# Unleashing the Potential of Noncanonical Amino Acid Biosynthesis to Create Cells with Precision Tyrosine Sulfation

Yuda Chen<sup>1</sup>, Shikai Jin<sup>2,3</sup>, Mengxi Zhang<sup>1</sup>, Yu Hu<sup>1</sup>, Kuan-lin Wu<sup>1</sup>, Anna Chung<sup>1</sup>, Shichao Wang<sup>1</sup>, Zeru Tian<sup>1</sup>, Yixian Wang<sup>1</sup>, Peter G. Wolynes<sup>1,2,3,4</sup>, and Han Xiao<sup>1, 3, 5\*</sup>

<sup>1</sup>Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas, 77005 <sup>2</sup>Center for Theoretical Biological Physics, Rice University, Houston, Texas, USA. <sup>3</sup>Department of Biosciences, Rice University, 6100 Main Street, Houston, Texas, 77005

<sup>4</sup>Department of Physics, Rice University, 6100 Main Street, Houston, Texas, U.S.A. <sup>5</sup>Department of Bioengineering, Rice University, 6100 Main Street, Houston, Texas, 77005, U.S.A.

\* To whom correspondence should be addressed. E-mail: han.xiao@rice.edu;

# Table of Contents

Supplemental Methods	13
Materials	13
Plasmids Construction	14
Expression and Purification of Proteins	15
Expression and Fluorescence Measurement of sfGFP	17
E. coli Intracellular sTyr Concentration Measurement	17
Exploration on evolutionary relationship of <i>Nn</i> SULT1C1	17
Kinetics Measurement of <i>Nn</i> SULT1C1	18
Protein Purification from Mammalian Cells	18
Mammalian Cell sTyr Concentration Measurement	18
Mass Spectra Methods For Proteins	19

## **Supplemental Figures**



**Supplementary Figure. 1: Screening reported sulfotransferases with GFP assay.** Data are plotted as means from n=2 independent samples. a.u. stands for arbitrary unit.



**Supplementary Figure. 2: Phylogenetic relationship of all sulfotransferases tested in Fig. 2d.** Phylogenetic tree was generated in MEGAX software with UPGMA method. A0A091VQH7 (https://www.uniprot.org/uniprotkb/A0A091VQH7/entry) was named *Nn*SULT1C1 and used for following experiments.

