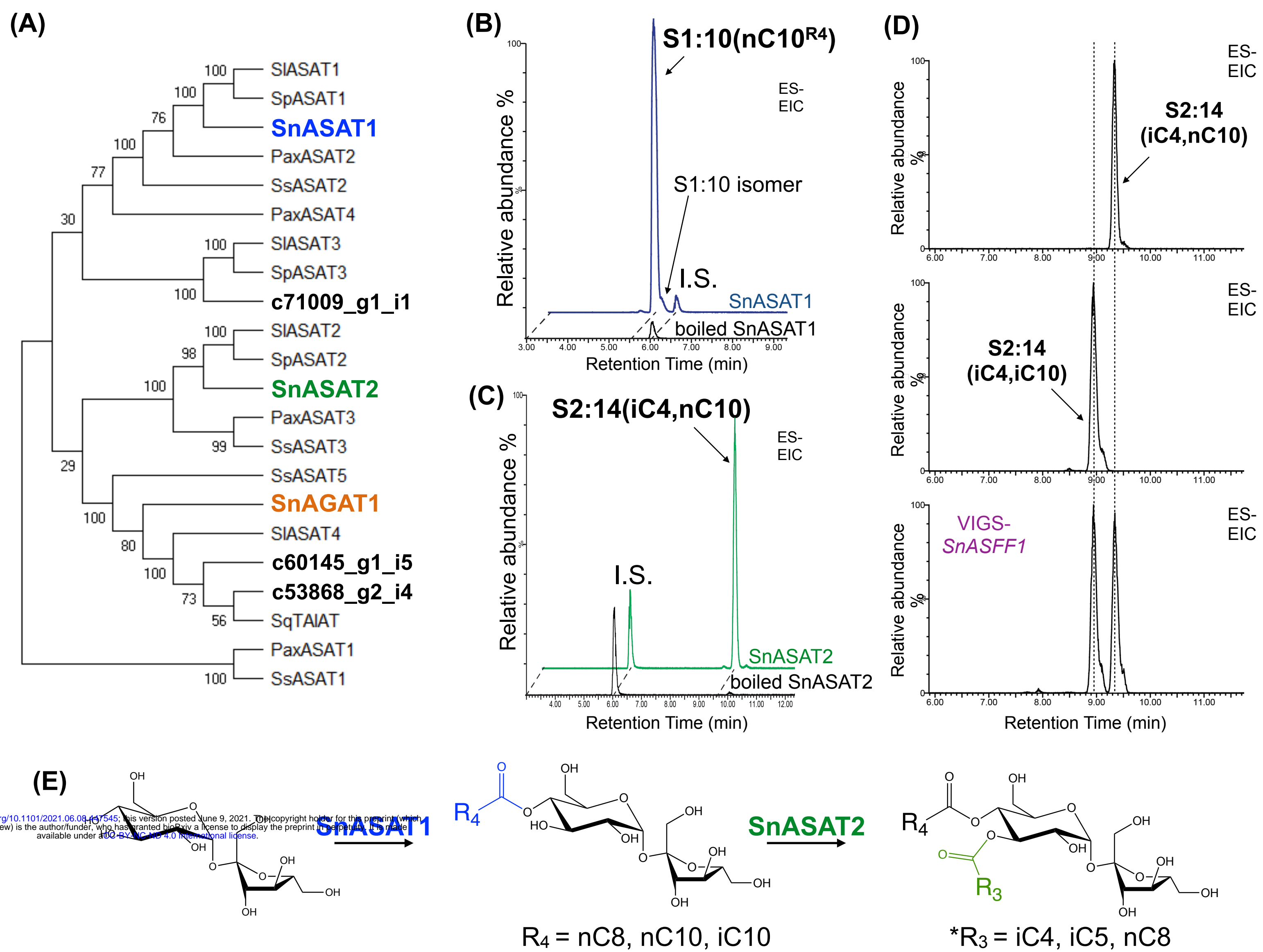


Figure 4

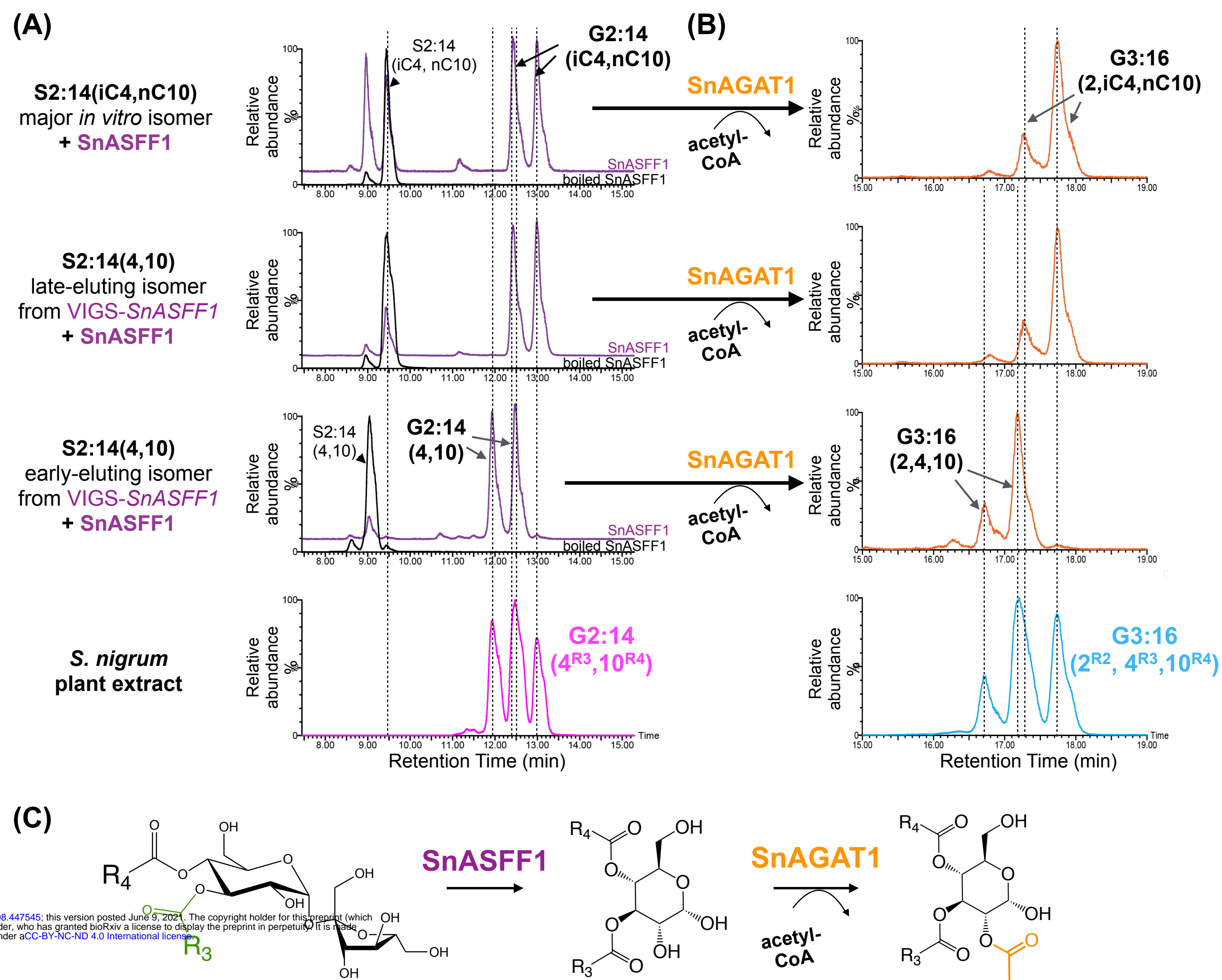
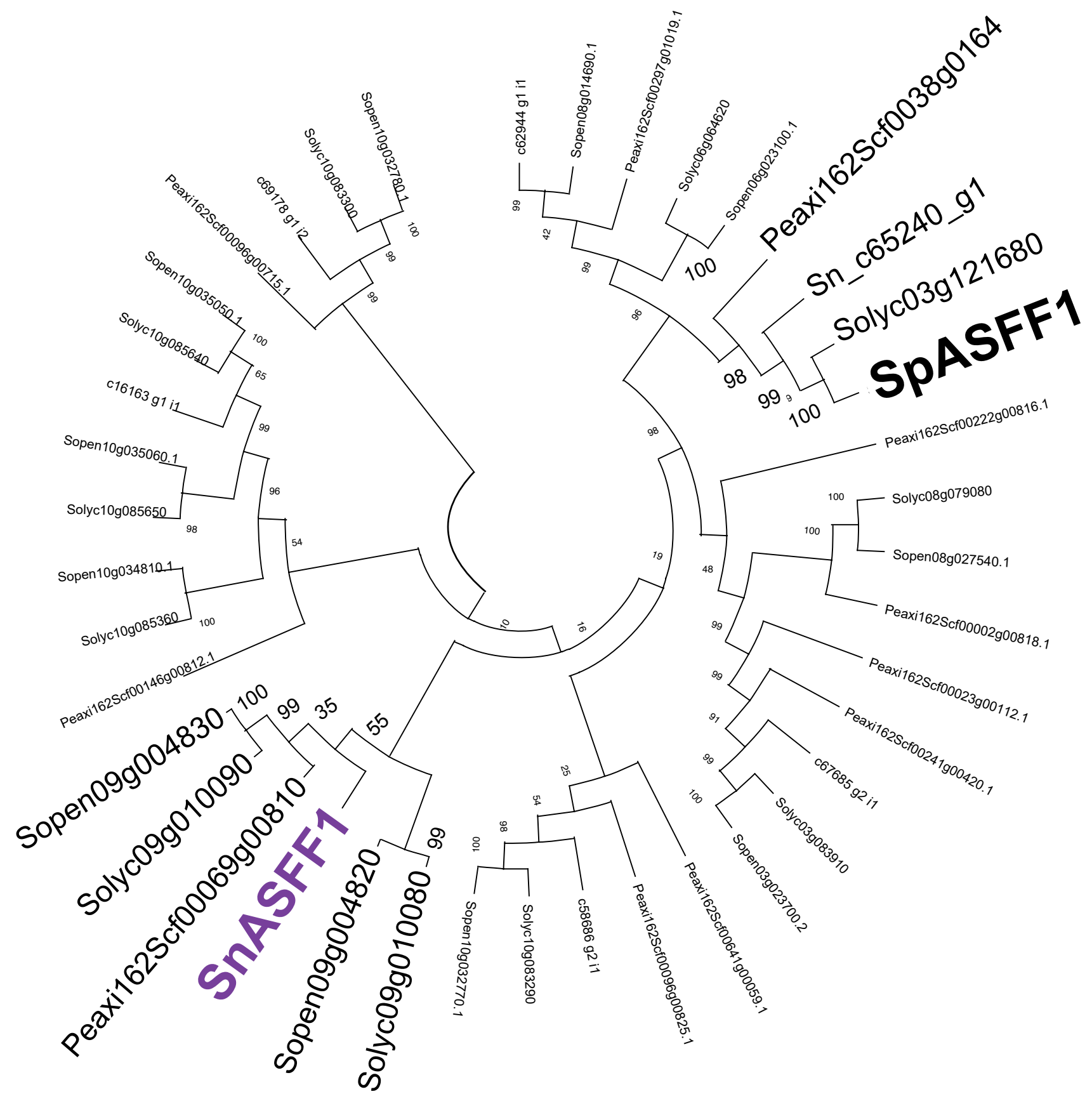
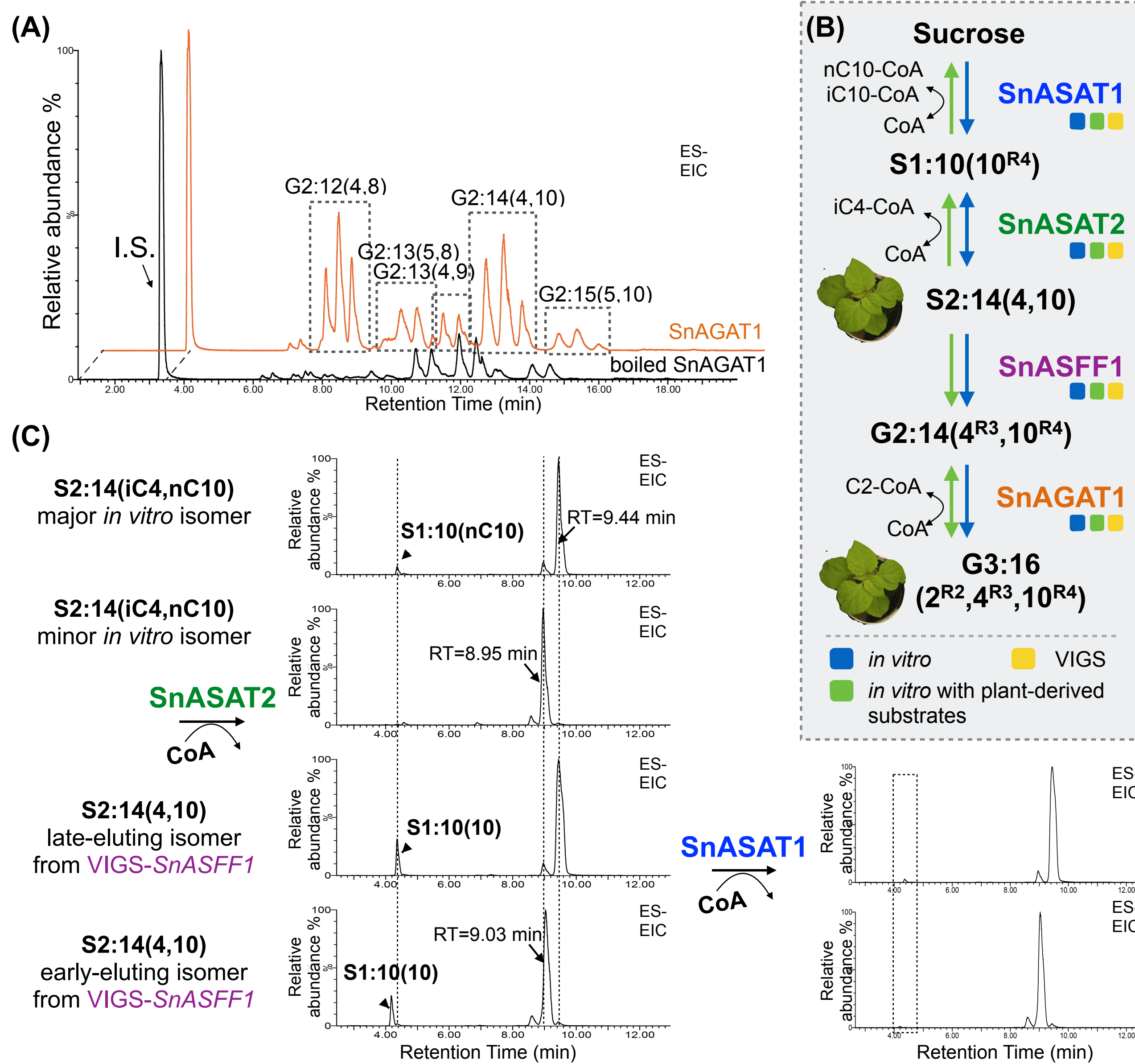


Figure 5



bioRxiv preprint doi: <https://doi.org/10.1101/2021.06.08.447545>; this version posted June 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6



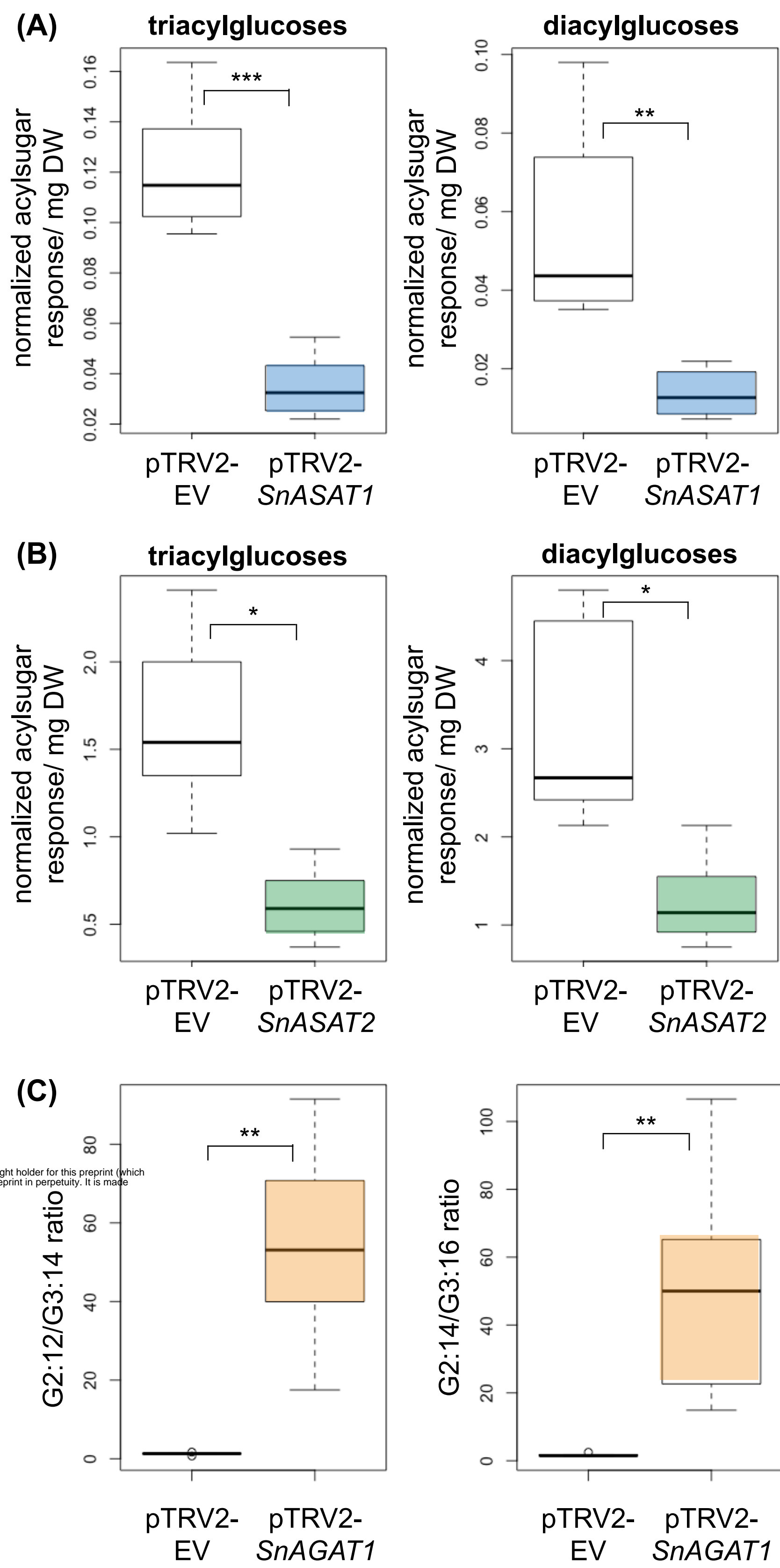


Figure 8

Supplementary Text

Acylsugar annotation

Acylsugar structures were inferred from positive and negative mode MS collision-induced dissociation (CID) as previously described (15, 29, 32, 34). In short, we employed two discrete negative mode CIDs, where the lower energy (15V) removes acyl chains from acylsugars in a stepwise manner and the higher energy (30V) creates fatty acid fragment ions from acylsugar acyl chains. For acylsucroses, the broken glycosidic linkage by positive mode CID (15 or 30V) provided additional information about acyl chain attachment: a neutral loss of the furanose ring resulting in $[M+NH_4-FRUC]^+$ accumulation would indicate that all acyl chains are attached to the glucopyranose ring. Ring acylation information is indicated in Table S11. Examples of typical fragmentation of *S. nigrum* di- and triacylglucoses and triacylinositols are depicted in Fig. S2, whereas examples of mass spectra of the mono- and di-acylsucrose intermediates on acylglucose biosynthetic pathways are depicted in Fig. S7C and S9.

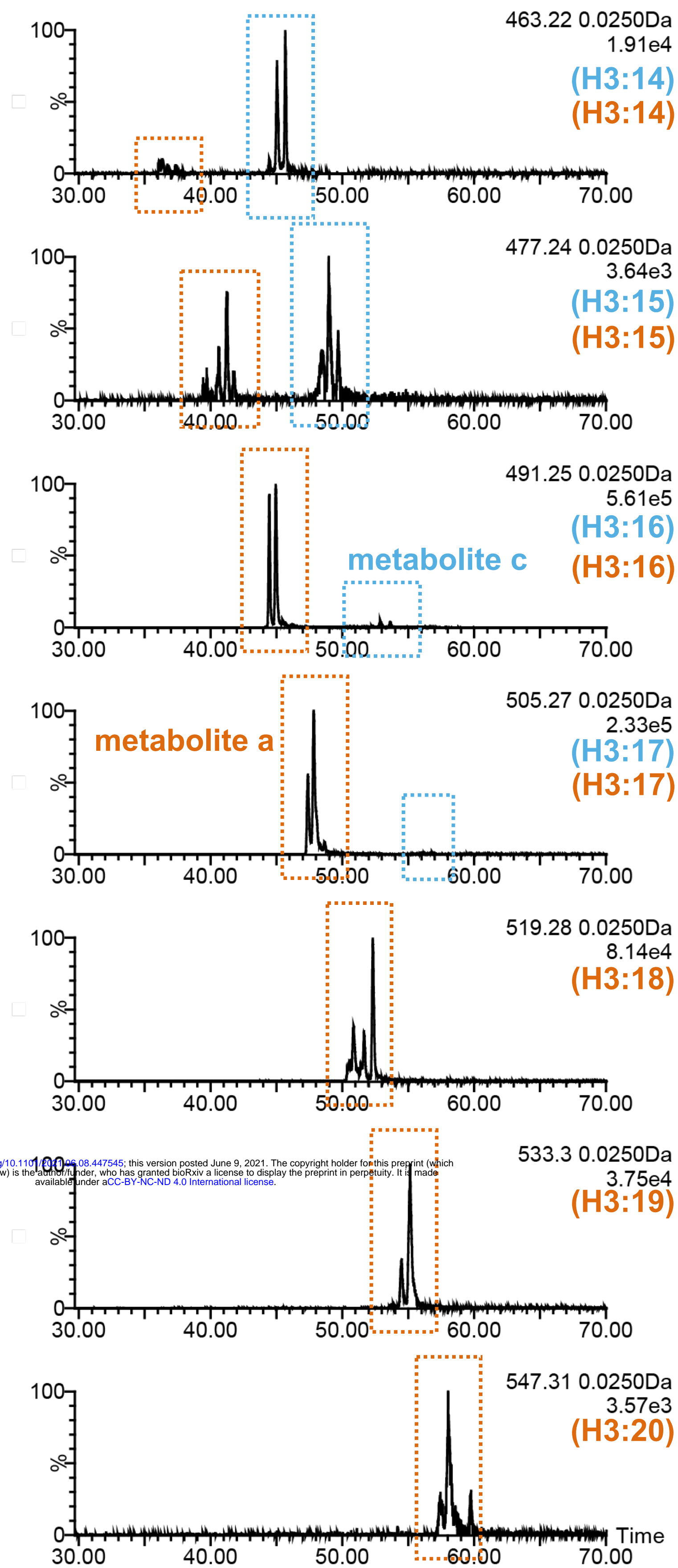
Below is the stepwise protocol with a list of criteria used in this manuscript to annotate and report acylsugars with different confidence levels. Putative acylhexoses that meet all the following criteria A-F were annotated with highest confidence in this manuscript. Those not meeting criteria E and/or F were reported with medium confidence, whereas potential acylhexoses that only meet criteria A were reported with low confidence. Confidence levels for each reported acylhexose are specified in Table S11.

- (A) **Exact mass.** Experimental exact masses of potential acylsugars (m/z listed in Table S11) were compared with theoretical exact masses of possible acylsugars. A mass difference within 10 ppm met this criterion. Then, we generated extracted ion chromatograms (EICs) of each potential acylsugar to compare with the EICs generated in B and C.
- (B) **Presence of co-eluting sugar core fragments.** For acylglucoses, EICs of characteristic glucose core structure (m/z 127.04 in 15 or 30V positive mode CID and m/z 143.03 in 0V negative mode CID) were generated. For acylsucroses, EICs for the sugar core structure on acylsucroses (m/z 341.10 and 323.09 in negative mode) were generated. As acylinositols fragment less well in negative mode, the characteristic *myo*-inositol core fragment ion (m/z 127.04) was used to generate EICs in positive mode CID (30V).
- (C) **Presence of co-eluting fatty acid carboxylate fragments in high CID potential (30V).** EICs of the carboxylate fragments of C4 (m/z 87.04), C5 (m/z 101.06), unsaturated C5 (m/z 99.05), C6 (m/z 115.07), C7 (m/z 129.09), C8 (m/z 143.10), C9 (m/z 157.12), C10 (m/z 171.14), C11 (m/z 185.15) and C12 (m/z 199.17) were generated in highest CID potential (30V) in negative mode.
- (D) **Presence of co-eluting fragments that correspond to stepwise loss of acyl chains.** The EICs generated in (A), (B) and (C) were closely examined for co-eluting peaks that indicate potential acylsugars. For each potential acylsugar, spectra from the top of the peak were then carefully examined for stepwise losses of acyl chains in negative mode CID at 15V. Acylinositol fragmentation for this criterion was observed in positive mode due to poor negative mode fragmentation^{5,6}.

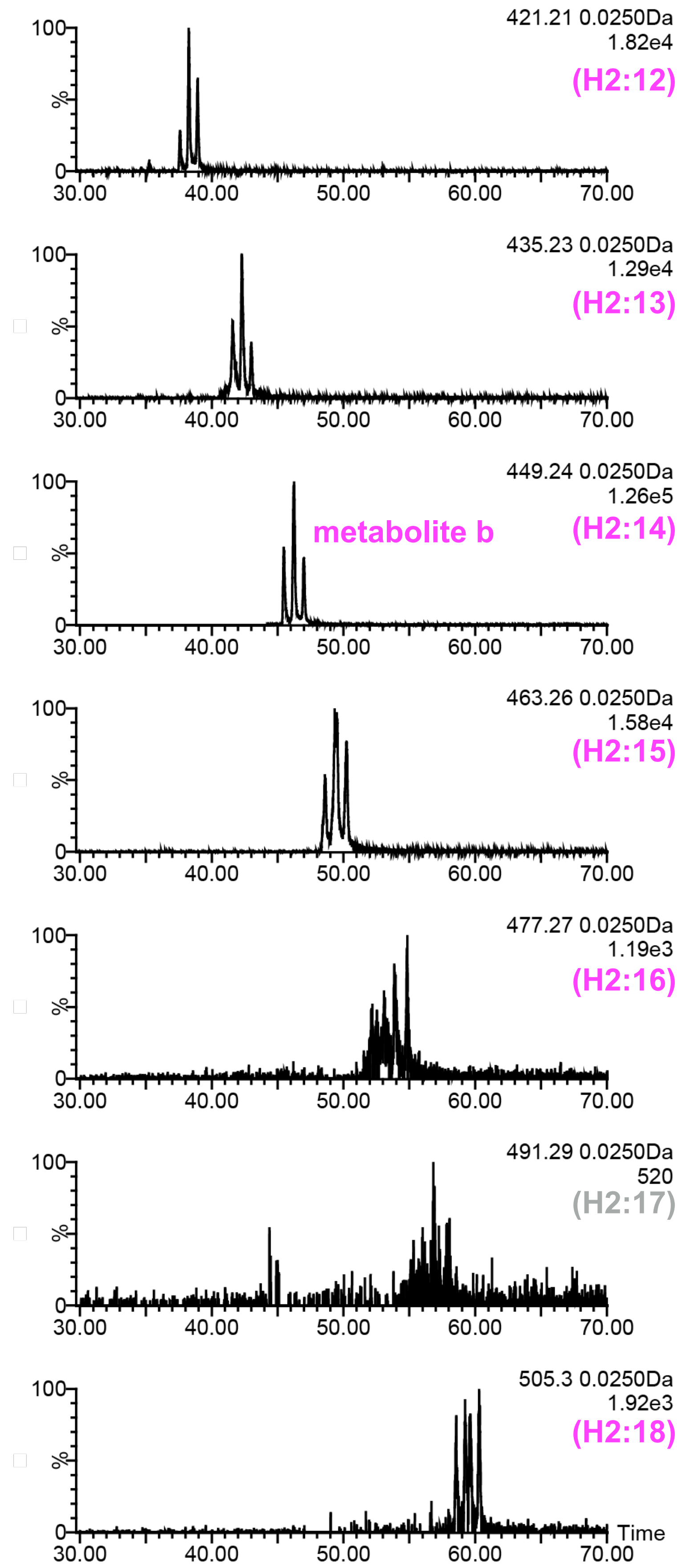
- (E) **Confirmed by MS/MS.** The above annotation were then confirmed with data-dependent (DDA) MS/MS when applicable.
- (F) **Consistency across samples.** Acylsugar presence was monitored across multiple metabolite extracts. Potential acylsugars met this criterion when they were detected across all samples analyzed.

