

# **A Novel 3-*O*-rhamnoside: 2''-*O*-xylosyltransferase Responsible for Terminal Modification of Prenylflavonol Glycosides in *Epimedium pubescens* Maxim.**

Yu Yao <sup>1</sup>, Jiajun Gu <sup>1,2</sup>, Yanjiao Luo <sup>1</sup>, Yixin Zhang <sup>1</sup>, Yuanyue Wang <sup>1</sup>, Yongzhen Pang <sup>3</sup>, Shangang Jia <sup>4</sup>, Chaoqun Xu <sup>1</sup>, Doudou Li <sup>1</sup>, Fengmei Suo <sup>1</sup>, Guoan Shen <sup>1,\*</sup> and Baolin Guo <sup>1,\*</sup>

<sup>1</sup> Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicines, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China

<sup>2</sup> School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

<sup>3</sup> Institute of Animal Sciences, The Chinese Academy of Agricultural Sciences, Beijing 100193, China

<sup>4</sup> College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China

\* Correspondence: gashen@implad.ac.cn (G.S.); blguo@implad.ac.cn (B.G.); Tel.: +86-10-57833172 (G.S. & B.G.)

## 1 Supplementary Figure Legends

Figure S1. Multiple sequence alignment of EpF3R2''XylT and other known plant flavonoid GT proteins. PSPG (the plant secondary product glycosyltransferase) domain was marked with red box.

Figure S2. Agarose gel electrophoresis analysis of DNA fragments amplified by PCR. Lane 1: DNA fragments amplified by first round PCR; Lane 2: DNA fragments amplified by second round PCR.

Figure S3. SDS-PAGE of recombinant EpF3R2''XylT proteins. Lane M: Protein ladder; Lane 1: Purified recombinant proteins.

Figure S4. Chemical structures of flavonol substrates used in the enzymatic assays of recombinant EpF3R2''XylT proteins.

Figure S5. Complete UPLC/ESI-MS spectra of flavonoid substrates and products in the enzymatic assays of recombinant EpF3R2''XylT proteins.

Figure S6. The UPLC chromatographs of reaction catalyzed by recombinant EpF3R2''XylT proteins with prenylflavonols as substrates, which showed no activity. Glc: glucose; Rha: rhamnose; Xyl: xylose. Prenylflavonols were marked with red arrow.

Figure S7. The UPLC chromatographs of reaction catalyzed by recombinant EpF3R2''XylT proteins with flavonols as substrates, which showed no activity. Glc: glucose; Rha: rhamnose; Xyl: xylose. Flavonols was marked with red arrow.

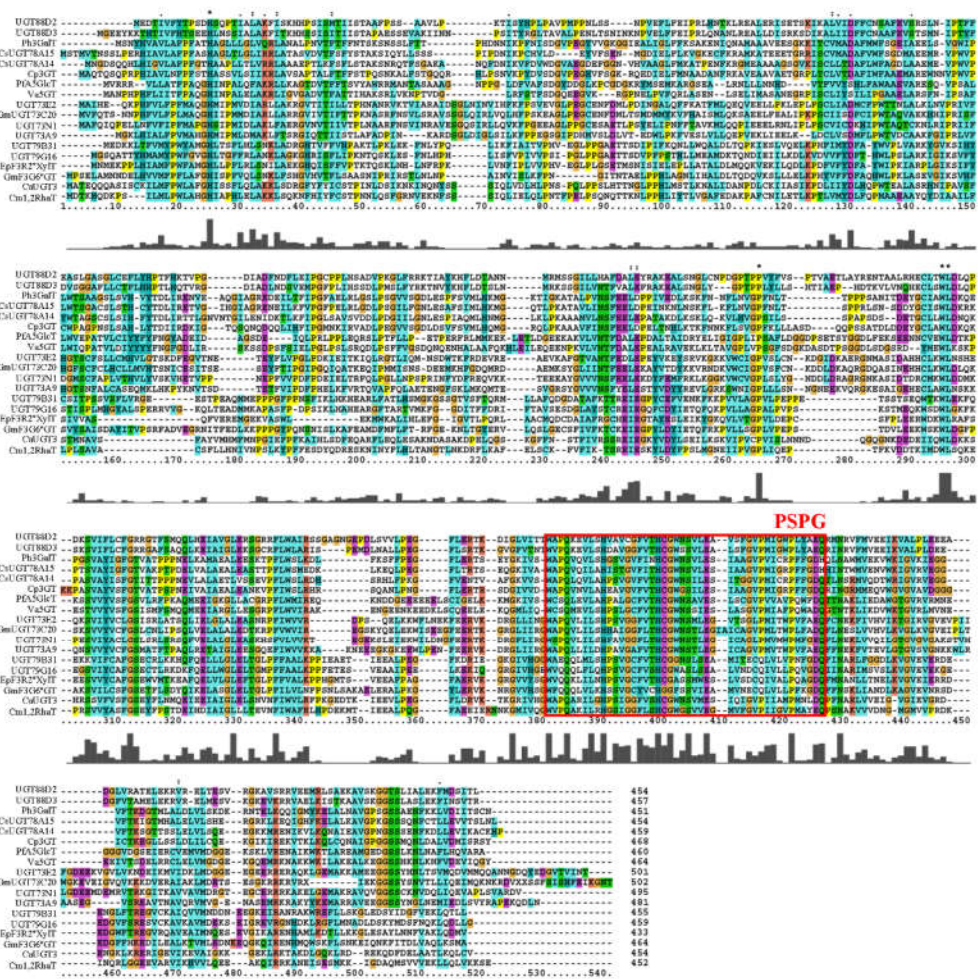
Figure S8. The UPLC chromatographs of reaction catalyzed by recombinant EpF3R2''XylT proteins with flavonol glycosides as substrates, which showed no activity. Glc: glucose; Rha: rhamnose; Xyl: xylose. Flavonol glycosides was marked with red arrow.