	Uniprot ID	Species																									
	NnSULT1C1	Nipponia nippon		RΗ	ΡF	۰L	ΕW	SI	IQE	PP	AA	۲ s	1 5	G	LE	L	ΑE	AI	МР	SF	R	ТΙ	K	ΤН	I L	P١	/ Q /
	A0A087QVZ4	Aptenodytes forsteri		RΗ	ΡF	۰L	ΕW	s١	/QE	S 5	A A	A S Y	1 5	5 G	LE	L	ΑE	AI	МР	SF	R	тι	K	ΤН	I L	P١	/Q
	A0A091M6P2	Cariama cristata		RΗ	ΡF	۰L	ΕW	ΤN	ЛР Е	P P	AN	۸ S ۱	1 5	5 G	LE	L	ΑE	AI	МР	SF	R	тι	K	ΤН	I L	ΡA	١Q
	A0A091NU80	Apaloderma vittatum		RΗ	ΡF	۰L	ΕW	FL	LPE	P P	S I	_ R 1	1 5	5 G	LE	L	ΑE	AI	МР	SF	R	тι	K	ΤН	I L	P١	/Q
	A0A091VNG6	Opisthocomus hoazin		RΗ	ΡF	۰L	ΕW	SN	A E A	S P	TL	S	1 5	G	LE	L	ΑE	AI	МР	FΡ	R	ТΙ	K	ΤН	I L	P١	/Q
	A0A1V4J451	Patagioenas fasciata		RΗ	ΡF	۰L	ΕW	ΥI	I <mark>P</mark> E	P S	5 S L	R	1 5	G	LE	L	ΑE	AI	МР	SF	R	ТΛ	ΛK	ΤН	I L	P١	/Q
	Q90WR6	Gallus gallus		RΗ	ΡF	۰L	ΕW	Y		) S S	S P I	G	1 5	G	LΚ	L	ΑE	AI	МР	SF	R	ТΛ	ΛK	ΤН	I L	P١	/Q
	A0A093Q5M0	Manacus vitellinus		RΗ	ΡF	۰L	ΕW	F۱	/ A E	PP	S	ER	1 5	G	VΕ	L	A N	AI	МР	SF	R	тι	K	ΙН	ΙL	P١	/ Q
<u> </u>	U3JLS0	Ficedula albicollis		RΗ	ΡF	۰L	ΕW	FI	I <mark>P</mark> E	PP	SN	A R	1 5	G	VΕ	L	ΑE	AI	МР	SF	R	тι	K	тн	ΙL	P١	/ Q
	H0ZHC5	Taeniopygia guttata		RΗ	ΡF	۰L	ΕW	YC	CAE	ΡP	SA	R	1 5	G	LE	L	ΑE	AI	МР	SF	R	ТΛ	٨ĸ	ΤН	I L	P١	/Q
0.02			4-																								
		Bits	2-	P	D			V			<b>D</b> _		V	$\mathbf{x}$	ÌF		Δ⊏	<b>-</b> ΔI		9	D	Τı	K	TI-		D١	IΛ
		-	11	<u>  N</u>		Ļ		Ę,		s	<u>5 è</u>	<u>85</u>	1					<u> </u>	V	2	<u> </u>		$\Lambda$	<u>  </u>		Ţ	V
				1		5			10		1	5			20			25			30			35	5		
NpSIII T1C1/1-308	, <sup>10</sup>	20		30	M	 	אאפ	40 = D I	51 6 6	TVV	50 (SPA	FIC	E V	FG	60	тк		ст	70 WDC		, ,	AP	80 80			vт	90 80
A0A087QVZ4/1-307					- M	ALC	O K M E	DL	SLKC	2 T - V	PRA	EIG	ΕV	EG	IPL	TT	тіс	ST	WDQ	vw	KFK	AR	PDD	) L L	IAT	YA	KA
A0A091M6P2/1-306 A0A1V4J451/1-320	ML	I P S L	 S I	FMO	•-М SST	A L C V L C	ОКМК ОКМА		SLQC	2Τ- L 2Τ- L	. S R A . G R A	K M G	EN	'EG IKG		тк	РІС АТҮ	DT	WDQ WDO	vwi	K F K K F K	AR	P D D P D D		1 A T 1 S T	' Y A ' Y A	KA
A0A091NU80/1-252										 																	MS
H0ZHC5/1-305	MCNVFQTITES	5 K 5 E K 1 Q E K N I E K V		HF51	- S L	- M D	) K M K ) K M K	DL	RLEE	ET - V	IRS	ELF	DV	DG	V A I V P L	AE	PIA	RI	WDQ	vwi	K F K	AR	P D L P D L	DV L		Y A Y A	KA
A0A093Q5M0/1-306						MVE			SLKE	ОТ-I		ELF	DI	EG	VPL	IK	PMO	i I T	WDQ	VQI	KFK	AR		OV F	IAT	YA	KA
Q90WR6/1-307					- M	A L C	D K M E	ENL	SLEE	EN - N	ALR S	EMG	ΕV	QG	IPV	ТК	РТС	DI	WDQ	vwi	NFK	AR	PDD		VAT	YA	KA
	100	110	_	12	0			130		_	140	_		_	150			_	160		4		170	)			180
NnSULT1C1/1-308 A0A087QVZ4/1-307	GTTWTQEIVDM GTTWTOEIVDM	I Q Q N G D V E K C R R A T I O O N G D V E K C R R A T	ТҮ	R R H I R R H I	P F L P F L	EW	5 I Q E 5 V O E	EPP		′SGL ′SGL	. E L A . E L A			PR	тік тік	ТН	LPV	QL	LPP VPP	S F \ S F \	N E Q N E H	INC	кіі кіі	YV	A R M A R M	I A K	D N D N
A0A091M6P2/1-306	GTTWTQEIVDM	I Q Q N G D V D K C R R A T	ТҮ	RRHI	PFL	EW	MPE	E P P	AMS	SGL	ELA	EAM	PS	PR	тік	ТН	LPA	QL	VPP	SFN	NEQ	NC	кіі	YV	ARM	I P K	DN
A0A1V4J451/1-320 A0A091NU80/1-252	GTTWTQEIVDM	I Q Q N G D V E K C R R A I I Q Q D G D V D K C R R D T	ТҮ	R R H I R R H I	PFL	EW	FLPE	= P S = P P	SLR	'SGL 'SGL	. E L A . E L A		IP S	PR	тик тик	ТН	LPV	QL	VPP	SFN	N E Q N E Q	NC	кіі кvі	YV	A R M A R M	I P K	D N
U3JLS0/1-339	GTTWTQEIVDM	I Q Q N G D T E K C K R E T	ТҮ			EW	FIPE	E P P	SMR	SGV	ELA	EAM	PS	PR	ТІК	ТН	LPV	QL	LPP	SF	NEQ	NS	кіі	YV			DN
A0A093Q5M0/1-306	GTTWTQEIVDM	I Q E N G D T E K C R R E T	ΤY	KRHI	PFL	EW	FVAE	E P P	SERY	' S G V	ELA	NAM	IP S	PR	тік	ÌН	LPV	QL	VPP	SIN	NEQ	NC	кіі	ΥV	ARM	AK	DN
A0A091VNG6/1-304	GTTWTQEIVDM GTTWTOEIVDM	I Q Q N G D V E K C R R A T I O O N G D I E K C R R A S	ТҮ	K R H I K R H I	PFL	EW	5 M E A 7 I P C	A S P D S S	T L S Y	′SGL ′SGL	. E L A . K L A	EAM		PR	тік тмк	TH	LPV	QL	LPP VPP	S F \ S F \	N E Q N E O	NC	кіі кіі	YV	A R M A R M	I A K	D N D N
	-		_															-				_					
	190	200		21	0			220			230	)			240				250	)			260	)			270
NnSULT1C1/1-308 A0A087QVZ4/1-307	LVSYYHFHRMS VSYYHFHRMN	K E M P D P G T W E E F M E K G M P D P G T W E E F M E	KF	M T G I M T G I	K V L K V L	WG ! WG !	S W Y E S W Y E	онv онv	KGWV	VKAK VKAK	(DRH (DKH	IRIL IRIL	. Y L . Y L	. F Y . F Y	E D N E D N	1 K E 1 K E	NPK	Q E Q E	IQK		K F L K F L	. E K . E K		NQE SQE	V L N V L N	1 K I 1 K I	LH
A0A091M6P2/1-306	. V S Y Y H F H R M N	K E L P D P G T W E F M E	KF	MTG	KVL	WG	SWYE	рну	KGW	VKAK			YL	FY	EDN	IKE		KE	TQK		KFL	E K		QE	VLN		νн
A0A1V4J451/1-320 A0A091NU80/1-252	VSYYHFHRMN	K G L P E P G T W E E F M E	KF	MTG	KVF	WG	SWYE	он V	KGWV	VKAK		IRIL	YL	. F Y	EDN	IKE	NPK	QE	IQK		KFL	. E K		E T	VLN	NK I	LR
U3JLS0/1-339	V S Y Y H F H R M N	ALP DP GT WEEFVE	KF	MTG	KVL	WG	SWYE	ЭΗУ	KGW	VKAK			YL	FY	EDN	IKE		RE	IQK		KFL	EK	DLS	S E E	VLN		VН
A0A093Q5M0/1-306	IVSYYHFHRMNI	KAMP EPGTWEEFVE	KF	MSG	Q V L	WG	SWYE	эну	KGWV	VKAK	DKF	IRIL	YL	. F Y	EDN	IKE	NPK	RE	IQK		KFL	. E K	DLS	SQ E	VLN	NK I	VН
A0A091VNG6/1-304	. V S Y Y H F H R M N	KAMP DP GT WEEFME	KF	MTGI	K V L F V I	WG	SWYE	онv онv	KGWV	VKAK		IRIL IRII	Y L	FY FY	EDN	IK E	NPK	QE	IQK		KFL	EK		NQ E		1 K I 1 K I	LD
0,507410/1-307	2.5										UD IN I				2011								0.00			••••	
	280	290		3(	00			310	)		32	0			33	0											
NnSULT1C1/1-308	NTSFEIMKDNP	MTNYTTEFQGIMD	H S I	SPF	M R K	GV	VGD	WKN	N Y F T	VAQ	NEK	FDE	DY	ккк	MA	DTS	LV	FRT	EL								
A0A087QVZ4/1-307	NTSFEIMKENP	MT NYT NE FQGI MD	HSI	SPF	MRK	GV	VGD	WKN	NY FT	VAQ	NKK	FDE	DY	K K K	MA		VV	FRT	EL								
A0A1V4J451/1-320	NTSFEIMKENP	MANYTKDFLGVMD	HSL	SP F	MRK	GV	VSD	WKN	NYFT	VAQ	NKK	FDE	DΥ	ккк	MA	DTS	LV	FRT	EL								
A0A091NU80/1-252	NT SFEAMKENP	MANYTKEFQGIMD	HSV	SPF	MRK	GV	VGD	WKN	NY FT	VAQ	NKK	FDE	DY	ккк	MA		LV	FRT	EL								
H0ZHC5/1-305	NTSFEVMKENP	MANYTKDFEGIMD	HSI	SPF	MRK	GI	VAD	WKN	NHFT	VAQ	NKK	FDE	DY	EKK	MS	DTS	LV	FRA	A E L								
A0A093Q5M0/1-306	NT S F E V M K E N P	MANYTKDFQGIMD	HSI	SPF	MRK	GV	VAD	WKN	NHFT	VAQ	NEK	FDE	DY	KRK	MS		LV	FRA	A E L								
Q90WR6/1-307	NTSFEIMKDNP	MTNYTKDFVGVMD	HSV	SPF	MRK	GS	VGD	WKN	NYFT	VAL	NKK	FDQ	DY	KKK	MA	DTS	LV	FRA	AEL								

Supplementary Figure. 3: Phylogenetic analysis and multiple sequence alignment of *Nn*SULT1C1 and

its 9 relatives. Phylogenetic tree was constructed by UPGMA method in MEGAX and multiple sequences were

aligned by ClustalW method.

