nature communications

Article

https://doi.org/10.1038/s41467-022-33111-4

Unleashing the potential of noncanonical amino acid biosynthesis to create cells with precision tyrosine sulfation

Received: 21 March 2022

Accepted: 1 September 2022

Published online: 16 September 2022

Check for updates

Yuda Chen[1], Shikai Jin[2,3], Mengxi Zhang[1], Yu Hu[1], Kuan-Lin Wu[1], Anna Chung[1], Shichao Wang[1], Zeru Tian[1], Yixian Wang[1], Peter G. Wolynes[1,2,3,4] & Han Xiao[1,3,5]

Despite the great promise of genetic code expansion technology to modulate structures and functions of proteins, external addition of ncAAs is required in most cases and it often limits the utility of genetic code expansion technology, especially to noncanonical amino acids (ncAAs) with poor membrane internalization. Here, we report the creation of autonomous cells, both prokaryotic and eukaryotic, with the ability to biosynthesize and genetically encode sulfotyrosine (sTyr), an important protein post-translational modification with low membrane permeability. These engineered cells can produce site-specifically sulfated proteins at a higher yield than cells fed exogenously with the highest level of sTyr reported in the literature. We use these autonomous cells to prepare highly potent thrombin inhibitors with site-specific sulfation. By enhancing ncAA incorporation efficiency, this added ability of cells to biosynthesize ncAAs and genetically incorporate them into proteins greatly extends the utility of genetic code expansion methods.

With the rare exceptions of pyrrolysine and selenocysteine, a standard set of 20 amino acid building blocks, containing a limited number of functional groups, is used by almost all organisms for the biosynthesis of proteins. The use of Genetic Code Expansion technology to enable the site-specific incorporation of noncanonical amino acids (ncAAs) into proteins in living cells has transformed our ability to study biological processes and provided the exciting potential to develop modern medicines[1–8]. The genetic encoding of ncAAs with distinct chemical, biological, and physical properties requires the engineering of bioorthogonal translational machinery, consisting of an evolved aminoacyl-tRNA synthetase/tRNA pair and a "blank" codon[1,6,9,10]. The high intracellular concentration of ncAA required to render this machinery operative has usually been achieved via chemical synthesis of the ncAA and its exogenous addition at high levels to the cell culture medium. Although most ncAAs could penetrate cell membrane for the genetic incorporation, ncAAs bearing negative charges or polar structures normally exhibit a low cell penetration efficiency[11–14]. The relatively low intracellular concentrations of these ncAAs greatly limit the efficiency of ncAA incorporation into proteins using the Genetic Code Expansion technology[12,13,15–18].

Strategies for engineering the structures of ncAAs or ncAA-binding proteins have been employed to improve the cellular uptake of ncAAs. In 2017, the Schultz group adopted a dipeptide strategy to enable the cellular uptake of phosphotyrosine. The phosphotyrosine-containing dipeptide can be synthesized and transported into cells via an adenosine triphosphate (ATP)-binding cassette transporter, followed by hydrolysis of the dipeptide by nonspecific intracellular peptidases[13]. In the same year, the Wang lab developed a two-step strategy for producing proteins with site-specific tyrosine phosphorylation[14]. This strategy utilized the incorporation of a phosphotyrosine anologue with a cage group, followed by chemical deprotection of the purified proteins. However, the synthesis and

[1]Department of Chemistry, Rice University, 6100 Main Street, Houston, TX 77005, USA. [2]Center for Theoretical Biological Physics, Rice University, 6100 Main Street, Houston, TX 77005, USA. [3]Department of Biosciences, Rice University, 6100 Main Street, Houston, TX 77005, USA. [4]Department of Physics, Rice University, 6100 Main Street, Houston, TX 77005, USA. [5]Department of Bioengineering, Rice University, 6100 Main Street, Houston, TX 77005, USA. e-mail: han.xiao@rice.edu