As an example, the three peaks with m/z 491.25 with retention time 52.66, 53.22 and 54.02 min on the 110-min share identical ($\delta_m = 0.1$ ppm) with the $[M+\text{formate}]^-$ of H3:16 (Fig. S3A). To confidently annotate these peaks, co-elution of $[M+\text{formate}]^-$ (m/z 491.25) and glucose fragment ion (m/z 143.03) was verified on EICs generated in 0V negative mode. Co-eluting carboxylate anion were identified for C10 carboxylate anion (m/z 171.14) and C4 carboxylate anion (m/z 87.04) in 30V negative mode. Presence of C4 and C10 acyl chains was further confirmed by ion abundance of $[M-C4-H_2O]^-$ (m/z 357.19) and $[M-C4-C10-2H_2O]^-$ (m/z 185.04) on CID at 15V negative mode. As these fragmentation were confirmed on MS/MS fragmentation of $[M+\text{formate}]^-$ (m/z 491.25), we report the three peaks between 52-55 min with m/z 491.25 as triacylhexoses H3:16(2,4,10) with high confidence level based on LC-MS data.

(A)



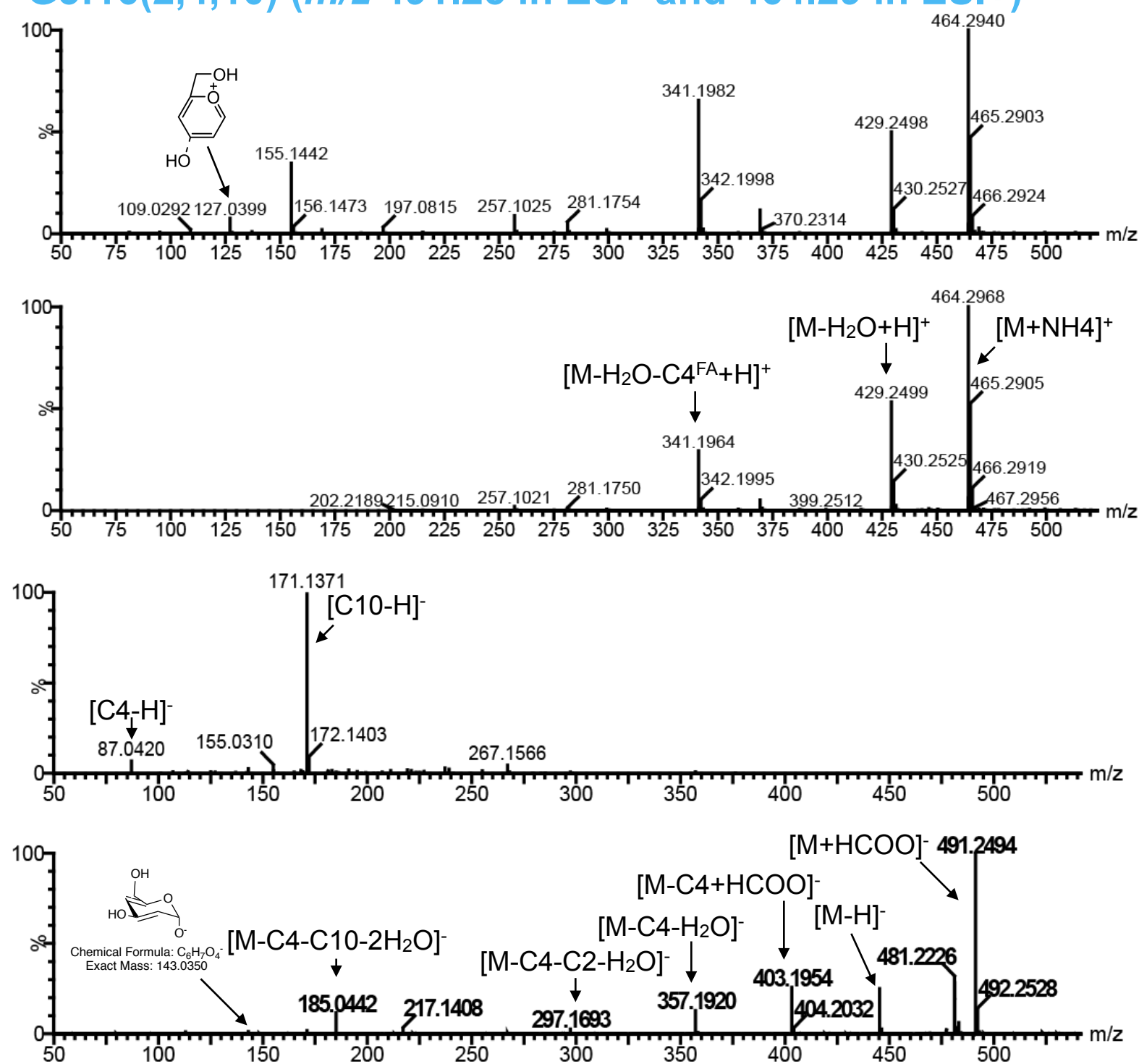
(B)



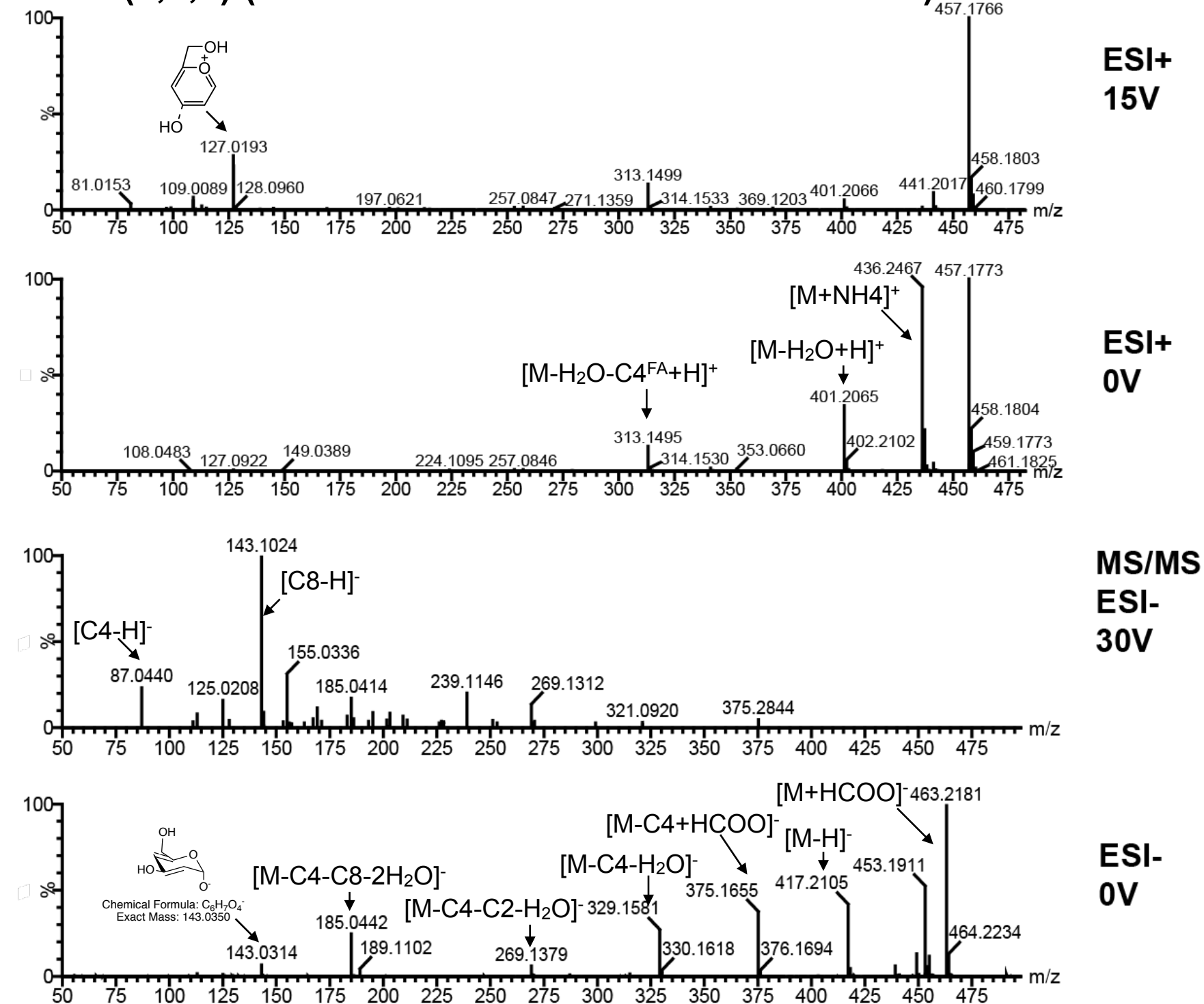
bioRxiv preprint doi: <https://doi.org/10.1101/2021.06.09.447545>; this version posted June 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S1. Chromatographs of abundant acylhexoses in *S. nigrum*. (A) Extracted ion chromatograms (EICs) from negative ion mode LC-MS analyses showing selected m/z values correspond to $[M+\text{formate}]^-$ of hexose triesters. Observed peaks in these chromatograms separate into two distinct retention time groups: peaks corresponding to triacylglucoses, such as metabolite c, are highlighted with cyan dash boxes, and those corresponding to triacylinositols, such as metabolite a, are highlighted in orange dash boxes. (B) Extracted ion chromatograms from negative ion mode LC-MS analyses showing selected m/z values correspond to $[M+\text{formate}]^-$ of hexose diesters. Numerous instances suggested common chromatographic overlap of isomeric acylhexoses was common. The corresponding m/z of these compounds are listed in Table S1 and Supplementary Text.

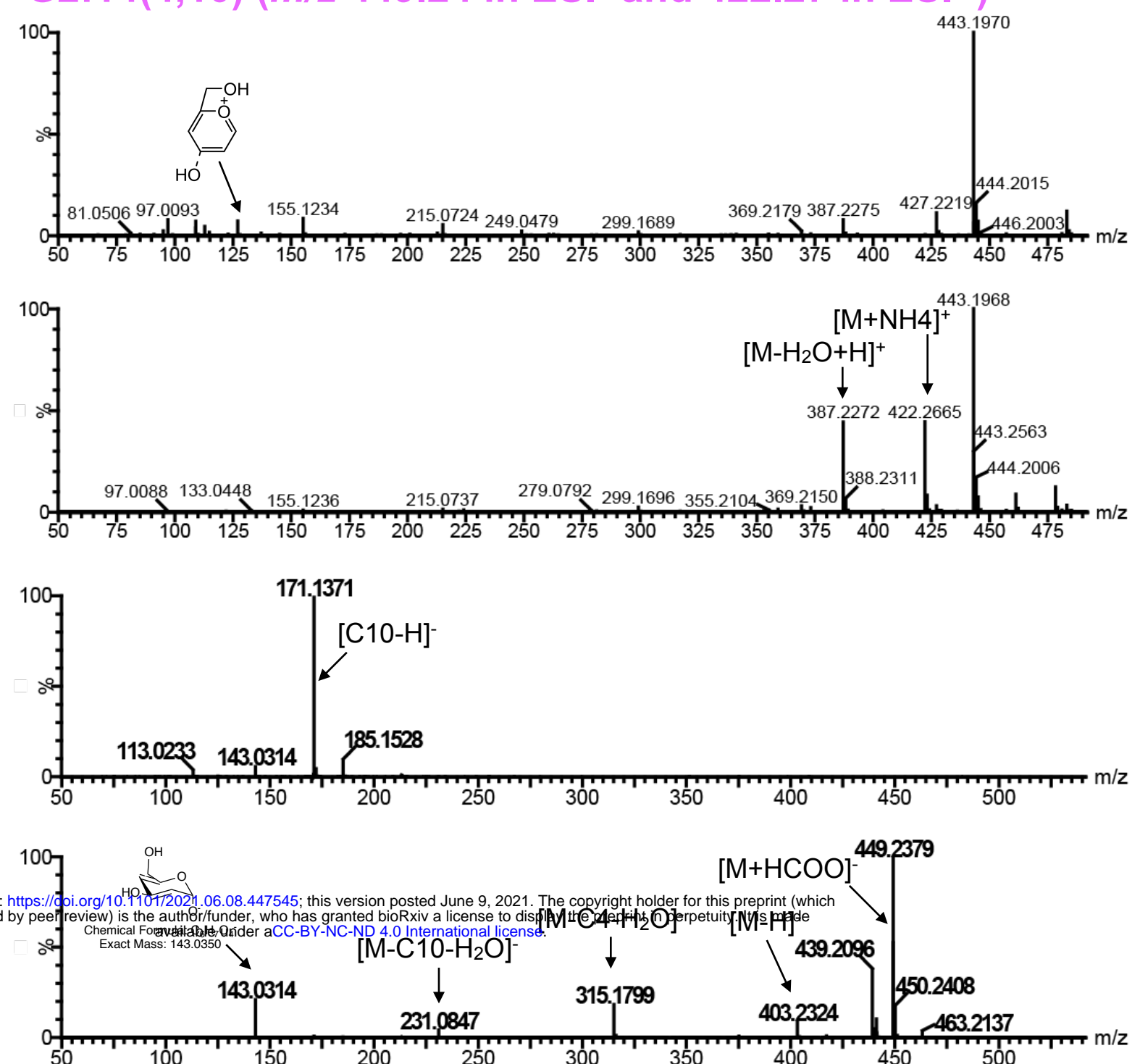
(A) Metabolite c
G3:16(2,4,10) (m/z 491.25 in ESI- and 464.29 in ESI+)



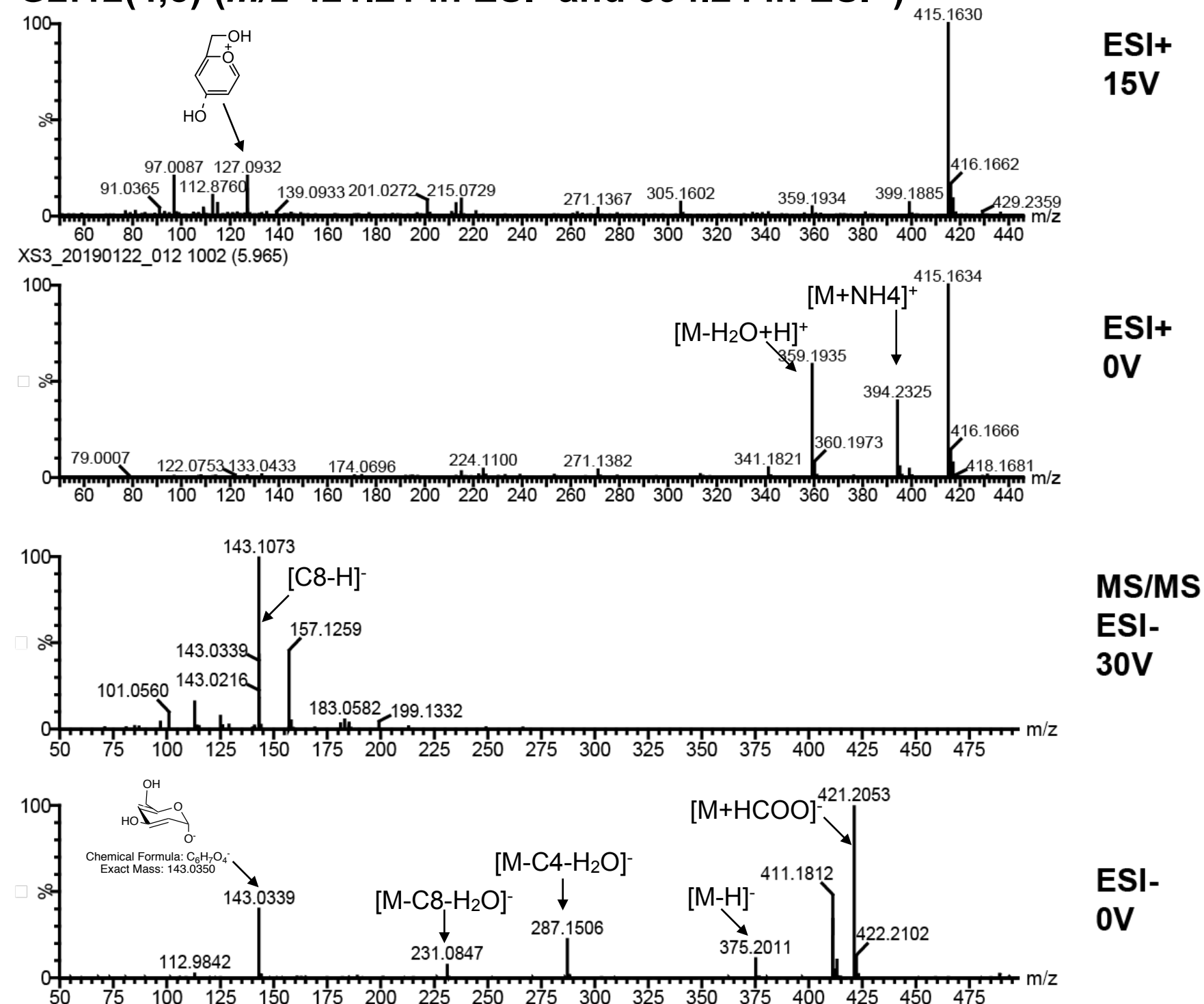
G3:14(2,4,8) (m/z 463.22 in ESI- and 436.25 in ESI+)



(B) Metabolite b
G2:14(4,10) (m/z 449.24 in ESI- and 422.27 in ESI+)



G2:12(4,8) (m/z 421.21 in ESI- and 394.24 in ESI+)



(C) Metabolite a
I3:16(4,4,8) (m/z 491.25 in ESI- and 464.28 in ESI+)

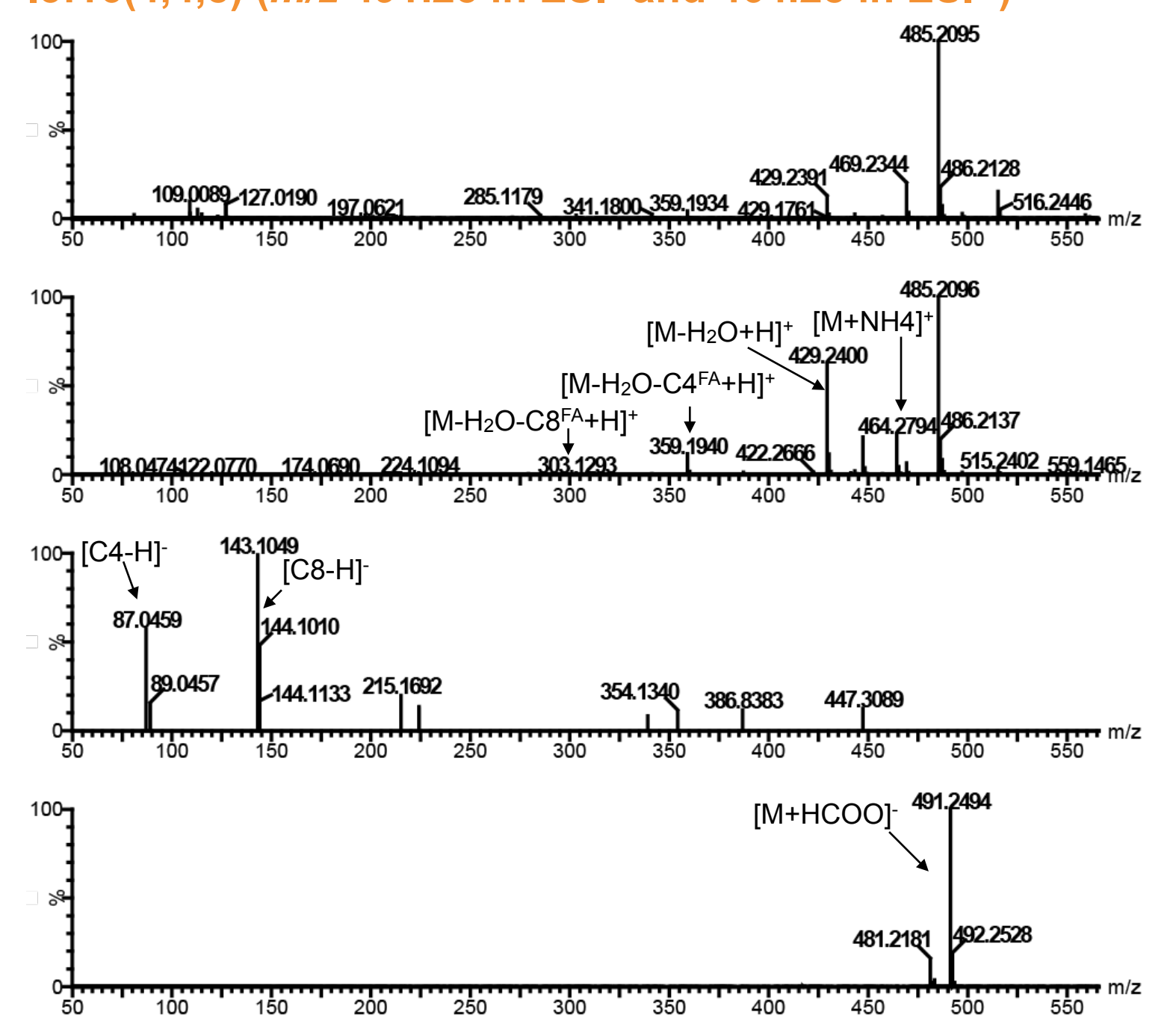
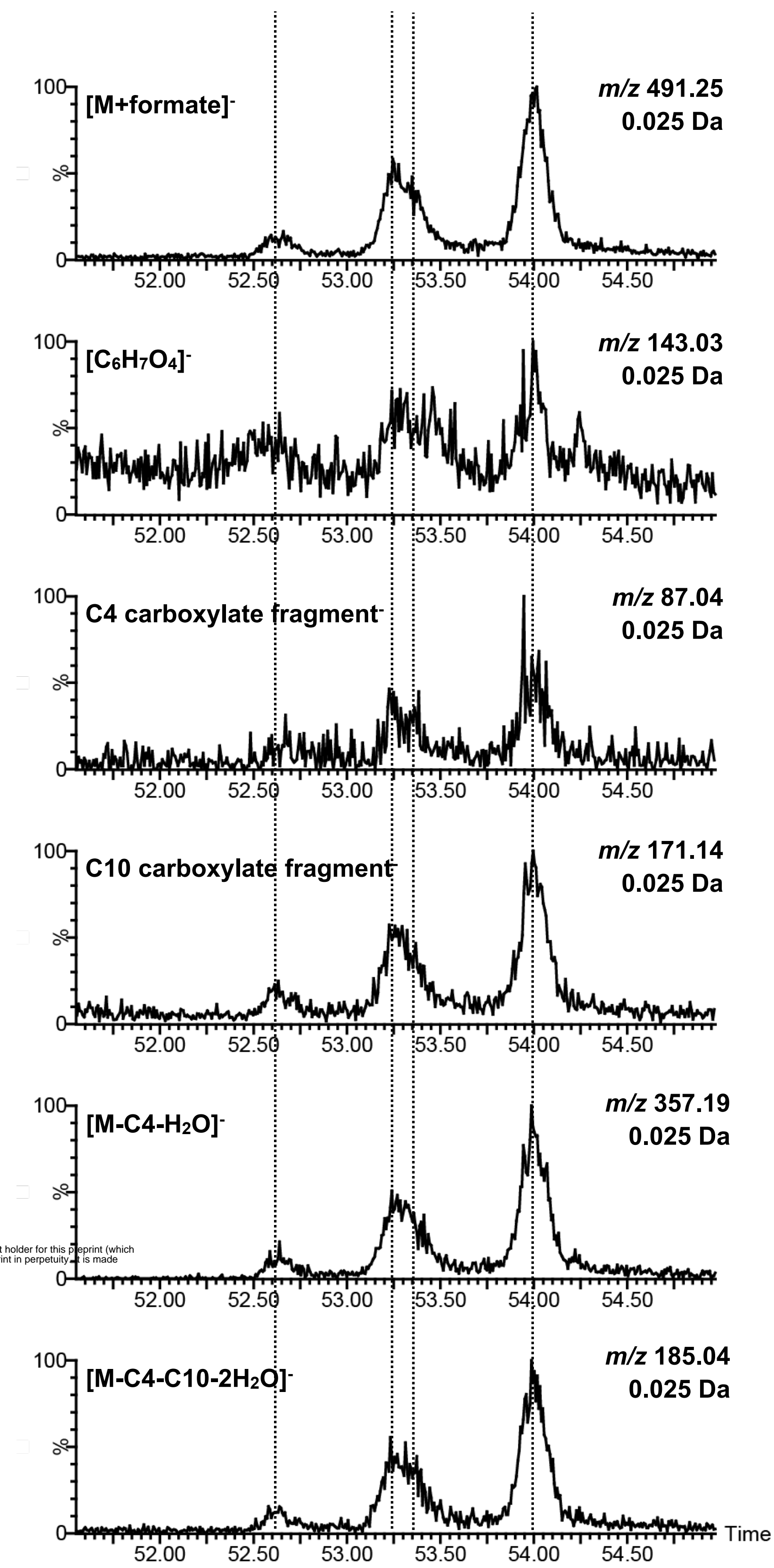
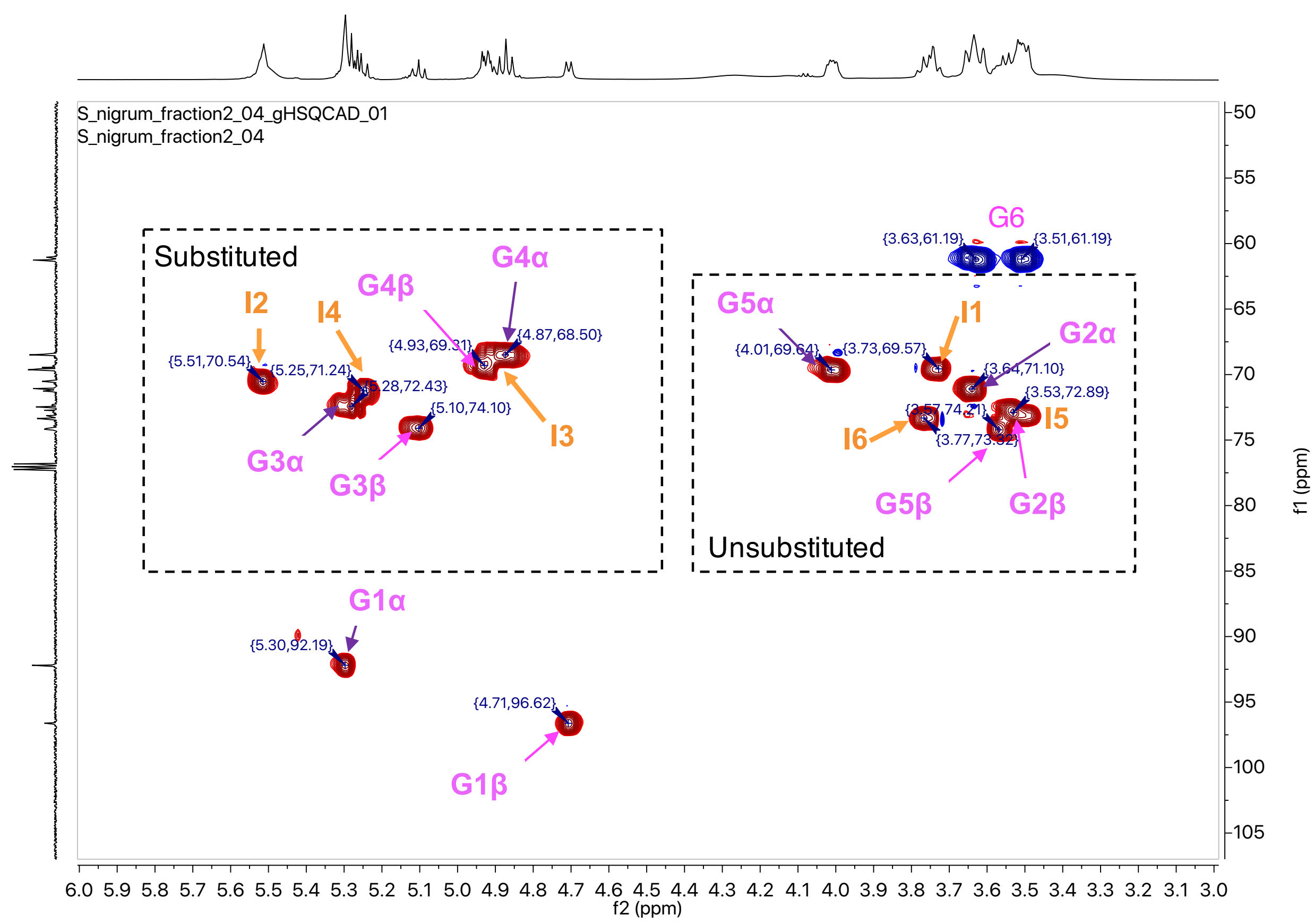