NnSULT1C1/1-308 SULT1C2/1-296 SULT1C3/1-304 SULT181/1-296 SULT1A1/1-295 SULT1A1/1-295 SULT1A2/1-295 SULT1A3/1-295 SULT1A1/1-294 SULT2A1/1-285	10   MALTS DK:   MALTS KN   MAKIE KN   MALSPK DI   MELIQ DI   MELIQ DI   MELIQ DI   MELIQ DI   MSEL DY	20 M E D L S L G K	30 K ET V V S R A E I C Q I K L K K P E L F N I M L R K D L K S R P P L E S R P P L E S R P P L E S R F P L	40 EVEGI PFT EVEGT LLQ EVDGV PTL VHGV PLI YVKGV PLI YVKGV PLI EVHGI LMY WFFGI AFPTM	50 K P I C S T W D Q V W- P A T V D N W S Q I Q - I L S K E W W E K V C - C A F A S N W E K I E - K Y F A E A L G P L Q - K Y F A E A L G P L Q - K Y F A E A L G P L Q - K Y F A E A L G P L Q - K G F K S T I K K V M D N Y E	60 2 K FK A R P D D L L I I S F E A K P D D L L I I N FQ A K P D D L L I I Q F H S R P D D L I I S FQ A R P D D L L I S FQ A R P D D L L I S FQ A R P D D L L I - A FQ A R P D D L L I D F E V I R D F D V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I D F E V I R D F V I I I D F E V I R D F V I I I D F E V I R D F V I I I I D F E V I R D F V I I I I D F E V I R D F V I I I I D F E V I R D F V I I I D F E V I R D F V I I I D F E V I R D F V I I I I I I I I D F E V I I I I I I I I I I I I I I I I I I	70 80 AT YT K A GT T WT Q E CT Y P K A GT T WI Q E AT Y P K S GT T WH Q E A T Y P K S GT T WV S Q ST Y P K S GT T WV S Q NT Y P K S GT T WV S Q AT Y P K S GT T WV S Q	90 I V DMI QQNGDV EK C R R I V DMI EN GDV EK C R R I L DMI L N DGDV EK C K R I L DMI V QGGD L EK C H R I L DMI V QGGD L EK C H R I L DMI V QGGD L EK C H R I L DMI V GGD L EK C K R I L DMI V EGDV EK C K E I L DMI V K EGDV EK C K E I L DMI V K EGDV EK C K E I C I MI V K EGDV EK C K E	100 A T A I A Q G F A P A P D V V P
SULT2B1/1-365 SULT4A1/1-284 SULT6B1/1-303	M D G P A E P Q I P G L W D T M A E S E A E M A D K S K F	Y EDDISEISQ TP STPG IEYIDEALEK	K L P G E Y F E F E S K Y F S K E T A L S H L F F	R Y K G V P F P V G E F H G V R L P T Y Q G I P Y P I T	L Y S L E S I S L A E M P F C R G K M E E I A - M C T S E T F Q A L D -	NTQDVRDDDIFI NFPVRPSDVWI TFEARHDDIVL	I T Y P K S G T T WM I E V T Y P K S G T S L L Q E A S Y P K C G S NW I L H	IICLILKEGDPSWIRS VVYLVSQGADPDEIGLM IIVSELIYAVSKKKYK	V P M N
	110	120	130	140	150	160	170 180	190	200
NnSULT1C1/1-308 SULT1C2/1-296 SULT1C3/1-304 SULT1C3/1-304 SULT1A1/1-295 SULT1A1/1-295 SULT1A2/1-295 SULT1A3/1-295 SULT1A1/1-295 SULT2A1/1-285 SULT2A1/1-284 SULT2A1/1-284 SULT2A1/1-303	T Y R R H P F L EW S ( Q E P I Q H R H P F L EW S R P T L D R H A F L E L K F T E K Y P M L E F K A I F M X P F L E F K Y I Y V K V P F L E F K Y I Y V K V P F L C R K I W E R S P W V E S E I W E R A P W C E T I I W E R A P W C E T I I D E Q L P V L E Y P Q Y P E F P V L E C D	P AA S Y S G L E L P Q P S G V E K P HK E K P D L E F P G L R T S G I E Q P G I - P S G M E T P G I - P S G M E T P G I - P S G L E T E N L - MN G V K Q I G Y T A P G L D I E K Y Q R	A EAMP S P R T I K AKAMP S P R I L K V L EM S S P Q L I K L EK N F S P R I V K L K D T P A P R L L K L K D T P P P R I I K L K D T P P P R I I K L D EM N S P R I V K L S ET E S P R L F S I K EL T S P R L M M K G F P S P R I L A	T H L P VQ L L P P T H L S T Q L L P P T H L P S H L I P P T H L P T D L L P K T H L P L A L L P Q S H L P L A L L P Q S H L P I Q L F P K S H L P I Q I F F K S H L P I Q I F F K S H L P Y R F L P S T H L H Y D K L P Q	S FWEQ NCK   I YY S FWENNCK F LYY S FWENNCK MI YY T LL DQ K VK VYY T LLDQ K VK VYY T LLDQ K VK VYY S FWEKDCK I I YI S FF S S KAK VI YY D LHNG D S K VI YY S I FENKAK I LY	/ A R NA K D N L V S Y / A R NA K D C L V S Y / A R NA K D V S V S Y / A R NA K D V S V S Y / A R NA K D V A V S Y / A R NA K D V A V S Y / A R NA K D V A V S Y L C R NA K D V A V S Y L M R NP R D V L V S Y M A R NP K D L V V S L I F R NP K D L V S Y	Y H H H M S K E M D P V Y H F O R M M H M L P P P Y H F D R M M A S F M P D P Y H F J H M A K V H P E P Y H F Y H M A K V Y P L P Y H F Y H M A K V Y P H Y H F H M A G L K D P Y Y F F L M Y A G L K D P Y Y S L A G L K D P Y O F H R S L R T M S Y R L H F H N D Y P D I P S Y	GTW E EFM EF FING KV LI GTW E EY FF FING KV VV QN LE EFY EK FMS GK VV GTW E EY LEK FM GE VS GTW E SY LEK FM GE VS GTW S SI LEK FM AG EV S GTW S SI LEK FM AG EV S GS FP EFP EK FM AG EV S GS FP EFP EK FM AG EV S GT D G L R D FL K GE V AJ GT P D G L R D FL K GE V AJ GT P D G L R D FL K GE V AJ	WG GG YG YG YG YG YG YG WG