purification of these dipeptides are challenging, and the required post-purification treatments limit the applicability of this methodology to efficiently incorporate phosphotyrosine in living cells. As an alternative approach, periplasmic binding proteins (PBPs) have been engineered to have improved affinities for specific ncAAs[16]. These mutant PBPs enhanced uptake of the respective ncAAs up to fivefold, as evidenced by elevated intracellular ncAA concentrations and the yield of ncAA-containing green fluorescent proteins[16]. Nevertheless, the engineered PBP species are only applicable to a subset of ncAAs, and exogenous feeding of high concentrations of the ncAAs is still required. The problem of ncAA uptake could potentially be bypassed by intracellular biosynthesis of the ncAAs from basic carbon sources[12,19–25]. For example, phosphothreonine (*p*Thr) cannot be detected intracellularly even when cells are incubated with 1 mM *p*Thr[12]. The Chin group overcame the membrane impermeability of *p*Thr by introducing the *Salmonella enterica* kinase, PduX, which converts *L*-threonine to *p*Thr intracellularly[12]. This biosynthesis of *p*Thr generated intracellular *p*Thr at levels >1 mM, sufficient for genetic incorporation of this amino acid[12]. A similar strategy was recently applied to the creation of autonomous bacterial cells that can biosynthesize and genetically incorporate *p*-amino-phenylalanine (*p*AF), 5-hydroxyl-tryptophan (5HTP) and dihydroxyphenylalanine (DOPA), although no autonomous eukaryotic cells have been reported[19,20,22,23,26]. We see there that additional biosynthetic pathways for producing polar or negatively-charged ncAAs would greatly expand the utility of genetic code expansion methods.

Tyrosine sulfation is an important post-translational modification of proteins that is essential for a variety of biomolecular interactions, including chemotaxis, viral infection, anti-coagulation, cell adhesion, and plant immunity[27–34]. Despite its importance and ubiquity, protein sulfation has been difficult to study due to the lack of general methods for preparing proteins with defined sulfated residues[31,35]. To circumvent this challenge, efforts have been previously made to site-specifically incorporate sulfotyrosine (sTyr) using the Genetic Code Expansion technology[36]. The resulting sTyr incorporation systems have enabled several applications, including generation of therapeutic proteins with defined sulfated tyrosines, evolution of sulfated anti-gp120 antibodies, and confirmation of tyrosine sulfation sites[35,37–40]. To achieve reasonsble expression levels of sulfated proteins in *E. coli*, however, most studies have required the exogenous feeding of 3–20 mM sTyr to compensate for low intracellular uptake of extracellular sTyr[37,40].

Here, we report the generation of metabolically modified prokaryotic and eukaryotic cells that can biosynthesize sTyr and incorporate it into proteins in a site-specific manner (Fig. 1a). sTyr is biosynthesized using a sulfotransferase discovered from a sequence similarity network (SSN). sTyr is subsequently incorporated into proteins in response to a repurposed stop codon. The molecular properties of this the sulfotransferase were explored using bioinformatics and computational approaches, revealing a loop structure and several residues in binding pocket within this enzyme responsible for its unique specificity for tyrosine. The further optimization of the genome and sTyr biosynthetic pathway of both prokaryotic and eukaryotic cells leads to greater expression yields of sulfated proteins than cells exogenously fed with sTyr. The utility of these sTyr autonomous cells is demonstrated by using them to produce highly potent thrombin inhibitors.