Figure S9.  $^1\text{H}$  NMR spectrum and  $^{13}\text{C}$  NMR spectrum of **1a** in DMSO- $d_6$ . (a)  $^1\text{H}$  NMR spectrum of **1a**. (b)  $^{13}\text{C}$  NMR spectrum of **1a**.

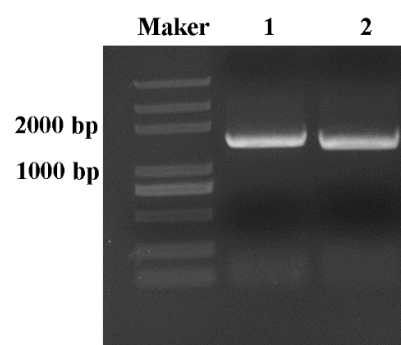
Figure S10.  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum for **1a**. The red arrow highlighted cross-peaks between H-2'' ( $\delta_{\text{H}}$  4.09) and C-2'' ( $\delta_{\text{C}}$  81.21).

Figure S11.  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum for **1a**. The red arrow highlighted cross-peaks between H-2'' ( $\delta_{\text{H}}$  4.09) and H-1''' ( $\delta_{\text{H}}$  4.22). The green line on the compound structure shows the relationship between the hydrogen bonds.

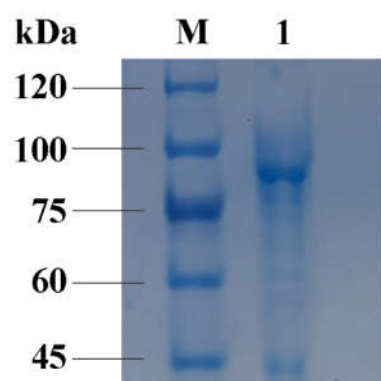
### Figure S1



**Figure S2**

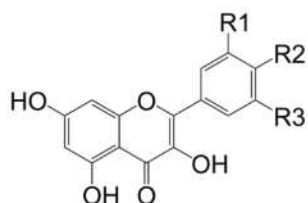


**Figure S3**



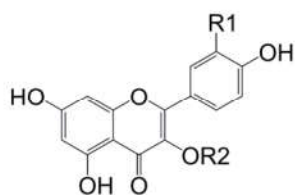
**Figure S4**

Flavonols



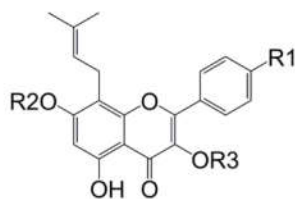
R1=H, R2=OH, R3=H, Kaempferol  
 R1=OH, R2=OH, R3=H, Quercetin  
 R1=OH, R2=OH, R3=OH, Myricetin  
 R1=H, R2=OCH<sub>3</sub>, R3=H, Kaempferide  
 R1=OH, R2=OCH<sub>3</sub>, R3=H, Tamarixetin  
 R1=OH, R2=OCH<sub>3</sub>, R3=OH, Mearnssetin  
 R1=OCH<sub>3</sub>, R2=OH, R3=H, Isorhamnetin

Favonol glycosides



R1= H, R2=Glc, Kaempferide-3-O-glucoside  
 R1= H, R2=Rha, Kaempferide-3-O-rhamnoside  
 R1= H, R2=Gal, Kaempferide-3-O-galactoside  
 R1= OH, R2=Glc, Quercetin-3-O-glucoside  
 R1=OH, R2=Rha, Quercetin-3-O-rhamnoside  
 R1=OH, R2=Gal, Quercetin-3-O-galactoside  
 R1= OCH<sub>3</sub>, R2=Glc, Isorhamnetin-3-O-glucoside  
 R1= OCH<sub>3</sub>, R2=Rha, Isorhamnetin-3-O-rhamnoside

8-prenylflavonol glycosides



R1= OCH<sub>3</sub>, R2=Glc, R3=Rha, Icaritin  
 R1= OCH<sub>3</sub>, R2=H, R3=Rha, Baohuoside I  
 R1= OH, R2=H, R3=Rha, Baohuoside II  
 R1= OH, R2=Glc, R3=Rha, Epimedeside A  
 R1= OCH<sub>3</sub>, R2=Glc, R3=H, Icariside I  
 R1= OCH<sub>3</sub>, R2=Glc, R3=Rha<sup>2</sup>-Glc, Epimedin A  
 R1= OCH<sub>3</sub>, R2=Glc, R3=Rha<sup>2</sup>-Xyl, Epimedin B  
 R1= OCH<sub>3</sub>, R2=Glc, R3=Rha<sup>2</sup>-Rha, Epimedin C

