

iScience, Volume 25

## **Supplemental information**

### **Glycosyltransferase engineering and multi-glycosylation routes development facilitating synthesis of high-intensity sweetener mogrosides**

**Jiao Li, Shicheng Mu, Jiangang Yang, Cui Liu, Yanfei Zhang, Peng Chen, Yan Zeng, Yueming Zhu, and Yuanxia Sun**

1 Supporting information

2 **Glycosyltransferases engineering and multi-glycosylation routes development**  
3 **facilitating biosynthesis of high-intensity sweetener mogrosides**

4 *Jiao Li<sup>1,2‡</sup>, Shicheng Mu<sup>1‡</sup>, Jiangang Yang<sup>1\*</sup>, Cui Liu<sup>1</sup>, Yanfei Zhang<sup>1</sup>, Peng Chen<sup>1</sup>,*  
5 *Yan Zeng<sup>1</sup>, Yueming Zhu<sup>1</sup>, and Yuanxia Sun<sup>1,2,3\*</sup>*

6 <sup>1</sup> National Engineering Laboratory for Industrial Enzymes, Tianjin Institute of  
7 Industrial Biotechnology, Chinese Academy of Sciences, 32 Xi Qi Dao, Tianjin  
8 Airport Economic Area, Tianjin 300308, China

9 <sup>2</sup> National Technology Innovation Center of Synthetic Biology, Tianjin 300308,  
10 China

11 <sup>3</sup> Lead contact

12 ‡ These authors contributed equally to this work

13 \* Correspondence: sun\_yx@tib.cas.cn; yang\_jg1@tib.cas.cn; Tel.: +862284861960

14

## 15 Content

16	1. Supporting tables.....	1
17	<b>Table S1.</b> Kinetic parameters of glycosyltransferases towards mogrosides. ....	1
18	<b>Table S2.</b> Enzymatic activity and product composition for MS1 and mutant M7 under	
19	different mogroside substrates.....	1
20	<b>Table S3.</b> Kinetic parameters of glycosyltransferase MG1 towards mogrol. ....	1
21	<b>Table S4.</b> Primers used in this study.....	1
22	<b>Table S5.</b> Strains and plasmids used in this study. ....	3
23	<b>Table S6.</b> <sup>13</sup> C and <sup>1</sup> H NMR data for <b>VI</b> (600 MHz). ....	4
24	2. Supporting figures .....	7
25	<b>Figure S1.</b> SDS-PAGE of recombinant protein purification fractions. ....	7
26	<b>Figure S2.</b> HPLC and MS analysis of MG1 enzymatic reaction products using mogrol as	
27	substrate.....	8
28	<b>Figure S3.</b> HPLC analysis of UGT72-269-1 and UGT94-289-3 enzymatic reaction	
29	products using mogrosides as substrates. ....	9
30	<b>Figure S4.</b> Glycosylations catalyzed by enzymes MS1 and MS2 using various substrates.	
31	.....	10
32	<b>Figure S5.</b> MS analysis of MS1 and MS2 enzymatic reaction products using mogrosides as	
33	substrates. ....	11
34	<b>Figure S6.</b> Multiple sequence alignment of 9 UGTs that show branched glycosylation	
35	activity to triterpenes. ....	12
36	<b>Figure S7.</b> Determination of kinetic parameters for recombinant glycosyltransferase MG1	
37	(on IE) and MS1 and MS2 (on mogroside IIE). ....	13
38	<b>Figure S9.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using	
39	mogroside IIE as substrate.....	15
40	<b>Figure S10.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using	
41	mogroside IIIA as substrate.....	16
42	<b>Figure S11.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using	
43	mogroside IIIIE as substrate. ....	17
44	<b>Figure S12.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using	
45	mogroside IVA as substrate.....	18
46	<b>Figure S13.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using	
47	mogroside IVE as substrate. ....	19

48	<b>Figure S14.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using mogroside Sia I as substrate. ....	20
49		
50	<b>Figure S15.</b> HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products using mogroside V as substrate. ....	21
51		
52	<b>Figure S16.</b> Conformation maps of M7-IIE and WT-IIE complexes during the 3*50 ns MD simulations. ....	22
53		
54	<b>Figure S17.</b> Optimization of reaction temperature, pH and divalent metal ions of MS1 and MS2. ....	23
55		
56	<b>Figure S18.</b> <sup>1</sup> H NMR spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz). ....	24
57	<b>Figure S19.</b> <sup>13</sup> C NMR spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz). ....	25
58	<b>Figure S20.</b> HMBC spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz) ....	26
59	<b>Figure S21.</b> HSQC spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz). ....	27
60	<b>Figure S22.</b> COSY spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz) ....	28
61	<b>Figure S23.</b> TOCSY spectrum of VI in Dimethyl sulfoxide-d <sub>6</sub> (600 MHz). ....	29
62	<b>Figure S24.</b> The gene sequences of MG1 and MS1 ....	30
63		

64

65 **1. Supporting tables**

66 **Table S1.** Kinetic parameters of glycosyltransferases towards mogrosides. **Related to**  
67 **Table1.**

Substrate	Enzyme	$K_m$ ( $\mu\text{M}$ )	$k_{cat}$ ( $\text{s}^{-1}$ )	$k_{cat}/K_m$ ( $\text{mM}^{-1}\cdot\text{s}^{-1}$ )
IIE	MS1	$34.31 \pm 0.5$	$0.014 \pm 0.002$	$0.41 \pm 0.01$
	MS2	$193.20 \pm 3.2$	$0.026 \pm 0.002$	$0.14 \pm 0.006$

68

69 **Table S2.** Enzymatic activity and product composition for MS1 and mutant M7 under  
70 different mogroside substrates. **Related to Figure 5.** To detect the product types and  
71 composition glycosylated by MS1 and its mutants, the glycosylation processes were  
72 controlled, in which 10-100  $\mu\text{g}$  of the purified enzyme at 40°C for 15-30 min  
73 (depending on the activity of mutants) and one glucose residue was transferred to  
74 substrates.

Substrates	III E			
Product types and composition	Conversion (%)		mU/mg	fold
	III A	III E		
WT	20	0	0.0085	1
M7	30	0	2.97	351

Substrates	III A			
Product types and composition	Conversion (%)		mU/mg	fold
	IV A	IV X		
WT	18	0	0.0074	1
M7	26	9	1.91	259

Substrates	III E			
Product types and composition	Conversion (%)		mU/mg	fold
	IV E	Sia I		
WT	6	17	0.0096	1
M7	3	32	3.09	322

Substrates	IV A			
Product types and composition	Conversion (%)		mU/mg	fold
	VA			
WT	-		0	-
M7	25		1.25	-

Substrates	IV E			
Product types and composition	Conversion (%)		mU/mg	fold
	V	VE		
WT	18	-	0.0072	1
M7	1	36	2.72	378

Substrates	Sia I			
Product types and composition	Conversion (%)		mU/mg	fold

	V	VX		
WT	15	-	0.0075	1
M7	14	11	1.04	139

Substrates	V		
Product types and composition	Conversion (%)		
	VI	mU/mg	fold
WT	-	0	-
M7	27	1.08	-

75

76 **Table S3.** Kinetic parameters of glycosyltransferase MG1 towards mogrol. **Related to**

77 **Table1.**

Substrate	Enzyme	$K_m$ ( $\mu\text{M}$ )	$k_{cat}$ ( $\text{s}^{-1}$ )	$k_{cat}/K_m$ ( $\text{mM}^{-1}\cdot\text{s}^{-1}$ )
Mogrol	MG1	$10.10 \pm 0.3$	$0.86 \pm 0.04$	$86.00 \pm 6.6$
IE	MG1	$61.27 \pm 0.7$	$0.04 \pm 0.005$	$0.63 \pm 0.04$

78



79 **Table S4.** Primers used in this study. **Related to STAR Methods.**

Primer name	Sequence (5' to 3')
MS1-F	GGTACCGACGACGACGACAAGATGGATGCTGCCCAACAAG
MS1-R	CAGTGGTGGTGGTGGTGGTGCATATTTTAAGCAAGAGAG
T11H-F	TGACACCACACATATTTTGATGCTTCCATGGCTCGGCTA
T11H-R	GCATCAAAATATGTGTGGTGTACCTTGTGGGCAGCAT
L13V-F	CACAACCATTGTGATGCTTCCATGGCTCGGCTATGGCCA
L13V-R	ATGGAAGCATCACAATGGTTGTGGTGTACCTTGTGGG
P16A-F	TTTGATGCTTGCCTGGCTCGGCTATGGCCATCTTTCAGC
P16A-R	AGCCGAGCCACGCAAGCATCAAAATGGTTGTGGTGTAC
W17A-F	GATGCTTCCAGCGCTCGGCTATGGCCATCTTTCAGCTTT
W17A-R	CATAGCCGAGCGCTGGAAGCATCAAAATGGTTGTGGTGT
L18A-F	GCTTCCATGGGCGGGCTATGGCCATCTTTCAGCTTTTCT
L18A-R	GGCCATAGCCCGCCCATGGAAGCATCAAAATGGTTGTGG
G19A-F	TCCATGGCTCGCGTATGGCCATCTTTCAGCTTTTCTCGA
G19A-R	GATGGCCATACGCGAGCCATGGAAGCATCAAAATGGTTG
Y20A-F	ATGGCTCGGCGCGGGCCATCTTTCAGCTTTTCTCGAGCT
Y20A-R	AAAGATGGCCCGCGCCGAGCCATGGAAGCATCAAAATGG
G21A-F	GCTCGGCTATGCGCATCTTTCAGCTTTTCTCGAGCTGGC
G21A-R	CTGAAAGATGCGCATAGCCGAGCCATGGAAGCATCAAAA
L23A-F	CTATGGCCATGCGTCAGCTTTTCTCGAGCTGGCCAAAAG
L23A-R	GAAAAGCTGACGCATGGCCATAGCCGAGCCATGGAAGCA
S24A-F	TGGCCATCTTGCCTGGCTTTTCTCGAGCTGGCCAAAAGCCT
S24A-R	CGAGAAAAGCCGCAAGATGGCCATAGCCGAGCCATGGAA
A25P-F	CCATCTTTCACCGTTTCTCGAGCTGGCCAAAAGCCTCTC
A25P-R	GCTCGAGAAACGGTGAAAGATGGCCATAGCCGAGCCATG
F26A-F	TCTTTCAGCTGCGCTCGAGCTGGCCAAAAGCCTCTCAAG
F26A-R	CCAGCTCGAGCGCAGCTGAAAGATGGCCATAGCCGAGCC
S34A-F	CAAAAGCCTCGCGAGGAGGAACTTCCATATCTACTTCTG
S34A-R	AGTTCCTCCTCGCGAGGCTTTTGGCCAGCTCGAGAAAAG
N37G-F	CTCAAGGAGGGGCTTCCATATCTACTTCTGTTCAACCTC
N37G-R	AGATATGGAAGCCCTCCTTGAGAGGCTTTTGGCCAGCT
F38H-F	AAGGAGGAACCATCATATCTACTTCTGTTCAACCTCTGT
F38H-R	AGTAGATATGATGGTTCCTCCTTGAGAGGCTTTTGGCCA
I40V-F	GAACTTCCATGTGTACTTCTGTTCAACCTCTGTTAATCT
I40V-R	AACAGAAGTACACATGGAAGTTCCTCCTTGAGAGGCTTT
S46P-F	CTGTTCAACCCCGTTAATCTTGACGCCATTAACCAAAA
S46P-R	CAAGATTAACCGGGGTTGAACAGAAGTACACATGGAAGT
D62P-F	TTCTTTCTCTCCGTCCATTCAATTTGTGGAGCTCCATCT
D62P-R	ATTGAATGGACGGAGAGAAAGAAGAAGGAAGCTTTGGTT
F77L-F	TTCTCCTGAGCTGCCTCCTCATCTTACACAACCAACGG
F77L-R	GATGAGGAGGCAGCTCAGGAGAAGAAGGGAGATGGAGCT
H82E-F	TCCTCATCTTGAAACAACCAACGGCCTTCCCCCTACCCT