Fig. S2. Mass spectra of abundant acylhexoses in *S. nigrum*. *S. nigrum* di- and triacylglucoses and triacylinositols fragment in either negative or positive ion mode by losing the short and medium-length acyl chains as neutral fatty acids, which can also be seen in negative ion mode. These observations allow for determination of the number and length of acyl chains attached to each acylglucose. For more detail, see Supplementary Text. **(A)** Fragmentation of *S. nigrum* triacylglucoses G3:16 and G3:14. Fragmentation of G3:16 in 0V ESI- mode is characterized by the loss of C4, C10 and C2 ketenes. ESI- MS/MS spectrum (30V) generated from [M+formate]⁻ of G3:16 reveals the presence of C4 (*m/z* 87.04) and C10 (*m/z* 171.14) fatty acids. Similarly, fragmentation of G3:14 in ESI- mode is characterized by the loss of C4, C8 and C2 ketenes under no collision energy (0V), while 30V ESI- MS/MS of product ions from [M+formate]⁻ of G3:14 reveals the presence of C4 (*m/z* 87.04) and C8 (*m/z* 143.10) fatty acids. Fragmentation of both triacylglucoses in ESI+ mode (0 and 15V) results in the loss of a C4 fatty acid, followed by loss of a C10 or C8 fatty acid, respectively. **(B)** Fragmentation of *S. nigrum* diacylglucoses G2:14 and G2:12. Fragmentation of G2:14 in ESI- mode is characterized by the loss of C4 and C10 ketenes under no collision energy (0V). ESI- MS/MS spectra (30V) of product ions generated from [M+formate]⁻ of G2:14 reveals the presence of C4 (*m/z* 87.04) and C10 (*m/z* 171.14). Similarly, fragmentation of G2:14 in ESI- mode is characterized by the loss of C4 and C8 ketenes under no collision energy (0V), while 30V ESI- MS/MS of product ions from [M+formate]⁻ of G2:12 reveals the presence of C4 (*m/z* 87.04) and C8 (*m/z* 143.10). ESI+ mode (15V) of both diacylglucoses reveal the loss of C4 and a C10 or C8, respectively, while 0V ESI+ results in little fragmentation. **(C)** Fragmentation of I3:16 from *S. nigrum*. Fragmentation of I3:16 in ESI+ mode is characterized by the loss of C4 and C10 fatty acids under no collision energy (0V). ESI- MS/MS spectra (30V) of product ions generated from [M+formate]⁻ of I3:16 reveals the presence of C4 (*m/z* 87.04) and C8 (*m/z* 143.10). In contrast to acylglucoses, 0V ESI- result in little fragmentation for acylinositols.



bioRxiv preprint doi: <https://doi.org/10.1101/2021.06.08.447545>; this version posted June 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S3. Co-chromatography of metabolite c and characteristic fragment ions under different collision-induced dissociations (CID). Extracted ion chromatograms (EICs) of (from top to bottom) $[M+\text{formate}]^-$ of metabolite c (m/z 491.25) in 0V ESI-, glucose core structure $[\text{C}_6\text{H}_7\text{O}_4]^-$ (m/z 143.03) in 0V ESI-, C4 (m/z 87.04) and C10 (m/z 171.14) carboxylate fragment under CID potential 30V, and $[M-\text{C4}-\text{H}_2\text{O}]^-$ (m/z 357.19) and $[M-\text{C4}-\text{C10}-\text{H}_2\text{O}]^-$ (m/z 185.04) fragment ions under CID potential 15V. For more details, see Supplementary Text.



Supplemental Figure 4

Fig. S4. HSQC spectrum reveals two type of acylsugars in the more-polar fraction of *S. nigrum* leaf surface extract. HSQC spectrum of fraction 2 shows signals from diacylglucoses (magenta) and triacylinositols (orange) that were confirmed after HSQC-TOCSY separating correlation signals.

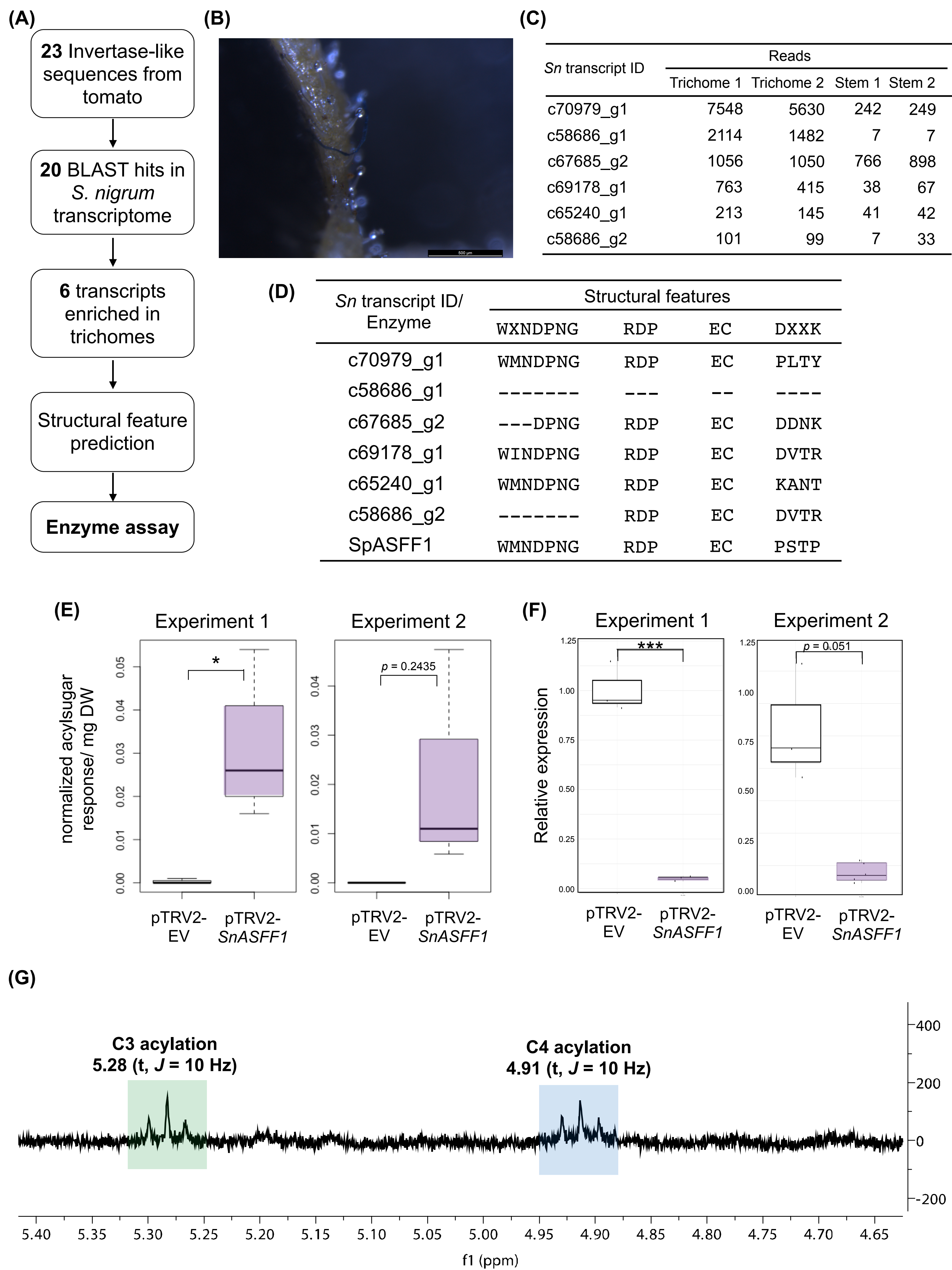
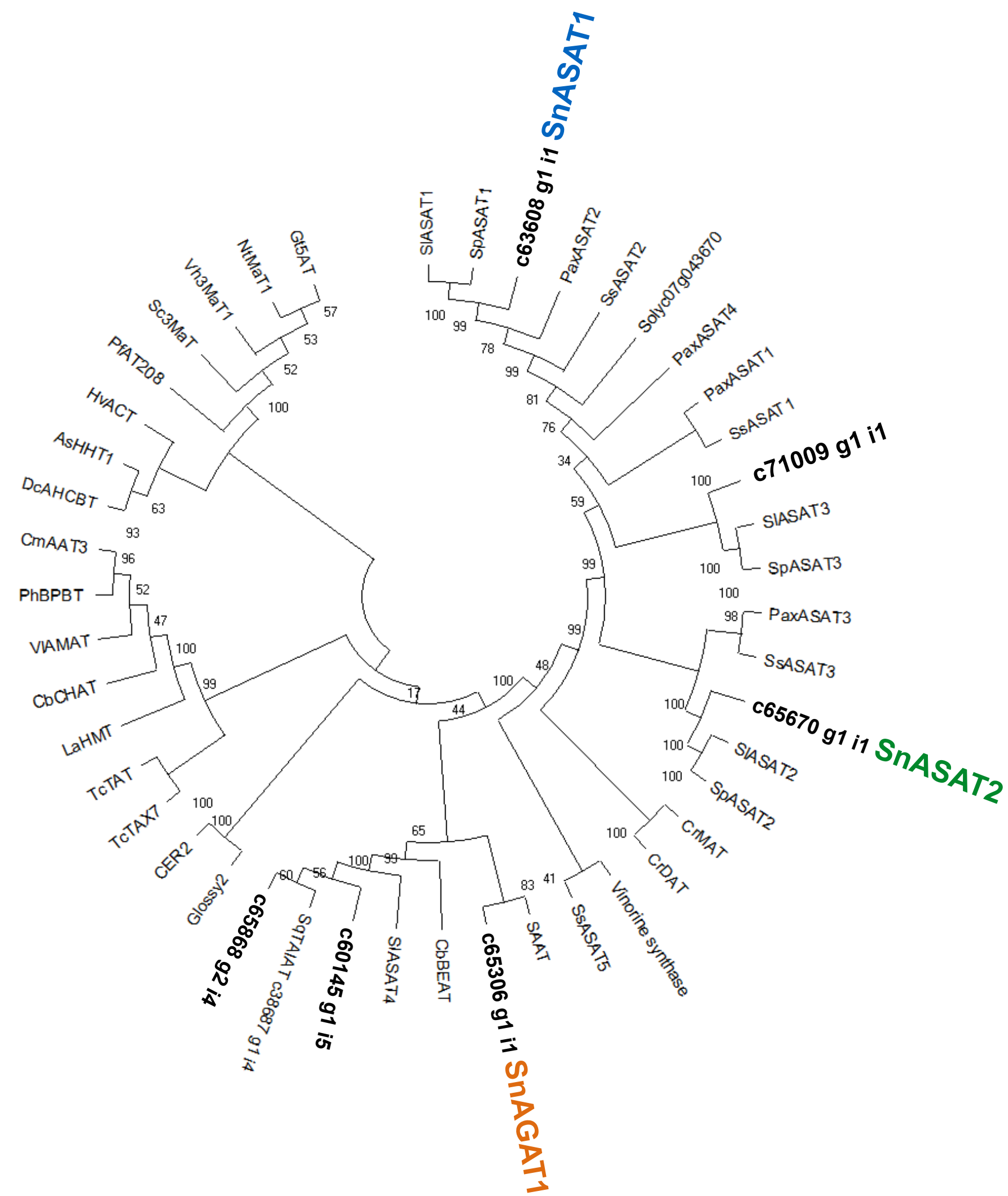


Fig. S5. Identification of a trichome-expressed ASFF candidate involved in *S. nigrum* acylglucose biosynthesis. (A) Flow chart of identifying ASFF candidate transcripts in *S. nigrum* transcriptome. (B) Close-up photo of a *S. nigrum* peduncle with acylsugar droplets accumulating on trichome tips. Excessive acylsugars produced on the peduncle appear to drip down and coat the trichome-free fruit. (C) Six invertase-like candidate transcripts are enriched in *S. nigrum* trichome RNA-Seq data from Moghe et al. (2017) (D) Structural features identified a transcript, *c70979_g1* (*SnASFF1*), with conserved catalytic sites and a non-canonical substrate binding site, as seen on SpASFF1. (E) Comparison of S2:14(4,10) accumulation in *SnASFF1*-targeted and empty vector VIGS plants from two independent experiments. Acylsugars were analyzed using LC-MS in ESI- mode. Peak areas of S2:14(4,10) (m/z 611.29) were integrated under negative mode and normalized to the internal standard telmisartan and dry leaf weights. See Table S5 for details. (F) qPCR analysis of *SnASFF1* transcript abundance in VIGS plants from two independent experiments. Relative transcript abundance was generated using the $\Delta\Delta C_t$ method (53) and normalized to the geometric mean of two actin gene transcript levels. Significant levels are shown (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; Welch's two sample t-test). (G) ^1H NMR of S2:14(4,10) purified from *SnASFF1*-targeted VIGS plants. Two triplet peaks at 5.28 (green box) and 4.91 ppm (blue box) indicate the 3- and 4-positions on sucrose, respectively, are acylated.

(A)



(B)

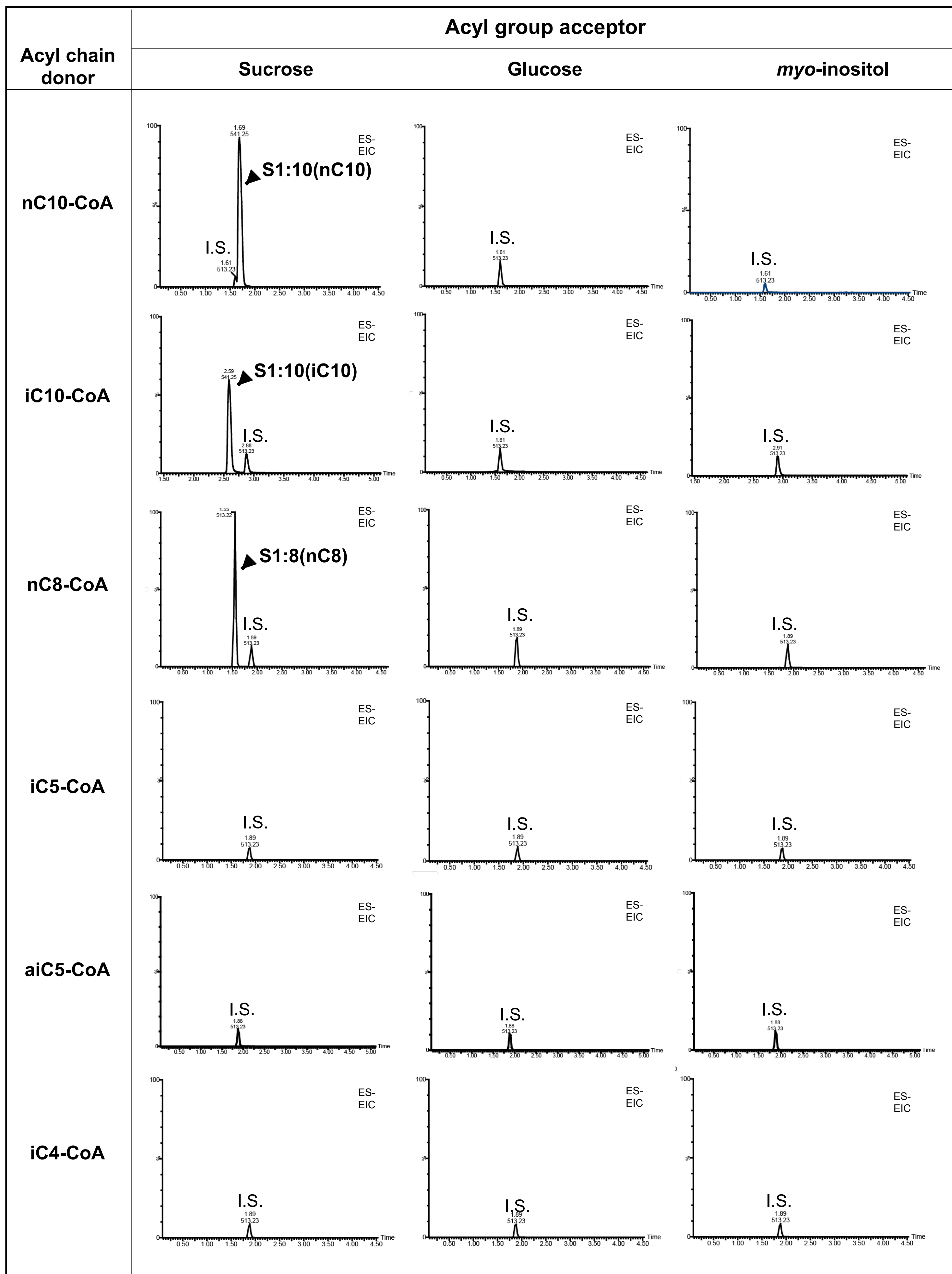
Sn transcript ID	Homolog	sequence length	Number of Reads			
			Trichome 1	Trichome 2	Stem 1	Stem 2
<i>c63608 g1</i>	SIASAT1	430	2750	2014	3	19
<i>c65670 g1</i>	SIASAT2	442	1574	1257	3	30
<i>c71009 g1</i>	SIASAT3	430	1993	1636	32	89
<i>c60145 g1</i>	SIASAT4	438	4904	3625	20	40
<i>c53868 g2</i>	SIASAT4	435	4928	3997	27	54
<i>c65306 g1</i>	SsASAT5	451	3765	3050	16	46

Fig. S6. Six ASAT homologs are preferentially expressed in *S. nigrum* trichomes. (A) Phylogenetic analysis of ASAT candidates and previously characterized BAHDs. The phylogenetic tree shows several previously characterized ASATs and other BAHD acyltransferases. BAHD candidates in *S. nigrum* are marked in bold, whereas characterized SnASATs are marked with respective colors (SnASAT1 in blue, SnASAT2 in green and SnAGAT1 in orange). **(B)** Sequence length and RNA-Seq data for homologs of characterized ASATs in *S. nigrum*. RNA-Seq data are derived from Moghe et al. (2017).

		1	10	20	30	40	50	60
SnASAT1	c63608_g1_i1							
SnASAT2	c65670_g1_i1	MSASKQLVLLSKKLIKPSPTPLSHRIHNL	SLMDQMGTHSYMAAMLFYPKQDTTSMPEP					
	c71009_g1_i2	MSVSRLVSSVCKKIKPYSPTPI	SHRCHKLSYLDQMIGGVYLPFAFFYPKLSNTWSNKPS					
	c53868_g2_i2	MAAS---TILSKMKIKLFSPTPPSLRRYNLSFMD	CINLPQYMPLAFFYKPKENCN----	K				
	c60145_g1_i5	-MANLEIQIESRKLKPSIPTPNHLRTLKLSLFDQLAPLAYVPLLFHYLPSSTEETI---						
SnAGAT1	c65306_g1_i1	----MEIIILSKEIVKPTPTPPHLRNHKL	CIFDQIALEICITPIFFYKNNNNIVD----					
SnASAT1	c63608_g1_i1	TKISQILEKSLSKVLTSYYPFAGHVK-DN-SFVECNDMGADLSQVRID-CPMSSIFDHRP						
SnASAT2	c65670_g1_i1	NIISDHLEKSLSKVLTNYYPLAGKLN-GN-ISVDCNDNGVEFFVTKID-CPMSEILNDPY						
	c71009_g1_i2	IQISQILENSLSKVLSSYYPFAGQIK-DNNTYVDCNDTGAEYLNQIN-RSMSEILNHPY						
	c53868_g2_i2	TERCDKIQKSLALTLTKFYPLAGRFB-EDEFSIHNDGVEYIETKVLNADLAEFLRQGT						
	c60145_g1_i5	---CDKLQKSLAETLTKFYPLVGRFS-EDEFSIHNDQGVYVETKVLNADLAEFLHHLG						
SnAGAT1	c65306_g1_i1	----NRLKSSLAETLTLFYPLAGELSS	EDSTFVDCNDNGVLYVEACVKDVCMD	DEFLRN-I				
SnASAT1	c63608_g1_i1	TDIDDLVLP--IDPWFPTD-----SLVATQLCHFECGGIALGACLS	HKVSDGYSMGN					
SnASAT2	c65670_g1_i1	SDKDNLAYPKGVPSAYSIEG-----SLAVFQLSHFNCGGIAVSICLS	HKIGDGYTVGN					
	c71009_g1_i2	NDVVDVVPQDLPWSSTLNR-----SPLVVQLSHFDCGGIAMSACMS	HKIVDGYFSK					
	c53868_g2_i2	NNIELLNDLIPIG--MQSCP-----LLRIQVNI	FNCGGLVMGIQMSHILADGFTLAT					
	c60145_g1_i5	PKNEPLVDLLPKMDHQSSP-----LLGVKVN	VFNCGGVVIGIHMSHIVADALTLAT					
SnAGAT1	c65306_g1_i1	DDLSQLINFLPQNASMCSKMNYVQ	TLETLSVIQVTRFDCGGVAIGGVLF	HRLLDGDTMSK				
SnASAT1	c63608_g1_i1	FLKDWAMVARDTEAKP----SPLFN--GASIFKPSNSS-SFQVVND----PPLNKNVSKR						
SnASAT2	c65670_g1_i1	FMNHWATIARNPMSLETYPISP	KFD--GAYFFPPAQDD-DLSNVSNNNIVPEREECVARN					
	c71009_g1_i2	FINDWAATARNMDFKP----SPQFN--ASTLFLTHDPPSFTPINVP---PLSQLRVSRM						
	c53868_g2_i2	FVKEWTRISQGT--TKD-CLPSFGHLPS--LFP	SRVVSQPYIS-PPSHN-RDTKIVTRR					
	c60145_g1_i5	FVDEWAHICKIGTTITKDSYLP	SFGHLSS--LFPTRVLSGTQFP-LPSPNTTGPKVVTRR					
SnAGAT1	c65306_g1_i1	FFNTWAKIARGDNHNDGVEKITPLDLTSSSLLFPAQNF	PQEFMETKKRMFFYEKPLVRR					
SnASAT1	c63608_g1_i1	YHFSASKLKTLS-LINSADSGSQIRPTTVETITAF	LSKCVNTP-----TFTPPLL					
SnASAT2	c65670_g1_i1	FSFTSSNLSALKARVINQSEVQN---PTDTEIVSAFIYQRAMATKKVMTSSGYSIRPSIL						
	c71009_g1_i2	YNFSSSLARLKNI IATNSQVQN---PTRVEVATSLLHKCCTAASMEKSG---LFKPSLL						
	c53868_g2_i2	FIFDALAI AKLKERI INSSGMLITR-PTRVVVMSLLWKVLMGISTAKHG--HSRD	SHLI					
	c60145_g1_i5	FVFDPAIANLKNNTLLDS-TDSRR-PTRVVIVMSLIWKVLARISSAKHG--HSRDSSFL						
SnAGAT1	c65306_g1_i1	FVFNKAIEDIKMKVSEN---VPN-PSRVEAL	TTFIWKHMILATRATKG--ISSRPTML					
SnASAT1	c63608_g1_i1	VQAVNLRGTSNDALVPADLAGNSILPFVSAANKEEMNLQRLVGELRKGKEKIQDMLKYI						
SnASAT2	c65670_g1_i1	HHAVNLHPP-----LPKNTMGNNFFSLLTKEEKEIDLPQVVSKLRKAKEELRKKYKNA						
	c71009_g1_i2	SHVMNLRPP-----IPLNTIGNATCLYSSIAMTEDIKLSNYVAQLRKAKQKLRDELKDL						
	c53868_g2_i2	FPVNLRGKSILPYVEHALGNFCLAGTATLEAS-QSRKELTEYVNMVGS	TARDTSATIGKA					
	c60145_g1_i5	FSINLRGKSNIP---HALGNCIMFGIAADLEA-SPKQELNYFVNLRNTIRETNL	GIGKA					
SnAGAT1	c65306_g1_i1	AHVVNLRSRIDPKLPNAFGNLAFAKSEVIEMSDHELELPHLLGIVREMLENMSEN	LKA					
SnASAT1	c63608_g1_i1	ESEELLCSEIAREMNERTSNND-VSIYRFTSLRRFPVND-MNFGWGRPRKVD---VA						
SnASAT2	c65670_g1_i1	KIEELLP-ITLEQYREANEMFSNNSCFDLYRFSSIIKFPFYD-VDFGWGKPEKVT---LA						
	c71009_g1_i2	DINQIVP-YALGKMKEVIDMI-EKDIFDIYVCTSLCNVGLYDKTNFGWGRPIRVT---HT						
	c53868_g2_i2	S-VNDITSMFATYGTQVINKLGLGNEIDIYLITSWCRFPWYE-ADFGWGKPFWVS----P						
	c60145_g1_i5	ESVDDISSLVVNNHIKAIKDFVQGDKMDVYPSTSWCGFTWYE-ADFGWGKPFWVS----S						
SnAGAT1	c65306_g1_i1	LEGDDALDSIVYMAIQLGILFP---QLDYNFSSWCNFGLYD-VDFGWGKPIFIAPFIEP						
SnASAT1	c63608_g1_i1	T-YPIN-MFVLMNDQNADGLEVI	VNLEEGEMSSFERNDELQFASPCSGL----					
SnASAT2	c65670_g1_i1	SDSPIKNMFI	LMDNKNRNGVEAISGMKEQDMLALERDEEFLQFASPSS-----					
	c71009_g1_i2	G-YRMNKSIVFFDDRSGDGIDLLITLPKDEMTIFQNNKOLLEFASPIVQSIK--						
	c53868_g2_i2	ISYYAIEGTL	LMETKDGDIAMVCLNENDMAEFERDPNLSSTSKV-----					
	c60145_g1_i5	VSIGDFESIVLIDTKNGDGI	EARIGLKENDMTEFERDPHILSSTSKVAFDSQI-					
SnAGAT1	c65306_g1_i1	ISSLNMHQIILVENGRNDGIEAWILRSNEEMIELEKDEEFLAYASPNPSVQYDP						

Supplemental Figure 7

Fig. S7. Multiple sequence alignment of candidate sequences. BAHD characteristic HXXXD motifs (yellow) and DFGWG-like motif (green) are highlighted. Hyphens indicate gaps in protein sequence.