Figure S5

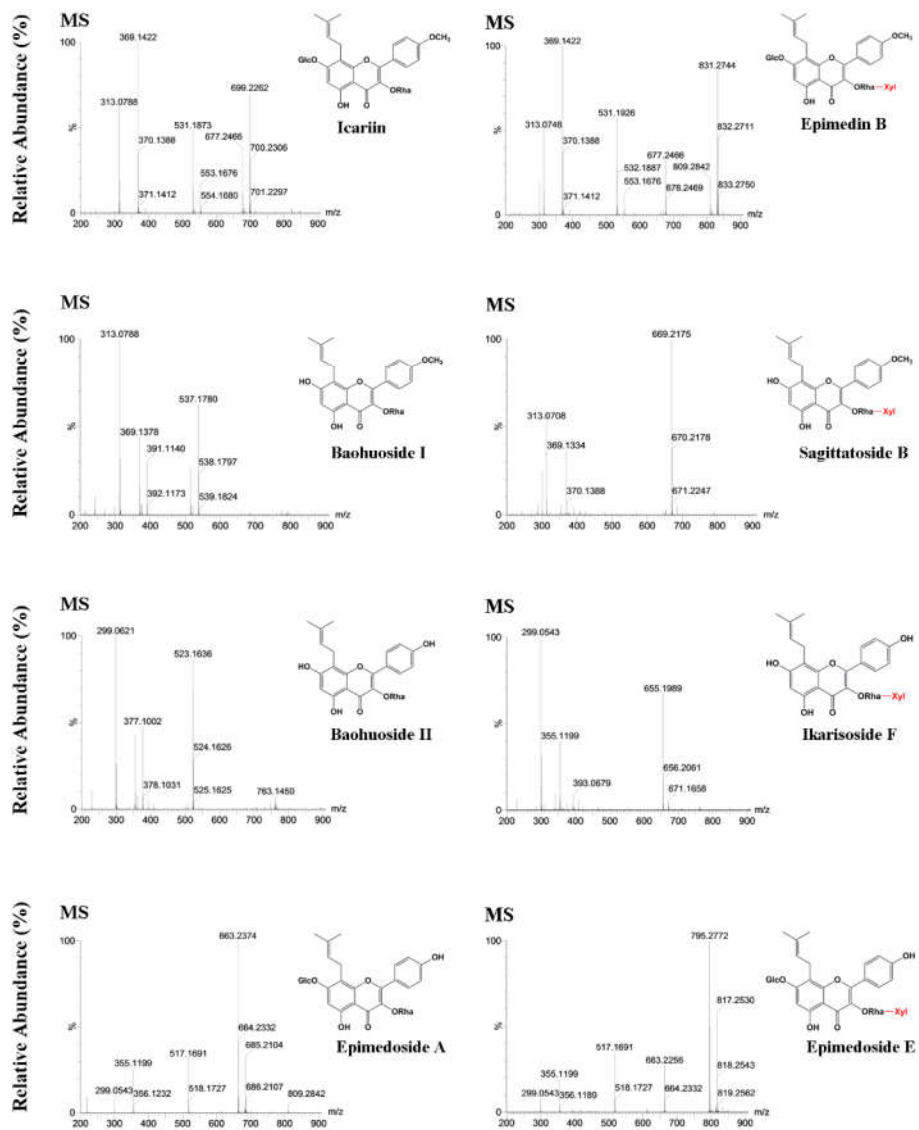


Figure S6