H82E-R	CGTTGGTTGTTTCAAGATGAGGAGGGAACTCAGGAGAAG
A102L-F	CTTCTCCATGCTGGCCCAGCACTTTGAGTCCATTTTACA
A102L-R	AGTGCTGGGCCAGCATGGAGAAGGCTTGGTGGAGAGCGG
E107S-F	CCAGCACTTTAGCTCCATTTTACAAACACTTGCCCCGCA
E107S-R	GTAAAATGGAGCTAAAGTGCTGGGCAGCCATGGAGAAGG
S108E-F	GCACTTTGAGGAAATTTTACAAACACTTGCCCCGCACCT
S108E-R	TTTGTA AAAATTTCTCAAAGTGCTGGGCAGCCATGGAGA
H116D-F	ACTTGCCCCGGATCTTCTCATTTATGACTCTCTTCAACC
H116D-R	AAATGAGAAGATCCGGGGCAAGTGTTTGTA AAAATGGACT
I139V-F	AATCCGGCCGTGA ACTTCAATACTACGGGAGTTTTTCGT
I139V-R	TATTGAAGTTCACGGCCGGAATTTTGAGGGATGAAGCTA
N142L-F	CATCAACTTCCTGACTACGGGAGTTTTTCGTCATTTCTCA
N142L-R	CTCCCGTAGTCAGGAAGTTGATGGCCGGAATTTTGAGGG
V146A-F	TACTACGGGAGCGTTCGTCATTTCTCAAGGGCTTCACCC
V146A-R	AAATGACGAACGCTCCCGTAGTATTGAAGTTGATGGCCG
N354A-F	CTGTGGATGGGCGTCCGGTGATGGAGAGCATGATGTTTGG
N354A-R	CCATCACCGACGCCCATCCACAGTGGCTCACGAATCCCC

---

81 **Table S5.** Strains and plasmids used in this study. **Related to STAR Methods.**

Strains and plasmids	description	Ref.
<b>plasmids</b>		
pET32a (+)	pBR322 ori with P <sub>T7</sub> ; Amp <sup>R</sup>	Novagen
pET32a-MG1	pET32 carrying MG1	This study
pET32a-MS1	pET32 carrying MS1	This study
pET32a-MS2	pET32 carrying MS2	This study
pYZ291	2MICRON, URA3	Novagen
pYZ291-MS1	TEF1p-MG1-CYC1t	This study
pYZ291-MS1-MS1	GPD1p-MS1-adh1t-TEF1p-MS1-CYC1t	This study
pYZ291-MS1-MG1	GPD1p-MS1-adh1t-TEF1p-MG1-CYC1t	This study
<b>Strains</b>		
CEN.PK2-1C	Saccharomyces cerevisiae strain CEN.PK2-1C (MATaleu2-3, 112 ura3-52 trp1-289 his3-Δ1 MAL2-8cSUC2)	Novagen
Mog1	CEN.PK2-1C containing pYZ291-MS1	This study
Mog2	CEN.PK2-1C containing pYZ291-MS1, with gene Exg1 deleted	This study
Mog3	CEN.PK2-1C containing pYZ291-MS1-MS1, with gene Exg1 deleted	This study
Mog4	CEN.PK2-1C containing pYZ291-MS1-MG1, with gene Exg1 deleted	This study

82

83 **Table S6.**  $^{13}\text{C}$  and  $^1\text{H}$  NMR data for **VI** (600 MHz). Related to Figure 5.

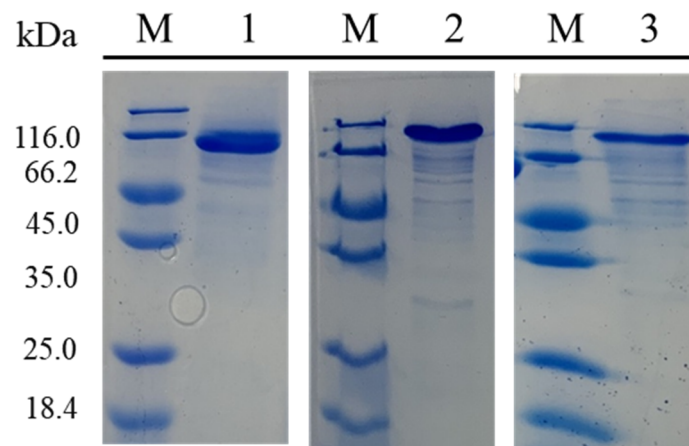
C	$\delta^{13}\text{C}$	$\delta^1\text{H}$
1	27.2	$\text{CH}_2$ , 1.49(m), 2.24(m)
2	29.3	$\text{CH}_2$ , 1.91(m)
3	88.6	CH, 3.49(br, s)
4	42.9	
5	144.7	
6	119.9	CH, 5.47(d, J=5.4Hz)
7	25.2	$\text{CH}_2$ , 1.79(m), 2.44(m)
8	44.7	CH, 1.67(m)
9	41.0	
10	37.1	CH, 2.50(d, J=12.4Hz)
11	79.6	CH, 3.85(m)
12	41.2	$\text{CH}_2$ , 1.81(m)
13	48.5	
14	50.4	
15	35.4	$\text{CH}_2$ , 1.14(m), 1.21(m)
16	29.5	$\text{CH}_2$ , 1.32(m), 1.96(m)
17	51.9	CH, 1.62(dd)
18	17.3	$\text{CH}_3$ , 0.92(s)
19	26.3	$\text{CH}_3$ , 1.15(s)
20	37.5	CH, 1.50(m)
21	19.3	$\text{CH}_3$ , 0.97(d)
22	34.1	$\text{CH}_2$ , 1.48(m), 1.56(m)
23	29.9	$\text{CH}_2$ , 1.40(m), 1.55(m)
24	93.2	CH, 3.40(m)
25	73.8	
26	26.8	$\text{CH}_3$ , 1.12(s)
27	24.2	$\text{CH}_3$ , 1.15(s)
28	28.1	$\text{CH}_3$ , 1.07(s)
29	26.6	$\text{CH}_3$ , 1.21(s)
30	20.0	$\text{CH}_3$ , 0.89(s)
3-O-Glc-1		
C-1	104.8	CH, 4.40(d, J=7.80Hz)
C-2	80.2	CH, 3.63(m)
C-3	78.2	CH, 3.54(m)
C-4	71.6	CH, 3.19(m)
C-5	77.1	CH, 3.44(m)
C-6	69.8	$\text{CH}_2$ , 3.80(dd, J1=5.8Hz, J2=11.4Hz), 4.06(d, J=11.4Hz)
3-O-Glc-2-Glc		
C-1	104.0	CH, 4.67(d, J=7.80Hz)
C-2	77.9	CH, 3.25(m)
C-3	78.6	CH, 3.63(m)

C-4	71.8	CH, 3.29(m)
C-5	75.7	CH, 3.27(m)
C-6	63.1	CH <sub>2</sub> , 3.67(dd), 3.85(d)
3-O-Glc-6-Glc		
C-1	104.8	CH, 4.42(d, J=7.80Hz)
C-2	77.9	CH, 3.19(m)
C-3	78.1	CH, 3.28(m)
C-4	71.6	CH, 3.32(m)
C-5	72.4	CH, 3.22(m)
C-6	62.7	CH <sub>2</sub> , 3.67(dd), 3.85(d)
24-O-Glc-1		
C-1	104.1	CH, 4.44(d, J=7.80Hz)
C-2	81.3	CH, 3.61(m)
C-3	78.0	CH, 3.37(m)
C-4	71.7	CH, 3.27(m)
C-5	76.5	CH, 3.51(m)
C-6	70.2	CH <sub>2</sub> , 3.63(dd, J1=5.8Hz, J2=11.4Hz), 4.24(d, J=11.4Hz)
24-O-1Glc-2-Glc		
C-1	104.6	CH, 4.77(d, J=7.80Hz)
C-2	78.3	CH, 3.28(m)
C-3	77.7	CH, 3.37(m)
C-4	75.2	CH, 3.20(m)
C-5	78.0	CH, 3.26(m)
C-6	63.6	CH <sub>2</sub> , 3.64(dd), 3.87(d)
24-O-1Glc-6-Glc		
C-1	104.4	CH, 4.29(d, J=7.80Hz)
C-2	75.2	CH, 3.23(m)
C-3	77.7	CH, 3.36(m)
C-4	71.5	CH, 3.27(m)
C-5	76.2	CH, 3.26(m)
C-6	62.8	CH <sub>2</sub> , 3.67(dd), 3.85(d)