NnSULT1C1/1-308 SULT1C2/1-296 SULT1C3/1-304 SULT1B1/1-296 SULT1A1/1-295 SULT1A2/1-295 SULT1A3/1-295 SULT1A3/1-295 SULT1E1/1-294 SULT24/1-295	210 SWY DHV K GWW KAK DR SW F DHV K GWW E MK DR SW F DHV K GWW A AK DM SW F THV K NWW KK E E SWY QHVQ E WWELS RT SWY QHVQ E WWELS RT SWY QHVQ E WWELS RT	220 H R I L Y L F Y E C H Q I L F L F Y E C H R I L Y L F Y E C H P V L Y L F Y E C H P V L Y L F Y E C H P V L Y L F Y E C	230 MK E N P K Q E I Q K I K R D P K H E I R K I K K D P K R E I E K MK E N P K R E I Q K MK E N P K R E I Q K MK E N P K R E I Q K	240 I L K F L E K D V I V MQ F M G K K V I I L K F L E K D I I R F L E K N L I I L E F V G R S L F I L E F V G R S L F	250 1 Q EV L NK I L H NT 5 ET V L DK I V Q ET 5 EE I L NK I I Y HT 1 D E I L DR I I H HT 6 EE T V D FV V Q HT 6 EE T V D LMV EHT	260 S F E I MK D N P MT N S F E K MK E N P MT N S F D V MK Q N P MT N S F E V MK D N P L V N S F K E MK K N P MT N S F K E MK K N P MT N	270 280 Y T T E F Q G I M D H R S T V S K S I L D Q Y T T L P T S I M D H Y T H L P T T V M D H Y T T V P Q E F M D H Y T T V R R E F M D H	290 - S I S P F M - R K G V V G D W - S I S S F M - R K G T V G D W - S I S P F M - R K G T A G D W - S K S P F M - R K G T A G D W - S I S P F M - R K G M A G D W - S I S P F M - R K G M A G D W	300 K N K N K N K T K T
SULT2B1/1-365 SULT2B1/1-365 SULT4A1/1-284 SULT6B1/1-303	SWFRHVRSWWERGRS SWFDHIHGWMPMREE SWFDHIKGWLRMKGK SWFEHVQEFWEHRMD RYFDFAINWNKHLDG	P R V L F L F Y E C K N F L L L S Y E E D N F L F I T Y E E S N V L F L K Y E C D N V K F I L Y E C	L K E D I R K E V I K L K Q D T G R T I E K L Q Q D L Q G S V E R M H R D L V T M V E Q L K E N L A A G I K Q	I L I H F L ER K P S I C Q F L G K T L I I C G F L G K P L C L A R F L G V S C I I A E F L G F F L 1	P E E T M D F M V Q H T S E E L V D R I I H H T P E E L N L I L K N S S K E A L G S V V A H S O K A Q L E A L T E H C I G E Q I Q T I S V Q S	S F K EMK K N P MT N S F Q EMK N N P S T N S F Q S MK E N K M S N T F S A MK A N T M S N H T F Q AMR A K S Q D T	Y T T V P Q E LM D H Y T T L P D E I M NQ Y S L L S V D Y V Y T L L P P S L L D H - Q L V DQ C C H G	- S I S P F M- R K G M A G D W - K L S P F M- R K G I T G D W /- D K AQ L L - R K G V S G D W - R R G A F L - R K G V C G D W N A E A L P V - G R G R V G L W - A V G P F L F R K G E V G D W	KN KN KN KN
SULT2B1/1-365 SULT2B1/1-365 SULT4A1/1-284 SULT6B1/1-303	SWFDHIKGWLPMRE SWFDHIKGWLRMKGK SWFEHVQEFWEHRMD RYFDFAINWNKHLDG 310 YFTVAQNEKFDEDYK	P R V L F L F Y E C K N F L L L S Y E E D N F L F I T Y E E S N V L F L K Y E C D N V K F I L Y E C 320 K K M A D T S L V V	L K ED I R K EV I K L K Q D T G R T I E K L Q Q D L Q G S V E R MH R D L V T M V E Q L K E N L A A G I K Q 330 F R T E L	LIHFLERKPS ICQFLGKTLI ICQFLGKTLI ICGFLGRPLC LARFLGVSCI IAEFLGFFL1	SEET MD FMVQ HT SEELVDR I I HHT PEELNLILKNS KEALGSVVAHS XAQLEALTEHCI GEQIQTISVQS	S F K EMK K N P MT N S F Q EMK N N P ST N S F Q SMK E N KM SN T F S AMK A N T M SN H T F Q AMR A K SQ DT 360	YTT LP DE LMDH YTT LP DE I MNQ YS LLS V DY VV YT LLP P S LL DH 	- S I S P F M - R K G M A G D W - K L S P F M - R K G I T G D W /- D K A Q L L - R K G V S G D W - R R G A F L - R K G V C G D W - R R G A F U - R K G V C G W - A V G P F L F R K G E V G D W	/K N /K N /K D /K N
SULT28/11-365 SULT28/11-365 SULT4A1/1-284 SULT6B1/1-303 SULT162/1-296 SULT162/1-296 SULT162/1-296 SULT1A1/1-295 SULT1A2/1-295	SW TO HIVK SWEEKGKS SW FOH IKGWLPMREGK SW FOH IKGWLPMRGK SW FEHVGEFWEHRMD RY FOFAL NWNKHLDG YFTVAONEK DEDVK HTTVAONEK DEDVK YFTVAONEK DEDV YFTVAONEK DALVK TFTVAONEK DALVK TFTVAONEK DADVA	P R V LF L F Y E E K N F L L S Y E E D N F L F I T Y E E S N V L F L K Y E E D N V K F I L Y E O X K M A D T S L V - R K M E G T S I N - K K M A G S L T E E K M A G S L S - E K M A G C S L S -	L K E D I R K E V I K L K Q D I Q G S V E R M H R D L V T W E Q L K E N L A A G I K Q 330 - F R T E L - F C M E L - F R T E L - F R T E L - F R T E L - F R S E L - F R S E L - F R S E L	ILE FUGRST ICC FLGKTLI ICG FLGRDLI IAR FLGRDL IAR FLG FFLT 340	PEETMDFMVQHT SELVDRIHHT PEELNLILKNS KEALGSVVAHS XKAQLEALTEHC GEQIQTISVQS	S F K E MK K N P M T N S F Q E MK N N P S T N S F Q S MK E N K M S N F S A MK A N T M S N H T F Q A MR A K S Q D T 360	YTT VPQ E LMDH YTT L PDE II MNQ YSLLS VDYV YTL LPPSLLDH QLVDQCC HG	- SISPFM-RKGMAGDW - KLSPFM-RKGITGOW - DKAQLL-RKGVGDW - RAGAFL-RKGVGDW :NAEALPY-GRGRVGLW - AVGPFLFRKGEVGDW	