Results

Discovery of tyrosine sulfotransferase using a sequence similarity network

In nature, sulfotransferases allow many organisms to utilize an active form of sulfate, 3′-phosphoadenosine-5′-phosphosulfate (PAPS), for biosynthetic purposes[41,42]. Based on their substrate preference and cellular location, sulfotransferases can be grouped into three major families, tyrosylprotein sulfotransferase (TPST), cytosolic sulfotransferase (SULT), and carbohydrate sulfotransferase[43,44]. To identify the enzyme responsible for sulfation of cytoplasmic tyrosine, we focused on SULTs. These enzymes catalyze sulfation of a wide variety of endogenous compounds, including hormones, neurotransmitters, and xenobiotics[43]. Based on their reported substrate specificities, we examined SULT1A1 and SULT1A3 from *Homo sapiens*, SULT1A1 from *Rattus norvegicus*, and SULT1C1 from *Gallus gallus*[43,45], all of which are known to recognize multiple phenolic substrates. To explore the activity of these sulfotransferases toward tyrosine, we used a green fluorescent protein assay[20,46]. These four sulfotransferase genes were codon-optimized for *Escherichia coli* and cloned into the pBad vector with DNA oligos in Supplementary Data 1. To generate a suppression plasmid for sTyr incorporation, we used pUltra-sTyr plasmid encoding the engineered *Methanococcus jannaschii* tyrosyl-tRNA synthetase (sTyrRS) and its corresponding *Mj*tRNA$^{Tyr}_{CUA}$[36,40]. The suppressor plasmid (pUltra-sTyr) was used to suppress the amber codon (Asp134TAG) within a sfGFP variant encoded by the pLei-sfGFP134TAG plasmid in the presence of sTyr. Expression of full-length sfGFP was carried out in LB medium for 16 h in parallel with controls BL21(DE3) harboring pUltra-sTyr, pLei-sfGFP134TAG and pBad-Empty in the presence and absence of exogenously fed 1 mM sTyr. As expected, sfGFP was expressed in the presence of 1 mM sTyr fed in controls cells (Supplementary Fig. 1). Unfortunately, none of these four sulfotransferases led to sfGFP expression, indicating the failure of the biosynthesis of sTyr. To circumvent the limited substrate range of the reported sulfotransferases, we accessed the full repertoire of protein sequence diversity in nature by using a sequence similarity network (SSN, Fig. 1b)[47]. SSNs provide an effective way to visualize and analyze the relatedness of massive protein sequences on the basis of similarity thresholds of their amino acid sequences[48]. We initially created an SSN with EFI-ESI based on SULT1A1 from *Rattus norvegicus* as an input sequence, since its cognate substrate *p*-coumaric acid is similar to tyrosine (Fig. 1b, c)[45]. An alignment score of 110 was set to limit the edges and a sequence identity of 80% was used to generate representative nodes, which resulted in a final SSN of 391 representative enzyme sequences. Interestingly, we found that human SULT1C2, whose substrate is tyramine, was in a different cluster of the SSN (Fig. 1b, c)[43]. We hypothesized that enzymes with high sequence similarity to SULT1A1 from *Rattus norvegicus* and SULT1C2 from *Homo sapiens* would be potential candidates to carry out the sulfation of tyrosine. To test this hypothesis, we selected 27 sequences from the SSN based on their similarity to both *Rn*SULT1A1 and *Hs*SULT1C2. These selected genes were cloned into the pBad vector and tested with the green fluorescent protein assay. To our delight, a 2.5-fold increase in fluorescence was observed for cells expressing A0A091VQH7 compared to cells not given exogenous sTyr, suggesting that sTyr was biosynthesized intracellularly and incorporated into sfGFP proteins (Fig. 1d). A0A091VQH7 is a putative sulfotransferase from *Nipponia nippon*, with over 90% sequence identity with SULT1C1 reported in other species. Thus, we name A0A091VQH7 as *Nn*SULT1C1 hereafter[49].

Molecular basis of *Nn*SULT1C1 action in the sulfation of tyrosine

To explore the origin of the unique tyrosine specificity of *Nn*SULT1C1 among all the sulfotransferases tested, we analyzed the phylogenetic relationships of the enzymes. Sulfotransferase amino acid sequences were used to generate a phylogenetic tree using the unweighted pair group method with arithmetic mean by MEGA X software package (Supplementary Fig. 2)[50]. The tree is subdivided into three major subfamilies, among which *Nn*SULT1C1 falls into subfamily I containing bird sulfotransferases. Most sequences from subfamilies II and III are derived from rodent and primate groups, respectively. To further analyze the molecular basis of the unique tyrosine specificity of *Nn*SULT1C1, we performed a multiple sequence alignment of all sequences within subfamily I of the phylogenetic tree (Supplementary