Supplemental Figure 8

Fig. S8. SnASAT1 activities with different acyl chain donors and acyl group acceptors.

Acylsugars were analyzed using LC-MS in ESI- mode. Combined extracted ion chromatogram showing internal standard telmisartan (m/z 513.23) and expected products as formate adducts: top two rows from left to right, S1:10 (m/z 541.25), G1:10 (m/z 379.21) and I1:10 (m/z 379.21); third row from left to right, S1:8 (m/z 513.22), G1:8 (m/z 351.17) and I1:8 (m/z 351.17); fourth and fifth rows from left to right, S1:5 (m/z 471.18), G1:5 (m/z 309.12) and I1:5 (m/z 309.12); bottom from left to right, S1:4 (m/z 457.16), G1:4 (m/z 295.11) and I1:4 (m/z 295.11).

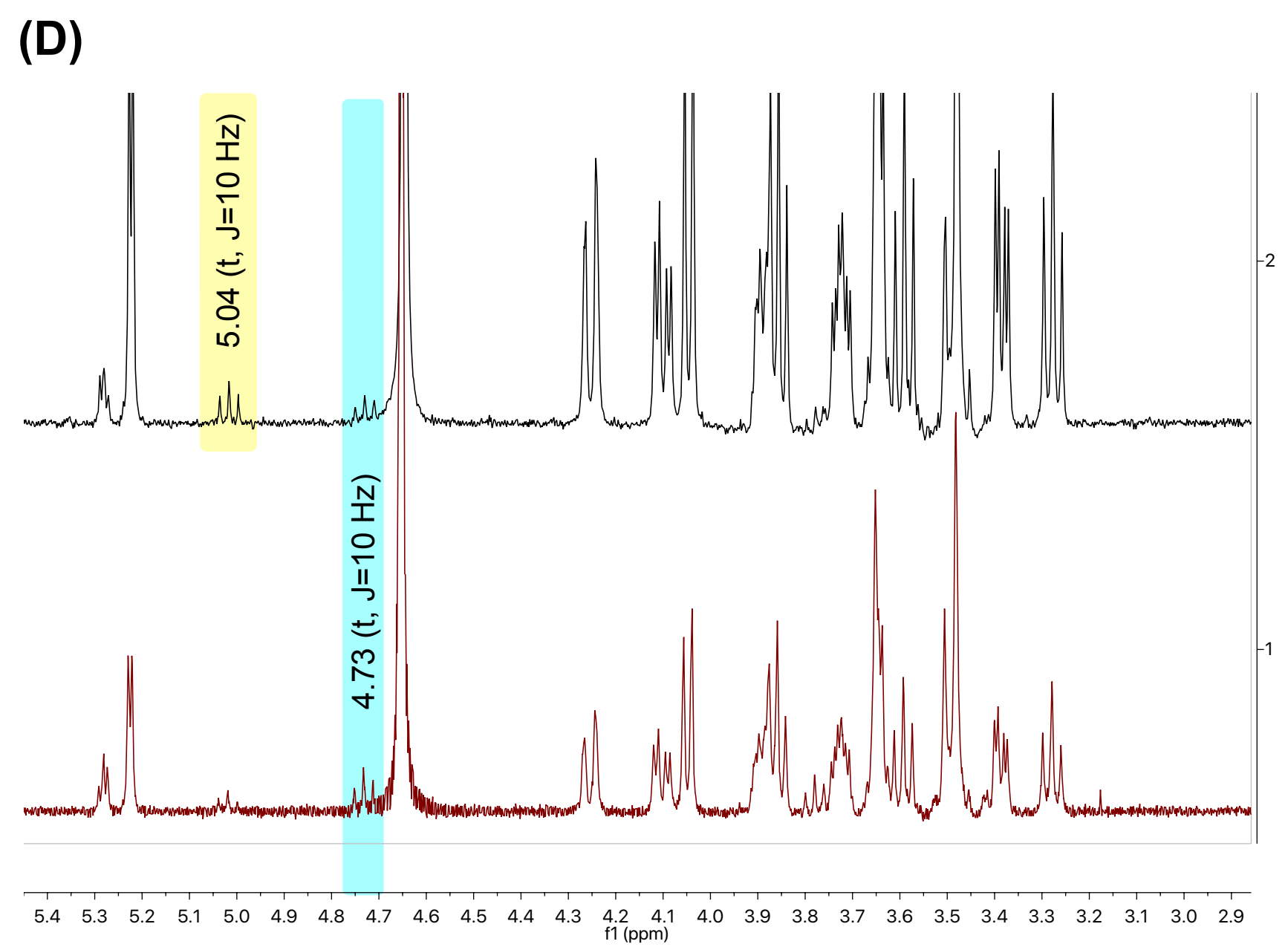
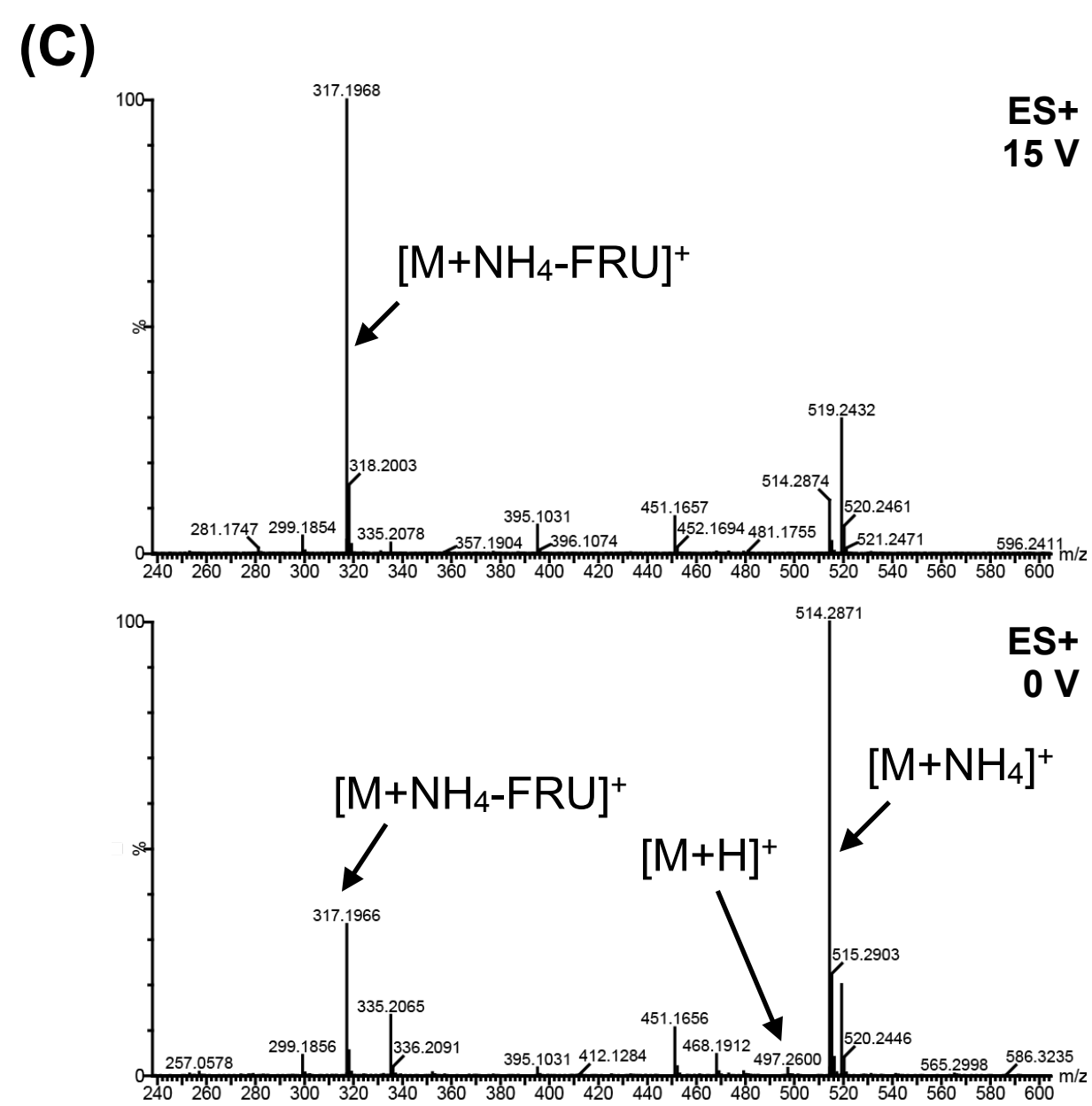
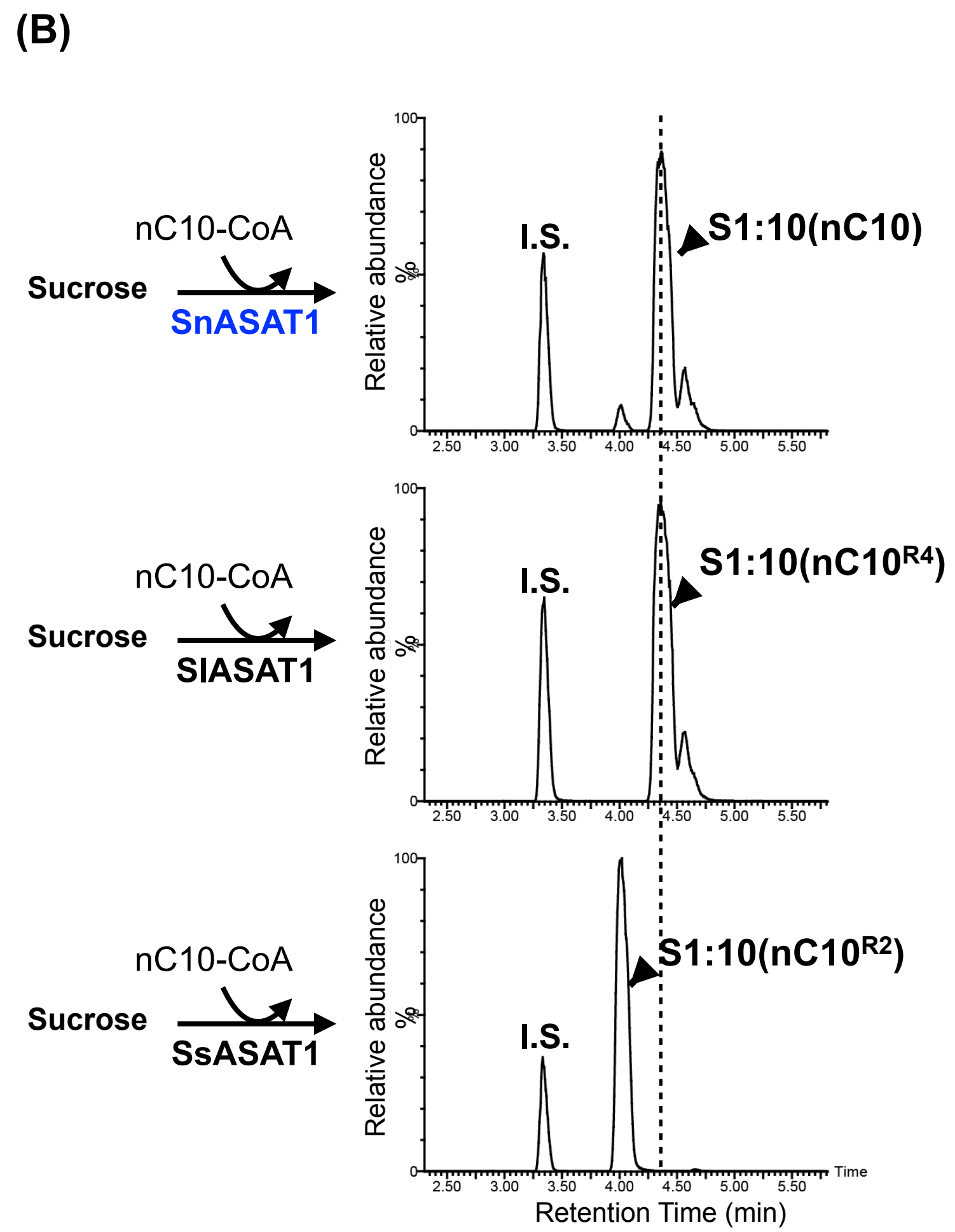
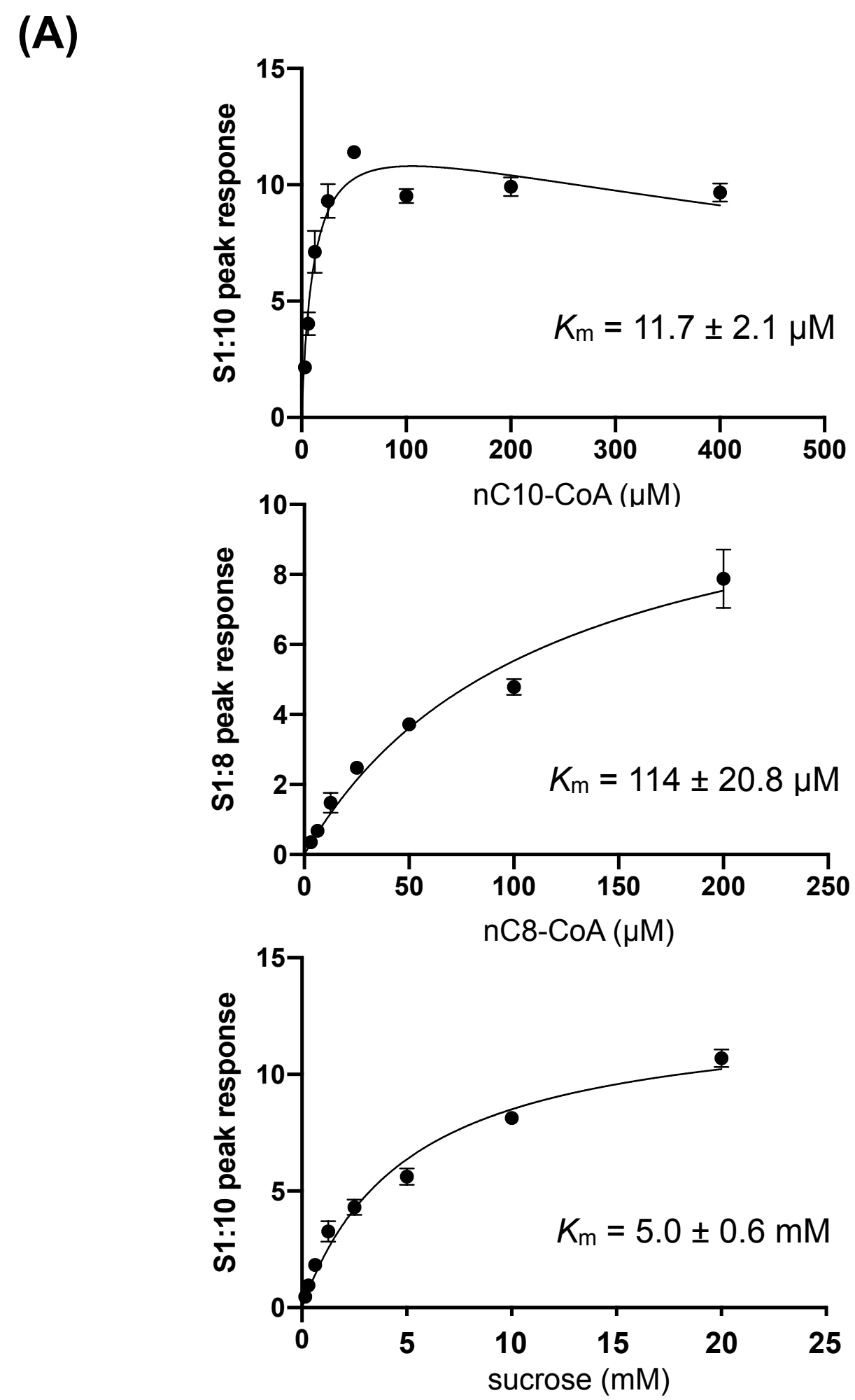
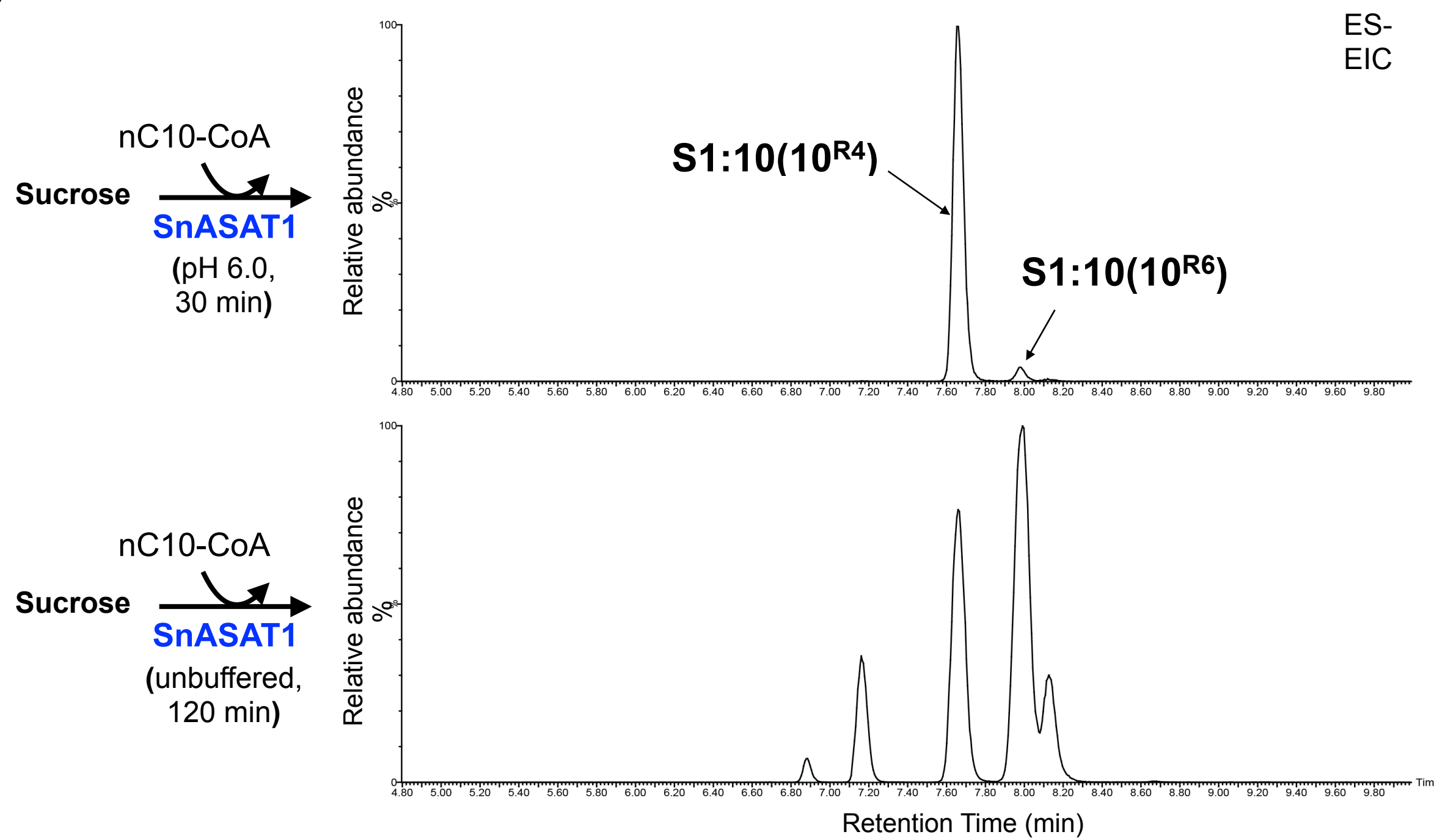
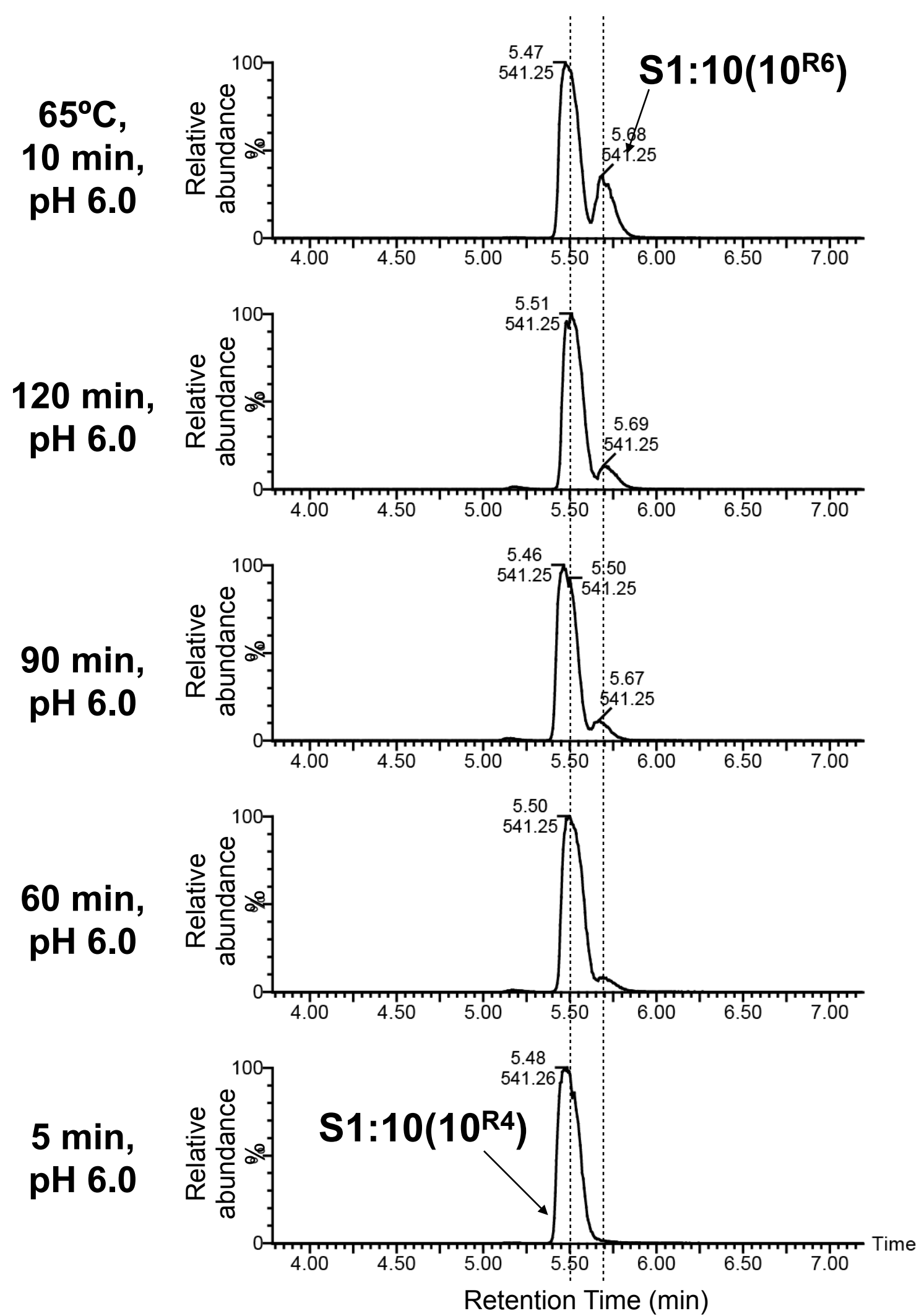


Fig. S9. SnASAT1 generates monoacylsucroses with a medium-length acyl chain on the R₄ position. (A) Kinetics analysis of different substrates for SnASAT1. Apparent K_m were calculated using the non-regression model in the GraphPad Prism 5 software. The enzyme assay products' peak area was normalized to the internal standard telmisartan peak area and plotted for each concentration of the varying substrate. Error bars indicate standard error with $n = 3$. (B) Co-chromatography of S1:10(10) produced by SnASAT1 (top), SIASAT1 (middle) and SsASAT1 (bottom). Acylation positions of SIASAT1- and SsASAT1-produced S1:10(10) were verified by NMR^{1,2}. (C) ESI+ spectra under CID potential 15V of product ions generated from $[M+NH_4]^+$ of SnASAT1-produced S1:10 results in neutral loss of unacylated fructose moiety. (D) Proton NMR determined the major S1:10(10) isomer is acylated at the R₄ position [4(CH) = 4.73 (t, J=10Hz); top and bottom spectrum]. A second, chromatographically separable S1:10 minor isomer with acylation at R₆ position [6(CH) = 5.04 (t, J=10Hz); top spectrum] accumulates after extended enzyme assay incubation time.