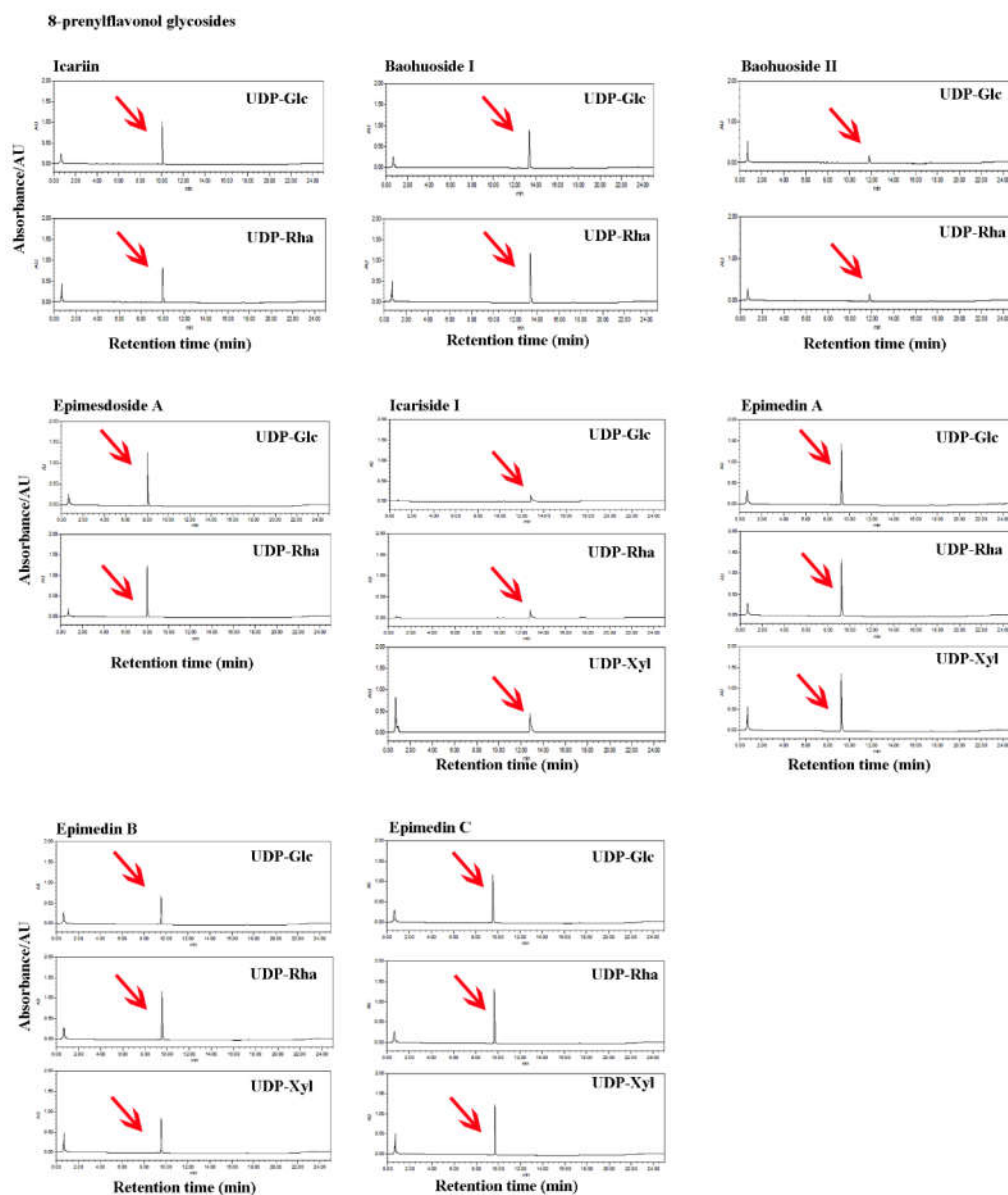


Figure S7

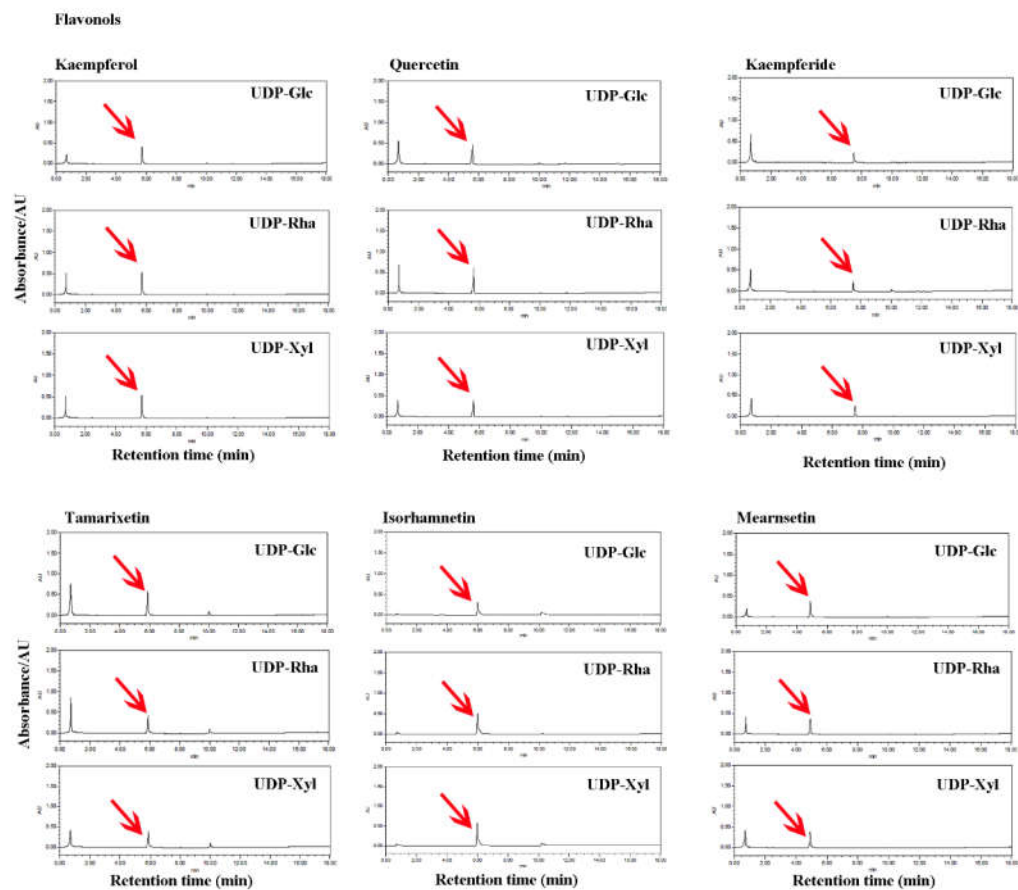