https://doi.org/10.1038/s41467-022-33111-4

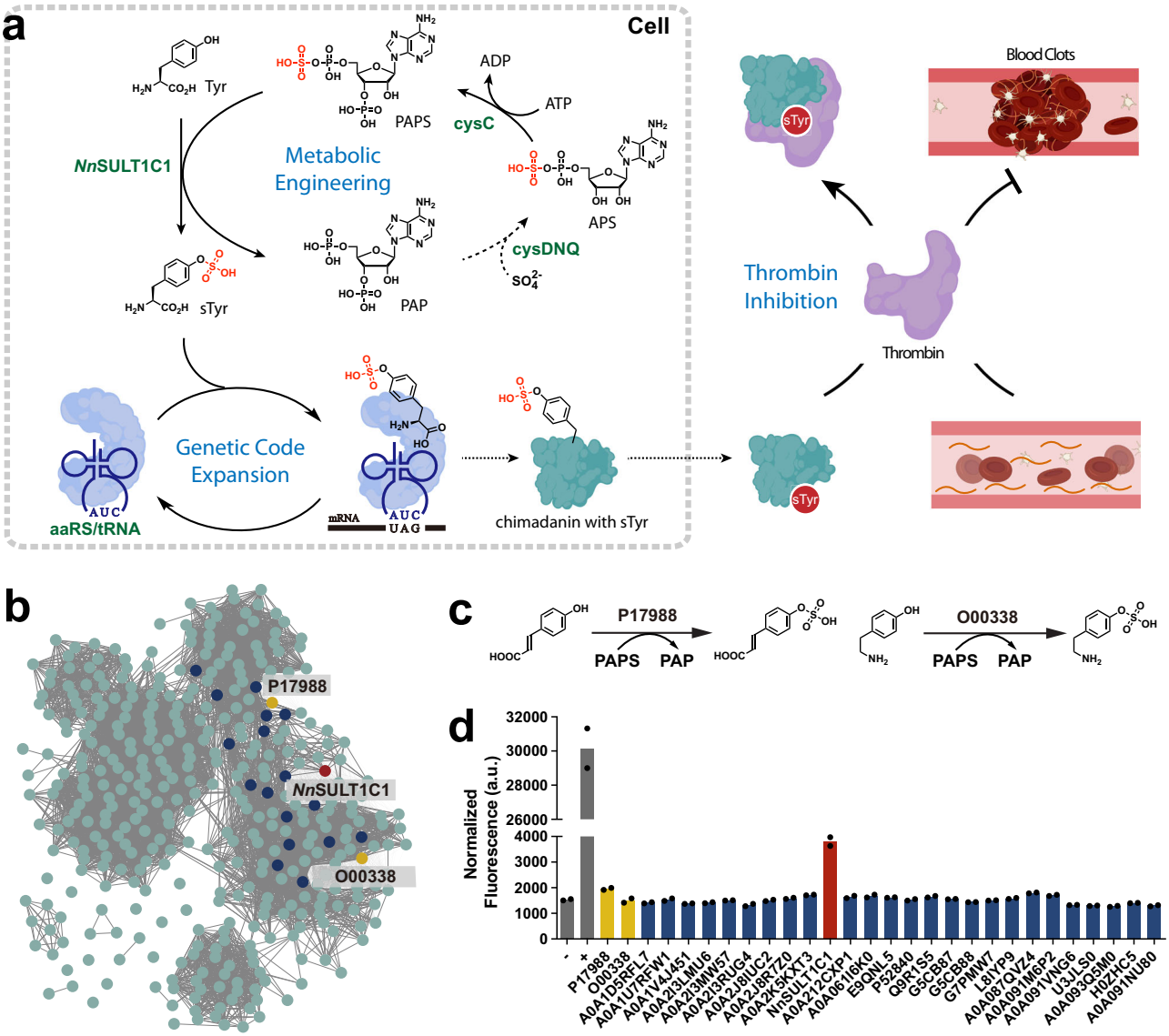
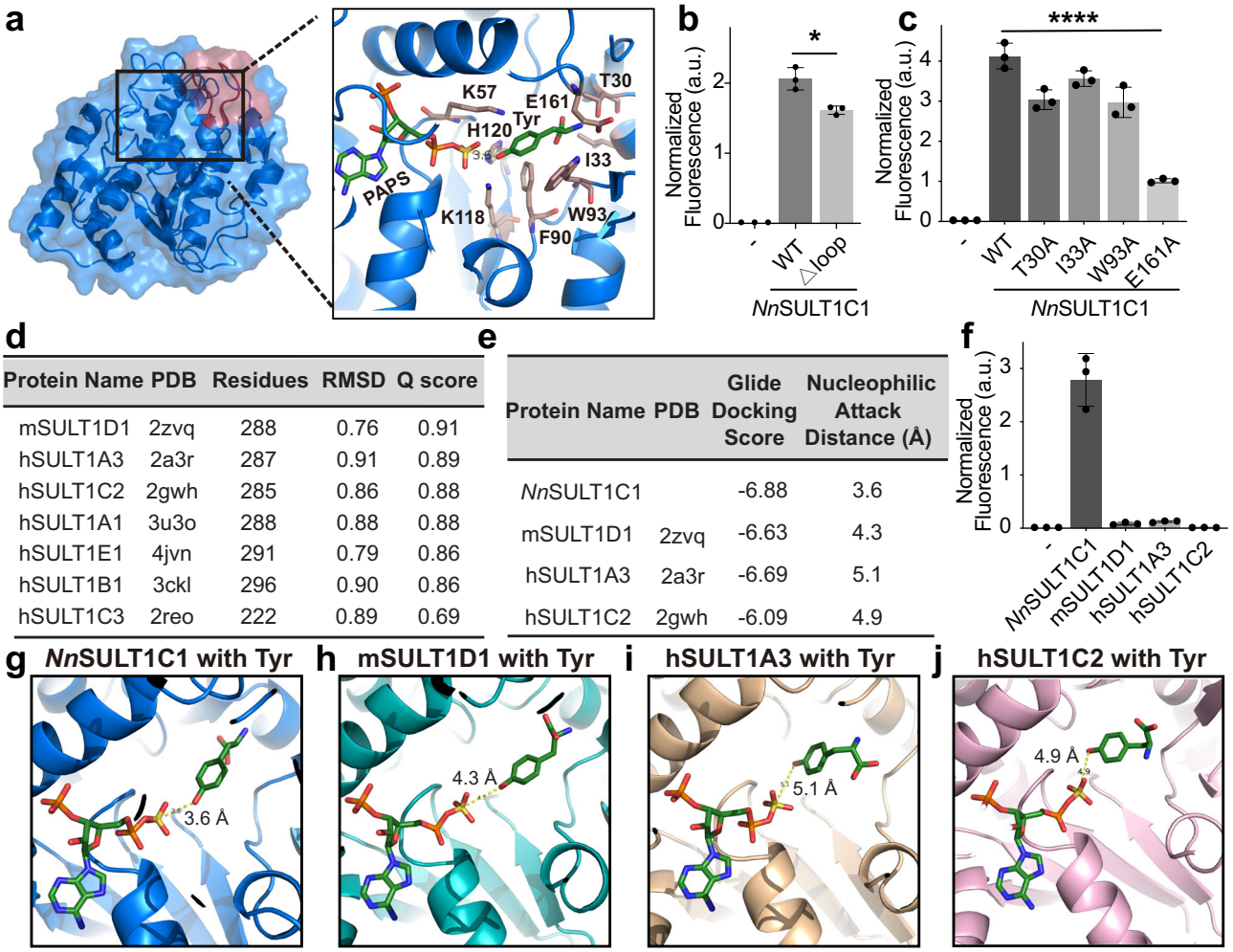


