Supplementary Information

UGT74B1 from *Arabidopsis thaliana* as a versatile biocatalyst for the synthesis of desulfoglycosinolates

Sami Marroun^{*a†*}, Sabine Montaut^{*b†*}, Stéphanie Marquès ^{*c†*}, Pierre Lafite^{*c*}, Gaël Coadou^{*a*}, Patrick Rollin^{*c*}, Guillaume Jousset^{*c*}, Marie Schuler^{*c*}, Arnaud Tatibouët^{*c*}, Hassan Oulyadi^{*a*} and Richard Daniellou^{*,*c*}

^a Normandie Univ, COBRA, UMR 6014 et FR 3038; Univ Rouen; INSA Rouen; CNRS, IRCOF, 1 rue Tesnière, 76821 Mont Saint Aignan Cedex, France ^b Department of Chemistry and Biochemistry, Biomolecular Sciences Programme, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada ^cUniversité d'Orléans, CNRS, ICOA, UMR 7311, Orléans, France.

Table S1: Percentages of sequence identities for UGT74B1 and its templates.

Table S2: Percentages of sequence identities for UGT74B1 and its templates in the active site region.

Figure S1 : Sequence alignment for UGT74B1 and its closest homologues.

Figure S2: ¹H and ¹³C NMR Spectra of compound **7**.

Figure S3 : HRMS Spectrum of compound 7.

Figure S4 : ¹H and ¹³C NMR Spectra of compound **9**.

Figure S5 : HRMS Spectrum of compound 9.

Figure S6 : ¹H and ¹³C NMR Spectra of compound **10**.

Figure S7 : HRMS Spectrum of compound 10.

	74B1	2VCE	3HBF	2C1Z	2PQ6	4WHM	2ACW
74B1	100	85.65	87.53	88.12	86.55	86.61	84.47
2VCE	25.83	100	83.7	84.01	84.99	82.56	87.09
3HBF	23.26	24.58	100	91.09	85.14	87.88	83.17
2C1Z	23.48	24.79	48.03	100	85.82	89.27	83.46
2PQ6	25.93	28.01	23.86	22.82	100	84.4	83.53
4WHM	26.09	22.71	39.65	42.98	21.99	100	82.99
2ACW	22.15	29.79	23.44	22.8	24.27	20.86	100

Table S1: Percentages of sequence identities obtained by a dual structure-sequence multiple sequence alignment (green) and structural similarity scores (blue) obtained using CATH-SSAP for UGT74B1 and its templates. The maximum possible structural similarity score is 100.

Active site	74B1	2VCE	3HBF	2C1Z	2PQ6	4WHM	2ACW
74B1	100	51.02	47.92	45.83	60.42	45.83	45.83
2VCE		100	46.94	51.02	44.9	40.82	55.1
3HBF			100	81.25	43.75	68.75	50
2C1Z				100	45.83	64.58	45.83
2PQ6					100	43.75	41.67
4WHM						100	39.58
2ACW							100

Table S2: Percentages of sequence identities obtained by a dual structure-sequence multiplesequence alignment for UGT74B1 and its templates in the active site region.