Supplementary Figure. 4: Sequence alignment of human cytosolic sulfotransferases (hSULTs) and *Nn*SULT1C1. The highly variable region we found (SIQEPPAAS) in NnSULT1C1 is aligned well with the reported residues of hSULTs important for substrate recognition.



**Supplementary Figure. 5: Superimposition of NnSULT1C1 and 2zvq.** *Nn*SULT1C1 was shown in blue colour and mouse SULT1D1 (PDB: 2zvq, https://www.rcsb.org/structure/2ZVQ) was shown in cyan colour. The variable region SIQEPPAAS of *Nn*SULT1C1 is shown in red.



**Supplementary Figure. 6: Expression condition screening for sfGFP-sTyr production.** The influence of expression medium, tyrosine addition, sulfate addition and glycerol addition on production of sfGFP-sTyr in bacterial cells containing sTyr biosynthesi and genetic incorporation machineries was evaluated with green fluorescent protein assay. Data are plotted as the mean +/- standard deviation from n=3 independent samples. a.u. stands for arbitrary unit.



**Supplementary Figure. 7: Kinetics measurment of tyrosine sulfation activity of** *Nn***SULT1C1.** (A) SDS-PAGE analysis of *Nn***SULT1C1-His6** expressed in LB medium. (B) Standard curve of authentic sTyr detected

by SIM mode of ESI-MS. Data are plotted as means of n=2 independent samples. (C) The effect of adding his6 tag to C terminal of *Nn*SULT1C1 on its activity. (D) Kinetics curve of *Nn*SULT1C1 with tyrosine as its substrate. Data are plotted as means of n=3 independent samples. Error bars represent standard deviations from n=3 independent samples. V<sub>max</sub> and K<sub>m</sub> were obtained by fitting the data to Michaelis-Menten equation in Prism. a.u. stands for arbitrary unit.



# Supplementary Figure. 8: Flow cytometry analysis of EGFP expression levels of HEK293T and HEK293T-*Nn*SULT1C1. (original data for Fig. 4C)

(a) Flow cytometry analysis of EGFP expression in HEK293T and HEK293T-NnSULT1C1 in the presence or absence of external sTyr. There are 3 independent biological replicates for one treatment group, as indicated by the numerical label. (b) Gating strategy used for (a).



**Supplementary Figure. 9: Cellular concentration of sTyr in HEK293T and HEK293T-***Nn***SULT1C1.** Indicated concentration of sTyr was added to the culture of HEK293T or HEK293T-*Nn*SULT1C1 for 2 hour. Data are plotted as means from n=2 independent groups.



Supplementary Figure. 10: ESI-MS analysis of EGFP39sTyr from HEK293T cells and HEK293T-*Nn*SULT1C1. The expected peak was calculated according to monoisotopic mass of EGFP39sY with N-terminal acetylation. Bottom left spectrum is identical to Fig. 4D. a.u. stands for arbitrary unit.



**Supplementary Figure. 11: SDS-PAGE analysis of thrombin inhibitors purified from LB medium.** sTyrcontaining inhibitors are expressed in LB medium with external addition of 3 mM sTyr.

	Madanin-1		Chimadanin					
Form	sTyr source	Protein Yield (mg/L)	Form	sTyr source	Protein Yield (mg/L)			
Wildtype		3.6	Wildtype		3.0			
32sTyr	Biosynthesis	0.3	28sTyr	Biosynthesis	0.11			
35sTyr		0.32	31sTyr		0.21			
32sTyr35sTyr		0.07	28sTyr31sTyr		0.08			
32sTyr	External	0.23	28sTyr	External	0.26			
35sTyr	addition	0.35	31sTyr	addition	0.55			
32sTyr35sTyr		0.04	28sTyr31sTyr		0.12			

Supplementary Figure. 12: Protein yields of all thrombin inhibitors used in this study.





Supplementary Figure. 13: ESI-MS analysis of all thrombin inhibitors used in this study. a.u. stands for arbitrary unit.

	Madanin-1			Chimadanin	
Form	sTyr source	Ki (nM)	Form	sTyr source	Ki (nM)
Wildtype		16.0 <u>+</u> 0.9	Wildtype		12.9 <u>+</u> 1
32sTyr	Biosynthesis	1.3 <u>+</u> 0.1	28sTyr	Biosynthesis	0.6 <u>+</u> 0.1
35sTyr		6.1 <u>+</u> 0.6	31sTyr		1.5 <u>+</u> 0.1
32sTyr35sTyr		0.5 <u>+</u> 0.1	28sTyr31sTyr		0.1 <u>+</u> 0
32sTyr	External	1.8 <u>+</u> 0.2	28sTyr	External	0.5 <u>+</u> 0.1
35sTyr	addition	6.0 <u>+</u> 0.3	31sTyr	addition	1.9 <u>+</u> 0.2
32sTyr35sTyr		0.3 <u>+</u> 0.1	28sTyr31sTyr		

Supplementary Figure. 14: Inhibition constants (Ki) of all thrombin inhibitors used in this study.  $K_i \pm$  standard error were calculated based on a tight-binding model, using Morrison equation in Prism. Standard errors were calculated from n=3 independent samples.



Supplementary Figure. 15: Original gel for Fig. 3F.