Fig. 1 | Discovery of tyrosine sulfotransferase from sequence similar network. **a** sTyr was biosynthesized from tyrosine and PAPS in the presence of sulfotransferase identified in this study. The resulting biosynthesized sTyr was site-specifically incorporated into thrombin inhibitors, yielding enhanced thrombin inhibition **b** Sequence similar network (SSN) generated by EFI-EST server with *Rn*SULT1A1 as an input sequence and *E* value of 5. Each circle stands for a representative node containing sequences with over 80% identity. Edges detection threshold was set at an alignment score of 110. The upper and lower yellow representative nodes are *Rn*SULT1A1 (P17988) and *Hs*SULT1C2 (O00338), respectively. **c** Schematic representation of reported sulfation reactions of P17988 and O00338. **d** Screening of tyrosine sulfotransferases with green fluorescent protein assay. All tested proteins are included in the representative nodes of **b** and *Nn*SULT1C1 is the protein with red label (**b**). **d** Data are plotted as the means from $n = 2$ independent samples. a.u. stands for arbitrary unit.

Fig. 3). This sequence alignment revealed that most regions of *Nn*SULT1C1, including the PAPS-binding site, are highly conserved except for a highly variable region corresponding to *Nn*SULT1C1 residues 94-102 (SIQEPPAAS) and residues likely involved in substrate binding pocket[51,52]. To explore the contribution of this highly variable region to substrate binding, the structure of *Nn*SULT1C1 was predicted via Alphafold 2. Alphafold 2 is a machine learning approach that has been shown to predict protein structure with a high degree of accuracy[53–56]. More than 90% of the residues in the predicted *Nn*SULT1C1 structure show Local Distance Difference Test values over 90, indicating they have a significant likelihood of predicting structure with a very high accuracy. Similar to the structures of other SULTs, the overall predicted structure of *Nn*SULT1C1 is composed of classical α/β motifs (Fig. 2a)[57,58]. This structure includes a β sheet surrounded by α-helices, giving rise to a narrow substrate-binding site (Fig. 2a)[59]. We found that the highly variable region (94–102 residue) of *Nn*SULT1C1 constitutes a loop for the substrate entry, which also aligns with the substrate entry loop of human SULT (Supplementary Fig. 4)[43]. The deletion of this loop on *Nn*SULT1C1, however, only results in 22% decrease of its activity to produce fluorescent protein with sTyr (Fig. 2b).