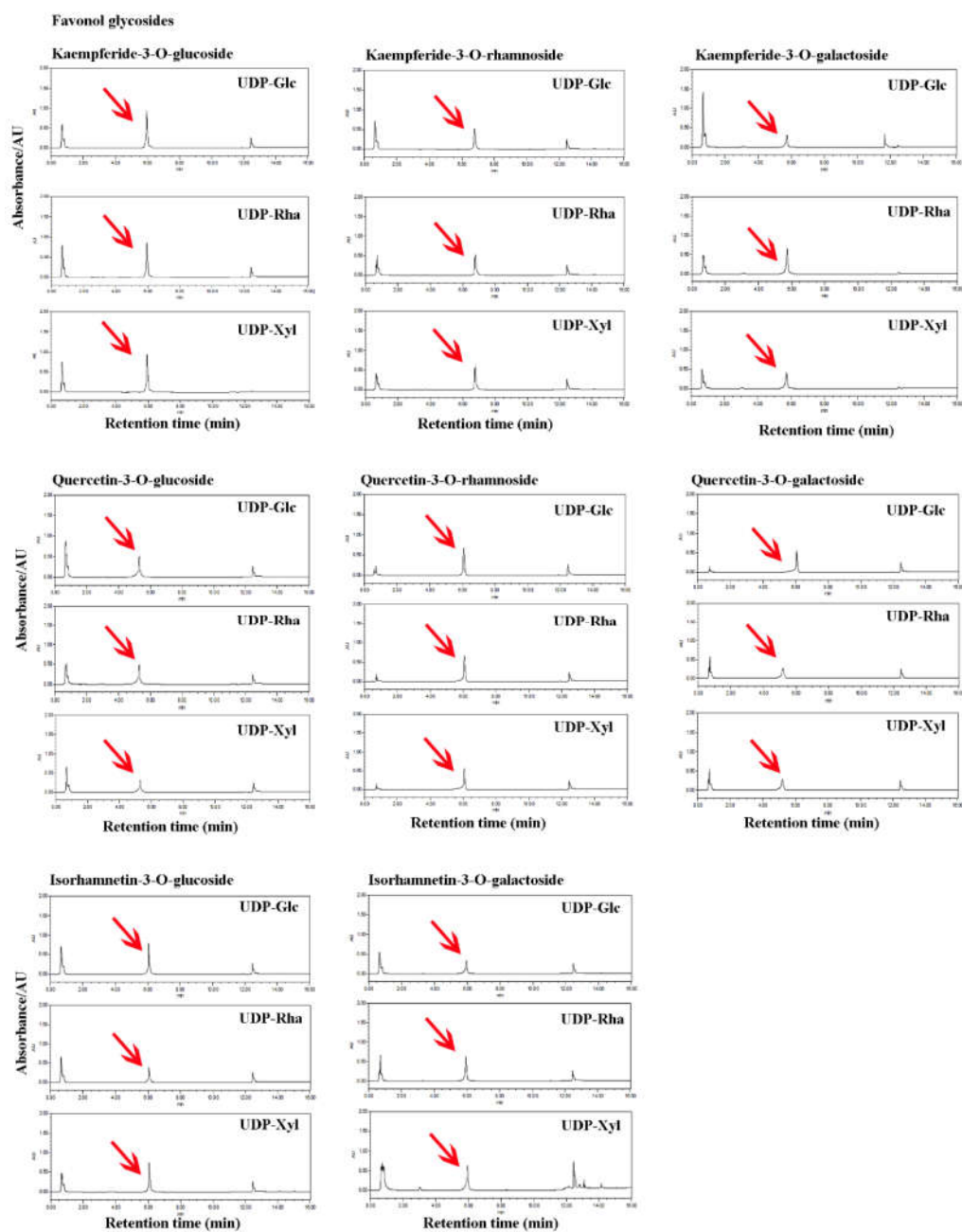
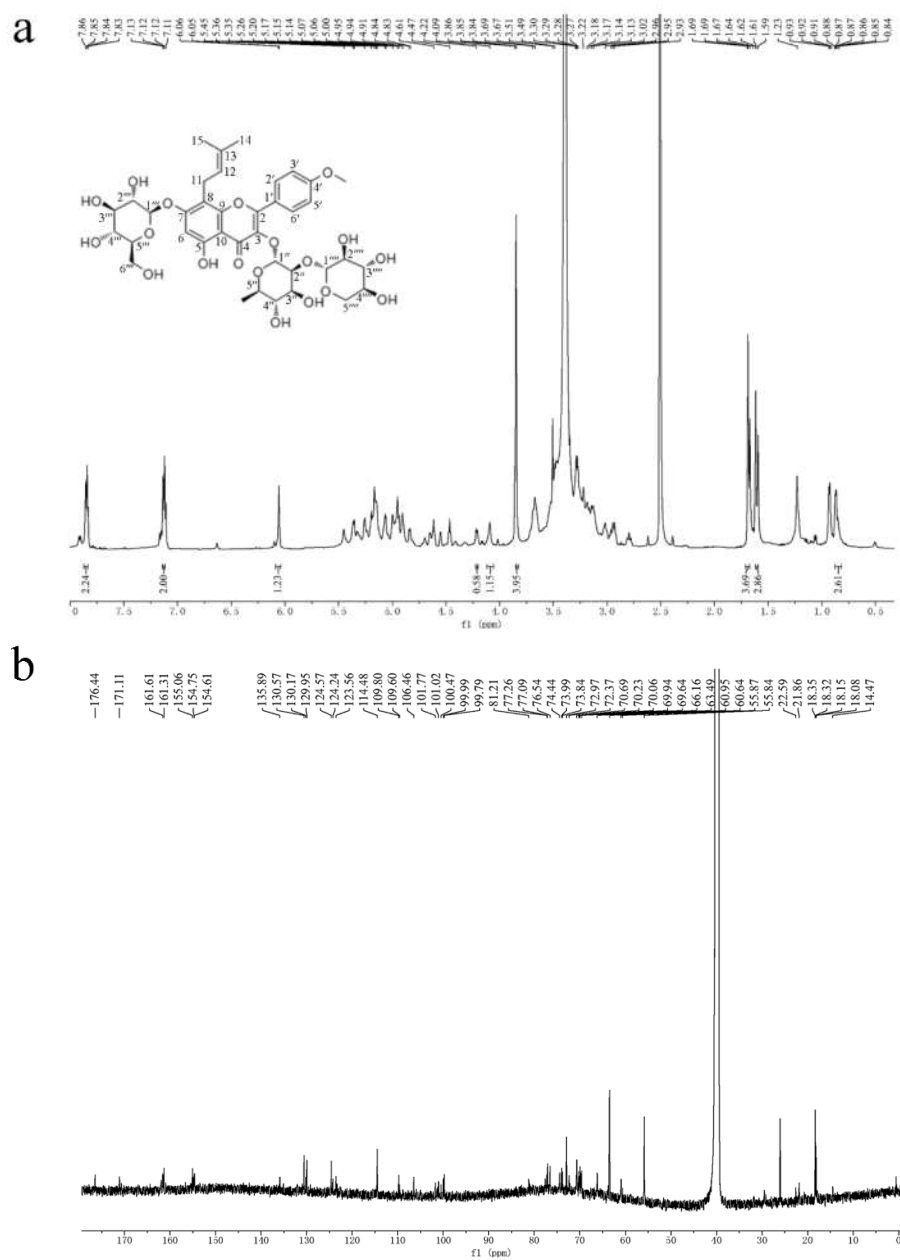