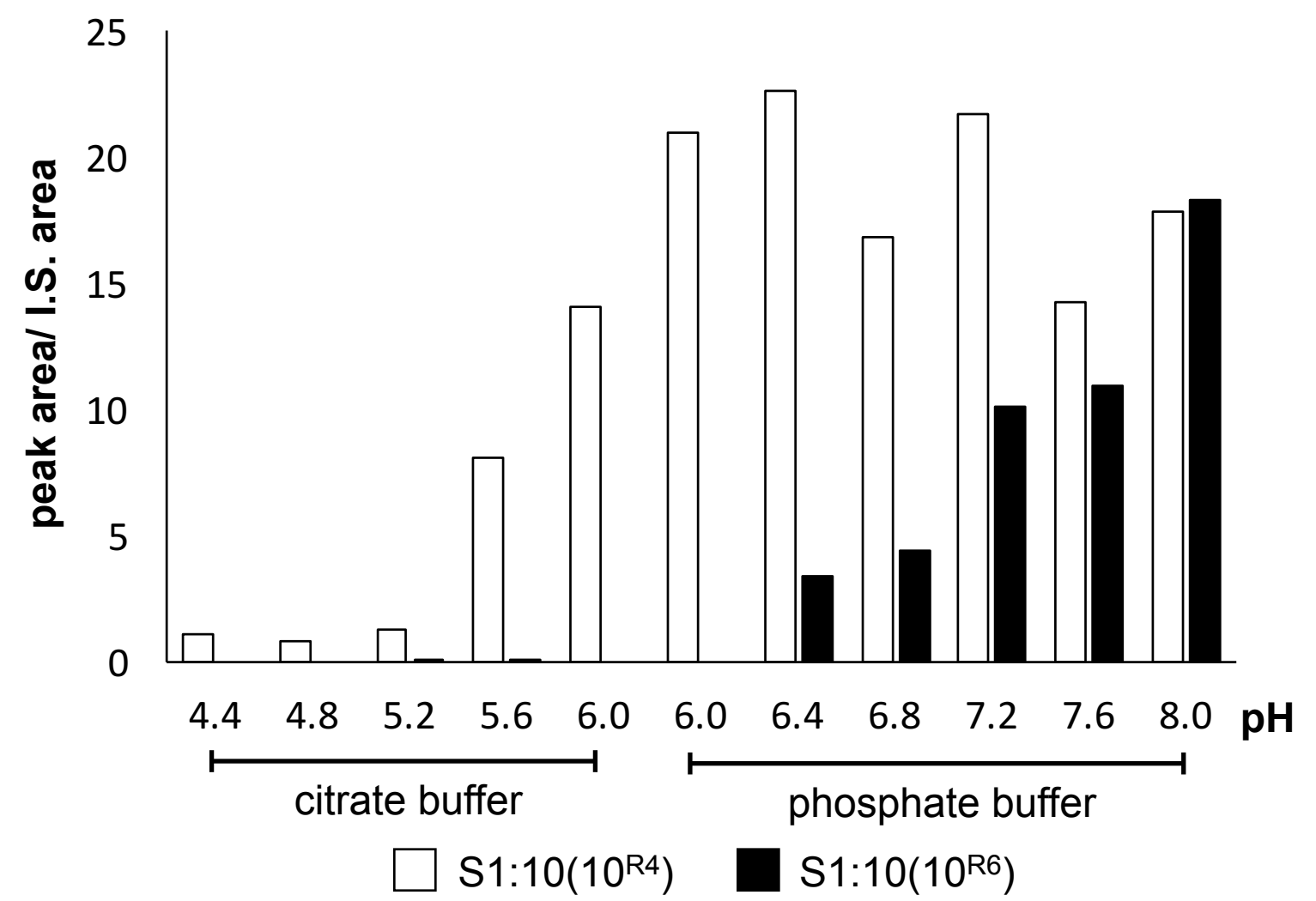
(A)



(B)



(C)



(D)

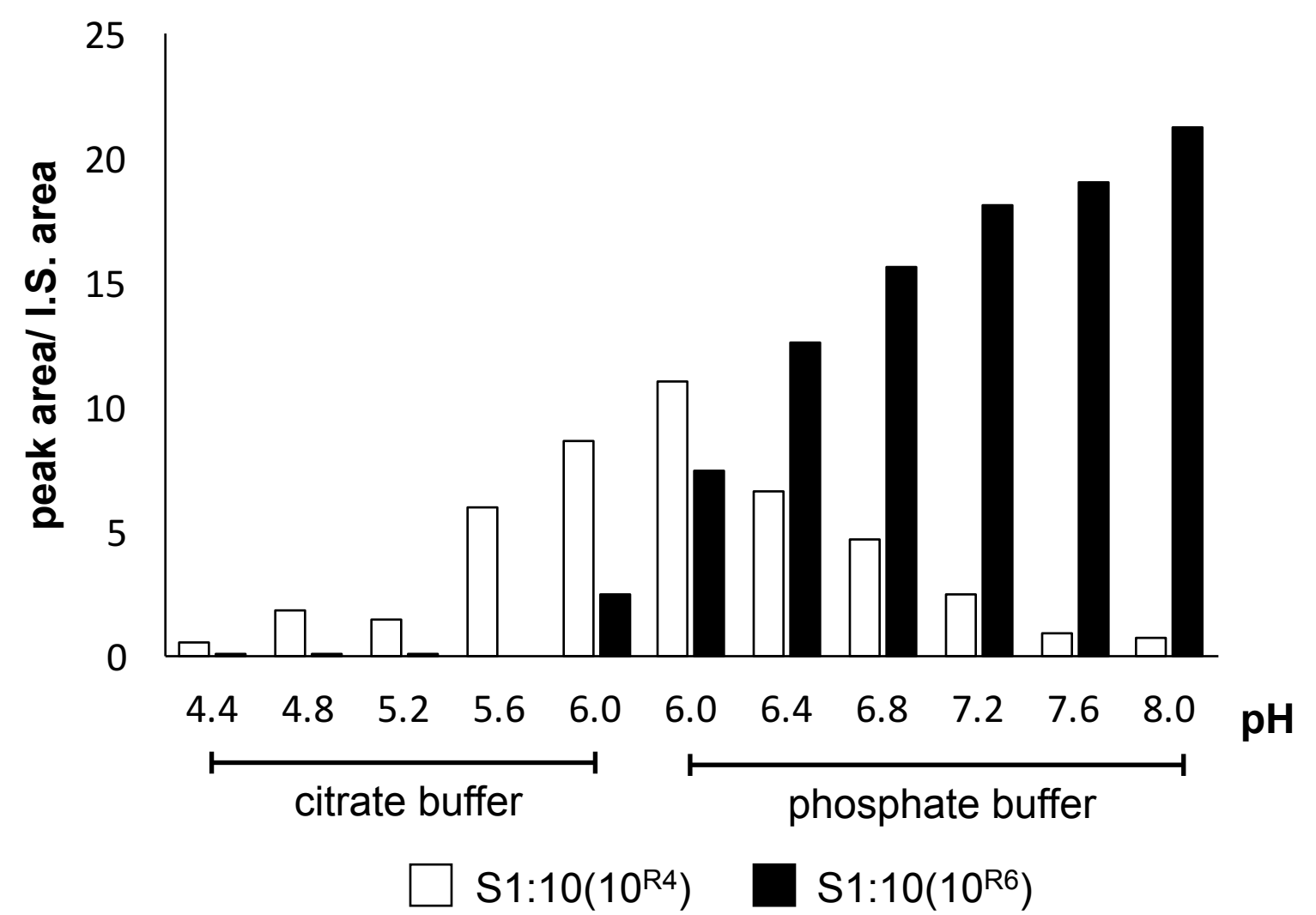
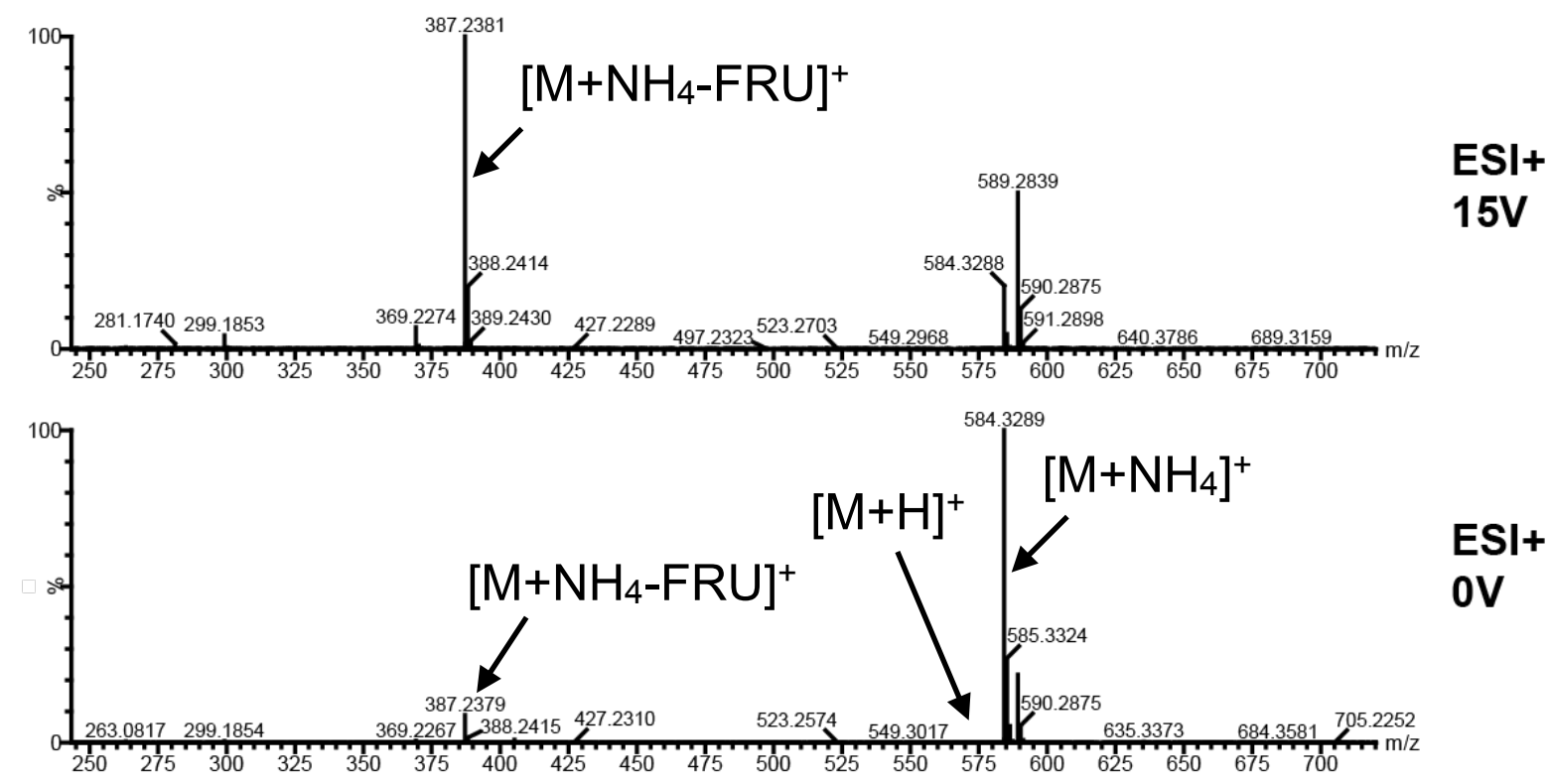
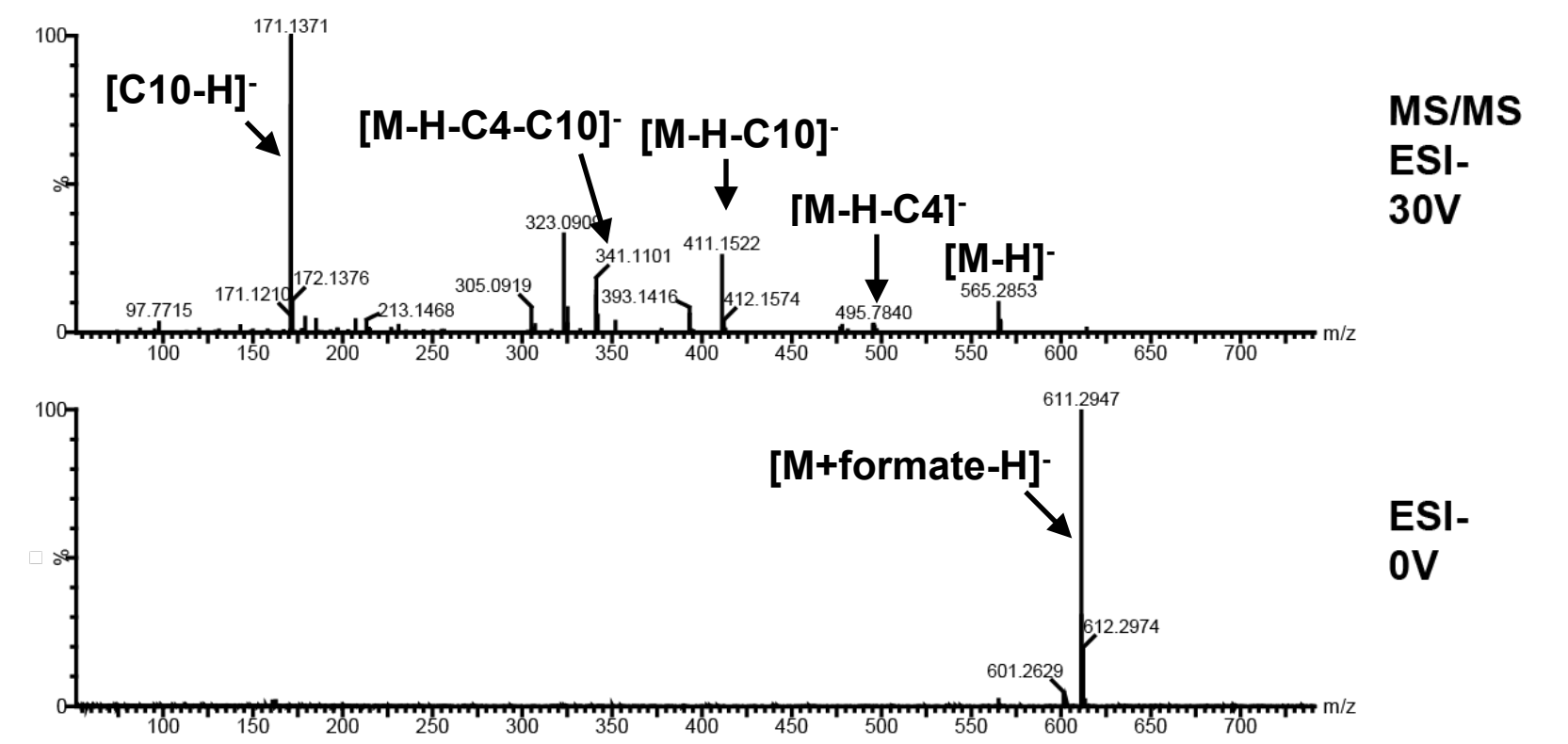


Fig. S10. R₆-acylated S1:10 isomer accumulates in SnASAT1 enzyme assays. (A) The co-chromatography of S1:10(10) isomers produced by SnASAT1 under pH 6.0 for 30 min (top) and unbuffered aqueous solution for 120 min (bottom). (B) LC-MS profile of enzyme assays carried out with different incubation conditions R₆-acylated S1:10 isomer accumulation increases after extended enzyme assay incubation time (bottom four) and brief exposure to elevated temperature (65°C) (top). (C) The concentration of the R₆-acylated isomer increases when enzyme assays were carried out in neutral-to-alkaline pH conditions, or (D) subjected to brief exposure to elevated temperature (65°C). Acylsugars were analyzed using LC-MS in ESI- mode. Extracted ion chromatograms are showing formate adducts of S1:10(10) (*m/z* 541.25). Peak areas of S1:10(10) were integrated under negative mode and normalized to the internal standard telmisartan.

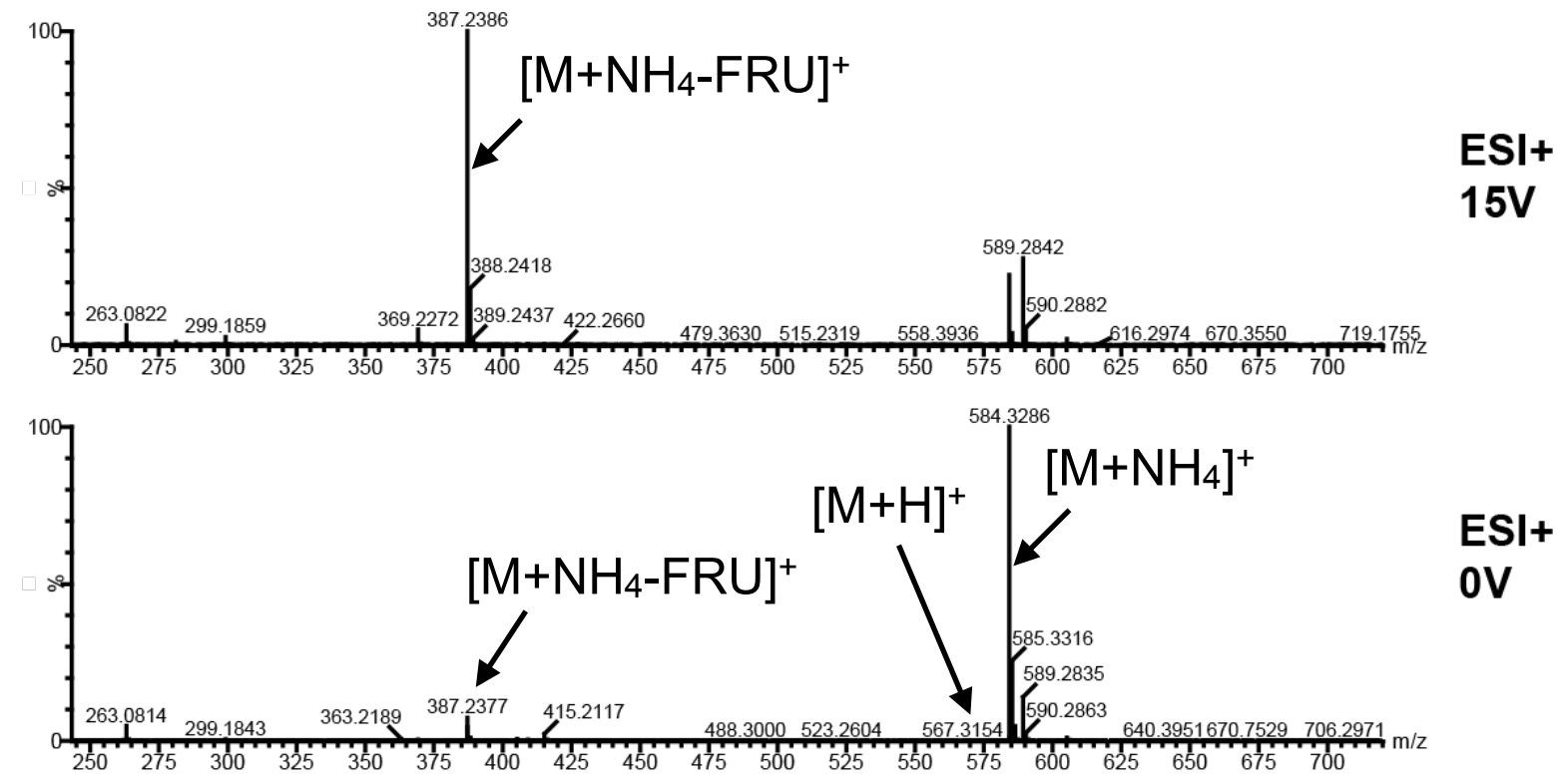
(A) *in vitro* S2:14(iC4, nC10)
(*m/z* 584.33 in ESI+) from **SnASAT1+SnASAT2** *in vitro* assay



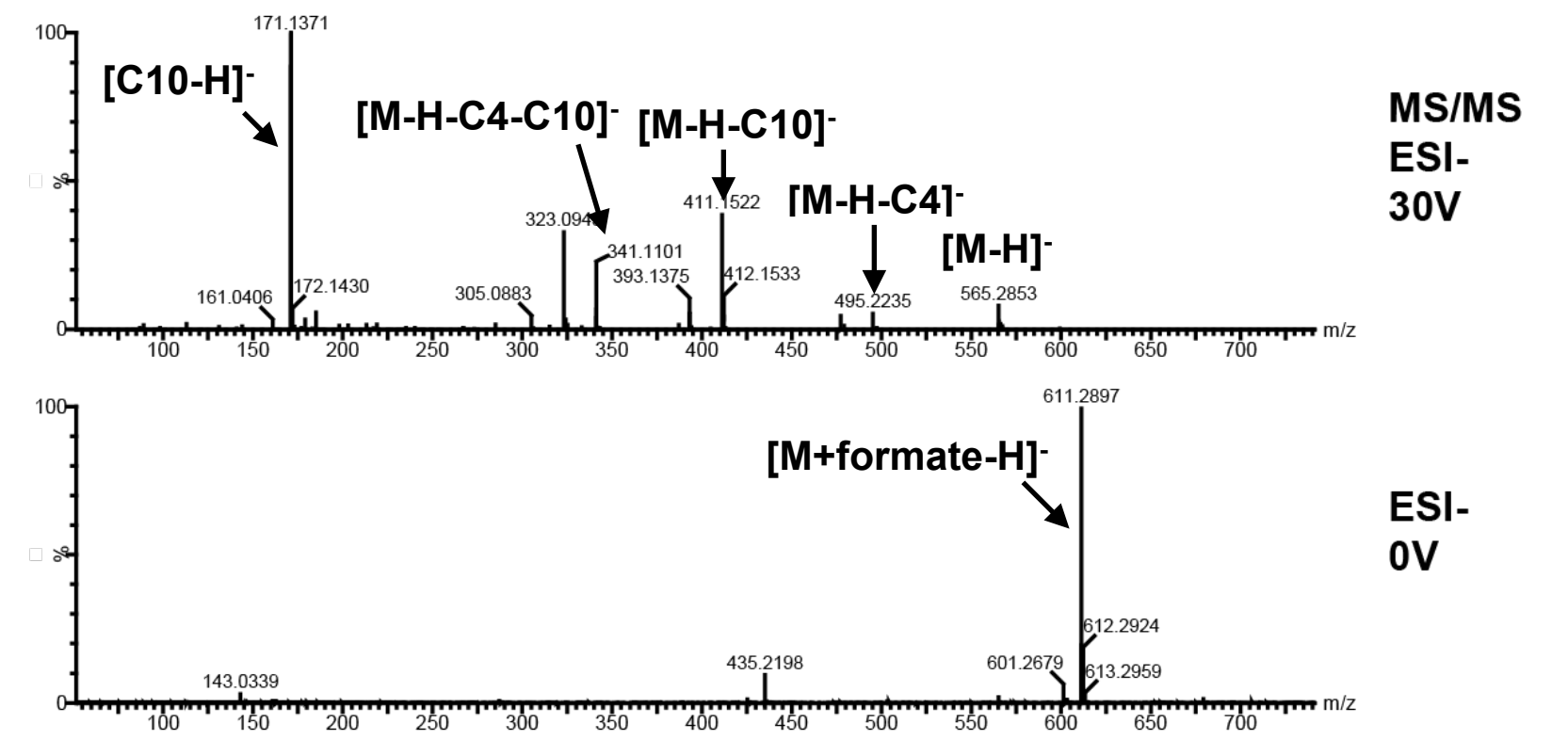
(B) *in vitro* S2:14(iC4, nC10)
(*m/z* 611.29 in ESI-) from **SnASAT1+SnASAT2** *in vitro* assay



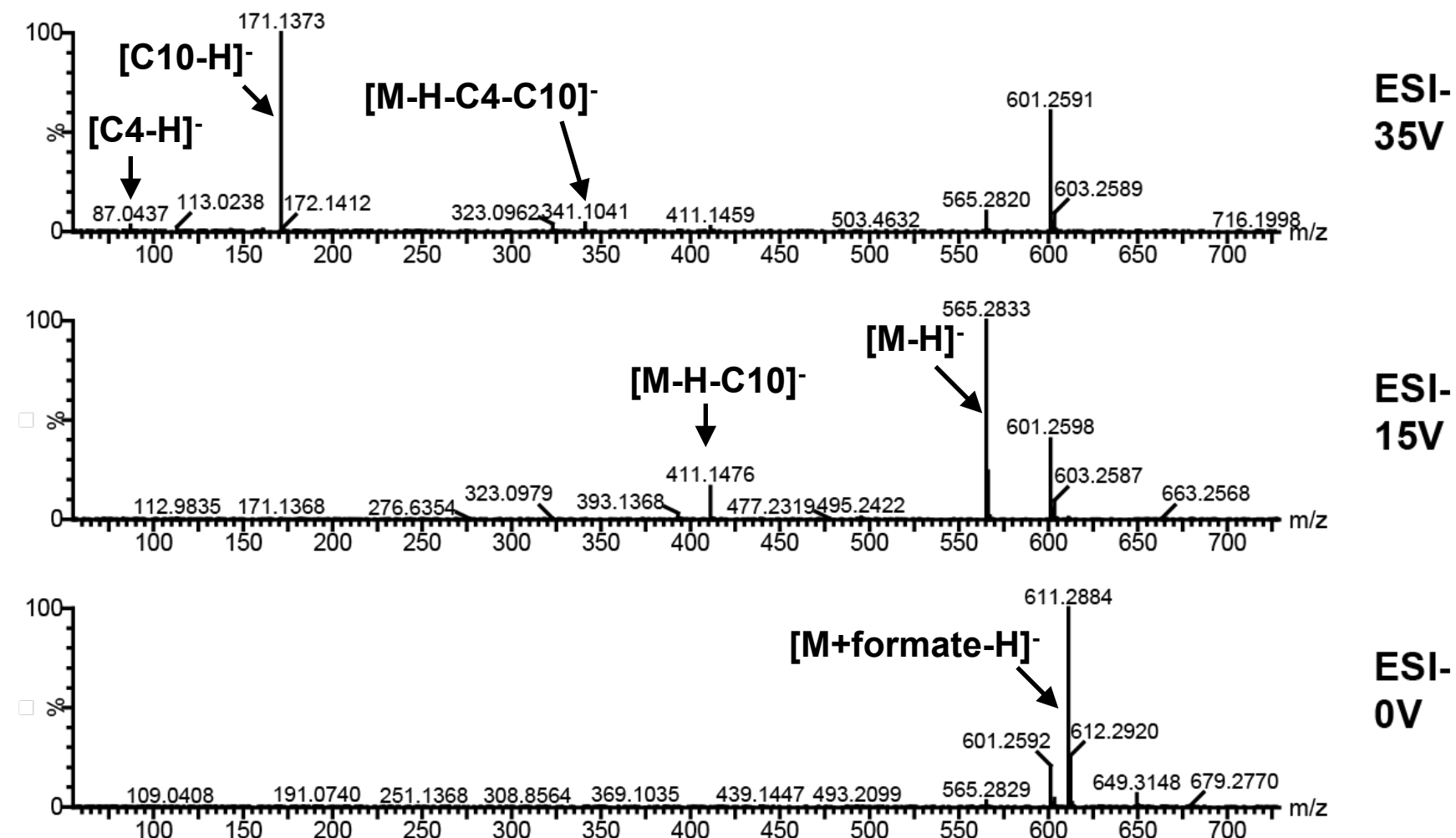
***in vivo* S2:14(4,10), late-eluting isomer**
(*m/z* 584.33 in ESI+) purified from **VIGS-SnASFF1** lines



***in vivo* S2:14(4,10), late-eluting isomer**
(*m/z* 611.29 in ESI-) purified from **VIGS-SnASFF1** lines



(C) *in vitro* S2:14(iC4, iC10)
(*m/z* 611.29 in ESI-) from **SnASAT1+SnASAT2** *in vitro* assay



***in vivo* S2:14(4,10), early-eluting isomer**
(*m/z* 611.29 in ESI-) purified from **VIGS-SnASFF1** lines

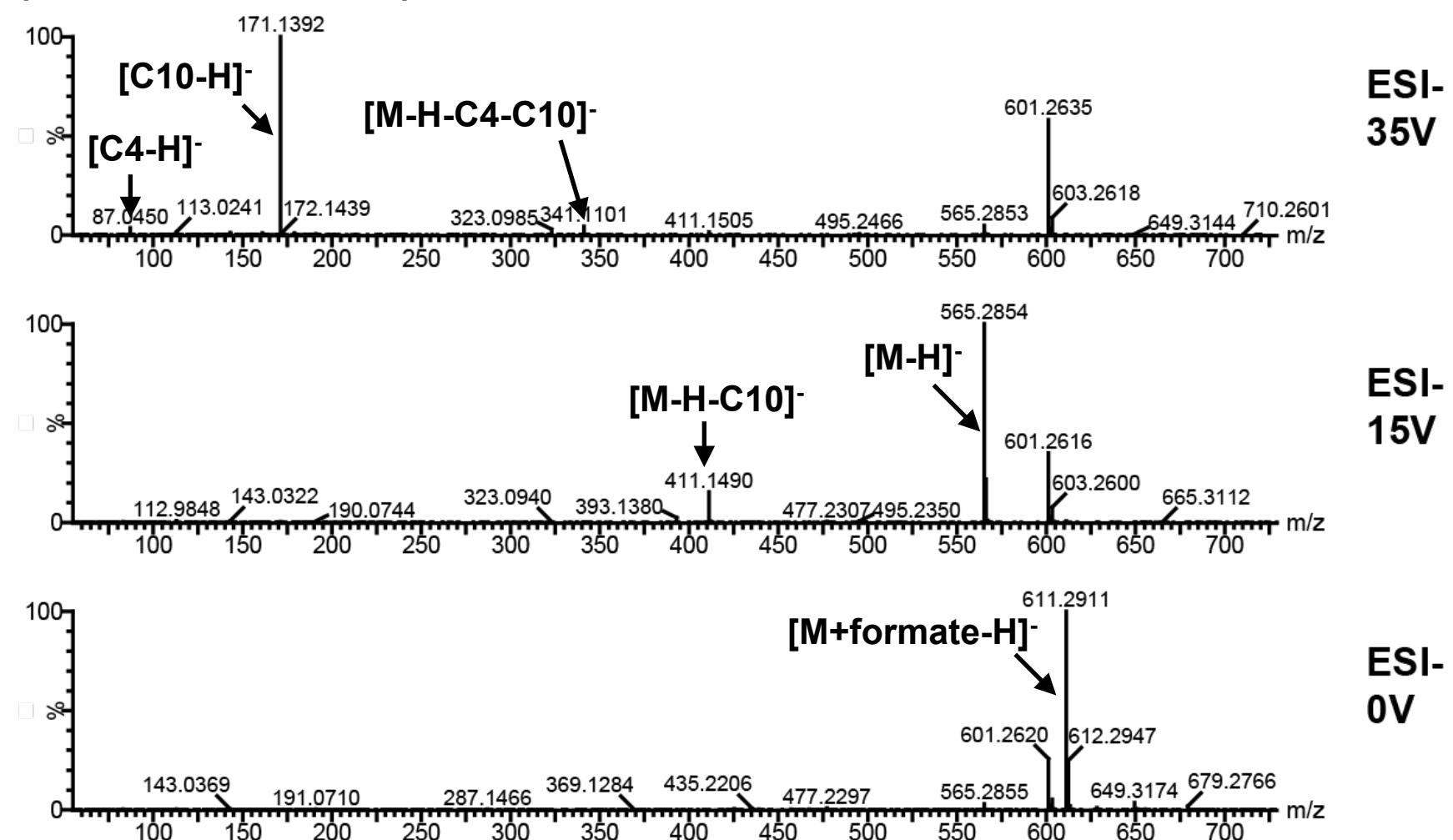


Fig. S11. Mass spectra of the co-eluting *in vivo* and *in vitro* S2:14 isomers. (A) ESI+ mode fragmentation of the co-eluting *in vitro* S2:14(iC4, nC10) and *in vivo* late-eluting S2:14(4,10) results in the cleavage of the glycosidic linkage. Fragment analysis of the acylsucroses using collision-induced dissociation reveals a fragment ion (m/z 387.24) consistent with the furanose ring of sucrose conjugated to a C10 acyl chain, indicating that both acylsugars possess a C10 acyl chain on the furanose ring. (B) ESI- MS/MS spectra (30V) of product ions generated from [M+formate]⁻ of the co-eluting *in vitro* S2:14(iC4, nC10) and *in vivo* late-eluting S2:14(4,10) isomer are both characterized by the loss of C4 and C10 ketenes and the presence of C10 fatty acid anions (m/z 171.14). (C) Comparable fragmentation of the co-eluting *in vitro* S2:14(iC4, iC10) and *in vivo* early-eluting S2:14(4,10) isomer in ESI- mode reveals the loss of C4 and C10 ketenes under high collision energy (35V). The acyl chain combination is validated by the presence of C4 (m/z 87.04) and C10 fatty acid anions (m/z 171.14).