To further explore the other residues involved in substrate binding of *Nn*SULT1C1, we performed protein-ligand docking using Glide v8.1 in Schrödinger software package v2018.4[60]. The Tyr was docked to the *Nn*SULT1C1 using OPLS_3 force field and the lowest energy pose was monitored[61]. For each docking experiment, 200 maximum output poses for each protein were set and Emodel energy was used for ranking the top 50 poses. The docking structure suggests that the α-amino group of Tyr is stabilized by *Nn*SULT1C1 residues Glu161, Thr30, Ile33, and Trp93. The π-π stacking interactions between Tyr and Phe90 are likely to improve the packing interaction (Fig. 2a). The phenolic hydroxy group of Tyr is in the proper Lys-Lys-His catalytic site to

https://doi.org/10.1038/s41467-022-33111-4



d

Protein Name	PDB	Residues	RMSD	Q score
mSULT1D1	2zvq	288	0.76	0.91
hSULT1A3	2a3r	287	0.91	0.89
hSULT1C2	2gwh	285	0.86	0.88
hSULT1A1	3u3o	288	0.88	0.88
hSULT1E1	4jvn	291	0.79	0.86
hSULT1B1	3ckl	296	0.90	0.86
hSULT1C3	2reo	222	0.89	0.69

e

Protein Name	PDB	Glide Docking Score	Nucleophilic Attack Distance (Å)
*Nn*SULT1C1		-6.88	3.6
mSULT1D1	2zvq	-6.63	4.3
hSULT1A3	2a3r	-6.69	5.1
hSULT1C2	2gwh	-6.09	4.9

Fig. 2 | Exploring the mechanism of unique tyrosine specificity of *Nn*SULT1C1. **a** *Nn*SULT1C1 structure (blue) predicted by AlphaFold2 and its active site consisting PAPS and Tyr. Tyr was docked into *Nn*SULT1C1 containing PAPS by Glide v8.1 in Schrödinger software. **b** Green fluorescent protein assay with wildtype *Nn*SULT1C1 (wt) or *Nn*SULT1C1 without the SIQEPPAAS (Δloop), $p = 0.01$. **c** Green fluorescent protein assay with wildtype *Nn*SULT1C1 (wt) or *Nn*SULT1C1 with alanine mutation at indicated resiudes, $p = 0.004$. **d** Structural similarity search of *Nn*SULT1C1 using the PDBeFold web server. **e** Characterization of Tyr docking with *Nn*SULT1C1 and its structurally similar sulfotransferases via docking score and nucleophilic attack distance. Docking scores were calculated using Glide v8.1 in Schrödinger software. Nucleophilic attack distance was defined as the distances between Tyr phenolic alcohol and PAPS sulfonate. **f** Comparison of tyrosine sulfation activity of *Nn*SULT1C1 and its structurally similar sulfotransferases using green fluorescent protein assay. Cells without any sulfotransferase (−) were used as control. **g**–**j** Tyr docking position with *Nn*SULT1C1 (**g**), mSULT1D1 (**h**), hSULT1A3 (**i**), and hSULT1C2 (**j**). PAPS and Tyr are shown as sticks with green carbon. Docking was performed by Glide v8.1 in Schrödinger software with the same parameters in **a**. **b**, **c**, **f** Data are plotted as the mean ± standard deviation from $n = 3$ independent samples. **b**, **c** Two-sided unpaired *t*-tests were performed with $^{*}p < 0.05$; $^{**}p < 0.01$. a.u. stands for arbitrary unit.

engage in sulfuryl transfer. The His120 residue serves as a catalytic base that can remove the proton from Tyr. The Lys57 and Lys118 residues interact with and stabilize the sulfuryl group of PAPS and the phenolic hydroxy group of Tyr, respectively. To validate the contribution of these residues interacting with Tyr on *Nn*SULT1C1 activity, Thr30, Ile33, Trp93, and Glu161 were mutated to alanine separately. Alanine mutation at Thr30, Trp93, or Glu161 significantly decreased the activity of *Nn*SULT1C1 (Fig. 2c). Among these residues, the E161A mutation exhibits the largest decrease in activity, confirming its important interaction with Tyr. To further explore whether other sulfotransferases may also carry out the tyrosine sulfation, we performed a structure similarity search using the PDBeFold (https://www.ebi.ac.uk/msd-srv/ssm/). Based on the Q score, the three proteins with structures most similar to *Nn*SULT1C1 are mouse SULT1D1 (pdb: 2zvq, https://www.rcsb.org/structure/2ZVQ), human SULT1A3 (pdb: 2a3r, https://www.rcsb.org/structure/2A3R) and human SULT1C2 (pdb: 2gwh, https://www.rcsb.org/structure/2GWH, Fig. 2d). The overall secondary structure of *Nn*SULT1C1 aligned well with 2zvq, which indicates its structural consistency with the other SULTs (Supplementary Fig. 5). To further illustrate the unique specificity of *Nn*SULT1C1 for Tyr, dockings of Tyr to the most similar sulfotransferases, including mSULT1D1, hSULT1A3, and hSULT1C2, were carried out using Glide v8.1 in the Schrödinger software package v2018.4 following the same method of Tyr docking used for *Nn*SULT1C1[60]. Docking of Tyr to *Nn*SULT1C1 exhibits the lowest Glide Docking score of −6.88 and the closest distance between the phenolic hydroxyl group and PAPS sulfonate (Fig. 2e). This result is consistent with the optimal ability of *Nn*SULT1C1, to generate sTyr-containing sfGFP in the green fluorescent protein assay among all tested sulfotransferases (Fig. 2f). The key step of the sulfotransfer reaction involves an S_N2-type nucleophilic attack on the PAPS sulfonate by the phenoxide of Tyr. Compared with mSULT1D1, hSULT1A3, and hSULT1C2, the docking of Tyr in *Nn*SULT1C1 results in the closest distance (3.6 Å) between the sulfur atom of PAPS and the phenolic hydroxyl group (Fig. 2g–j). Furthermore, the acceptor phenolic hydroxyl group of Tyr lies on the backside of the S-O bond of PAPS in the Tyr docking structure with *Nn*SULT1C1, indicating a more proper orientation for the nucleophilic attack (Fig. 2g).