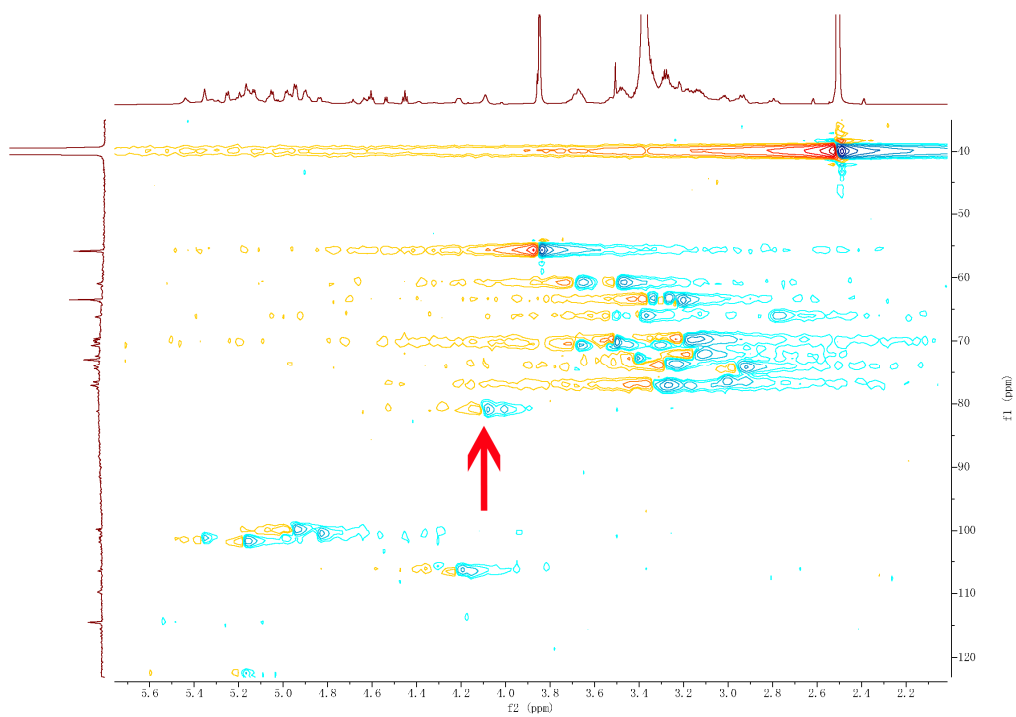
Figure S8



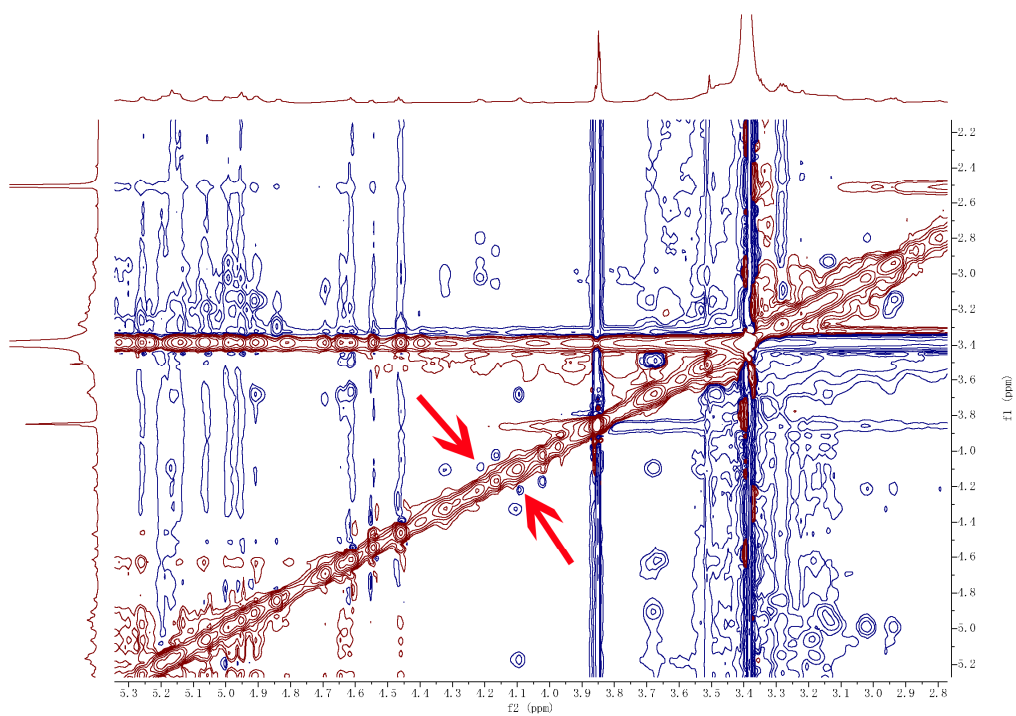
**Figure S9**



**Figure S10**



**Figure S11**



## 2 Supplementary Tables

**Supplementary Table S1. GTs used for phylogenetic tree**

Protein name	Length	Plant species	Accession ID	Modification site	Substrates
GmF3G6"Gt	464	<i>Glycine max</i>	NP_001345940.1	Glc-6"	Flavonol 3- <i>O</i> -glucoside
CaUGT3	454	<i>Catharanthus roseus</i>	AB443870	Glc-6"	Flavonol 3- <i>O</i> -glucoside
UGT79B31	455	<i>Petunia hybrida</i>	BBE29003.1	Glc-2"	Flavonol 3- <i>O</i> -glycoside
UGT79G16	459	<i>Ipomoea purpurea</i>	AB192315.1	Glc-2"	Anthocyanidin 3- <i>O</i> -glucoside
Cm1,2RhaT	452	<i>Citrus maxima</i>	AY048882	Glc-2"	Flavone 7- <i>O</i> -glucoside
AtUGT79B1	467	<i>Arabidopsis thaliana</i>	NP_200217.1	Glc-2"	Anthocyanidin 3- <i>O</i> -glucoside
UGT73A9 AmF7GlcT	481	<i>Antirrhinum majus</i>	BAG31950.1	7-OH	Flavone
UGT73E2 Am7GlcT	501	<i>Antirrhinum majus</i>	BAG16513.1	7-OH	Flavone
UGT73N1 Am7GlcT	495	<i>Antirrhinum majus</i>	BAG16514.1	7-OH	Flavone
GmUGT73C20	502	<i>Glycine max</i>	XP_003554403	7-OH	Isoflavone
Ph3GalT	451	<i>Petunia hybrida</i>	Q9SBQ8.1	3-OH	Flavonol
CsUGT78A14	459	<i>Camellia sinensis</i>	KP682360.1	3-OH	Flavonol
CsUGT78A15	454	<i>Camellia sinensis</i>	KP682361.1	3-OH	Flavonol
Cp3GT	468	<i>Citrus x paradisi</i>	ACS15351.1	3-OH	Flavonol
PfA5GlcT	460	<i>Perilla frutescens</i>	AHL68667	5-OH	Anthocyanidin 3- <i>O</i> -glucoside
Va5GT	464	<i>Vitis amurens</i>	AHL68667.1	5-OH	Anthocyanidin
UGT88D2 LvC4'GlcT	454	<i>Linaria vulgaris</i>	BAE48240.1	4'-OH	Chalcone
UGT88D3 AmC4'GlcT	457	<i>Antirrhinum majus</i>	Q33DV3.1	4'-OH	Chalcone