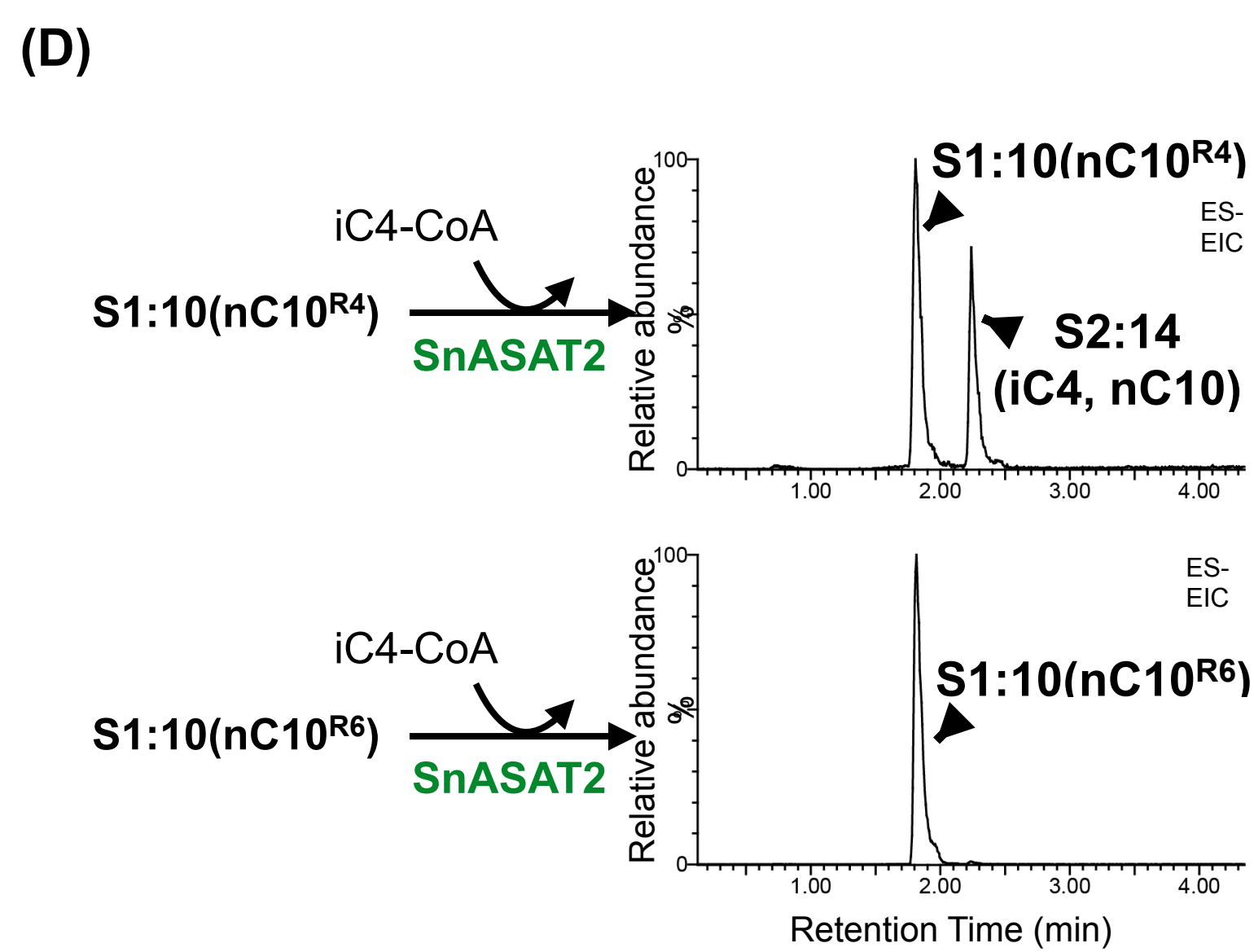
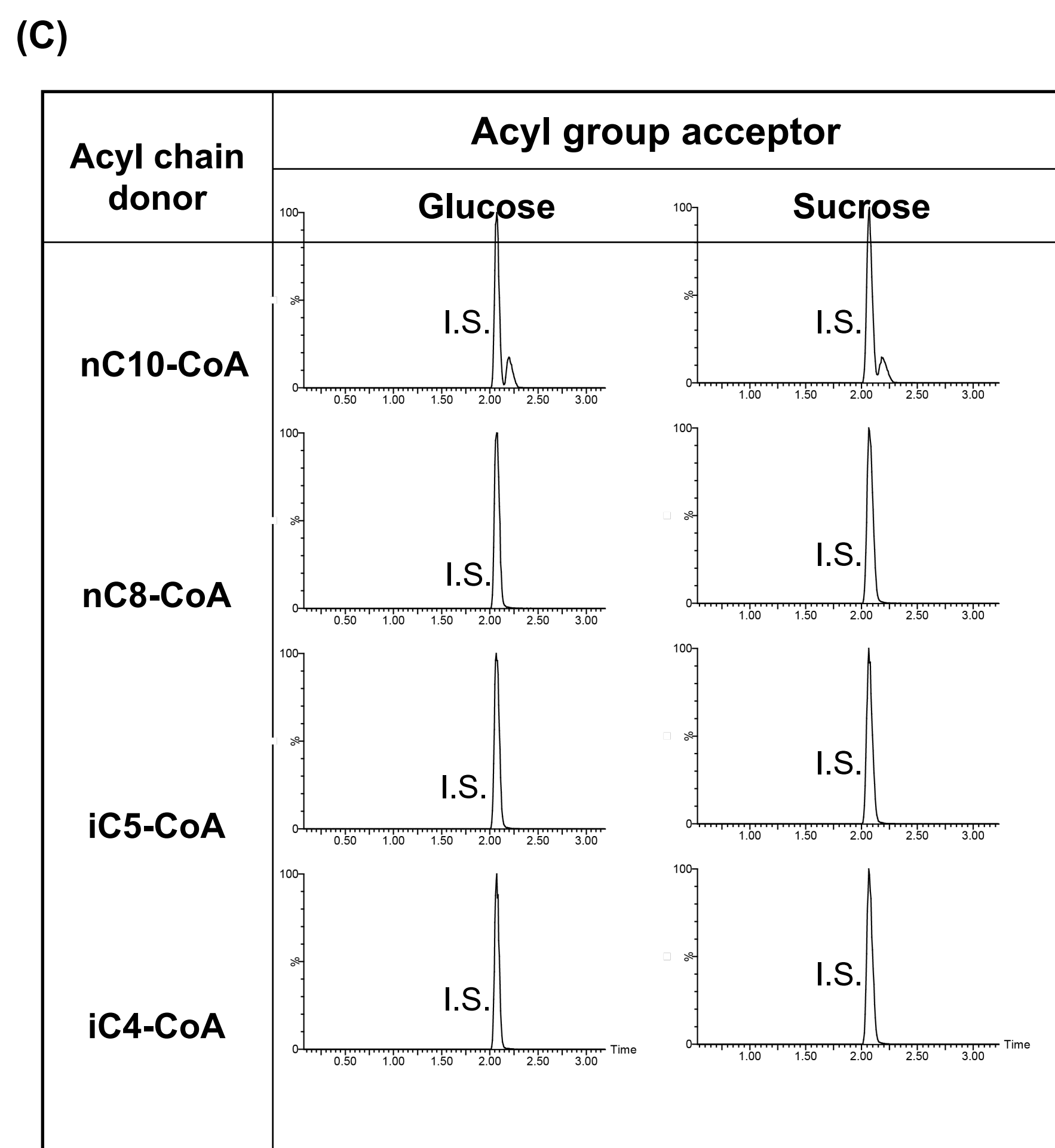
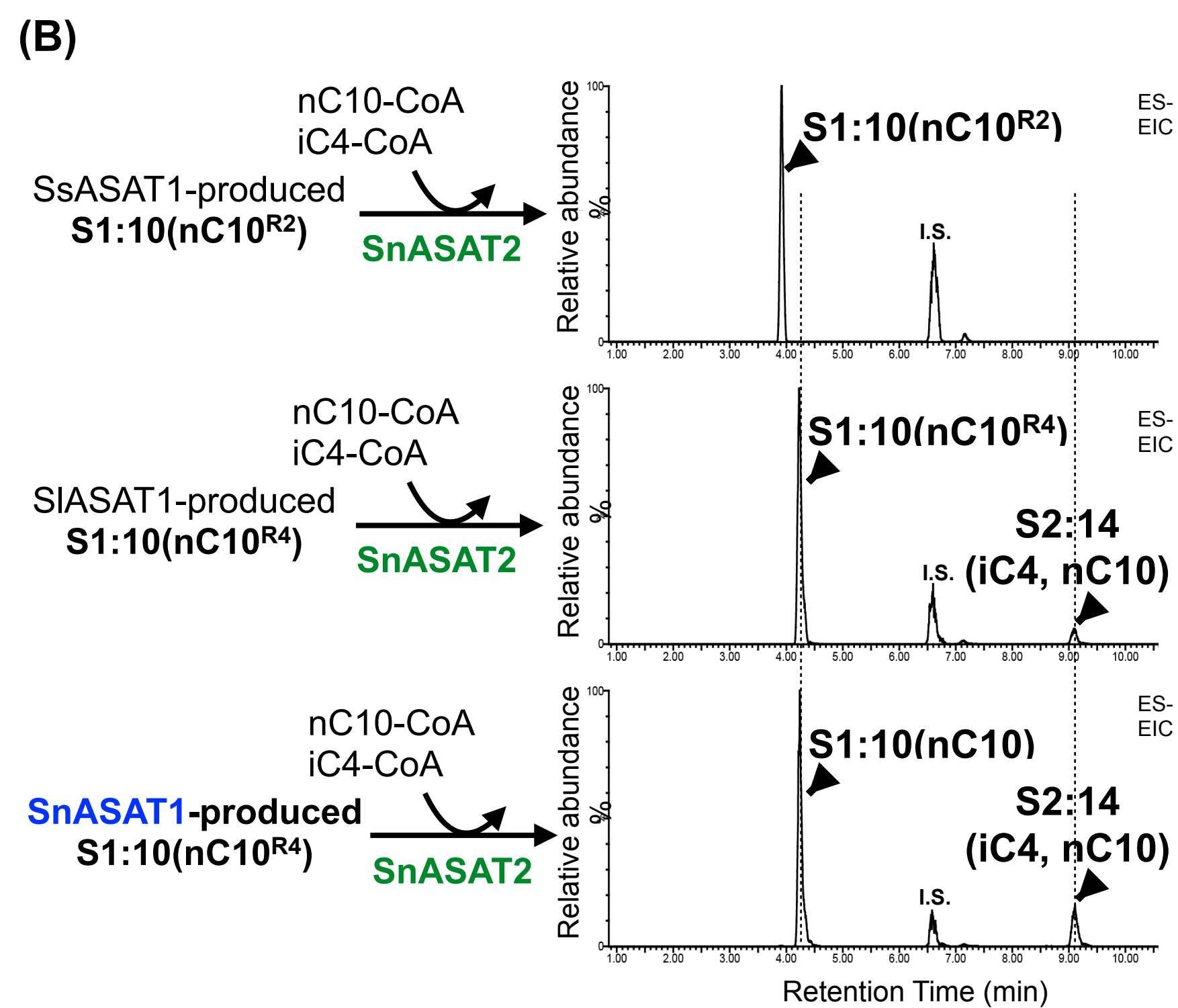
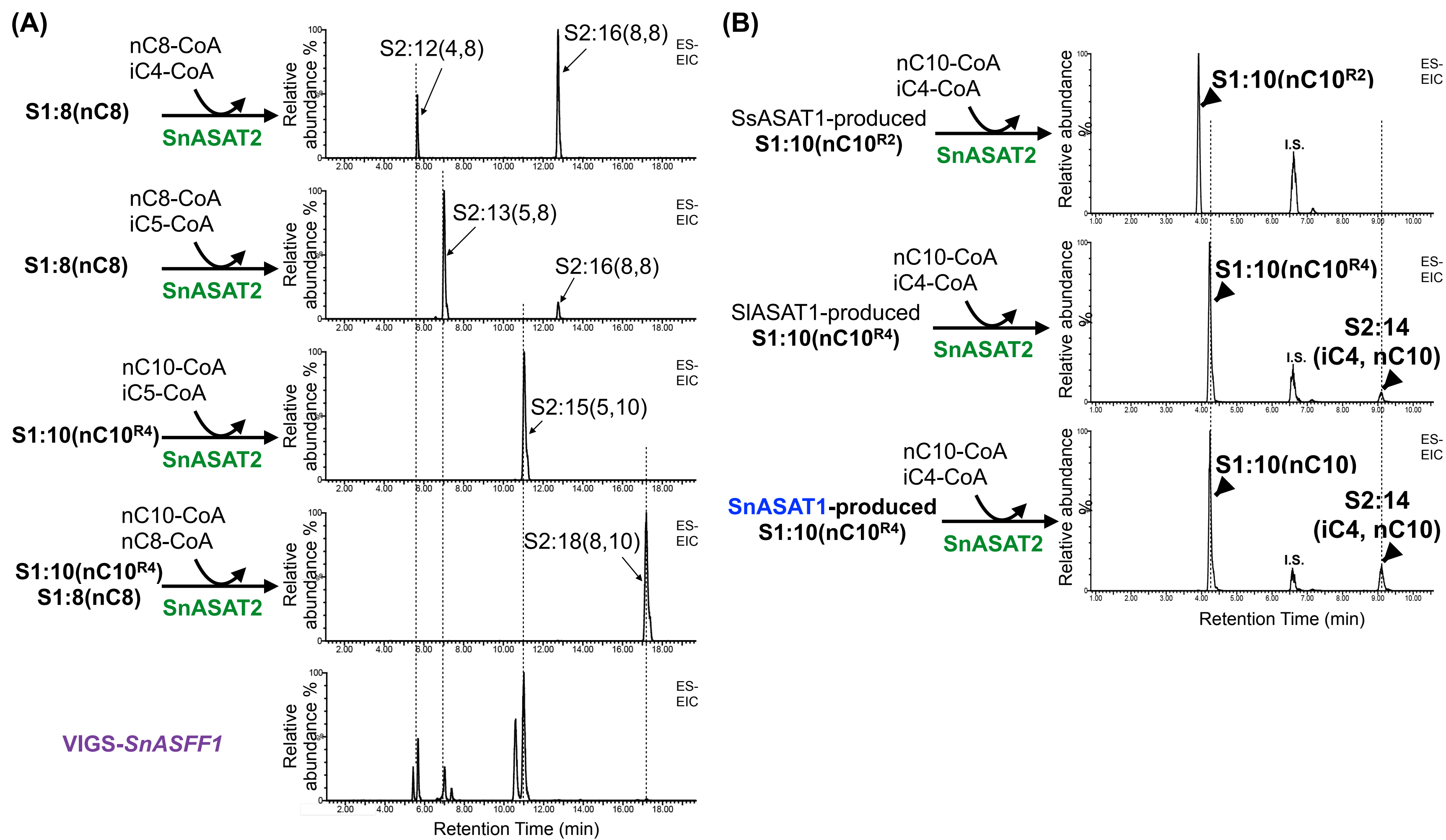


Fig. S12. SnASAT2 activity with different acyl chain donors and acyl group acceptors. (A) LC-MS analysis of *in vitro* enzyme assay products indicates SnASAT2 takes SnASAT1-produced monoacylsucroses and iC4-, iC5 and nC8-CoAs as substrates to produce *in vitro* diacylsucroses (top four) that co-elute with *in vivo* diacylsucroses accumulating in VIGS-*SnASFF1* plants (bottom). All acylsugars were analyzed using LC-MS in ESI- mode. Combined extracted ion chromatogram showing enzyme assay products as formate adducts: S2:12 (*m/z* 583.26), S2:13 (*m/z* 597.28), S2:14 (*m/z* 611.29), S2:15 (*m/z* 625.31), S2:16 (*m/z* 639.33), S2:18 (*m/z* 667.36). (B, C, D) LC-MS analysis of *in vitro* enzyme assay products indicates SnASAT2 takes R₄-acylated S1:10(10) as substrates, but not R₂- or R₆-acylated S1:10(10), unmodified glucose or unmodified sucrose. The S1:10 isomers used in panel D were purified from SnASAT1 enzyme assays and verified by NMR (Fig. S7D). All acylsugars were analyzed using LC-MS in ESI- mode. Combined extracted ion chromatogram showing internal standard telmisartan (*m/z* 513.23) and formate adducts of S1:10 (*m/z* 541.25) and S2:14 (*m/z* 611.29) in panel B and D, and S1:10 (*m/z* 541.25), G1:10 (*m/z* 379.21), S1:8 (*m/z* 513.22), G1:8 (*m/z* 351.17), S1:5 (*m/z* 471.18), G1:5 (*m/z* 309.12), S1:4 (*m/z* 457.16) and G1:4 (*m/z* 295.11) in panel C.

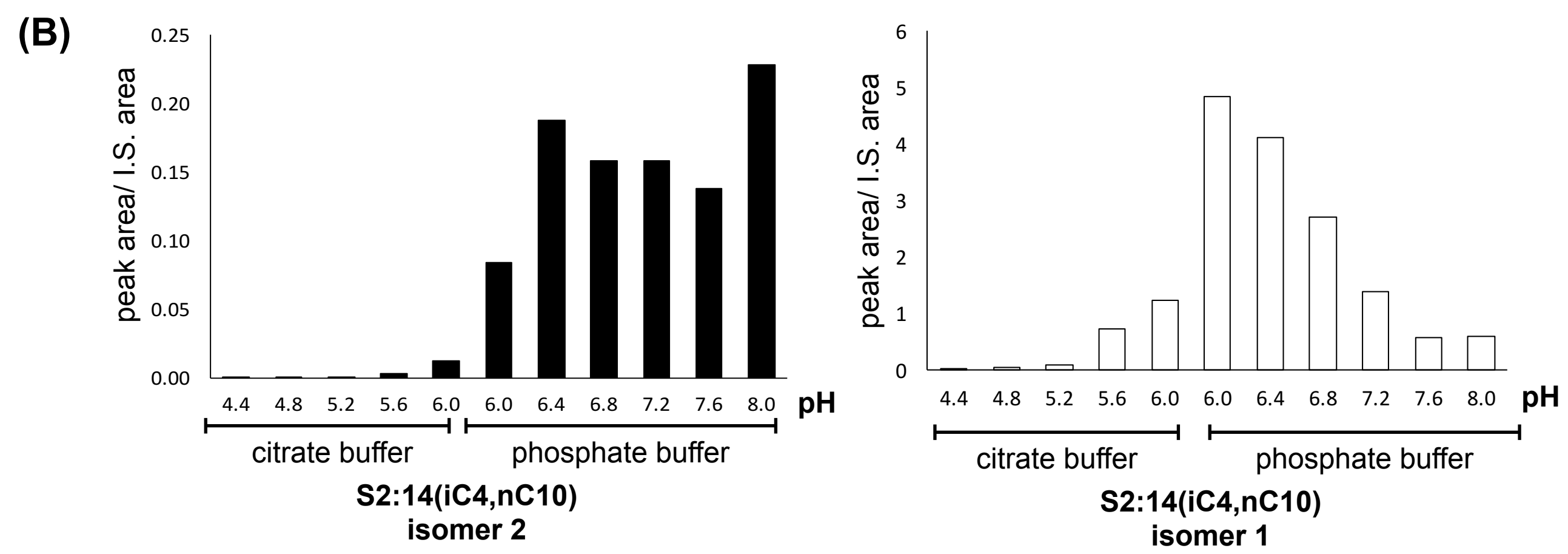
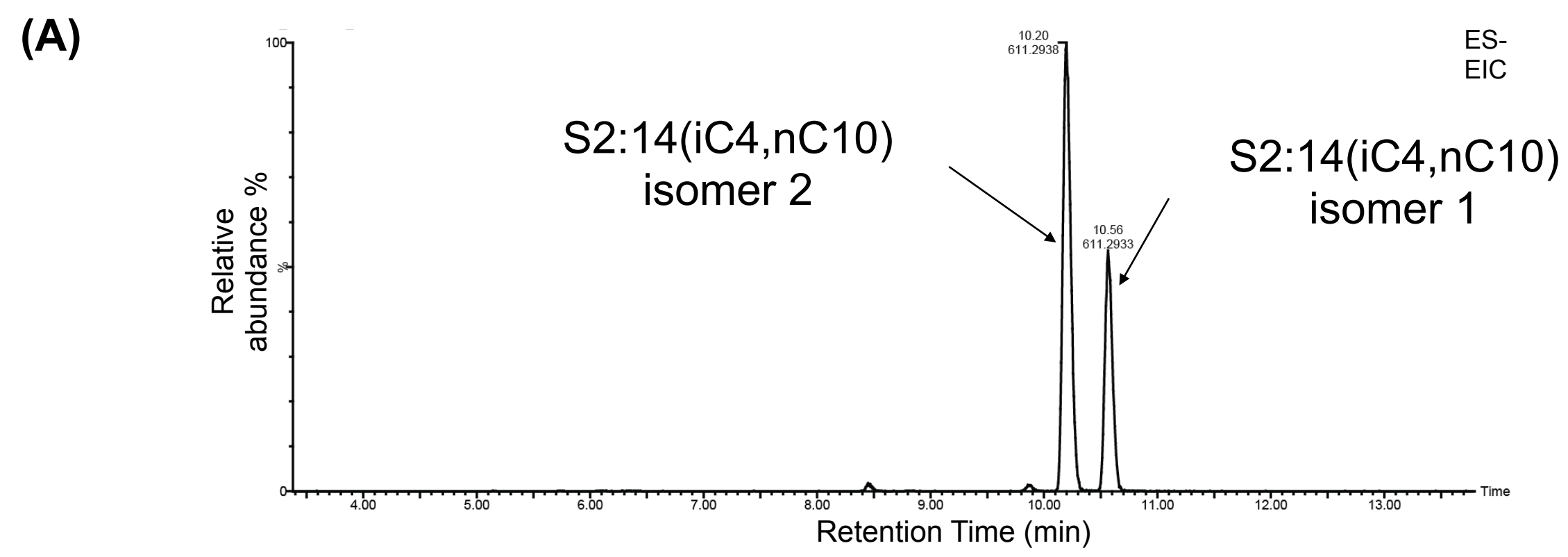
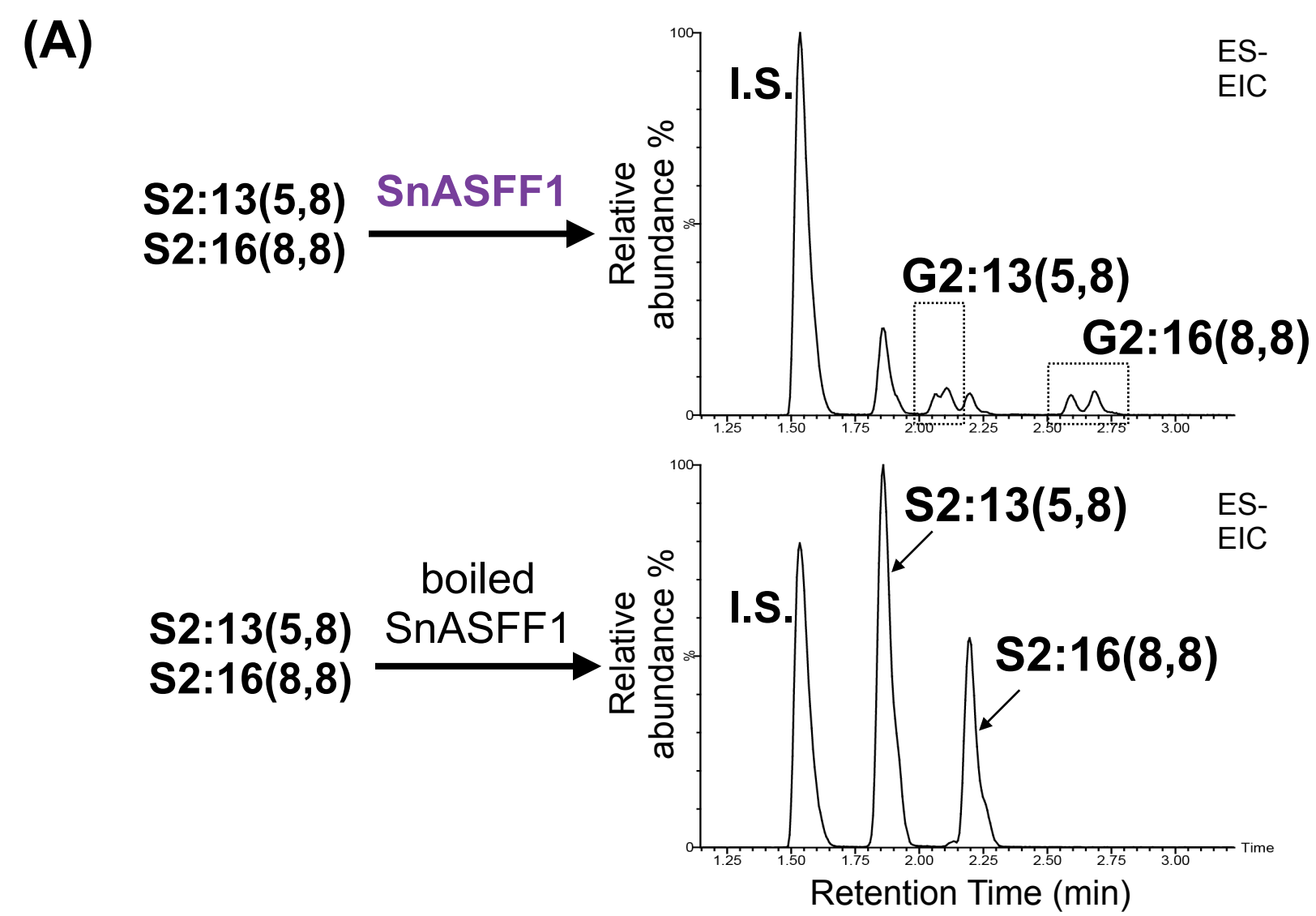
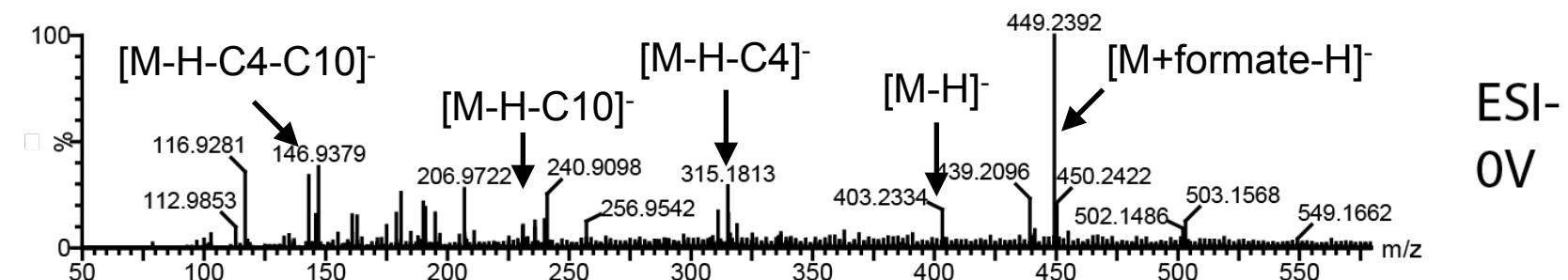
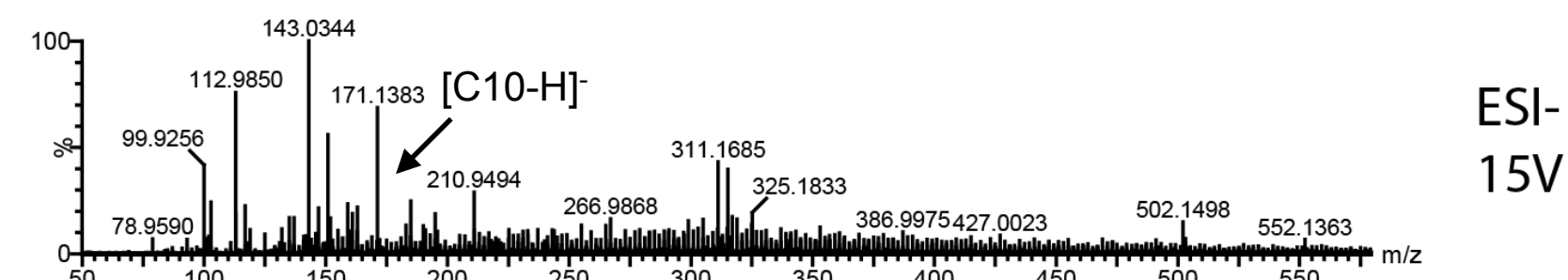