---



86 **2. Supporting figures**

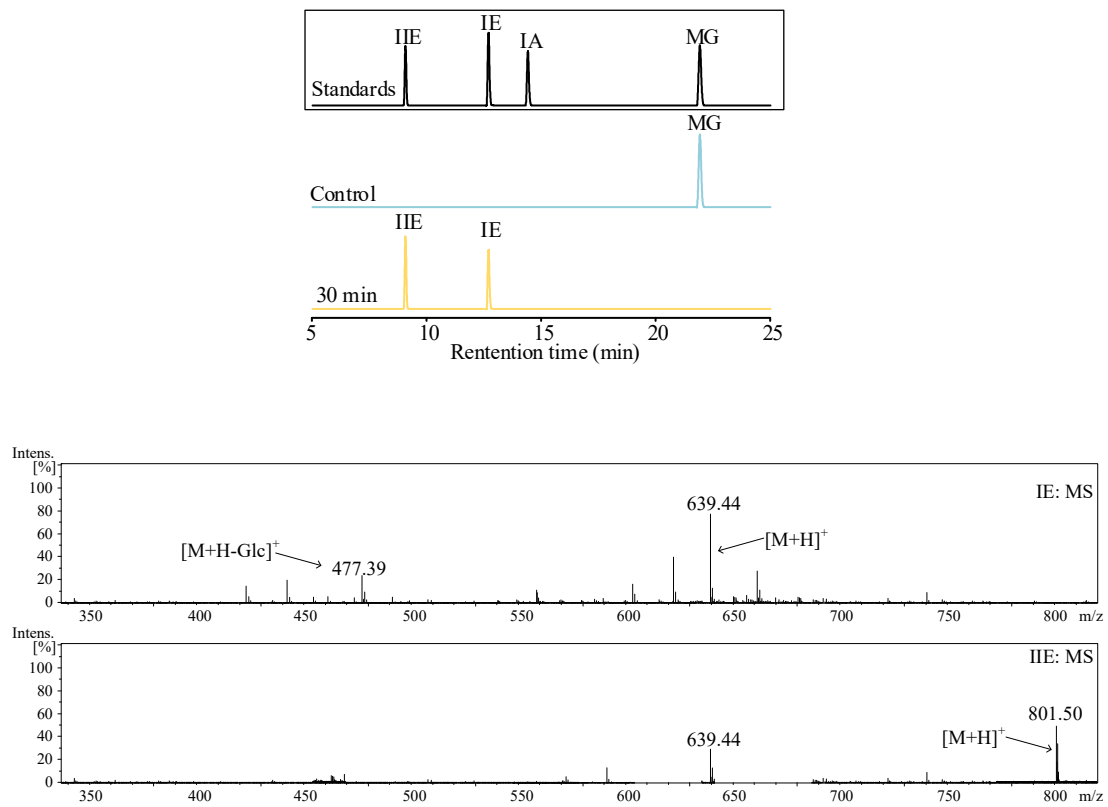


87

88 **Figure S1.** SDS-PAGE of recombinant protein purification fractions. **Related to Figures**

89 **2 and 3.** Lane M: protein marker; lane 1: purified MG1 (54 kDa); lane 2: MS1 (72 kDa);

90 lane 3: MS2 (71 kDa).

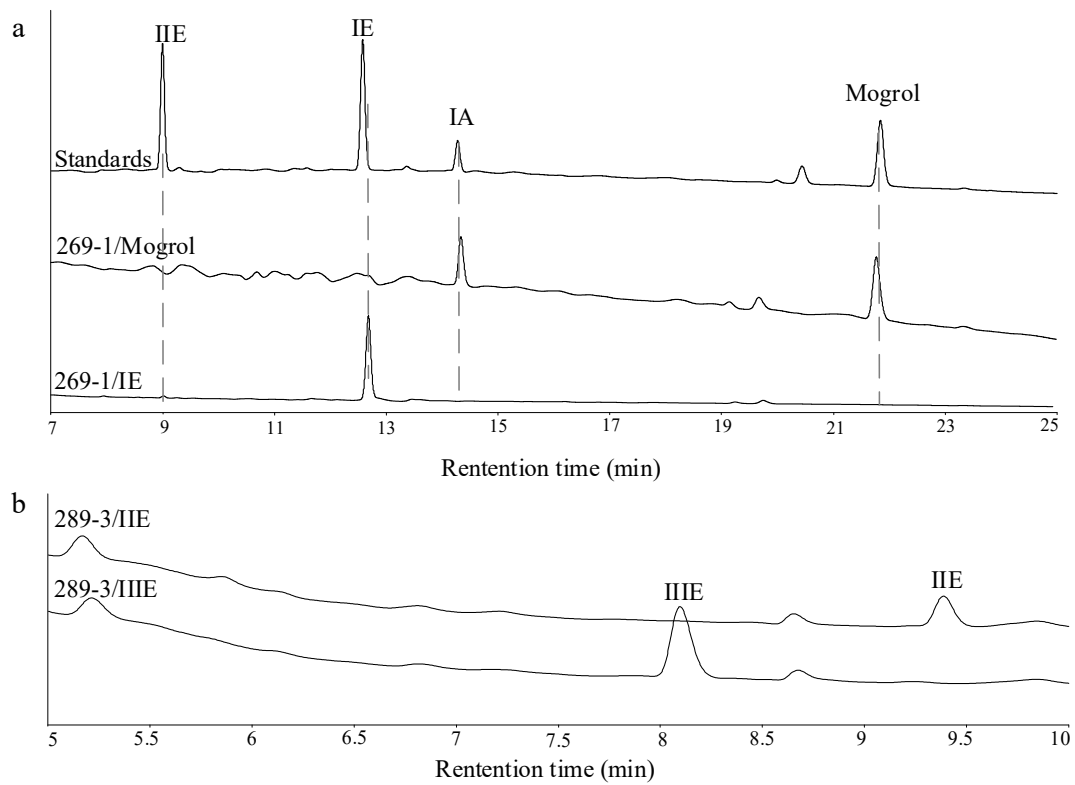


91

92 **Figure S2.** HPLC and MS analysis of MG1 enzymatic reaction products using mogrol

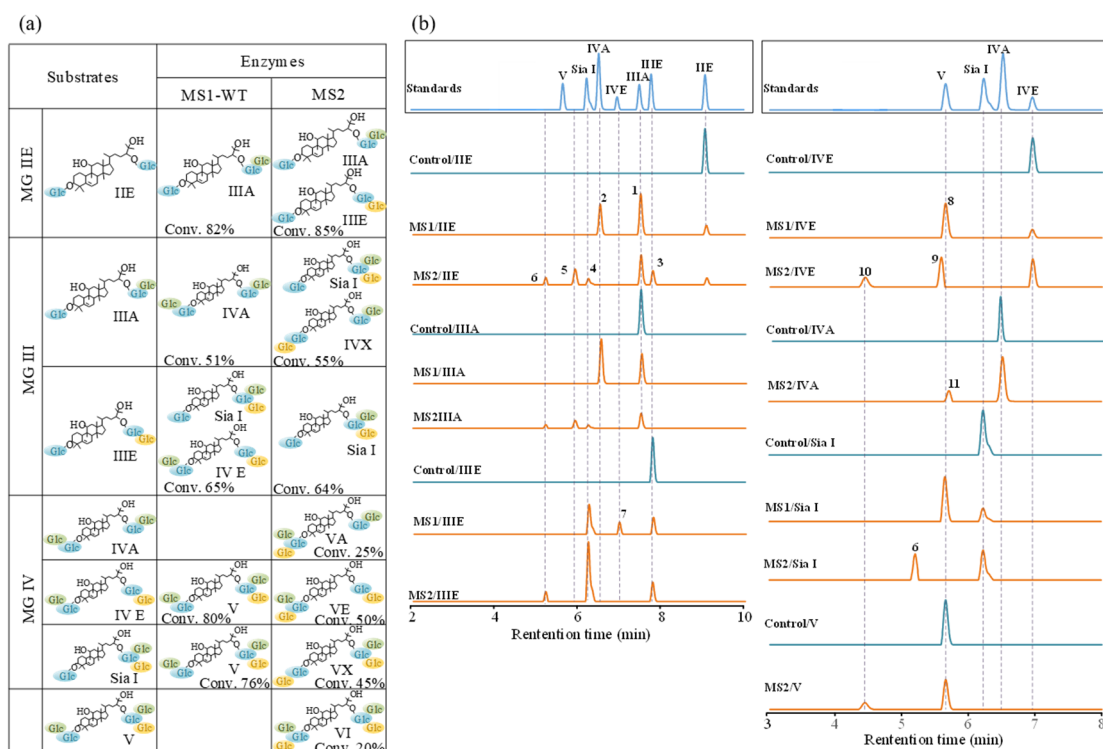
93 as substrate. **Related to Figure 2.**





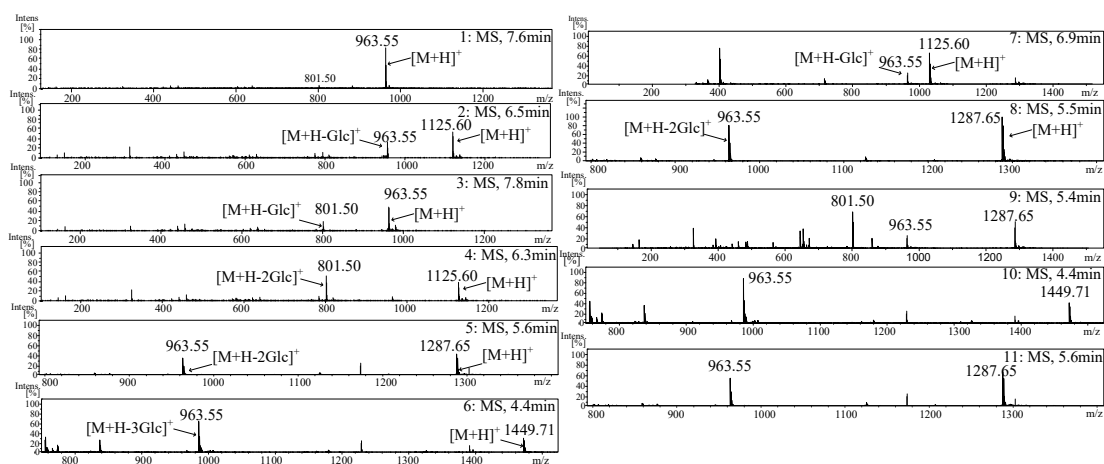
94

95 **Figure S3.** HPLC analysis of UGT72-269-1 and UGT94-289-3 enzymatic reaction  
 96 products using mogrosides as substrates. Related to Figure 3.



98

99 **Figure S4.** Glycosylations catalyzed by enzymes MS1 and MS2 using various  
 100 substrates. **Related to Figure 5.** (a) Schematic summary of glycosylations by enzymes  
 101 MS1 and MS2 using various substrates. (b) HPLC analysis of MS1 and MS2 enzymatic  
 102 reaction products using mogrosides as substrates. The HPLC methods as follow:  
 103 Mobile phase A was ddH<sub>2</sub>O with 0.1% formic acid, and mobile phase B contained  
 104 CH<sub>3</sub>CN and 0.1% formic acid. The gradient are as follows: 0-25 min, 25%-80% pump  
 105 B for mogrol, mogroside IE, mogroside IIE, and mogroside III; 0-25 min, 25%-55%  
 106 pump B for mogroside IV, V, and VI, and the flow rate was 1 mL/min.



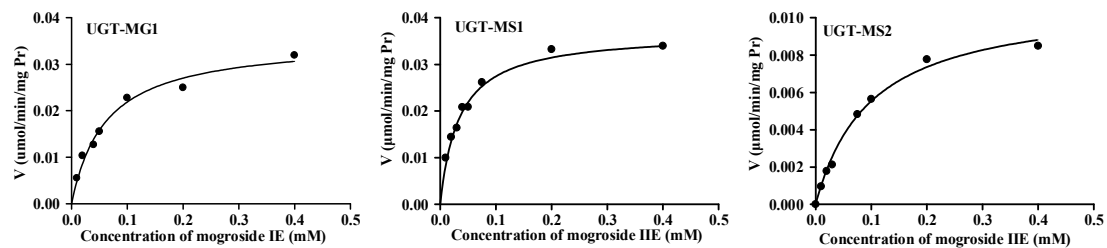
107

108 **Figure S5.** MS analysis of MS1 and MS2 enzymatic reaction products using  
 109 mogrosides as substrates. **Related to Figure 5.** Flight-high-resolution electrospray  
 110 ionization-mass spectrometry (ESI-MS) parameters were as follows: the scan range was  
 111 100–1,500 m/z in positive ion mode, spray voltage was 4,500 V, capillary temperature  
 112 was 400°C, dry gas was 6 ml/min, dry temperature was 180°C and nebulizer pressure  
 113 was 1 bar.