EpF3R2"XylT	433	<i>Epimedium pubescens</i>	ON569258	Rha-2"	Prenylflavonol 3- <i>O</i> -rhamnoside

**Supplementary Table S2. Primers used in this study**

Gene ID	Primers	Sequences (5'-3')
EpF3R2"XylT (first round)	EpF3R2"XylT-F1	TTAGACCTGATAATCTCCAGTTG
	EpF3R2"XylT-R1	TAAATGTTGGATGGGTCACACAG
EpF3R2"XylT (second round)	EpF3R2"XylT-F2	ATGAACATGGAAAAGCCTCCTT
	EpF3R2"XylT-R2	TCAGACCATATCTTGCAGCTTTG
EpF3R2"XylT (for pMAL-c2X)	EpF3R2"XylT-EcoRI	CGAGGGAAGGATTT <u>CAGAATTC</u> ATGAACATGGAAAAGCCTCCTT
	EpF3R2"XylT-XbaI	CTTGCCTGCAGGTCGACT <u>TCTAGAT</u> CAGACCATATCTTGCAGCTTTG
EpF3R2"XylT (for pCAMBIA1302)	EpF3R2"XylT-NcoI	AGAACACGGGGGACTCTTG <u>ACCATGGT</u> AATGAACATGGAAAAGCCT
	EpF3R2"XylT-SpeI	GTGAAAAGTTCTTCTCCTTT <u>ACTAGT</u> GACCATATCTTGCAGCTTTG
EpF3R2"XylT (qPCR)	EpF3R2"XylT-qPCR-F	ATGACATCAGTAGAGGAAG
	EpF3R2"XylT-qPCR-R	GACAATCACTCACTAACG
Actin (qPCR)	Actin-qPCR-F	TCAGAAGGAAGAACATACAC
	Actin-qPCR-R	AAGAGAAGGACGGATAGC

**Note:** Enzyme digestion sites were underlined.

**Supplementary Table S3. Flavonoids used in the assays of EpF3R2''XylT enzymatic activity**

Substrate type	Substrate names	EpF3R2''XylT
<b>8-prenylflavonol glycosides</b>	Icaritin	++
	Baohuoside I	+
	Baohuoside II	+
	Epimedoside A	+
	Icariside I	-
	Epimedin A	-
	Epimedin B	-
	Epimedin C	-
<b>Flavonols</b>	Kaempferol	-
	Quercetin	-
	Kaempferide	-
	Tamarixetin	-
	Isorhamnetin	-
	Mearnsetin	-
<b>Favonol glycosides</b>	Kaempferide-3- <i>O</i> -glucoside	-
	Kaempferide-3- <i>O</i> -rhamnoside	-
	Kaempferide-3- <i>O</i> -galactoside	-
	Quercetin-3- <i>O</i> -glucoside	-
	Quercetin-3- <i>O</i> -rhamnoside	-
	Quercetin-3- <i>O</i> -galactoside	-
	Isorhamnetin-3- <i>O</i> -glucoside	-
	Isorhamnetin-3- <i>O</i> - galactoside	-