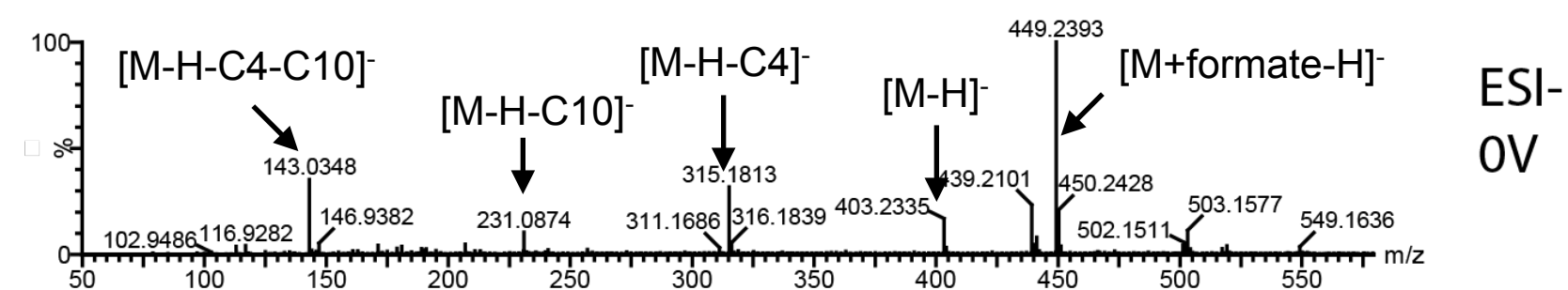
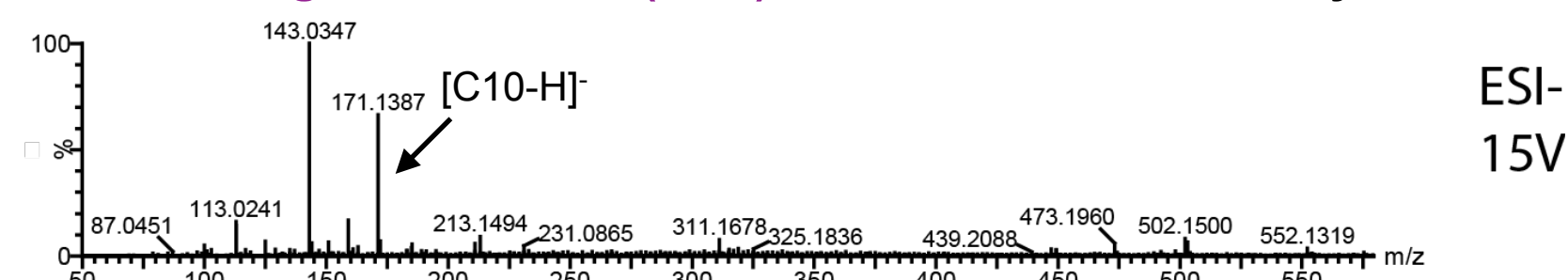
Fig. S13. A second S2:14 isomer accumulates in SnASAT2 enzyme assays. (A) LC-MS analysis of SnASAT2 *in vitro* assay products showed a second isomer (isomer 2) accumulating in unbuffered water based solution for 120 min. **(B)** The concentration of the second isomer increases when enzyme assays were carried out in neutral-to-alkaline pH conditions. Acylsugars were analyzed using LC-MS in ESI- mode. Extracted ion chromatogram showing formate adducts of S2:14 (*m/z* 611.29). Peak areas of S2:14 were integrated under negative mode and normalized to the internal standard telmisartan.



**(B) G2:14(iC4, nC10) (m/z 449.24 in ESI-) from
SnASAT1+SnASAT2+SnASFF1 *in vitro* assay**



**G2:14(4,10) (m/z 449.24 in ESI-) from
late-eluting *in vivo* S2:14(4,10)+SnASFF1 *in vitro* assay**



**G2:14(4,10) (m/z 449.24 in ESI-) from
early-eluting *in vivo* S2:14(4,10)+SnASFF1 *in vitro* assay**

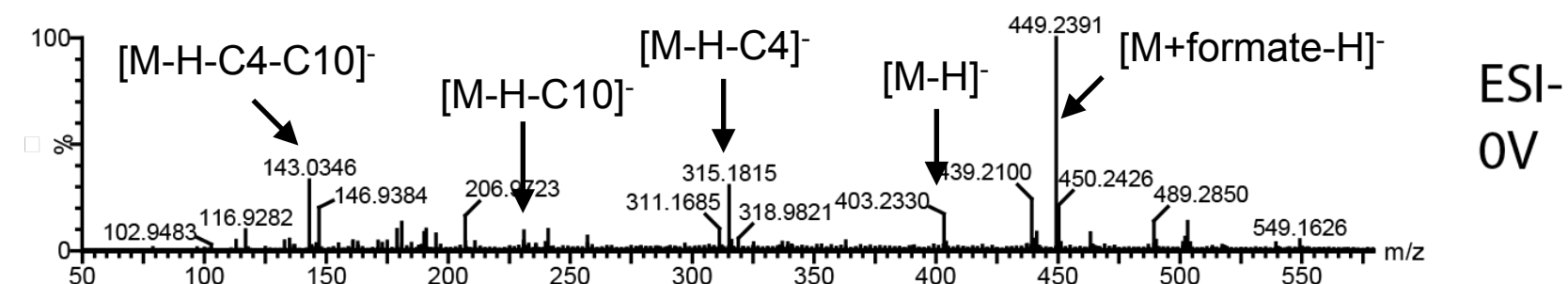
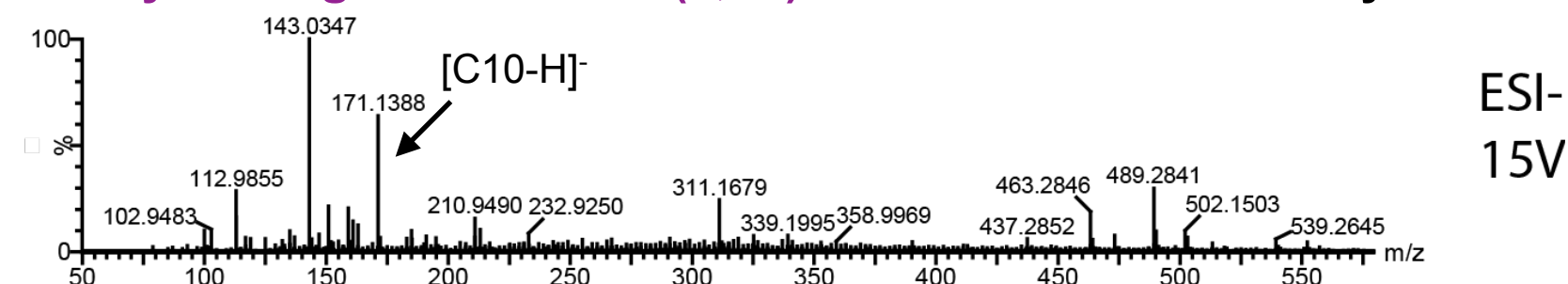


Fig. S14. *In vitro* validation of involvement of SnASFF1 in *S. nigrum* acylglucose biosynthesis. (A) ESI- mode LC-MS analysis of *in vitro* enzyme assay products indicates SnASFF1 hydrolyzes S2:13(5,8) and S2:16(8,8) to produce G2:13(5,8) and S2:16(8,8). (B) The G2:14 isomers produced from *in vitro* (top) and *in vivo* (bottom two) S2:14 isomers generate comparable mass spectra that reveals the loss of C4 and C10 ketenes under no collision energy (0V) and the presence of C10 fatty acid anions (m/z 171.14) at elevated collision energy (15V).

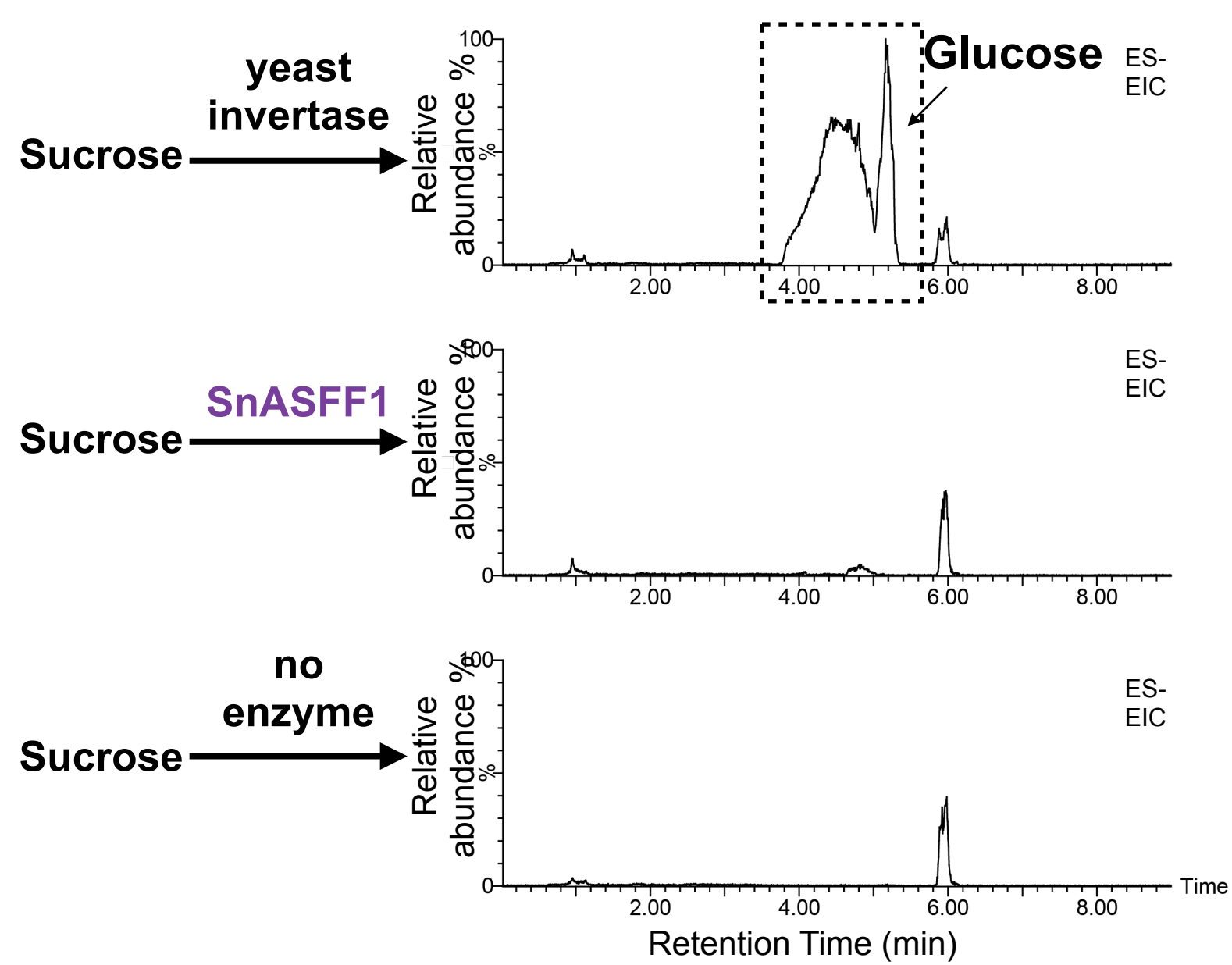
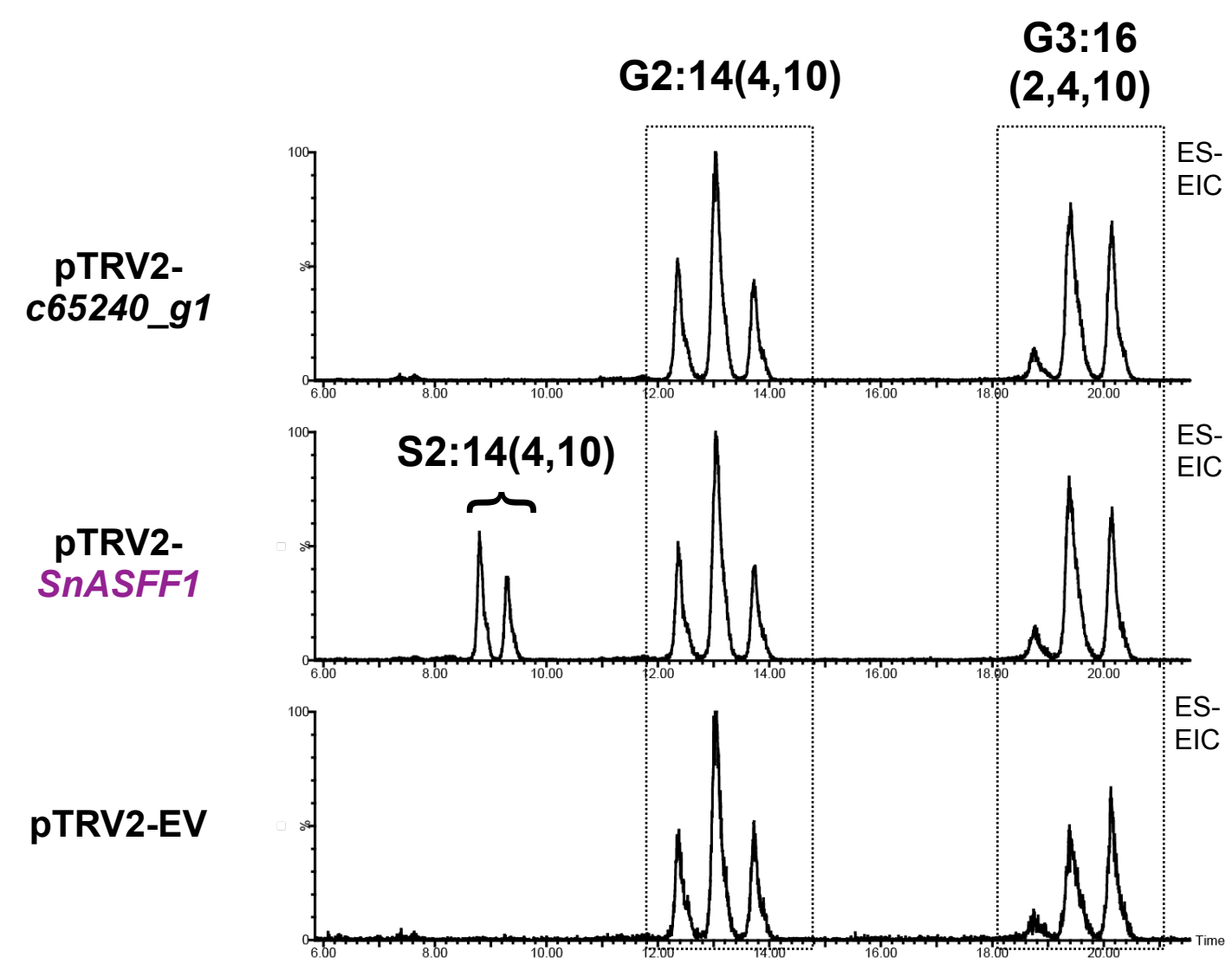
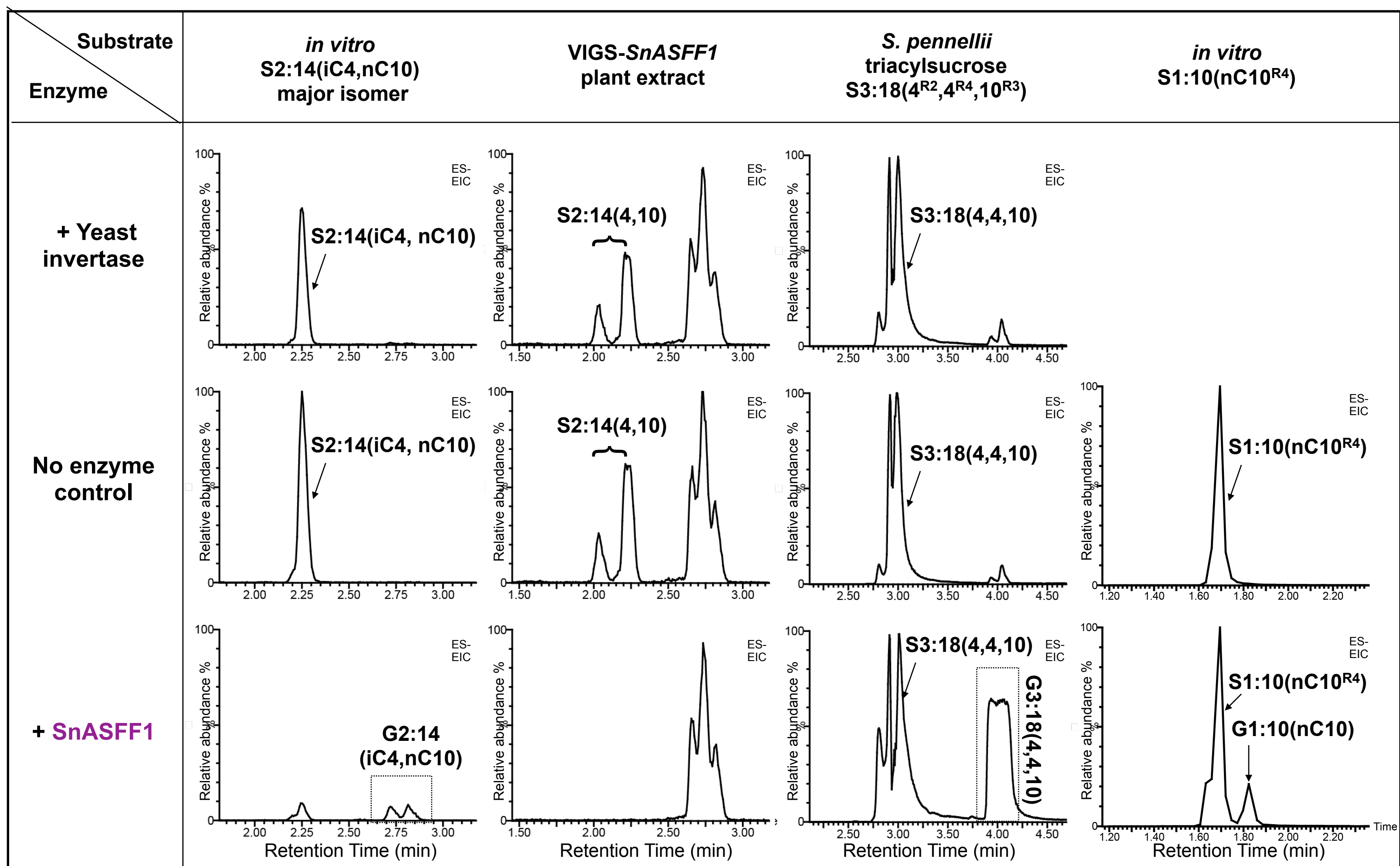
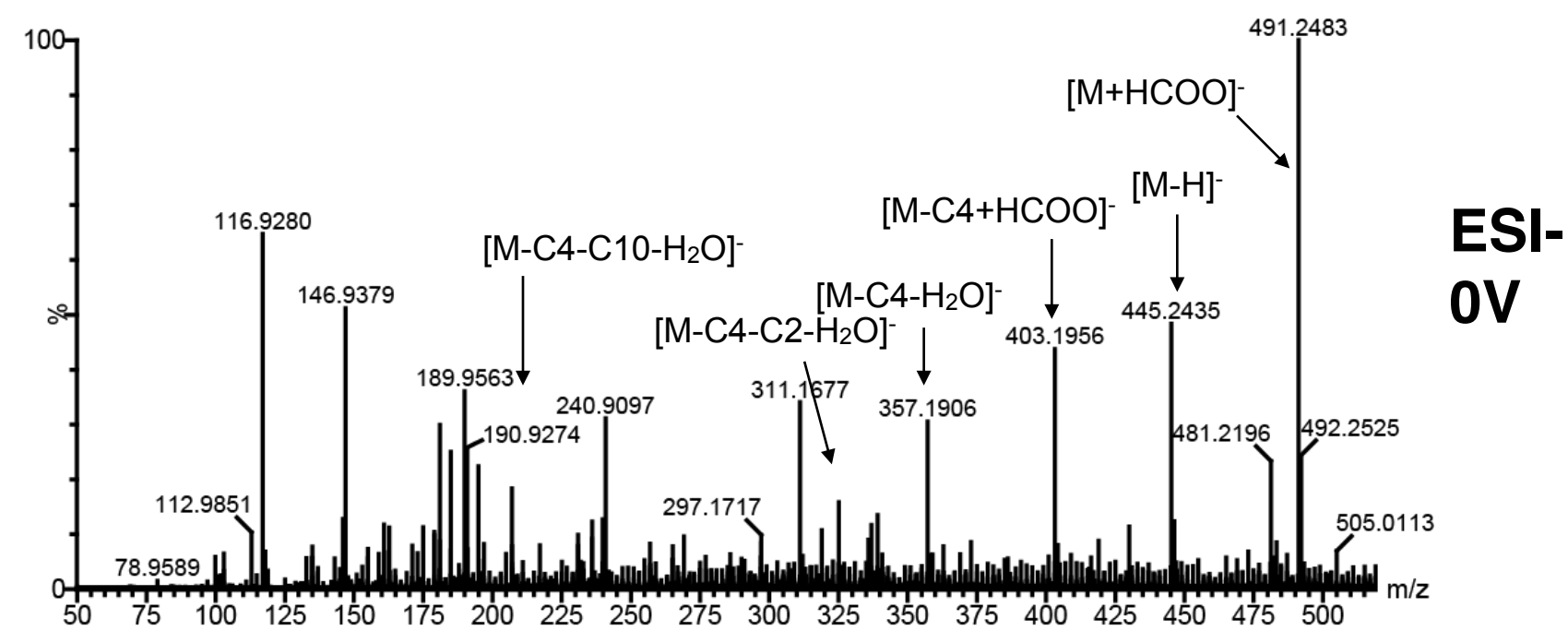
(A)**(C)****(B)**

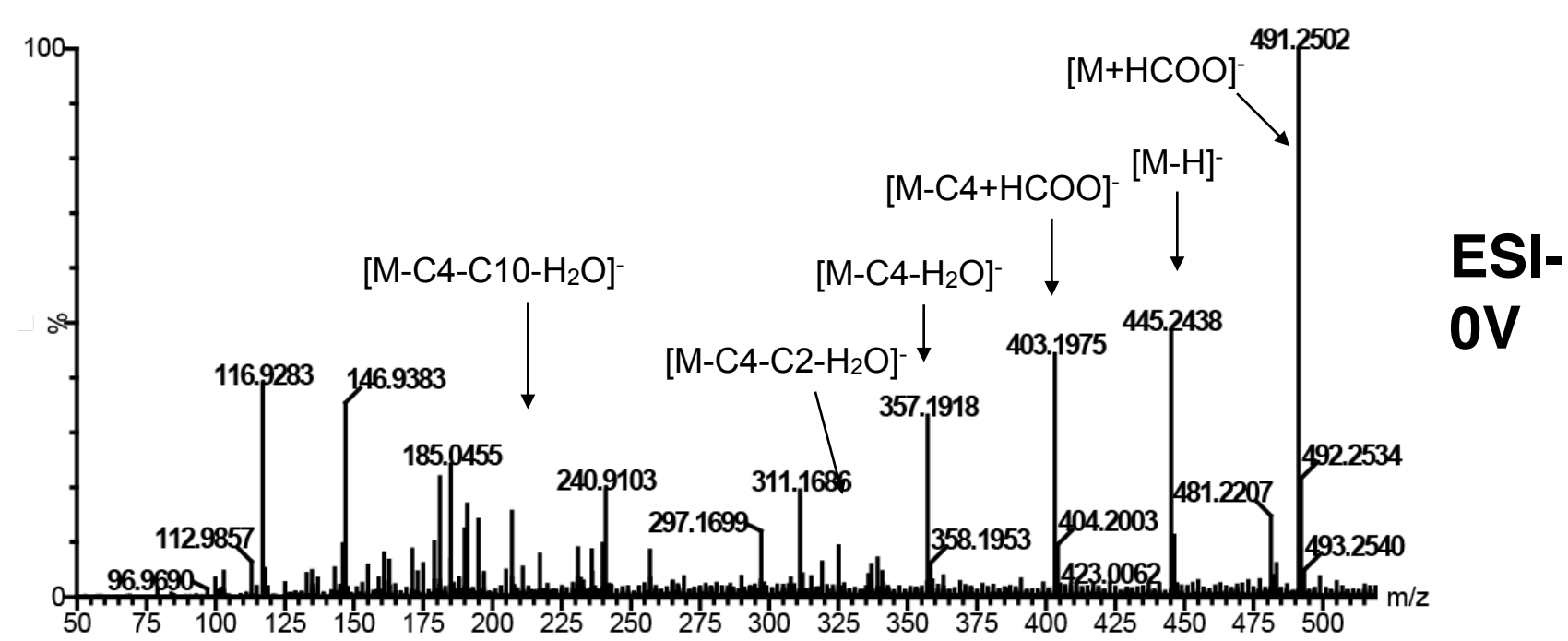
Fig. S15. SnASFF1 cleaves mono-, di- and triacylated sucroses but not unmodified sucrose while yeast invertase cleaves unmodified sucrose but not acylsucroses. (A) LC-MS analysis of *in vitro* assays with unacylated sucrose showing abundant glucose accumulating in enzyme assay with yeast invertase (top) but not with SnASFF1 (middle). Extracted ion chromatogram (ESI-) showing glucose products (m/z 179.06). (B) Representative LC-MS of *SnASFF1*-targeted, *c65240_g1*-targeted and empty vector VIGS plants showing diacylsucrose accumulation in VIGS-*SnASFF1* but not in VIGS-*c65240_g1* plants. Combined extracted ion chromatogram showing formate adducts of S2:14 (m/z 611.29), G2:14 (m/z 445.26) and G3:16 (m/z 491.25). (B) ESI- mode LC-MS analysis of *in vitro* enzyme assay products indicates that SnASFF1 hydrolyzes *in vitro* and *in vivo* diacylsucrose S2:14(4,10), R₄-acylated monoacylsucrose S1:10 and R₂, R₃, R₄-acylated triacylsucrose S3:18(4^{R2},4^{R4},10^{R3}) purified from *S. pennellii*. Chemical structure of S3:18(4^{R2},4^{R4},10^{R3}) was verified by NMR³. Combined extracted ion chromatogram showing formate adducts of substrates and products: column 1 and 2, S2:14 (m/z 611.29) and G2:14 (m/z 445.26), column 3, S3:18 (m/z 654.37) and G3:18 (m/z 492.31), column 4, S1:10 (m/z 541.25) and G1:10 (m/z 379.2). Combined extracted ion chromatogram showing formate adducts of S2:14 (m/z 611.29), G2:14 (m/z 445.26) and G3:16 (m/z 491.25).