Supplementary Figure. 16: Original gel for Madanin-1. (Fig. 3F left and S11 left)



Supplementary Figure. 17: Original gel for Chimadanin. (Fig. 3F right and S11 right)



Supplementary Figure. 18: Original gel for Fig. S7. (The lane next to ladder is *Nn*SULT1C1, proteins in the rest lanes are not relevant to this study)

## **Supplemental Methods**

#### Materials

LB agar and 2YT were obtained from BD Difco<sup>™</sup>. Isopropyl-β-D-thiogalactoside (IPTG) was ordered from Anatrace. 4-12% Bis-Tris gels for SDS-PAGE was purchased from Invitrogen. Oligonucleotide primers were purchased from Integrated DNA Technologies and Eurofins Genomics (Supplementary Table S1 lists the oligonucleotides used in this report). Plasmid DNA preparation was carried out with the GenCatch<sup>™</sup> Plasmid

DNA Miniprep Kit and GenCatch<sup>™</sup> Advanced Gel Extraction Kit. BugBuster<sup>™</sup> Protein Extraction Reagent was obtained from Novagen (Cat. 70584). Pierce<sup>™</sup> universal nuclease was purchased from Thermo Scientific (Cat. 88700). Ni<sup>2+</sup>-NTA Agarose was obtained from Qiagen (Cat. 30230). M9-glucose minimal medium contain M9 salt (6.78 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L KH<sub>2</sub>PO<sub>4</sub>, 1 g/L NH<sub>4</sub>Cl, 0.5 g/L NaCl), heavy metal solution (1 µg/L CuSO<sub>4</sub>·5H<sub>2</sub>O, 4 µg/L MnCl<sub>2</sub>·4H<sub>2</sub>O, 4 µg/L ZnCl<sub>2</sub>, 1.2 µg/L FeSO<sub>4</sub>·5H<sub>2</sub>O), 1 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 5 µg/mL Thiamine, 300 µM Leucine 4 µM D-Biotin, Glucose (4 g/L).

Unless otherwise mentioned, all solvents and chemicals for synthesis were purchased from Alfa Aesar and Fisher Chemical and used as received without further purification, unless otherwise specified.

#### **Plasmids Construction**

pUltra-sTyrRS and pAcBac2.tR4-OMeYRS/GFP\* were obtained from addgene. pLei-sfGFP134TAG and pEvol-Mj are generous gifts from Dr. Peter Schultz. Piggybac vector and Piggybac transposase plasmids are kind gifts from Dr. Caleb Bashor. The sequences of all DNA oligos used in this study were shown in supplementary table file.

Sulfotransferase-containing pBad plasmids for initial screening were constructed by Gibson Assembling of pBad vector amplified from pBad-HER2-ScFv with Da343&344 and sulfotransferase fragment amplified from synthetic DNA. pEvol-*Nn*SULT1C1 was acquired by Gibson Assembling of pEvol-Mj vector amplified by Da443&444 and *Nn*SULT1C1 amplified from pBad-*Nn*SULT1C1with Da441&442. To generate pEvol-cysDNC, cysDNC cassette was amplified from *E.coli* genome with Da446 and Da447 and inserted into pEvol vector amplified from pEvol-Mj with Da443&445. To generate pEvol-cysDNCQ, cysDNC cassette and cysQ cassette amplified separately from *E.coli* genome with Da446&448 and Da449&450, respectively, were overlapped and Gibson Assembled into pEvol vector amplified from pEvol-Mj with Da443&445. pEvol-*Nn*SULT1C1-cysDNC(Q) were generated by inserting *Nn*SULT1C1 fragment amplified from pBad-*Nn*SULT1C1 into the vector of pEvol-cysDNC(Q) amplified with Da444&454.

To generate the plasmid for sTyr genetic incorporation in mammalian cells, OMeYRS of pAcBac2.tR4-OMeYRS/GFP\* was substituted with sTyr-selective EcTyrRS mutant (L71V, D182G, L186M), which was achieved by Gibson Assembling of synthetic sTyrRS amplified with Da462&463 and pAcBac2.tR4-OMeYRS/GFP\* vector digested by Xho1 and Nhe1. To integrate *Nn*SULT1C1 into genome of HEK293T, PB-*Nn*SULT1C1 was constructed by Gibson assembling synthetic *Nn*SULT1C1 amplified by Da655&656 and Piggybac vector digested with BsrG1.

pET22b-T5-chi was constructed by Gibson Assembling of synthetic chimadanin sequence amplified with Da556&557 and pET22b-T5-sfGFP151TAG vector digested with Hind3 and Nde1. pET22b-T5-chi-28TAG and pET22b-T5-chi-31TAG were made according to the protocol on NEBaseChanger with Da664&665 and Da584&585, respectively, using pET22b-T5-chi as their template. pET22b-T5-chi-28TAG31TAG were made by following the same protocol with Da666&Da664 with pET22b-T5-chi-31TAG as a template.

pET22b-T5-mad was constructed by Gibson Assembling of synthetic madanin-2 sequence amplified with Da558&559 and pET22b-T5-sfGFP151TAG vector digested with Hind3 and Nde1. pET22b-T5-mad-32TAG and pET22b-T5-mad-35TAG were made according to the protocol on NEBaseChanger with Da586&587 and

Da661&662 with pET22b-T5-mad as a template. pET22b-T5-mad-32TAG35TAG were made by following the same protocol with Da663&Da661 using pET22b-T5-mad-32TAG as a template.

To explore whether sulfotransferase with similar structure with *Nn*SULT1C1 could catalyze tyrosine sulfation, pEvol-mSULT1D1/hSULT1C2-cysDNCQ was constructed by assembling synthetic sulfotransferases amplified by Da832&833/ Da838&839 with pEvol-*Nn*SULT1C1-cysDNCQ vector amplified with Da802&803.

To test the importance of *Nn*SULT1C1 loop (SIQEPPAAS) and residues (T30, I33, W93, E161) involved in substrate binding, their corresponding mutants was obtained by Gibson Assembling pEvol-*Nn*SULT1C1-cysDNCQ fragments amplified with Da858&859, Da871&872, Da873&874, Da875&876, Da877&878.

To express NnSULT1C1 with his6 tag at C terminal for its kinetics measurement, *Nn*SULT1C1 was amplified from pEvol-*Nn*SULT1C1-cysDNCQ with Da687&688 and Gibson Assembled with pET22b-T5-sfGFP151TAG digested with Hind3 and Nde1, which yields pET22b-T5-*Nn*SULT11-His6.