**Note:** + and - indicate if the enzyme activity is detectable or not, respectively.

**Supplementary Table S4. Mass spectrometry of enzymatic product of EpF3R2"XylT**

NO.	Compounds	Molecular formula	Molecular weight	[M+H] <sup>+</sup>	Fragment ions
<b>1</b>	Icariin	C <sub>33</sub> H <sub>40</sub> O <sub>15</sub>	676.7	677.2584	531.1926 [M+H-Rha] <sup>+</sup> , 369.1378 [M+H-Rha-Glc] <sup>+</sup> , 313.0708 [M+H-Rha-Glc-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>1a</b>	Epimedin B	C <sub>38</sub> H <sub>48</sub> O <sub>19</sub>	808.8	809.2971	677.2466 [M+H-Xyl] <sup>+</sup> , 531.1873 [M+H-Xyl-Rha] <sup>+</sup> , 369.1334 [M+H-Xyl-Rha-Glc] <sup>+</sup> , 313.0708 [M+H-Xyl-Rha-Glc-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>2</b>	Baohuoside I	C <sub>27</sub> H <sub>30</sub> O <sub>10</sub>	514.5	515.1958	369.1422 [M+H-Rha] <sup>+</sup> , 313.0708 [M+H-Rha-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>2a</b>	Sagittatoside B	C <sub>32</sub> H <sub>38</sub> O <sub>14</sub>	646.6	647.2325	515.1906 [M+H-Rha] <sup>+</sup> , 369.1334 [M+H-Rha-Glc] <sup>+</sup> , 288.2591 [M+H-Rha-Glc-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>3</b>	Baohuoside II	C <sub>26</sub> H <sub>28</sub> O <sub>10</sub>	500.5	501.1778	355.1241 [M+H-Rha] <sup>+</sup> , 299.0543 [M+H-Rha-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>3a</b>	Ikariside F	C <sub>31</sub> H <sub>36</sub> O <sub>14</sub>	632.6	633.2184	501.1778 [M+H-Xyl] <sup>+</sup> , 355.1199 [M+H-Xyl-Rha] <sup>+</sup> , 299.0543 [M+H-Xyl-Rha-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>4</b>	Epimedoside A	C <sub>32</sub> H <sub>38</sub> O <sub>15</sub>	662.6	663.2374	517.1691 [M+H-Rha] <sup>+</sup> , 355.1199 [M+H-Rha-Glc] <sup>+</sup> , 299.0543 [M+H-Rha-Glc-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>
<b>4a</b>	Epimedoside E	C <sub>37</sub> H <sub>46</sub> O <sub>19</sub>	794.7	795.2772	663.2256 [M+H-Xyl] <sup>+</sup> , 517.1691 [M+H-Xyl-Rha] <sup>+</sup> , 355.1199 [M+H-Xyl-Rha-Glc] <sup>+</sup> , 299.0543 [M+H-Xyl-Rha-Glc-C <sub>4</sub> H <sub>8</sub> ] <sup>+</sup>

### 3 Sequences of *EpF3R2*"XylT

#### 3.1 DNA sequence

ATGAACATGGAAAAGCCTCCTTTACATATTGCTATGTTTCCATGGTTTGGCATGGGCCATCTGCTTCCCTTTCTCCGCCTTTC  
AAACATCTTAGCAGAGAAAAGGCCACCAAATCTCCTTTTGTACCCACAAAAACCCAGTCCAAGTTAAACCATCTCAATTTTC  
GTCCCAAACTCGTC AATTTCAATTCCTTGGTTGCTCATGTAGAAGGCCTTCCACTTGGTTCTGAAACCATGTCGAACATC  
TCAATCGAACTTGAACCCCTCCTTGCAACTGCCTTAGACCTTATGCAACAAAAGGTTGAAAAAATCTTCAAGATCTAAAACC  
TGATTTTGTCTTCTACGACTTCGCCTACTGGATACCAAAAATTGCTCGTCCCTTGGGATCAAGTCCATATTCTACTCAATTG  
TGGTTGCATCACAATTTGTTGAACGTGAAATGGGCGAAAAGGTCGCATCTTGGCTTGAGAAAATGTGGAAGCACTCATCCAT  
CTTGAATTTGGGATTGGGGTAACATTGCCTCAAAGACTTGCTGCCTGCATGCAGGACTGCGATGCCATTGCCTTCAGAGGGTG  
TCATGAAATCGAGGGAACAGCCTATGAATCCCTTGAGATAAAGTATGGGAAACAAGTACTCGTAACTGGTCCAGTTTGGGATG  
AACCATGTAGTTTCCCTTGGGAAGAGCGTTGGGATAAGTGGTTAAGGGCATTTTCAGAGGAATCTGTAGTTTACTGTGCGTTT  
GGGAGCGAGTGGGTATGACTAAAGAAGCATTTAGGAATTGGTCTTAGGTTTGGAGTTTACTGGATTGCCATTCTTTGTGGC  
ACTTAAACCACCACATGGGATGACATCAGTAGAGGAAGCATTCCTGGCTGGGTTTCGCGAAAGGGTGAAGGGAAGAGGGGTTG  
TTTATTCGGGCTGGGTACAACAGAAGCTCATCCTAAACCACCCATCGGTGGGATGTTTTGTGACCCATTGCGGGGCTTCGTCA  
ATGTGGGAATCGTTAGTGAGTGATTGTCAGATAGTAGCCCTGCCACAAGCAGGGGATCAGTTTATGAATGCTAATTTGTTGAC  
AAACGAATCAAGTTGGTGTGGAATTTAGCGGAGGGATGAGGATGGGTGGTTCACAAGGAGGGAGTGCCTCAGGCTGTTG  
AGGCTATAATGAACCAAGAGAGCGAAGTTGGTATAAAAGCCAGGGAACCATGCTATGTTGAAGGATACTTTGTTGAAGAAA  
GGACTAGAATCGGCTTACCTGAACAATTCGTTGCAAAGCTGCAAGATATGGTCTGA

#### 3.2 Protein sequence

MNMEKPLHIAMFPWFAMGHLLPFLRLSNILAEKGHQISFFVPTKTQSKLNHLNFRPKLVNFIPLVVPHEGLPLGSETMSNI  
SIELEPLLATALDLMQQKVEKILQDLKPDFVFYDFAYWIPKIARPLGIKSI FYSIVVASQFVEREMGEKVASWLEKMWKALIH  
LEFGIGVTLPQRLAACMQDCDAIAFRGCHEIEGTAYESLEIKYGKQVLVTGPVLDEPCSFPLEERWDKWLRAFSEESVYCAF  
GSEWVMTKEAFQELVLGLEFTGLPFFVALKPPHGMTSVEEAFPAERVKGRGVVYSGWVQQKLILNHPSVGC FVTHCGASS  
MWESLVSDCQIVALPQAGDQFMNANLLTNELKVGEIERRDEGWF TREGVRQAVEAIMNQSEVGIKARENHAMLKDTLLKK  
GLESAYLNNFVAKLQDMV