MS1	... MDAQQ... GETTTI LMLPLWYGYGLS AFLELAKSLSRRNFHI YFCSTSVNLDAI KP KLP... SSFSDSI QFVELHLPLSSPEEPHHLHTT NGLPPTL	91
MS2	... MDSGYSSS YAAAAGHVVI CPWAF GELLPCLDLQRLASRGHRVSVSTPRNI SRLPPVR... PALAPLVAFAVLP LPRVEGLDGAETNDVPHDR	95
UGT91D2	... NATSDSI VDDRK QLVHATFPWAFGHILPYLQSKLIEKGHKVFLSTTRNI QRLS... SHI SPLI NVVQLTLP RVELEDAEATTDVHPED	91
UGT73P12	MDSFVGEVGDHQADTTVLKAVFLPFI SKSLEL RVVDKARI FAVHGVDVTI I TTPANA AAFQTSI DHDSSRSRSI KTHI WPPFPQVPLQOGFERL DADTPOH	100
Ole_D	... MTQTTP... AHI AMFSI AAHG VNP SLEVI RELVARGHRTYAI PPFVADKVA... ATGPRPVLVHSTLPGPDADAEAWGSLLDNRRRT	84
UGT91G16	... NAASASRP... LHLVI CPWAF GELLPCLEL AHRLASRGHRVSVSTPRNI SRLPPLR... PAVAPLVNFAVLP LAQVPLDGAETNDVGDGK	89
HV1	... MDGNSSS... PLBVI CPWAL GELLPCLDI AERLASRGHRVSVSTPRNI ARLPPLR... PAVAPLVDFVAVLP LPHVDGLDGAETNDVGYDK	90
Pn3-31	... MDI EK... GRI SI ALLPFLAHGHISPFLELAKLAKRRNCNIFLCSTPI NLSSI KNRVSD... KDSASIKLVELHLPLSSPEEPHHTT NGLPSHL	96
Pn3-32	... MDNQK... GRI SI ALLPFLAHGHISPFLELAKLAKRRNCNIFLCSTPI NLSSI KN... KDSASV KLV EHLPLSSPEEPHHTT NGLPSHL	90
Consensus	... h p	
MS1	... MPALHQAFA SAAQHEFESI C T L A P H... LLI YFSLQWVPRVSSSLKI PAI NFNTTGVFVI SQG... LHPI HYPHSKFFPSEFV	169
MS2	PDMVELHRRAFDGLAAPSEFIFGTACAD... WVI VVVFHWA AALAEHKVPCAMLLGS AHMI ASI... ADDRLEAETESPAAGGRR AAAP	184
UGT91D2	... I PYLKKASDGLQEVTRFLEQHS PD... WI IYTYHYMLPSI AASLGI SRAHFSVTTWAI AYGPSADAMI NGSDRRTTVE LDTTPKWF PF	181
UGT73P12	... LLPKI YQGLSI LQEQQQLREMRPD... FI VITMYPSPVDA AELGI PRLVNCNGS YFAQSAVN... SI ELFSQAKVDSNTETFTLLPGL	186
Ole_D	... FLNDAI QALPQADAYADDI P. D... LVLHDI TSYPARVLRVGVPAVLSPLNVAWKGYE... SIELSPQAKVDSNTETFTLLPGL	144
UGT91G16	... AELRLRASDGLAEPSEFIRAACKP... DWLI VDI VNNWAAA AAEHKVPCVLLQCAARVFTPL... AGP... VSEHAAA AVGKERPAEAP	155
HV1	... FELHKAFDGLAEPSEFIRAACAEAGGSRPDWLI VTFHFWAAA AVENKVPVCLLLGAATVI AGF... ARG... VSEHAAA AVGKERPAEAP	180
Pn3-31	... MVLKNAFETVGTTFSEI LKTLDPD... LLI YEFNPSWAPEI ALSHNI PAVYFLTSAATSSVA... LRALKNPGEKYPFPDFY	168
Pn3-32	... MVLKNAFETVGTTFSEI LKTLNDP... LLI YEFNPSWAPEI ASSHNI PAVYFLTSAATSSSI G... LHAFKNPGEKYPFPDFY	164
Consensus	... d a	
MS1	LHNHWKAMYSTADGAS. TERTKRGEAFLYLCHASCSVI LINSFRELE GKMYDLYSVLL... NKKVVPVGLVYEP... NQDGEDEGYSI KNWLD	258
MS2	TFEVARMKLI RTKG... SSGMSLAERFSLT LRSLSLVVGRSCVEFPETVP L LSTLR... GKPI TFLGLMPP LH... EGRR. EDGEDATVRWLD	268
UGT91D2	PTKVCKWRKHLARLVK. YKAPGI SDGYRMLV LKGS DCLLSKCYHEFGTQWLP LLET LH... QVPVVPVGLL PPEI P... GDEKDET VVSI KKWLD	270
UGT73P12	PHEVENTRQLQPDWLRGAPNEYTYLAKMI KDSERKS YGSLFNSYELGTYEEHYKKAAGTKS WS VGPVSLWNQDASDKACRGGVKEKGGDGVLLTWLD	286
Ole_D	VAEPPWREPRQTE... RGRAYARFEAWLKENGIT EHPDTFASHP PRSLVLI PKALQP. HADR VDE D VYTFV G... ACCGDRAEEGGQR	227
UGT91G16	SHNDWMKTYTVES.P... ASGVS I AERCAI TLKACKLAALRSCL EWPDAVPLVKTHIQ. SGT P V V T L G L L P P P P P... SADTRGKDDDDATVRWLD	246
HV1	SFETERRKLVTTQN... ASGMTVAERYFLTL MRS DLVAI RS CAWEPESVAAL T T L A... GKP V V P L G L L P P S P... EGGRGVS KEDA VRWLD	265
Pn3-31	DNSNI TPEPPSAD... KMLHDFVACFKRS CDI I L I K S F R E L E G K Y I D L S T L S... K K T L V P V G L V Q D P... L G H D E P K T G H I N W L D	251
Pn3-32	DNSNI TPEPPSAD... NKMLHDFI AC FERS CDI I L I K S F R E L E G K Y I D L S T L S... D K T L V P V G L V Q D P... M G H N E D P K T E Q I I N W L D	247
Consensus	... w	
MS1	KKEPSS TVVFSFCSYFPSKEMEELAHGLE. ASEVNF I WVVRFPQDNTSG. I EDALPKGFLERAGE. RGVVKGWAPQAKI KHVSTGGFVSLGCMN	354
MS2	AQPAKS VVYVALCSYVPLGVEKVFELALGLE. LAGTRFLMALRKP TGVS... DADLLPAGFEERTRG. RGVVATRVVQVSI AHAAGVAF LTHCGMN	361
UGT91D2	GKQKGS VVYVALCSYVPLGVEKVFELALGLE. LSGPLFVMA YRKP TGPAK. SDSVELPDGFVTRTRD. RGLVWTSWAPQLRI SHESVCGFLTHCGSG	365
UGT73P12	SKTEDSVLYVFSYVNFKNFKPTQVFI AHALE. DSGHDFI WYVGI EGE GGG... ADELREFEKKVKEKNGYLI VGVAPQLLI I EHPAVGAVTTCGMN	381
Ole_D	PAGAEKVVYVALCSYVPLGVEKVFELALGLE. LSGTRFLWALRRPDAADADADDVLP GFEEERTRG. RGLVVFVGVVQVSI LAHGAVAAFLTHCGMS	343
UGT91G16	AQPAKS VVYVALCSYVPLRAEQVHVALGLE. LSGARFLWALRKP TDAP... DAADVLP GFEEERTRG. RGLVVTGVPQI GVLAHGAVAAFLTHCGMN	358
HV1	KRAESTVVFVCSYFPSNEELVVALGLE. I SVVNF I LAVRFL EGEKGG... VLPEGFVORVGD. RGLVVEGVVAPQARI GHSS TGGFVSLTHCGMS	343
Pn3-31	KRAESTVVFVCSYFPSNEELVVALGLE. I SVVNF I LAVRFL EGEKGG... VLPEGFVORVGD. RGLVVEGVVAPQARI GHSS TGGFVSLTHCGMS	343
Pn3-32	KRAESTVVFVCSYFPSNEELVVALGLE. I SVVNF I WVRVLI EGEKGG... VLPEGFVORVGD. RGLVVEGVVAPQARI GHSS TGGFVSLTHCGMS	339
Consensus	... g s c w p q i h g	
MS1	SVMSVMFGVPI I GVPVH. VDQFNAGLVEEAGVGVAEKRDP... DGKI QRDEVAKLI KEVVVEKTR. EDVRKKARENSEI LRSKGEK... 438	
MS2	STIEGLMFGVPI I MLP I F. GDQGP NARLI EAKNAGLQVARNDG... DGSFDREGVAAAI RAVAVEEESKVFAKAKLQEI VADVACHERY... 449	
UGT91D2	SIIEGLMFGVPI I MLP I F. GDQPLNARLLEDKCVGI EI PRNEE... DGCLTKESVARSLSVVEKEG. EI YKANARELSKI YNDTKVEKEY... 452	
UGT73P12	TVMSVNASLPLATVPLFAEYFQFNEKLVVDVVKI GVPVGVKEWRNNE... FGDVVKRREI GKAI AFLVGGDESELEVRKRVKLSGATKKAIVQDGSY... 407	
Ole_D	GSQGLATATPM AVPQA. VDQFGNADMLQGLGVARKLATEEATADLLRETALALVDDPEVARRLRRI QAEMAOEGGT RRAADLI EAETPARHERQEP... 479	
UGT91G16	STIEGLRFGVPI I VMLP I ATGDQWP NARLMEERGVL RVP R DGN... DGSFHREGLAATVRAVTA DQGG... TFAANAGKLVVADRECHERC... 430	
HV1	STIEGLLFGVPI I MLP I S. SDQGP NARLMEGRKVGWQVRDES... DGSFRREDVAATVRAVVEEDGRRVFTANAKM QEI VADGACHERC... 446	
Pn3-31	SIIEGSMKFGVPI I AVARH. LDQPLNAKLA AEVGVGMEVVRDE... NGKYKREAI AEVI RAVVMEKNG. EVI RRRARELSSEKVKETGQE... 427	
Pn3-32	SIIEGSMKFGVPI I AVARH. LDQPLNAKLA AEVGVGMEVVRDE... NGKYKREAI AEVI RAVVMEKSG. EVI RRRARELSSEKMKKEGQE... 423	
Consensus	... e p	
MS1	... FDENVAEI SLLLKI... 452	
MS2	... I DGF I QQLRSYKD... 462	
UGT91D2	... VSQFVDYLEKNARAVAI DHES... 473	
UGT73P12	TKLKELI EELKSI KLQKVNNKLMEAV... 505	
Ole_D	... VGD RPNVGD RPA G VRS DRQRSAL... 430	
UGT91G16	... I DGF I QQLTSYMD... 443	
HV1	... I DGF I QQLRSYKA... 459	
Pn3-31	... I GRAVEELVQI CKMKKDAQY... 447	
Pn3-32	... I DRAVEELVQI CKMKKDAQ... 442	
Consensus		

114

115 **Figure S6.** Multiple sequence alignment of 9 UGTs that show branched glycosylation  
116 activity to triterpenes. **Related to Figure 4.** The following UGTs were used in this  
117 analysis: MS1, MS2 (GenBank: XP\_015629141), UGT91D2 (GenBank: ACE87855.1),  
118 UGT73P12 (GenBank: BBN60799.1), OleD (GenBank: ABA42119.2), UGT91G16  
119 (GenBank: QHG10987.1), HV1 (GenBank: BAJ98242.1), Pn3-31 (GenBank:  
120 QOJ43868.1) and Pn3-32 (GenBank: QOJ43866.1).