#### Expression and Purification of Proteins.

*E. coli* BL21(DE3) cells, transformed with pUltra-sTyrRS, pLei-sfGFP134TAG, and pBad-Empty/pBad-*Hs*SULT1A1/pBad-*Hs*SULT1A3/pBad-*Rn*SULT1A1/pBad-*Gg*SULT1C1/pBad-*Cs*SULT1C2, were grown in 2YT medium at 37°C. The protein expression was carried out in Luria-Bertani (LB) medium with or without 1 mM sTyr addition. When the OD600 of the cell culture reached 0.6, protein expression was induced by the addition of IPTG and *I*-arabinose to a final concentration of 1 mM and 0.2%, respectively. After growth overnight at 30 °C. Cells were harvested by centrifugation at 4,750 × g for 10 min and used for GFP fluorescence and cell optical density measurements. (**Fig.S1**)

BW25113, ΔtrpE BW25113, ΔtyrA BW25113, ΔackA BW25113, ΔptsH BW25113, ΔcysH BW25113 cells transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-*Nn*SULT1C1/pEvol-Empty, were grown in 2YT medium at 37°C. The protein expression was carried out in Luria-Bertani (LB) medium with or without 1 mM sTyr addition. When the OD600 of the cell culture reached 0.6, protein expression was induced by the addition of IPTG and *l*-arabinose to a final concentration of 1 mM and 0.2%, respectively. After growth overnight at 30 °C. Cells were harvested by centrifugation at 4,750 × g for 10 min and used for GFP fluorescence and cell optical density measurements. (**Fig. 3B**)

 $\Delta$ cysH BW25113, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-Empty/pEvol-*Nn*SULT1C1/pEvol-*Nn*SULT1C1-cysDNC/pEvol-*Nn*SULT1C1-cysDNCQ, were grown in 2YT medium at 37°C. The protein expression was carried out in Luria-Bertani (LB) medium with or 1 mM sTyr addition. When the OD600 of the cell culture reached 0.6, protein expression was induced by the addition of IPTG and I-arabinose to a final concentration of 1 mM and 0.2%, respectively. After growth overnight at 30 °C. cells were harvested by centrifugation at 4,750 × g for 10 min and used for GFP fluorescence and cell optical density measurements. (**Fig. 3C**)

ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-Empty/pEvol-*Nn*SULT1C1-cysDNCQ, were grown in 2YT medium at 37°C. The protein expression was carried out in LuriaBertani (LB) medium. When the OD600 of the cell culture reached 0.6, *Nn*SULT1C1 expression was induced by indicated concentration of *l*-arabinose and grown at 30°C for 6 h. Then the cells were diluted 5 times to OD 0.6. Expression of reporter sfGFP and sTyrRS were induced with 1 mM IPTG and indicated concentration of sTyr was added at same time. Additional *l*-arabinose was also added to maintain its indicated concentration. After growth at 30 °C for 18 hours, cells were harvested by centrifugation at 4,750 × g for 10 min and used for GFP fluorescence and cell optical density measurements. (**Fig. 3D**) Proteins were purified on Ni-NTA resin (Qiagen) following the manufacturer's instructions. The purified protein was used for SDS-PAGE and ESI-MS analysis. (**Fig. 3F-H**)

ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-Empty/pEvol-*Nn*SULT1C1-cysDNCQ, were grown in 2YT medium at 37°C. The protein expression was carried out in Luria-Bertani (LB) medium. When the OD600 of the cell culture reached 0.6, *Nn*SULT1C1 expression was induced by 15 mg/L *l*-arabinose and grown at 30°C for 6 h. Then the cells were diluted 5 times to OD 0.6. Expression of reporter sfGFP and sTyrRS were induced with 1 mM IPTG and indicated concentration of sTyr was added at same time. Additional *l*-arabinose was also added to maintain its final concentration of 15 mg/L. After growth overnight at 30 °C for 18 hours, cells were harvested by centrifugation at 4,750 × g for 10 min and used for measuring cellular sTyr concentration. (**Fig. 3E**)

To express wildtype thrombin inhibitors, BL21(DE3) cells transformed with either pET22b-T5-chi or pET22b-T5mad were grown in 2YT medium at 37°C. The protein expression was carried out in LB medium. When the OD600 of the cell culture reached 0.6, protein expression was induced by the addition of 0.4 mM IPTG. After growth overnight at 18 °C for 18 hours, cells were harvested by centrifugation at 4,750 × g for 10 min. Proteins were purified on Ni-NTA resin (Qiagen) following the manufacturer's instructions. The purified protein was used for SDS-PAGE and ESI-MS analysis. (**Fig. 5C**)

To express thrombin inhibitors containing sTyr, ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-inhibitor-X-TAG, and pEvol-*Nn*SULT1C1-cysDNCQ, were grown in 2YT medium at 37°C. In the control group, ΔcysH BW25113 cells were transformed with pUltra-sTyrRS, pET22b-T5-inhibitor-X-TAG, and pEvol-Empty. When the OD600 of the cell culture reached 0.6, *Nn*SULT1C1 expression was induced by 15 mg/L concentration of *I*-arabinose and grown at 30°C for 6 h. Then the cells were diluted 5 times to OD 0.6. Expression of inhibitor and sTyrRS were induced with 1 mM IPTG and 3 mM sTyr was added to only control cells at same time. Additional *I*-arabinose was also added to maintain its final concentration of 15 mg/L. After growth overnight at 18 °C for 18 hours, cells were harvested by centrifugation at 4,750 × g for 10 min. Proteins were purified on Ni-NTA resin (Qiagen) following the manufacturer's instructions. The purified protein was used for SDS-PAGE and ESI-MS analysis. (**Fig. 5C and S11**)

To test the importance of the variable loop and residues in binding pockets, ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-*Nn*SULT1C1-cysDNCQ with indicated sequence mutations were grown in 2YT medium at 37°C. To test the tyrosine sulfation activity of the top 3 sulfotransferases with similar structure to *Nn*SULT1C1, ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-X sulfotransferase-cysDNCQ were grown in 2YT medium at 37°C. The protein expression was carried out in LB medium. When the OD600 of the cell culture reached 0.6, sulfotransferase expression was induced by 15 mg/L *I*-arabinose and grown at 30°C for 6 h. Then the cells were diluted 5 times

to OD 0.6. Expression of sfGFP and sTyrRS were induced with 1 mM IPTG. Additional *l*-arabinose was also added to maintain its final concentration of 15 mg/L. After growth overnight at 30 °C for 18 hours, cells were harvested by centrifugation at  $4,750 \times g$  for 10 min and used for GFP fluorescence and cell optical density measurements. (Fig. 2B, C, F)

To express *Nn*SULT1C1 for kinetics measurment, BL21(DE3) cell transformed with pET22b-T5-*Nn*SULT11-His6 was grown in 2YT medium at 37°C. The protein expression was carried out in Luria-Bertani (LB) medium. When the OD600 of the cell culture reached 0.6, protein expression was induced by the addition of 0.3 mM IPTG. After growth overnight at 30 °C for 18 hours, cells were harvested by centrifugation at 4,750 × g for 10 min. Proteins were purified on Ni-NTA resin (Qiagen) following the manufacturer's instructions. The purified protein was used for SDS-PAGE and kinetic assay. (**Fig. S7**)

#### Expression and Fluorescence Measurement of sfGFP

After sfGFP expression with the methods described above, 0.5 mL cells were harvested by centrifugation at 4,750 × g for 10 min and then suspended with 0.5 ml PBS (pH 7.4). Fluorescence of cells was measured using excitation/emission wavelengths of 395/509 nm. Optical Density at 600 nm was also obatined. The sfGFP fluorescence/OD600 was used as the normalized fluorescence. The error bars represent the standard deviations of 3 independent protein expression trials.