	-	15	15	41	12
7481 sp)(048676)U7481_ARATH	MAETT P K V K G H V I	ILP Y P V Q G H L N P M	V Q F A K R L V S K	N V · · K V T I A T	T T Y T A S S · · ·
2VCEIsplQ9M156U72B1 ARATH	MEESK TPHV	IIP SPGMGHLIPL	VEFAKRLVHL	HG-LTVTFVI	A G E G P P S K A -
3HBFlsplA6XNC6IUGFGT_MEDTR	MSTFKNEMNG N LLHV	VLAFPFGTHAAPL	LSLVKKIATE	APKVTFSFFC	T T T T N D T L F S
2C1ZlsplP51094UF0G_VITVI	<u>MSQT</u> H N P	VLAFPFSTHAAPL	LAVVRRLAAA	A P H A V F S F F S	T S Q S N A S I F H
2PQ6ltrjA6XNCSJA6XNC5 MEDTR awhatitaastaalaastaa ci tts	M G N F A · · · · · N R K P H V ·				T E Y N H K R L L -
2ACW[tr]QSIFH7]QSIFH7_MEDTR	M S M S D I N K N S E L			D K N L Y I T V F C	- K F P G M P - F -
	61	8	16	101	E
7481 [sp]048676[U7481_ARATH	L I I P S L	VEP I SDGFDFIPI	GIPG.FSVDT	Y S E S F K L N G S	ETLTLIEK.
2VCE[sp]Q9M156[U72B1_ARATH	- Q R T V L D S L P S S I	S V F L P P V D L T D L S	• • 5 5 • T R I E S	R I S L T V T R S N	PELRKVFDS.
3HBFlsplA6XNC6JUGFGT_MEDTR	- R S N E F L P N I	YYN VHDGLPKGYV	S S - G N P R E	PIFLFIKAMQ	ENFKHVIDEA
2C1ZIspIP51094UFOG VITM		SYDISDCVPECYV		DIELFTRAAP	E S F R Q G M V M A
2PQ6triA6XNCSIA6XNCS MEDTR avviuntetaarspatiatera of the	X S R G P C X A F D G F T D C				
2ACW[tr]QSIFH7 QSIFH7_MEDTR	A D S Y I K S V L A S Q P Q I C			Y I L T F L E S L I	PHVKATIKT -
	101			5	10
7481 [colO48676]17481 ARATH					
2VCEIsplQ9M156U72B1 ARATH				A N V L S F F L H L	
3HBF[sp]A6XNC6[UGFGT_MEDTR	VAET GKNITCLVTD	FFW FGADLAEEMH	A K W V P L W T A G	Р Н S L L T Н V Y T	DLIREKT GSK
2C1Z splP51094 UFOG_VITM	VAET GRPV SCLVAD	FIW FAADMAAEMG	LAWLPFWTAG	P N S L S T H V Y I	DEIREKIGVS
2PQ6[tr]46XNC5]A6XNC5_MEDTR	LNHSTNVPPV TCLVSD	CMSFTIQAAEFE	L P N V L Y F S S S	ACSLLNVMHF	R S F V E R G I I P
4WHMItrIA4F1R4IA4F1R4 CLITE	VAETKQSVTCIIAD	FVT SSLLVAQTLN	V P W I A F W P N	SCSLSLYFNI	D L I R D K C S K D
2ACW[tr]Q5IFH7]Q5IFH7_MEDTR		F C V S M I D V G N E F G	I P S Y L F L T S N	V G F L S L M L S L	K N R Q I E E V F D
	181 191	201	211	221	231
7481 [sp]048676 U7481_ARATH	A - S N A	FRIRG-LPSL <mark>SYD</mark>	ELPSFVGRHW	LTH.PEHGRV	LLNQFPNHEN
2VCElsplQ9M156U72B1_ARATH		L M L P G - C V P V A G K	DFLDPAQDR-	K - D D A Y K W	LLHNTKRYKE
3HBF[sp]A6XNC6/UGFGT_MEDTR	E V H D V K - S	DVL PG-FPEL <mark>KAS</mark>	D L P E G V I K - D	- I - D V P F A T M	LHKMGLELPR
2C1ZlsplP51094[UFOG VITM				- L - N S L F S R M	
2PQ6[tr]A6XNC5]A6XNC5_MEDTR	FKDESYLTNG CLET-K	DWIPG-LKNFRLK			FIEVADRVNK
4WHMItriA4F1R4/A4F1R4_CLITE					
ZACW[tr]Q2[FH7[Q2]FH7_MED1K		L N L P G L S N Q V P S N	V L P U A C F N K -		Y Y N L A E H F K D
	241 251	261	271	281	291
7481 sp 048676 U7481 ARATH	ADWLFVNGFE GLEETOI	C - E N G E S D A	M K A T L I G P M I	P S A Y L D D R M E	
2VCEIsplQ9M156U72B1 ARATH	A E G I L V N T F F E L E P N -	A - I KALQEPGLDK	Ρ Ρ Υ Υ Ρ Υ G Ρ L Υ	N 1 6 K Q · · · ·	· · · E · A · · ·
3HBF[sp]A6XNC6[UGFGT_MEDTR	A N A V A I N S F A T I H P L	I - E N E L N S K F	K L L L V G P F N	L T T P Q	a
2C1Zlsplp51094lUFOG_VITVI	A T A V F I N S F E E L D D S	L - T N D L K S K L	K T Y L N I G P F N	L I T P P	<u>م</u>
2PQ6ltrl/a6XNC5/A6XNC5_MEDTR	DTTLLNTFN ELESD	V - I N A L S S T I	P S I Y P I G P L P	SLLKQTPQIH	
4WHMItriA4F1R4IA4F1R4 CLITE	AKAVVVNFFAELDPP-	LFVKYMRSKL	Q S L L Y V P L P	C P Q L L	
2ACWItrIO5IFH7IO5IFH7 MEDTR	TKGIIVNTFS DLEOS-	S I D A L Y D H D E K I	P P I Y A V G P L L	D L K G O · · · · ·	P . N P K L

Figure S1 -

3



Figure S1 (continued): Sequence alignment for UGT74B1 and its closest homologues using the TCOFFEE accurate mode. Secondary structures are highlighted in green for sheets and in yellow for helices. Secondary structure assignment for UGT74B1 is based on the homology model obtained using the MODELLER program.



Figure S2: 1 H (top) and 13 C (bottom) NMR Spectra of compound **7**.



Figure S3 : HRMS Spectrum of compound 7.



Figure S4: ¹H (top) and ¹³C (bottom) NMR Spectra of compound 9.



Figure S5 : HRMS Spectrum of compound 9.



Figure S6 : 1 H (top) and 13 C (bottom) NMR Spectra of compound 10.



FigureS7 : HRMS Spectrum of compound 10.