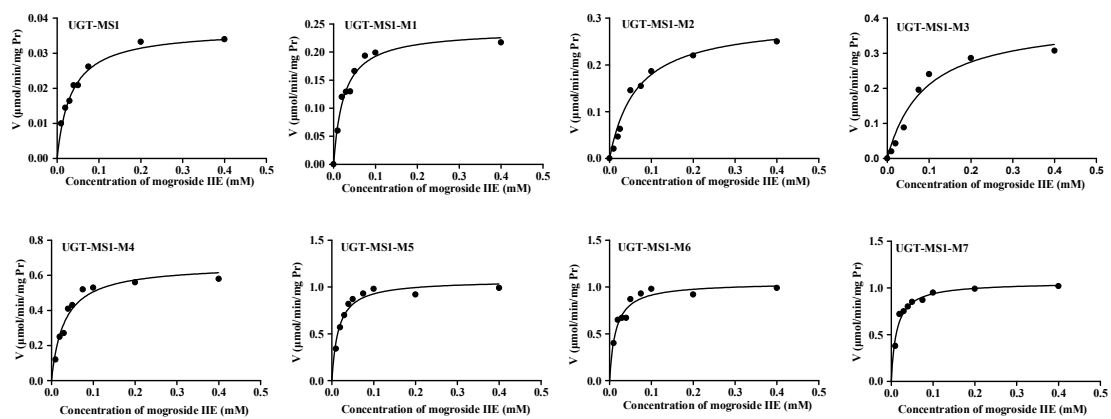


121

122 **Figure S7.** Determination of kinetic parameters for recombinant glycosyltransferase

123 MG1 (on IE) and MS1 and MS2 (on mogroside IIE). **Related to Table 1.** The catalytic

124 constants were obtained using GraphPad Prism 5 software.

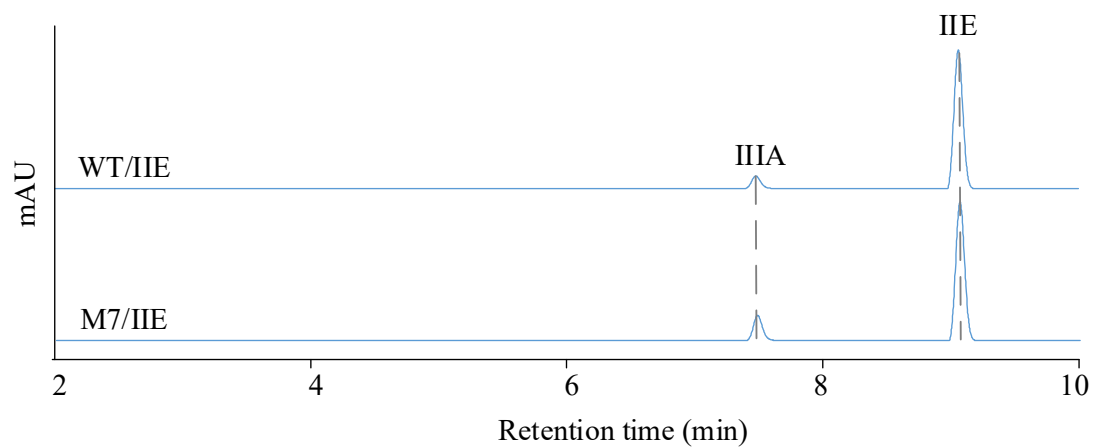


125

126

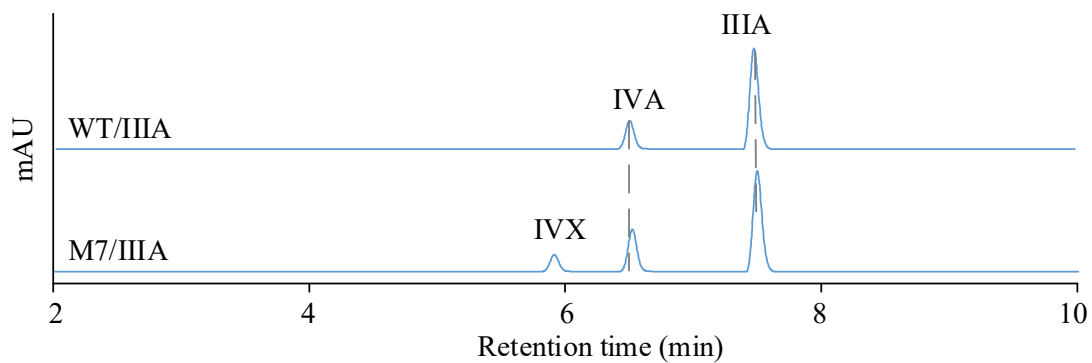
127

**Figure S8.** Determination of kinetic parameters for recombinant glycosyltransferase MS1, and its mutants on mogroside IIE. **Related to Table 1.**



128

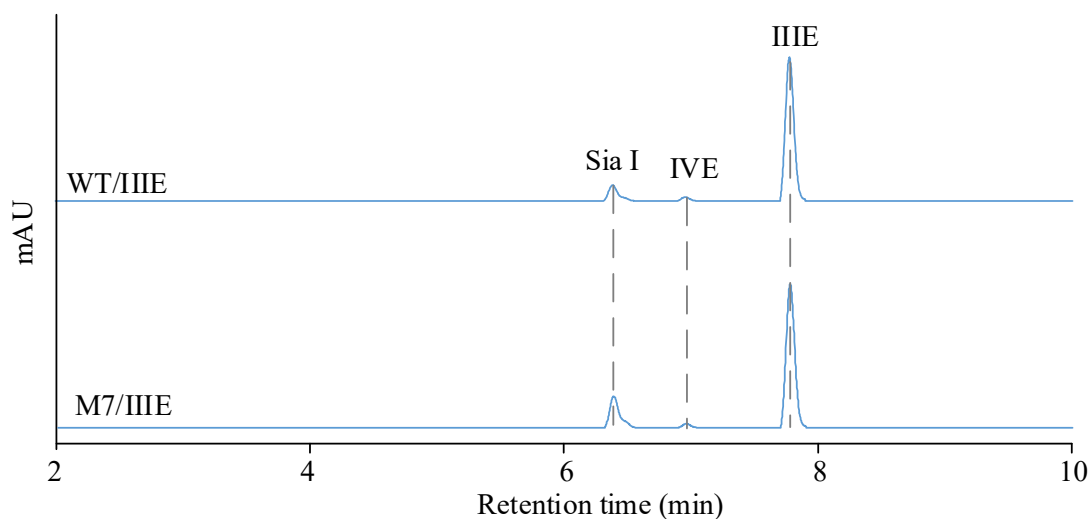
129 **Figure S9.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products  
130 using mogroside IIE as substrate. **Related to Figure 5.**



131

132 **Figure S10.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products  
133 using mogroside IIIA as substrate. **Related to Figure 5.**

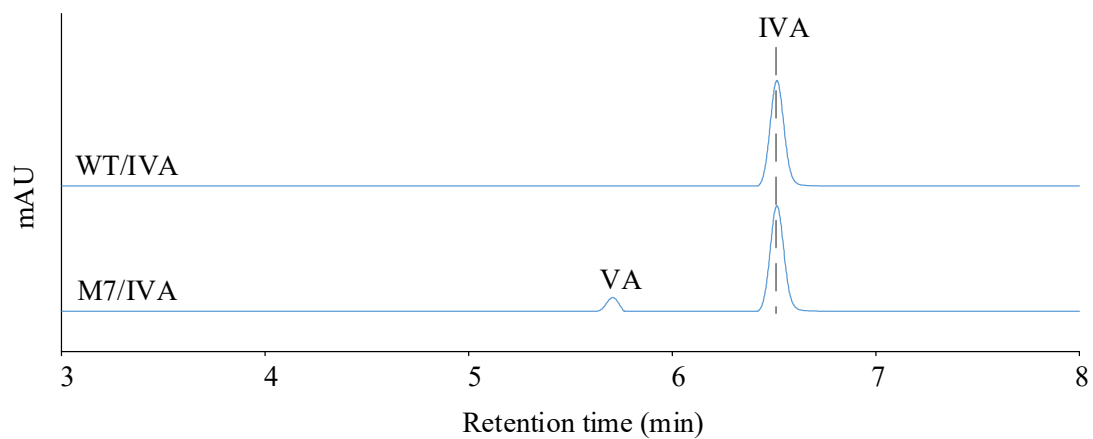




134

135 **Figure S11.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products

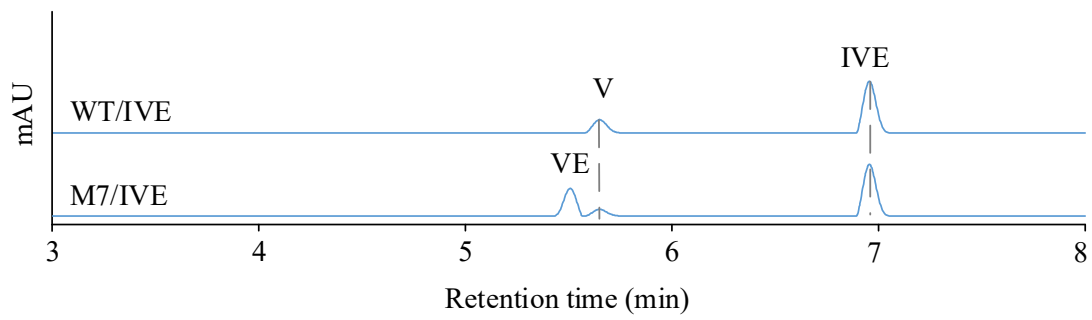
136 using mogroside III E as substrate. **Related to Figure 5.**