(A)

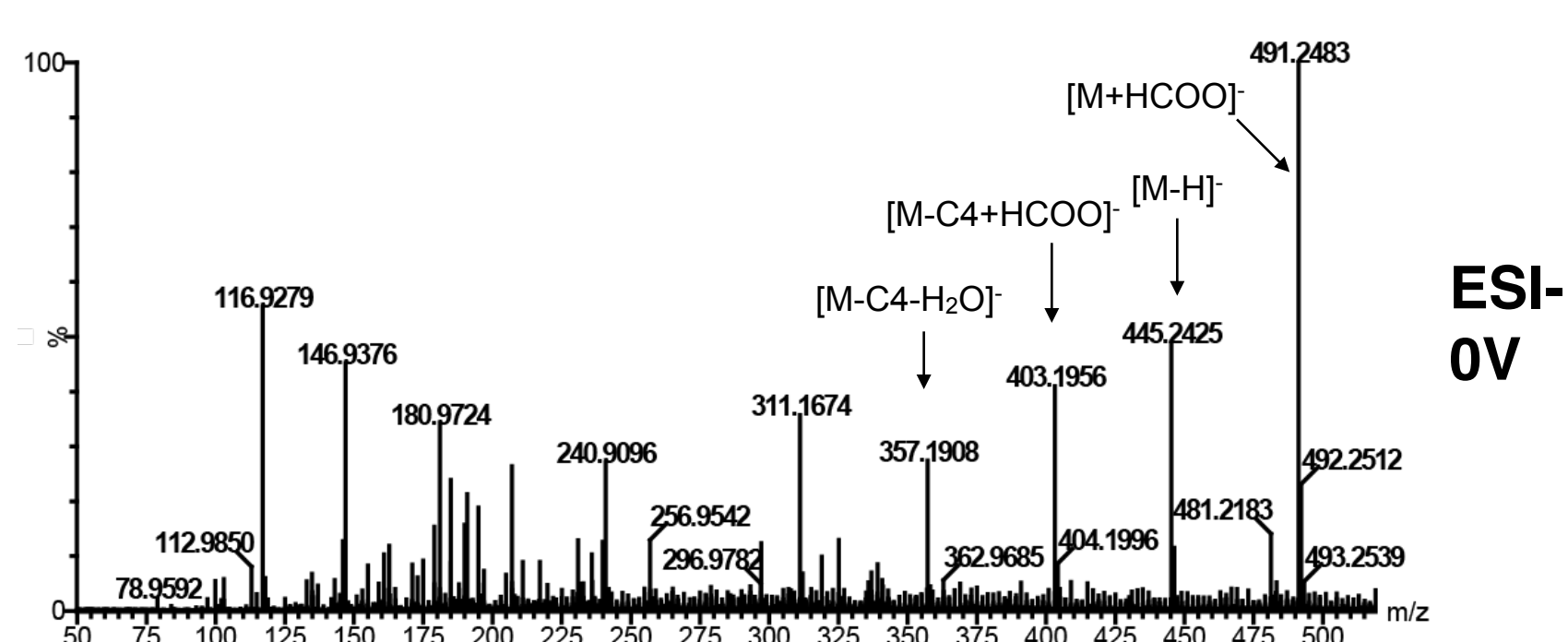
G3:16(2,iC4,nC10) (m/z 491.25 in ESI-) from
SnASAT1+SnASAT2+SnASFF1+SnAGAT1 *in vitro* assay



G3:16(2,4,10) (m/z 491.25 in ESI-) from
late-eluting *in vivo* S2:14(4,10)+SnASFF1+SnAGAT1 *in vitro* assay

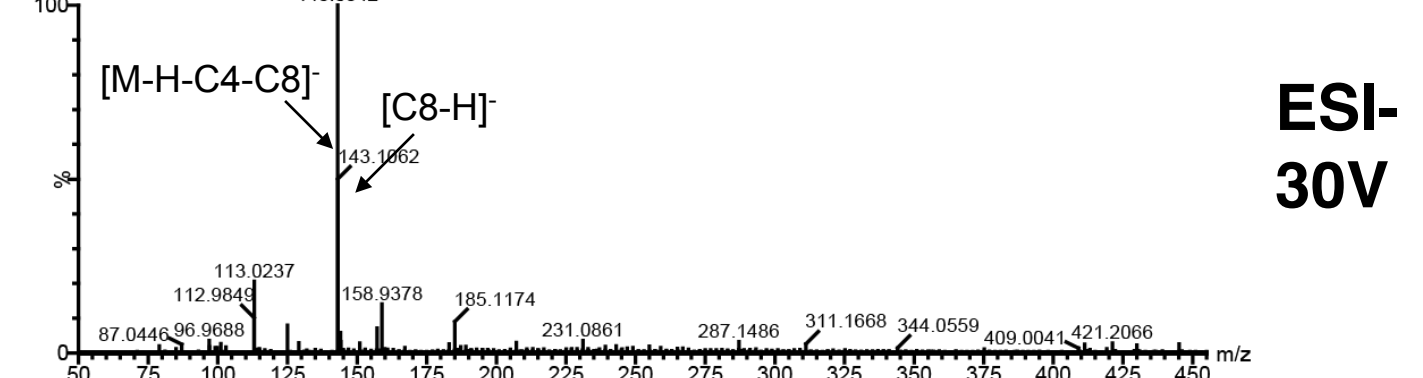


G3:16(2,4,10) (m/z 491.25 in ESI-) from
early-eluting *in vivo* S2:14(4,10)+SnASFF1+SnAGAT1 *in vitro* assay

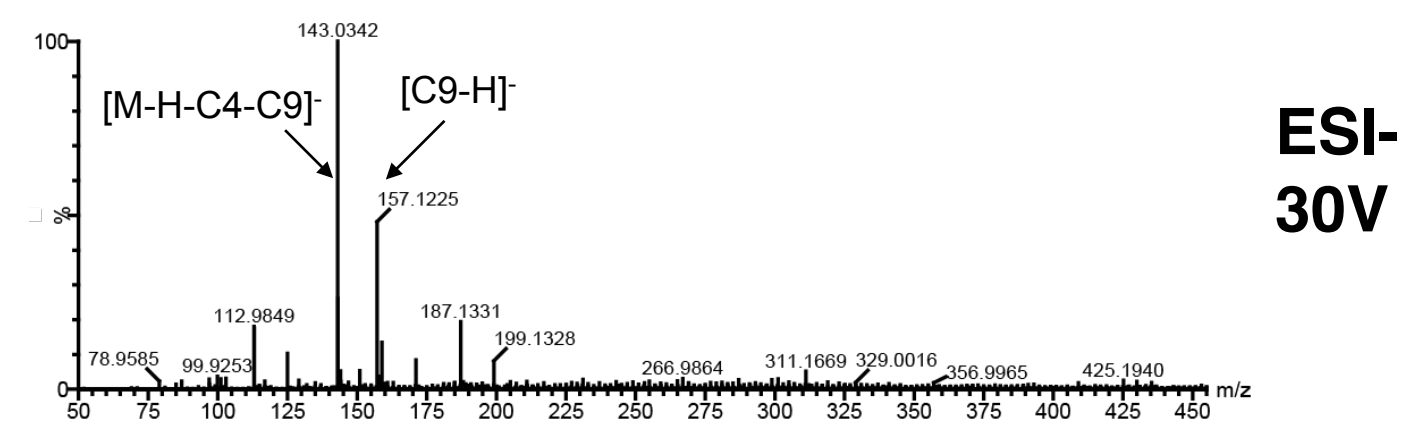


(B)

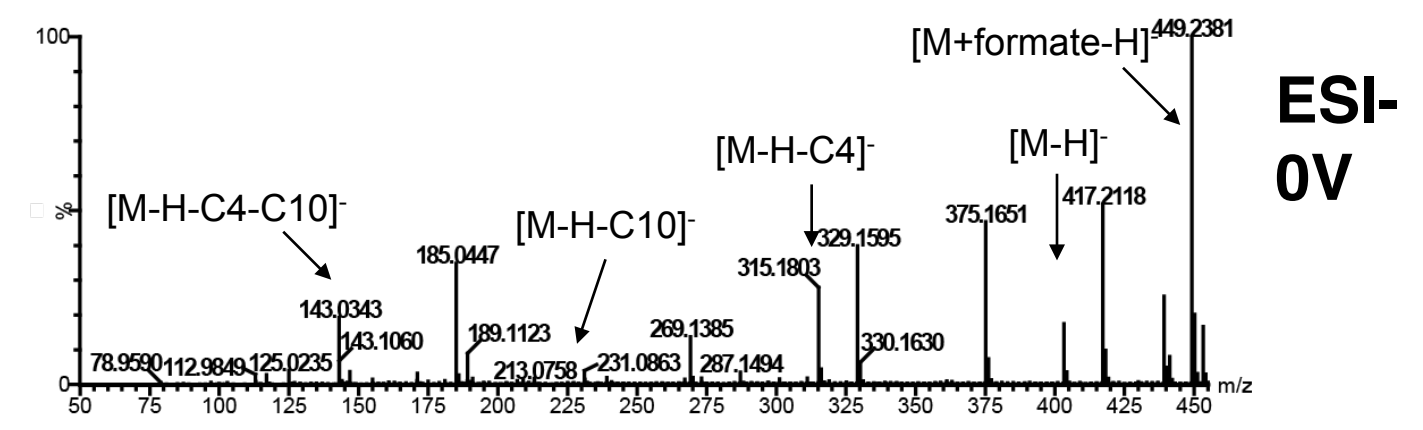
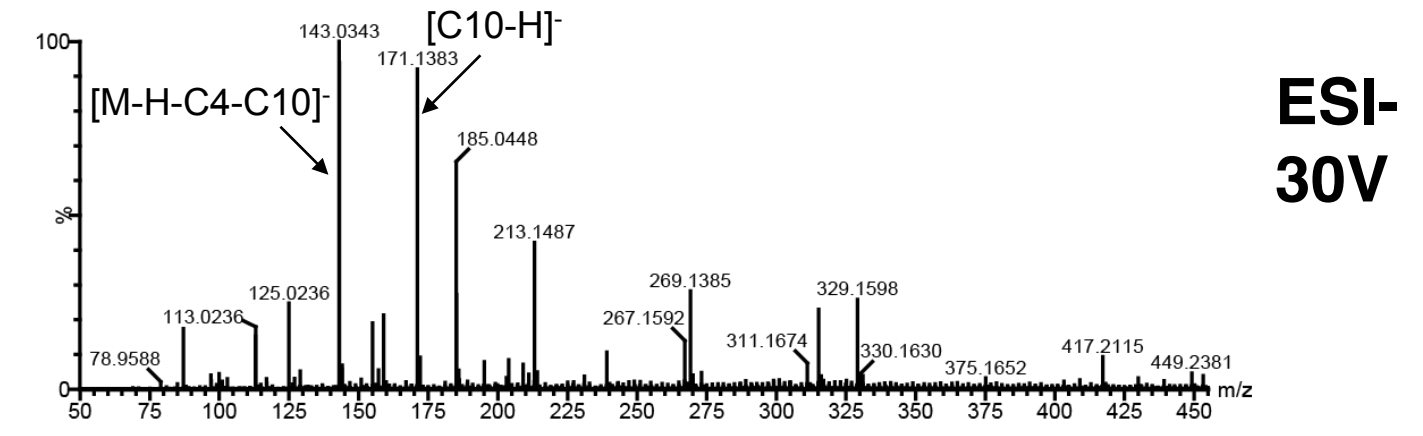
G2:12(4,8) (m/z 421.21 in ESI-) from
S. nigrum frac 1 + SnAGAT1 *in vitro* assay



G2:13(4,9) (m/z 435.22 in ESI-) from
S. nigrum frac 1 + SnAGAT1 *in vitro* assay



G2:14(4,10) (m/z 449.24 in ESI-) from
S. nigrum frac 1 + SnAGAT1 *in vitro* assay



(C)

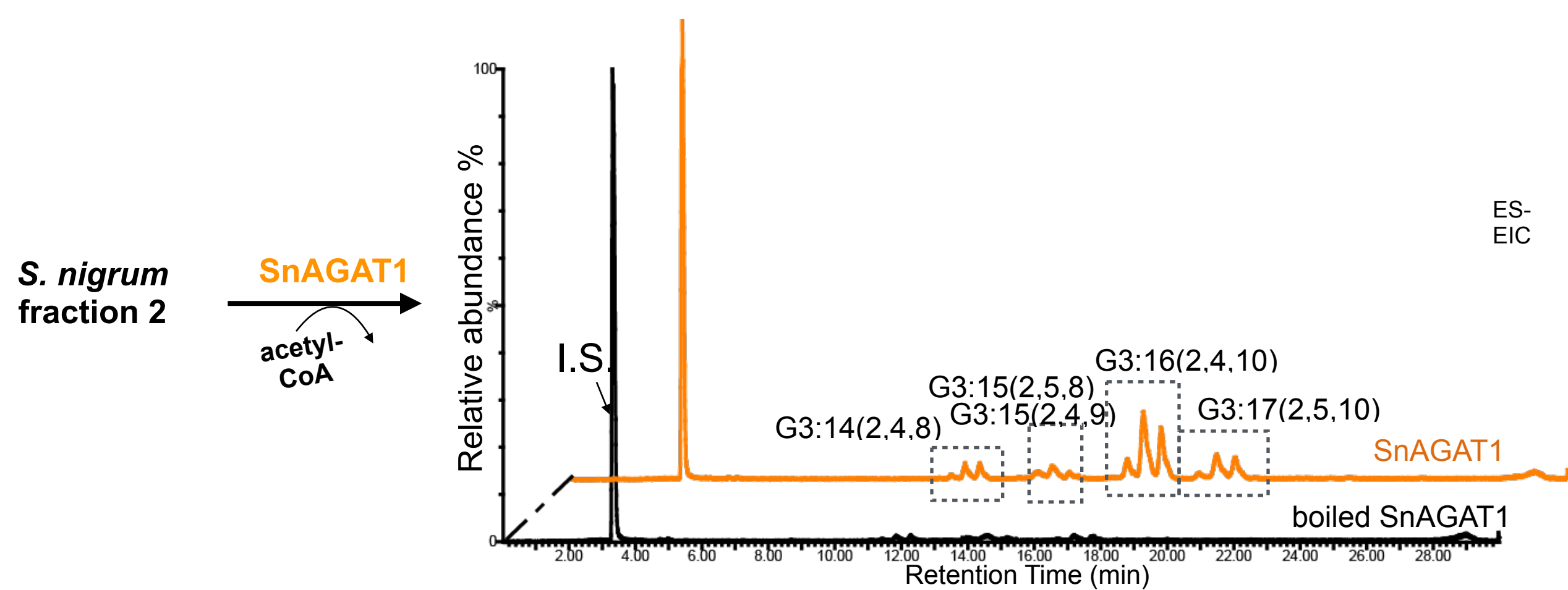


Fig. S16. SnAGAT1 forward and reverse enzyme assay products agrees with the acetyltransferase converting diacylglucoses to triacylglucoses. (A) The G3:16 isomers produced from SnASFF1 + SnAGAT1 with *in vitro* (top) and *in vivo* (bottom two) S2:14 isomers all generate mass spectra (0V) that are comparable to *S. nigrum* G3:16 (Fig. S1C). Fragmentation of these G3:16 in 0V ESI- mode are characterized by the loss of C4, C10 and C2 ketenes. (B) Mass spectra of G2:12, G2:13 and G2:14 diacylglucoses produced from SnAGAT1 reverse enzyme assay with *S. nigrum* extract fraction 1 all agrees with losing an acetyl group. (C) LC-MS analysis of enzyme assay products from SnAGAT1 activity (orange trace) shows SnAGAT1 acetylating *S. nigrum* fraction 2 diacylglucoses in the presence of acetyl-CoA. Combined extracted ion chromatogram (EIC) under negative electrospray ionization (ESI-) showing telmisartan as internal standard (I.S.) (m/z 513.23) and formic adducts of G3:14, G3:15, G3:16 and G3:17. The corresponding m/z 's of these compounds are listed in Table S1 under otherwise specified.

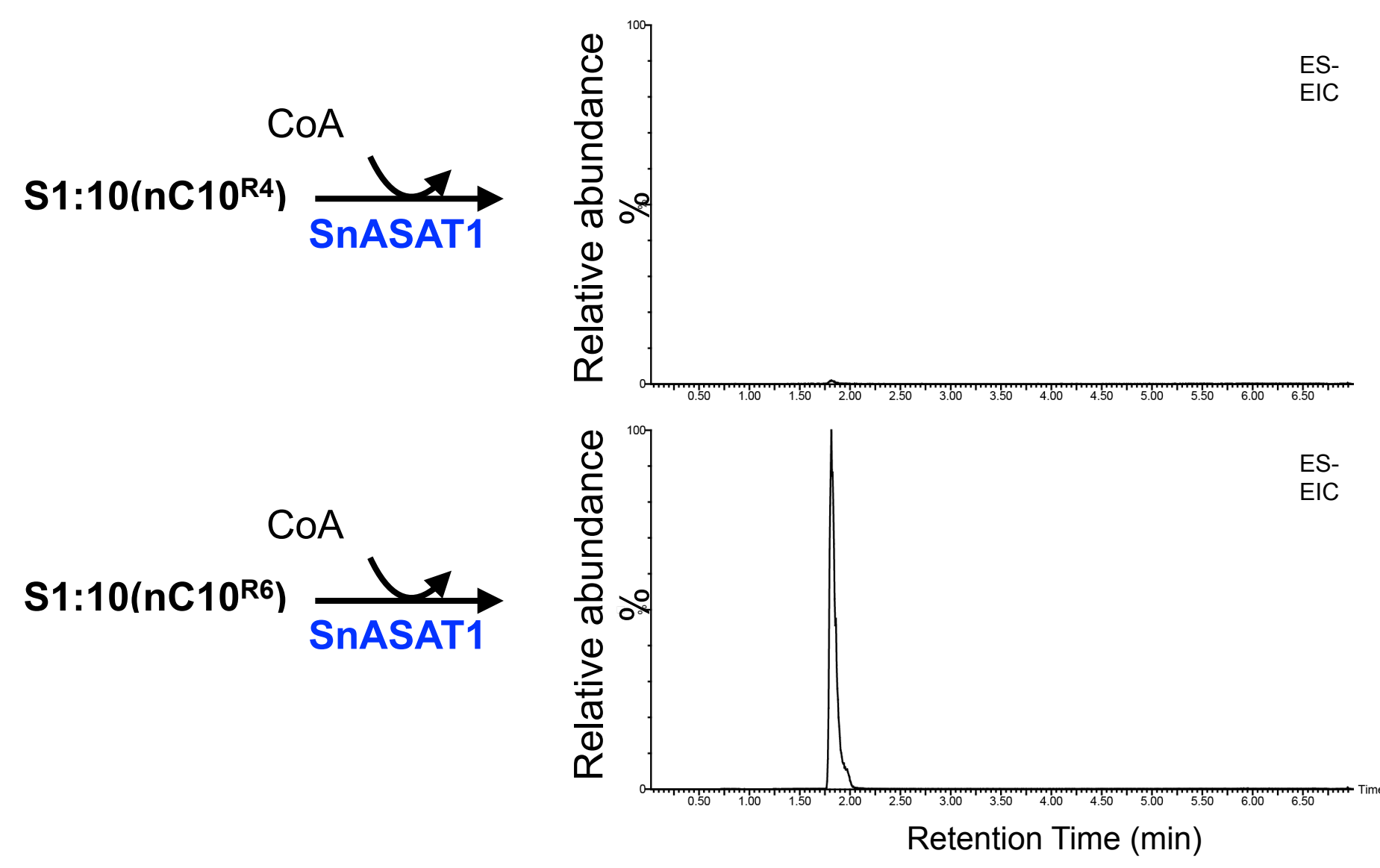


Fig. S17. SnASAT1 reverse activity deacylates S1:10(10^{R4}) but not the rearranged S1:10(10^{R6}) isomer. LC-MS analysis of *in vitro* enzyme assays with purified S1:10(10^{R4}) and S1:10(10^{R6}) showing SnASAT1 reverse activity diminishes S1:10(10^{R4}) accumulation (top) while S1:10(10^{R6}) accumulation (bottom) is unaffected. Extracted ion chromatogram showing formate adducts of S1:10 (*m/z* 611.29).

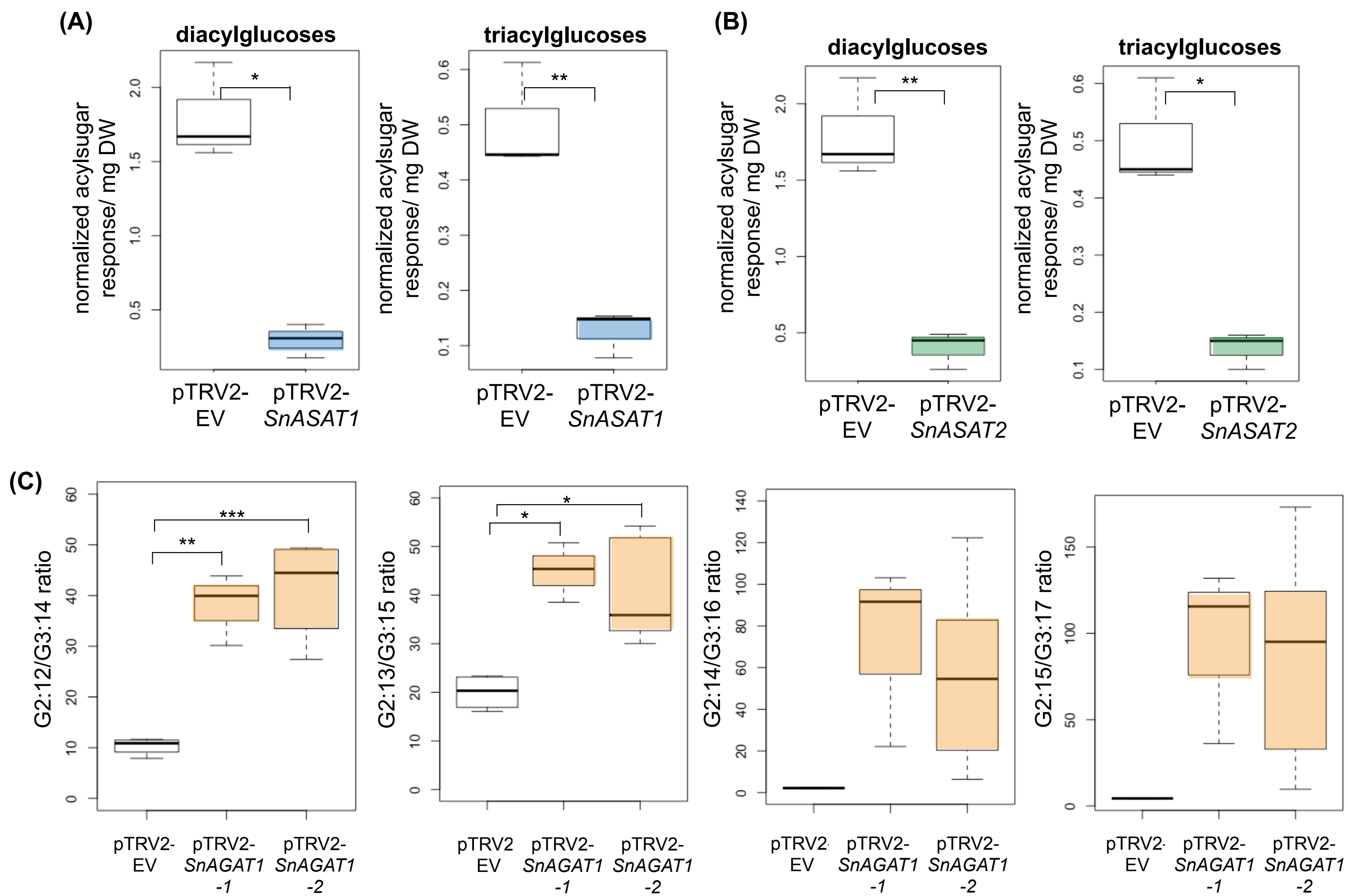


Fig. S18. Independent VIGS validation of the BAHD enzymes on *S. nigrum* acylglucose biosynthetic pathway. Comparison of acylglucose accumulation in (A) *SnASAT1*-targeted, (B) *SnASAT2*-targeted and empty vector VIGS plants. Independent experiments using different constructs from data shown in Fig. 8. (C) Comparison of the ratio of G2:12/G3:14, G2:13/G3:15, G2:14/G3:16, and G2:15/G3:17 in *SnAGAT1*-targeted and empty vector VIGS plants. Two constructs targeting different region on *SnAGAT1* were used. Data for pTRV2-*SnAGAT1-1* shown here are biological replicates of data presented in Fig. 8. Acylsugars were analyzed using LC-MS in ESI- mode. Diacylglucose quantities were measured by integrating peak areas of G2:12, G2:13, G2:14 and G2:15, whereas triacylglucose quantities were measured by integrating peak areas of G3:14, G3:15, G3:16 and G3:17. The integrated peak areas were normalized to the internal standard telmisartan (for all) and dry leaf weights (for *SnASAT1*- experiment and *SnASAT2*-silencing experiments). The ratio between tri- and diacylglucoses were independently calculated. The corresponding *m/z* of analyzed acylsugars are listed in Table S1 unless otherwise specified. Significant levels are shown (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; Welch's two sample t-test).