#### E. coli Intracellular sTyr Concentration Measurement

Cells were harvested by centrifugation at 4,750 × g for 10 min and washed with PBS 7.4 for three times. The cell pellets were re-suspended in 300  $\mu$ L of bugbuster lysis buffer : toluene (80: 20) solution and shaken at 30 °C for 1 h. The resulting lysate was centrifuged at 21000 g for 30 min at 4 °C. 200  $\mu$ l supernatant was transferred to a new tube and re-centrifuged at 21000 g for 2 h. 50  $\mu$ l supernatant from the top was then analyzed using the LC-MS. An Agilent 1260 Infinity II LC System coupled with Single Quadrupole ESI-MS System was used for analysis of all samples. To measure the sTyr ions, ions detected were set to selected ion monitoring (SIM) mode (262 m/z) to detect positive ions of sTyr. Standards of 1  $\mu$ M, 25  $\mu$ M, 50  $\mu$ M, 100  $\mu$ M, 200  $\mu$ M, and 400  $\mu$ M of authentic sTyr (Bachem) were also prepared and analyzed by the same method. Using LC-MS data, a linear standard curve was generated based on peak areas corresponding to sTyr ions and the concentration of sTyr in standards. The standard curve was then used to calculate the concentration of sTyr from different cell lysates. Each sample was carried out in n=3 independent groups. The intracellular concentration of sTyr in cells was calculated based on the following equation.

$$[sTyr intracellular] = \frac{sTyr \ concentration \ in \ lysate \ \times \ volumn \ of \ lysate}{total \ cell \ numbers \ \times \ E. \ coli \ cell \ volumn}$$

Total cell numbers were calculated with the approximate values:  $8 \times 10^8$  cells per OD600. 0.6 fL was used as an average *E. coli* cell volume.

#### Exploration on evolutionary relationship of NnSULT1C1

The rooted phylogenetic tree was inferred using the UPGMA method in MEGA X software. The UPGMA algorithm constructs the tree that reflects the genetic distance between protein sequences present in a pairwise similarity matrix. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson

correction method and are in the units of the number of amino acid substitutions per site. 10 sequences from bottom branch were used for Multiple sequence alignment (MSA) on https://www.ebi.ac.uk/Tools/msa/ with Maftt method. The alignment result was visulized in Jalview software. The sequence consensus was analyzed at https://weblogo.berkeley.edu. (**Fig. S2**)

#### Kinetics Measurement of NnSULT1C1

The purified *Nn*SULT1C1 was buffer exchanged to 10 mM NH<sub>4</sub>OAc buffer pH 8 via PD-10 column. The concentration of enzyme was calculated based on its Absorption at 280 nm. The enzymatic reaction was performed in 100 µL 10 mM NH<sub>4</sub>OAc buffer pH 8 including 1 µM *Nn*SULT1C1, 20 µM PAPS, 5 mM CaCl<sub>2</sub> and variable concentration of tyrosine. The reactions were incubated at 37°C and quenched with 100 µL ACN at 5 minutes. The supernatants of these mixtures were used for sTyr quantification via ESI-MS. To improve the sensitivity of sTyr detection, Selected Ion Monitoring (SIM) mode was used to detect the negative ion (m/z=260) under 50 °C drying gas temperature and 2400 V capillary voltage. The standard curve of authentic sTyr was prepared under the same condition, which yields the linear relationship between area under the curve and its sTyr concentration. To obtain the concentration of sTyr produced in enzymatic reaction, its area under the curve was used to calculate sTyr concentration based on the equation obtained from standard curve. Each sample was carried out in n=3 independent samples. The data was fitted to Michaelis-Menten equation in Prism.

#### **Protein Purification from Mammalian Cells**

To confirm the genetic incorporation of sTyr from either biosynthesis or external addition, HEK293T and HEK293T-*Nn*SULT1C1 cells were transfected with pAcBac2.tR4-sTyrRS/GFP\* with Polyjet In Vitro DNA Transfection Reagent (SignaGen Laboratories) in the presence or absence of 3 mM sTyr addition. Mediums were changed at 12-16 hour after transfection. After 48 hours of transfection, cells were harvested with trypsin and subsequently washed by DPBS for 3 times. Cells were lysed using the Mammalian Cell PE LB reagent (G-Bioscience) according to its manual. The cell lysates were centrifuged at 15,000 rpm for 10 minutes. The protein in the supernatant was purified from the supernatant using Ni-NTA resin (Qiagen) following the manufacturer's instruction. The purified protein was used for ESI-MS analysis.

#### Mammalian Cell sTyr Concentration Measurement

HEK293T and HEK293T-*Nn*SULT1C1 were detached from plate with trypsin and washed with DPBS for 3 times. The number of cells was counted by hemocytometer. Cells were resuspended in 0.5 mL methanol-water (2:3) and lyzed by six freeze-thaw cycles. The resulting cell lysates were centrifuged at 21000 × g for 1h at 4°C. The resulting supernantants were injected to LC-MS for the quantification of sTyr with selected ion monitoring (SIM) mode. An Agilent 1260 Infinity II LC System coupled with Single Quadrupole ESI-MS System was used for the analysis of all samples. To measure the sTyr ions, ions detected were set to Selected Ion Monitoring (SIM) mode (262 m/z) to detect positive ions of sTyr. Standards of 1  $\mu$ M, 25  $\mu$ M, 50  $\mu$ M, 100  $\mu$ M, 200  $\mu$ M, and 400  $\mu$ M of purchased sTyr (Bachem) dissolved in methanol-water (2:3) were also prepared and analyzed by the same method. A linear standard curve was generated based on peak areas corresponding to sTyr ions and the concentration of sTyr in standard samples. The standard curve was then used to calculate the concentration of

sTyr from different cell lysates. The intracellular concentration of sTyr in cells was calculated with the following equation.

$$[sTyr intracellular] = \frac{sTyr \ concentration \ in \ lysate \ \times \ volumn \ of \ lysate}{total \ cell \ numbers \ \times \ cell \ volumn}$$

2 pL was used as an average volume of mammalian cells.

### Mass Spectra Methods For Proteins

A single quadrupole mass spectrometer (Agilent: G7129A) coupled with 1260 infinity II Quaternary Pump (Agilent: G7111B) was used for all the protein samples with PLRP-S (1000A, 5 μm) column. Water with 0.1% formic acid and ACN with 0.1% formic acid were the organic and aqueous mobile phase, respectively. Flow gradient was initially set at 5% ACN, 15% ACN at 0.1 min, 55% ACN at 4.5 min and then back to 10% ACN at 5 min. Spectra were deconvoluted using the Maximum Entropy deconvolution algorithm in the software BioConfirm.