137

138 **Figure S12.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products

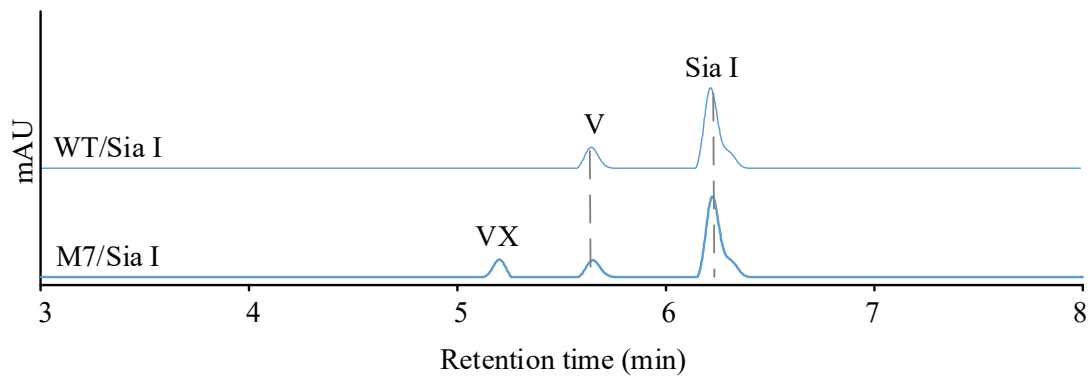
139 using mogroside IVA as substrate. **Related to Figure 5.**



140

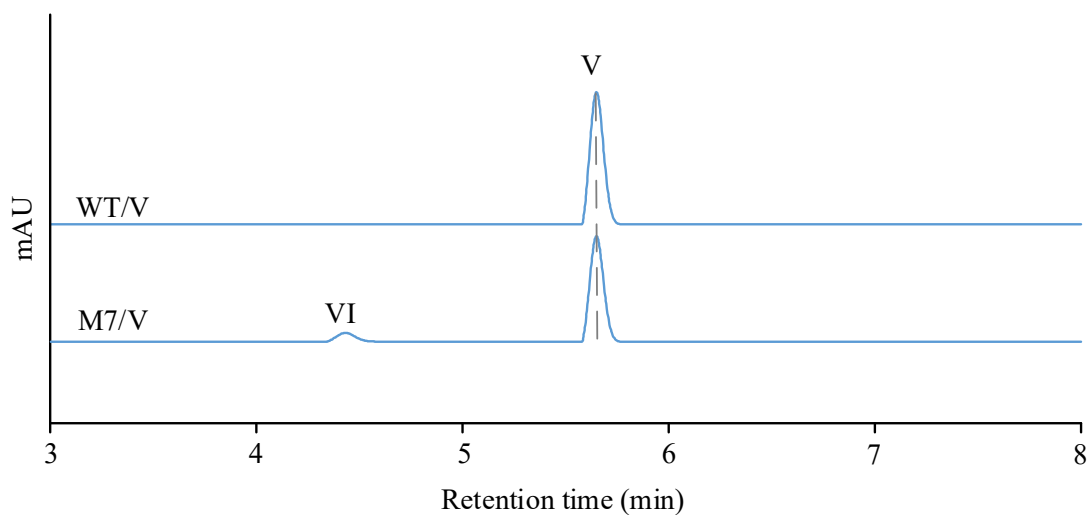
141 **Figure S13.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products

142 using mogroside IVE as substrate. **Related to Figure 5.**



143

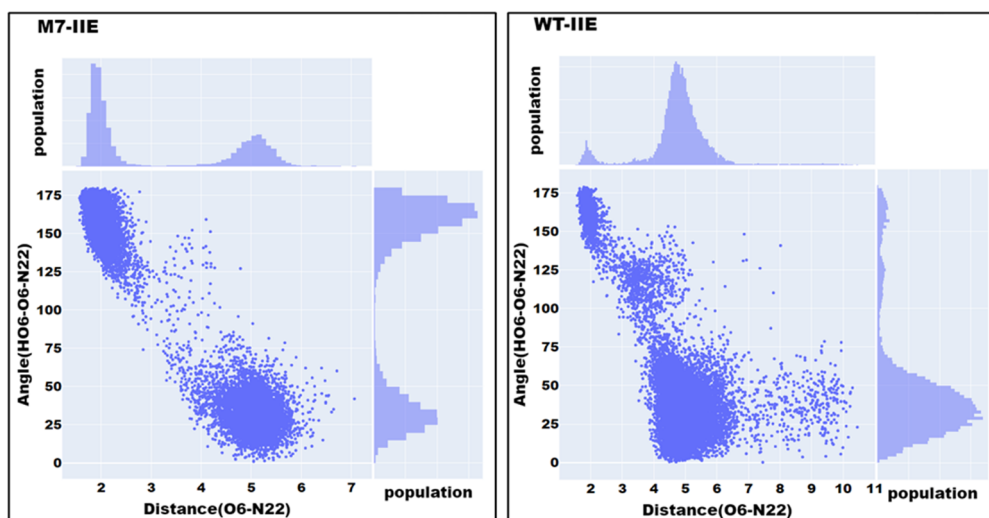
144 **Figure S14.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products  
145 using mogroside Sia I as substrate. **Related to Figure 5.**



146

147 **Figure S15.** HPLC analysis of MS1 WT and mutant M7 enzymatic reaction products

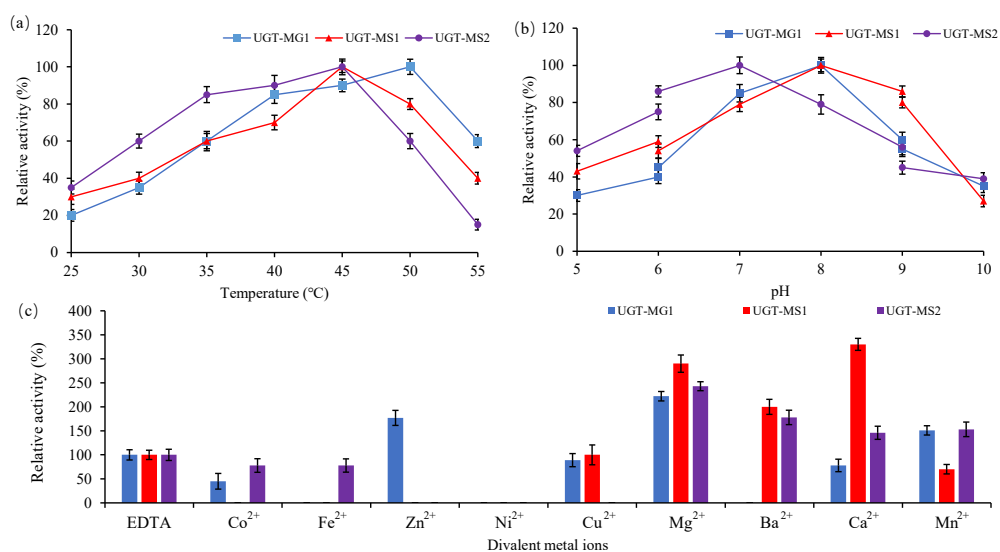
148 using mogroside V as substrate. **Related to Figure 5.**



149

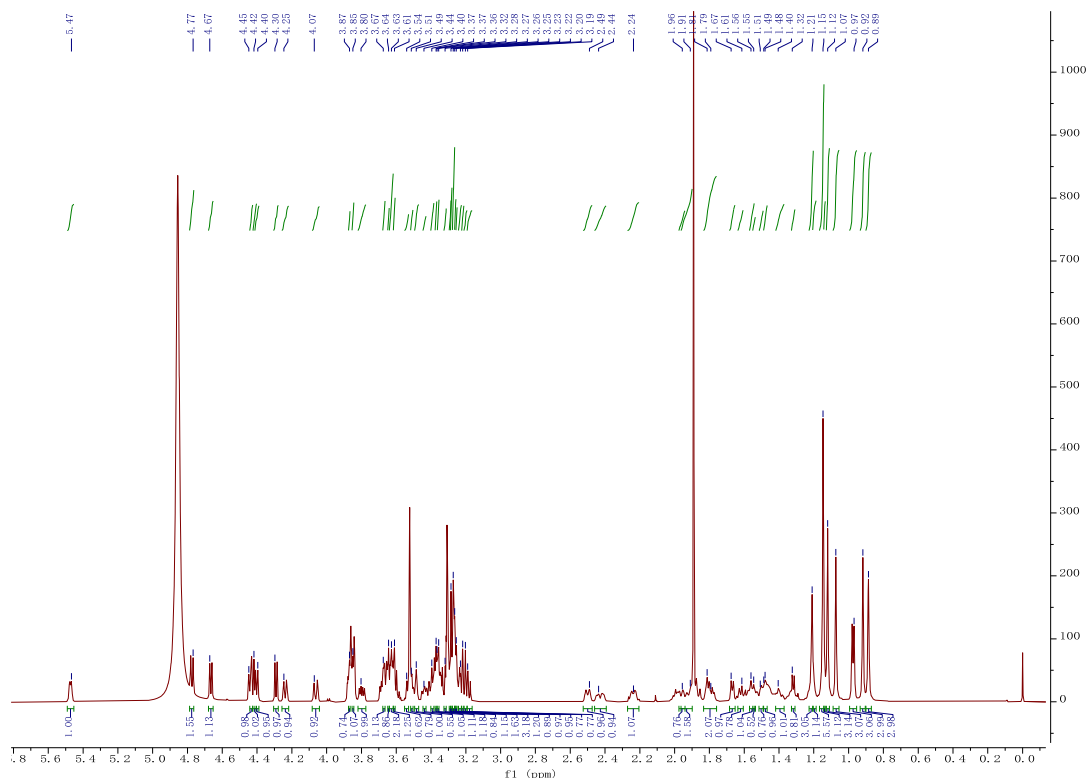
150

151 **Figure S16.** Conformation maps of M7-IIE and WT-IIE complexes during the 3\*50 ns  
 152 MD simulations. Related to Figure 6. According to reaction mechanism, a catalytic  
 153 conformation, which supports catalysis, occurs when the distance between the O<sub>6</sub> atom  
 154 of C24-sugar and the N<sub>22</sub> nitrogen of the catalytic residue His22 is less than 3.6 Å and  
 155 the angle of H<sub>o6</sub>-O<sub>6</sub>-N<sub>22</sub> is larger than 135°



156

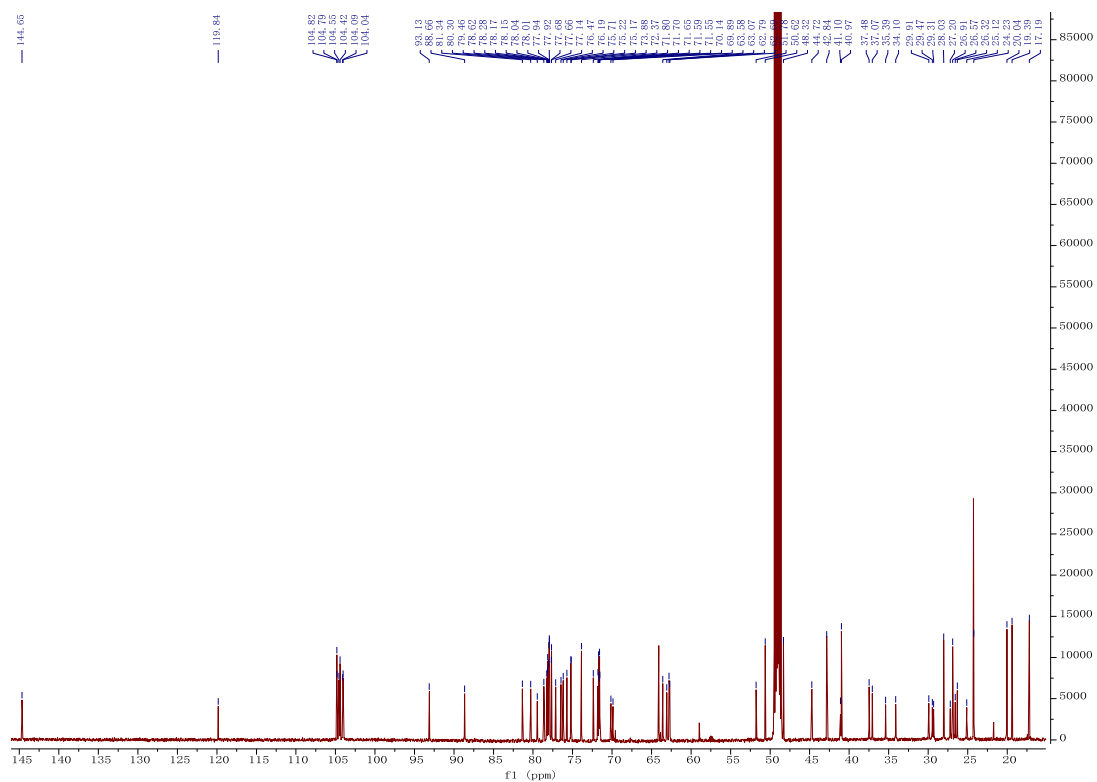
157 **Figure S17.** Optimization of reaction temperature, pH and divalent metal ions of MS1  
 158 and MS2. **Related to Figure 7.** (a) Relative activity of MS1 and MS2 at various  
 159 temperatures (25–60°C) at pH 8.0. (b) Relative activity of MS1 and MS2 at different  
 160 pH conditions (pH 5.0–10.0). (c) Relative activity of MS1 and MS2 with divalent metal  
 161 ions.



162

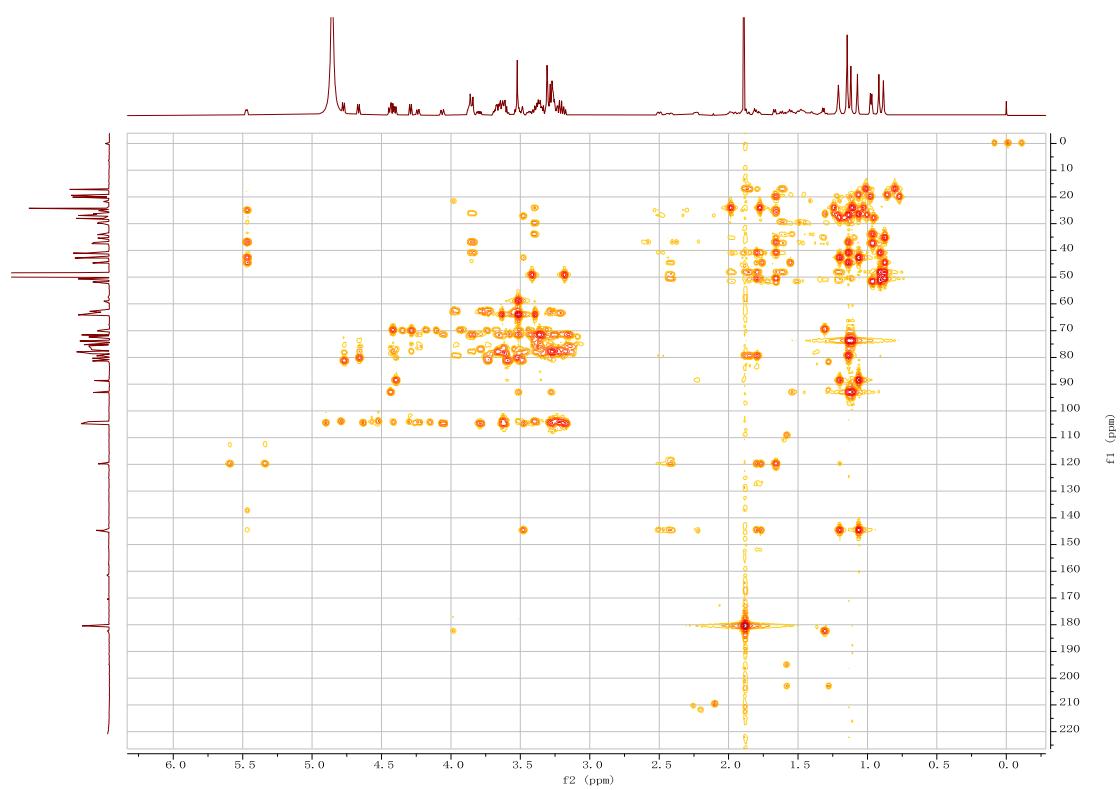
163 **Figure S18.**  $^1\text{H}$  NMR spectrum of VI in Dimethyl sulfoxide- $d_6$  (600 MHz). **Related**  
 164 **to Figure 5.** Glycosides VI was purified by an Agilent 1260 preparative HPLC system  
 165 with semi-preparative C18 reverse-phase column ( $21.2 \times 250$  mm,  $5 \mu\text{m}$  particles, Welch,  
 166 Shanghai, China). The fractions with the same mogroside glucoside were collected and  
 167 concentrated on rotary evaporator. The purified product was dissolved in deuterated  
 168 methanol, after vacuum freeze drying. 1D NMR and 2D NMR spectroscopies of the product,  
 169 including  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR, correlation spectroscopy (COSY), total correlation  
 170 spectroscopy (TOCSY), heteronuclear single quantum coherence (HSQC), and heteronuclear  
 171 multiple bond correlation (HMBC), were recorded on an Advance DMX-600 NMR  
 172 spectrometer (Bruker, Karlsruhe, Germany), to identify the chemical structure of the  
 173 compound.





174

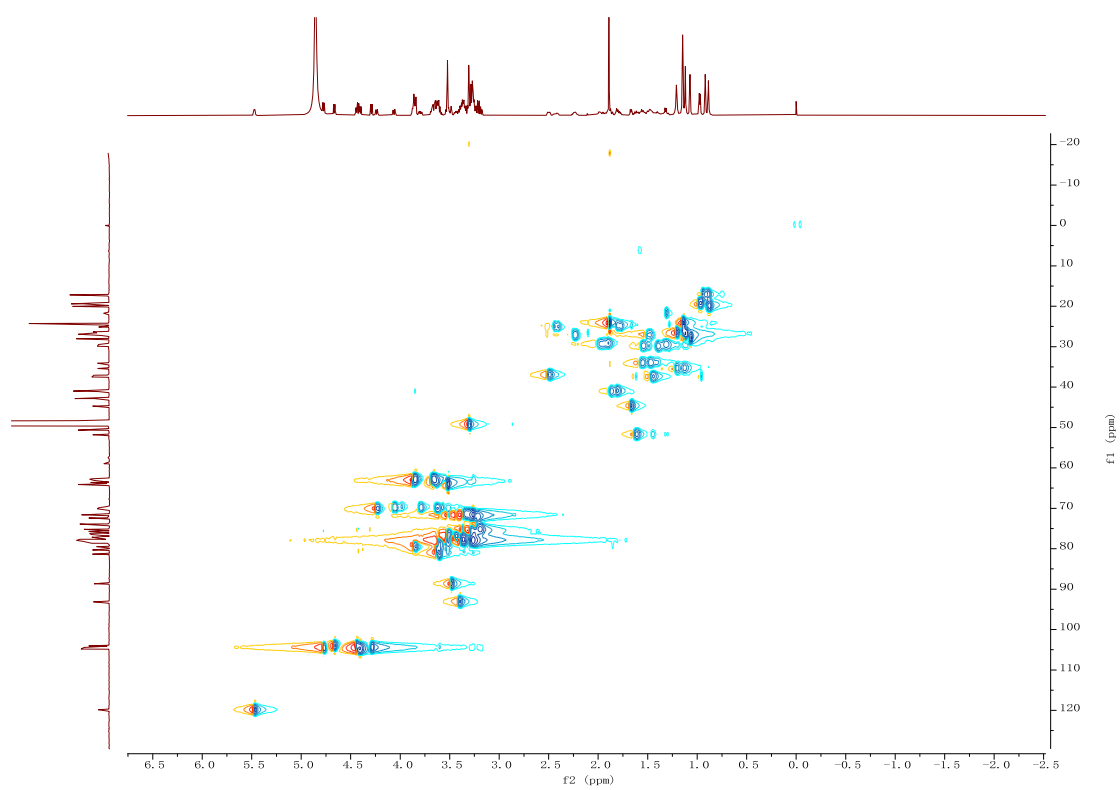
175 **Figure S19.**  $^{13}\text{C}$  NMR spectrum of VI in Dimethyl sulfoxide- $d_6$  (600 MHz). **Related**  
 176 **to Figure 5.**



177

178 **Figure S20.** HMBC spectrum of VI in Dimethyl sulfoxide-d6 (600 MHz). Related to

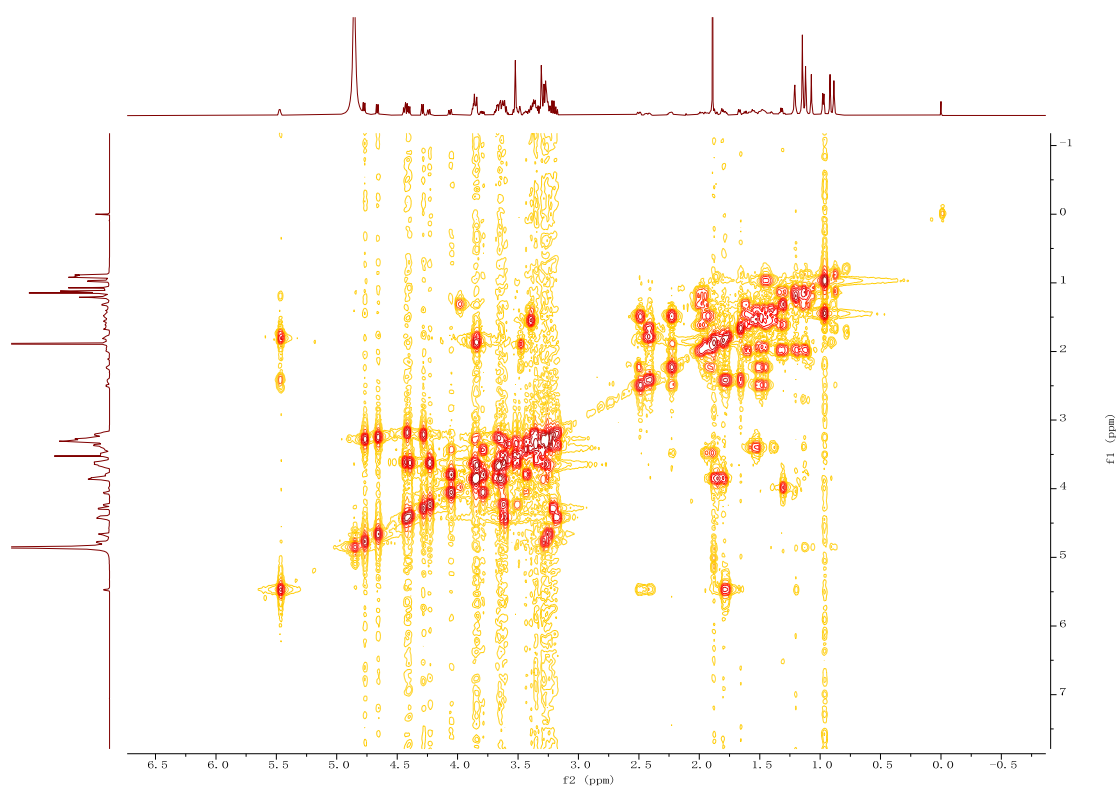
179 **Figure 5.**



180

181 **Figure S21.** HSQC spectrum of VI in Dimethyl sulfoxide-d<sub>6</sub> (600 MHz). Related to

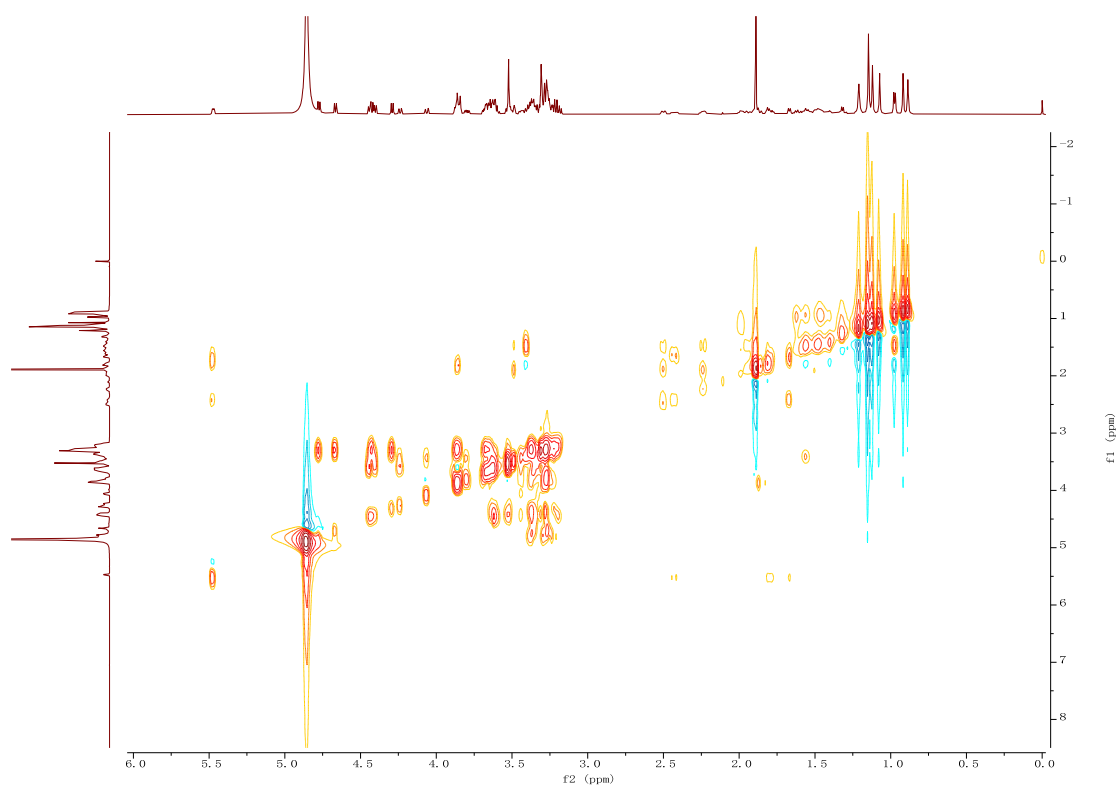
182 **Figure 5.**



183

184 **Figure S22.** COSY spectrum of VI in Dimethyl sulfoxide-d<sub>6</sub> (600 MHz). Related to

185 **Figure 5.**



186

187 **Figure S23.** TOCSY spectrum of VI in Dimethyl sulfoxide-d6 (600 MHz). **Related to**

188 **Figure 5.**

```

>MG1
GGATCCATGGAAAAGGGTGACACTCACATCTTGGTTTTCCCATCCAGCTCAAGGTCACATTAACCCATTGTTGCAATTGTCTAAGCAC
TTGATCGCTAAGGGTATTAAGGTTTCTTTGGTCACTACCTTGCACGCTCTAACAGAATGCAATTGCAAGGTGCTTACTCTAACTCTGTT
AAGATCGAAGTCATTTCTGACGGTCTGAAGACAGATTGGAACACTGACACCTTGAGACAATACTGGACAGATTGACAAAAAGATGACC
AAGAACTTGGAAAGACTTCTTGCAAAAAGGCTATGGTTTTCTTAACCCACAAAGTTCATCATCTACGACTCTACTATGCCATGGGTTTTG
GAAGTCGCTAAGGAATTCGGTTTTGGACAGAGCTCCATTCTACACCCAATCTTGTGCTTTGAACCTATCAACTACCACGTTTTGCACGGT
CAATTGAAGTTGCCACCAGAACTCCAACCATTTCTTTGCCATCTATGCCATTGTTAAGACCATCTGACTTGGCAGCTTACGACTTCGCAC
CCAGCTTCTACTGACACCATCATTGACTTGTGACTTCTCAATACTCTAACATCCAAGACGCTAACTTGTGTTCTGTAACACTTTGCAC
AAGTTGGAAGGTGAAATCATTCAATGGATGGAACCTTGGGTAGACCAAGTTAAGACTGTCGGTCCAACCGTCCATCTGCTTACTTGGAC
AAGAGAGTCGAAAACGACAAGCACTACGGTTTGTCTTTGTTCAAGCCAAACGAAGACGTTTGTGTAAGTGGTTGGACTTAAGCCATCT
GGTCTGTTTTGTACGCTCTTACGGTCTTTGGTCAAATGGGTGAAGAACAATTGAAGGAATTGGCTTTGGGTATTAAGGAAACTGGT
AAATCTTCTGTGGGTTGTGACAGACACCGAAGCTGAAAAGTTGCCACCAAACTTCGTTGAATCTGTCGCTGAAAAGGGTTTTGGTTGTC
TCTTGGTGTCTCAATTGGAAGTTTTGGCTCACCCATCTGTCGGTTGTTTCTCACTACTGTGGTTGGAACCTACCTTGGAAAGCTTTG
TGTTTTGGGTGTTCCAGTTGTCGCTTTCCCAATGGGCTGACCAAGTCACTAACGCTAAGTCTTGGAAAGACGTTTGGAAAGTCGGTAAA
AGAGTTAAGAGAAACGAACAAAGATTGGCTTCTAAGGAAGAAGTTAGATCTTGTATCTGGGAAGTCATGGAAGGTGAAAAGAGCTTCTGAA
TTCAAGTCTAACTCTATGGAATGGAAGAAGTGGGCTAAGGAAGCTGTTGACGAAGGTGGTCTTCTGACAAGAACATTGAAGAATTCGTC
GCTATGTTGAAGCAAACtaaCTCGAG
>MS1
GGATCCATGGACGCTGCTCAACAAGGTGACACTACCCTATTTGATGTTGCCATGGTTGGGTTACGGTCACTGTCTGCTTTCTTGAA
TTGGCTAAGTCTTTGTCTAGAAAGAACTTCCACATCTACTTCTGTCTACCTCTGTTAACTTGGACGCTATTAAGCCAAAGTTGCCATCT
TCTTTCTCTGACTCTATCCAATTCGTCGAATTGCACCTTGGCATCTTCTCCAGAATTGCCACCACACTTGCACACCCTAACGGTTTTGCCA
CCAACTTGTATGCCAGCTTTGCACCAAGCTTTCTCTATGGCTGCTCAACACTTCGAATCTATCTTGCAAACTTTGGCTCCACACTTGTG
ATCTACGACTCTTTGCAACCATGGGCTCCAAGAGTTGCTTCTTCTTGAAGATCCAGCTATTAACCTTCAACACCCTGGTGTCTTCGTT
ATCTCTCAAGGTTTGACCCCAATTACTACCCACACTCTAAGTTCCTTCTGCAATTCGTTGCAACAACCCTGGAAGGCTATGTAC
TCTACCGCTGACGGTGCTTCTACCGAAAGAAGTGAAGAAGAGAGGTGAAGCTTTCTTGTACTGTTTGCACGCTTCTTGTCTGTTATCTTG
ATTAACCTTTTCAAGAAATTGGAAGGTAAATACATGGACTACTTGTCTGTTGTTGAACAAGAAGGTTGTCACGCTGGTCCATTGGTT
TACGAACCAACCAAGACGGTGAAGACGAAGGTTACTTCTTATTAAGAAGTGGTTGGACAAGAAGGAACCATCTTCTACTGTCTTCGTT
TCTTTCCGTTCTGAATACTTCCCATCTAAGGAAGAAATGGAAGAAATGCTCACGGTTTGGAAAGCTTCTGAAGTAACTTCAITCGGGT
GTCAGATTTCCACAAGGTGACAACCTCTGGTATTGAAGACGCTTTGCCAAAAGGGTTTCTTGGAAAGAGCTGGTGAAAAGAGGTATGGTT
GTCAAGGGTTGGGCTCCACAAGCTAAGATCTTGAAGCACTGGTCTACTGGTGGTTTCGCTCTCACTGTGGTTGAACTCTGTTATGGAA
TCTATGATGTTCCGTGTCCCAATCATTGGTGTCCCAATGCAGGTTGACCAACCATTCAACGCTGGTTTGGTTGAAGAAGCTGGTGTCCGT
GTTGAAGCTAAGAGAGATCCAGACGGTAAAATCCAAAGAGACGAAGTTGCTAAGTTGATTAAGGAAGTTGTCGTTGAAAAGACTAGAGAA
GACGTCAGAAAAGAAGGCTAGAGAAATGCTGAAATCTTGAGATCTAAGGGTGAAGAAAAGTTTCGACGAATGGTTGCTGAAATCTCTTTG
TTGTTGAAGATTtaaCTCGAG

```

189

190 **Figure S24.** The gene sequences of MG1 and MS1, which were synthesized by  
 191 GenScript (Nanjing, China) with codon optimization according to *S. cerevisiae*. **Related**  
 192 **to Figures 1 and 2.**

193