https://doi.org/10.1038/s41467-022-33111-4

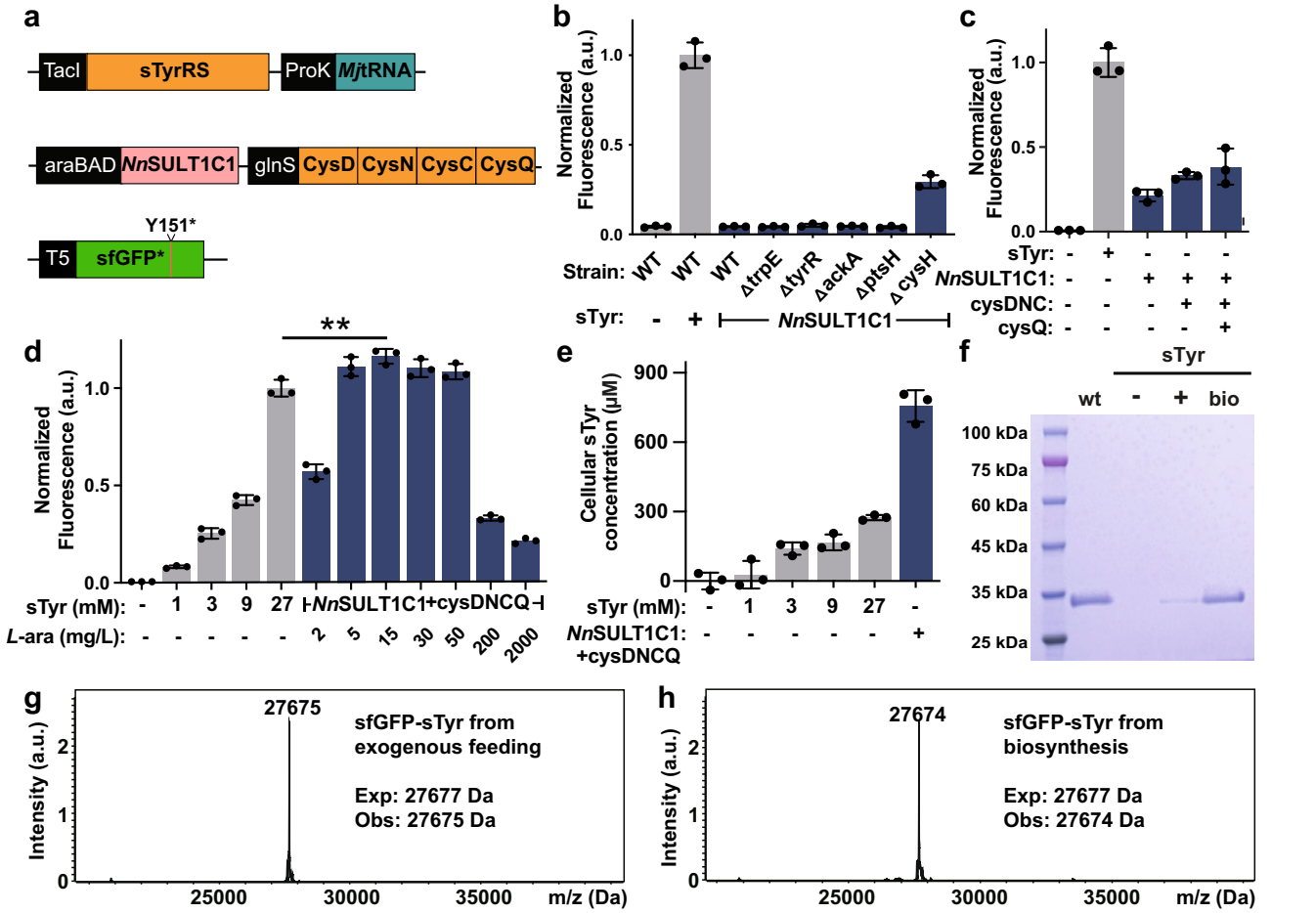


Fig. 3 | Generation of completely autonomous sTyr synthesizing *E. coli*. **a** Schematic representation of genetic circuits used for generating completely autonomous sTyr synthesizing *E. coli*. **b** Screening of the knockout strains for sfGFP-sTyr production after the expression of *Nn*SULT1C1. **c** The roles of PAPS recycling enzymes in producing sfGFP-sTyr using ΔcysH BW25113 strain. **d** Production of sfGFP-sTyr from cells with the addition of chemically synthesized sTyr or the biosynthesized Tyr. The effect of *Nn*SULT1C1 expression level on producing sfGFP-sTyr was screened by altering the concentration of inducer, *L*-arabinose (*L*-ara), $p = 0.008$. **e** Cellular concentrations of sTyr of cells with the addition of chemically synthesized sTyr or the biosynthesis of sTyr. **f** SDS-PAGE analysis of sfGFPs expressed in LB in the presence (+) or absence (−) of exogenous 1 mM sTyr or when inducing *Nn*SULT1C1 expression (bio). **g**, **h** ESI-MS analysis of sfGFP-sTyr proteins expressed in cells with the addition of 1 mM chemically synthesized sTyr or the biosynthesis of sTyr. **b**–**e** Data are plotted as the mean + standard deviation from $n = 3$ independent samples. **d** Two-sided unpaired *t*-tests were performed with $^{**}p < 0.01$. a.u. stands for arbitrary unit.

Biosynthesis and genetic encoding of sTyr in *Escherichia coli*

Having identified *Nn*SULT1C1 as a functional tyrosine sulfotransferase, we explored whether the biosynthesized sTyr can be genetically incorporated into proteins in *E. coli* in response to the amber codon. As an initial goal, we wanted to increase sTyr production in these cells in order to optimize its availability for incorporation into proteins. Since *Nn*SULT1C1 utilizes tyrosine and PAPS for producing sTyr, we quantified sTyr production in five knockout *E. coli* cell lines in which the gene knockout has been shown to improve the yield of either tyrosine or PAPS in *E. coli*[62–65]. To evaluate the effect of knocking out these genes on the biosynthesis of sTyr, we transformed the suppression plasmid pUltra-sTyr, reporter plasmid pET22b-T5-sfGFP151TAG, and the biosynthesis plasmid pEvol-*Nn*SULT1C1 into wildtype *E. coli* BW25113 or knockout strains (Fig. 3a). The expression of sfGFP with sTyr at position 151 (sfGFP-sTyr) was carried out in LB medium for 18 h. To our delight, we found that knockout of the cysH gene significantly improved the production of sTyr-containing sfGFP, compared to that seen in the wildtype BW25113 strain (Fig. 3b). CysH encodes the PAPS sulfotransferase responsible for degradation of PAPS to 3′-phosphoadenosine-5′-phosphate (PAP). This observation of enhanced sfGFP-sTyr production in BW25113ΔcysH is consistent with the previous report that knockout of cysH gene can increase cellular PAPS concentration and the production of sulfated products in *E. coli*[65,66]. Next, we examined whether manipulation of PAPS synthetic and recycling pathways in *E. coli* could further enhance intracellular PAPS levels. We amplified the gene cysDNC encoding adenosine-5′-triphosphate (ATP) sulfurylase and adenosine 5′-phosphosulfate kinase to increase the intracellular level of PAPS, followed by the introduction of the gene cycQ encoding adenosine-3′,5′-diphosphate (PAP) nucleotidase for PAP recycling[45,66–68]. We found that cells expressing all these genes exhibited the largest increase in fluorescence, suggesting a higher expression level of sfGFP-sTyr (Fig. 3c). The *Nn*SULT1C1 expression level has a significant influence on the production of sfGFP-sTyr, since we found that the concentration of *Nn*SULT1C1 inducer is important. Among all *L*-arabinose concentrations tested, *Nn*SULT1C1 expression induced by 15 mg/L *L*-arabinose yielded the highest production of sfGFP-sTyr, even higher than cells with 27 mM external sTyr addition[35–40]. Thus, the addition of 15 mg/L *L*-arabinose was used in future experiments. (Fig. 3d). We also screened other conditions for sfGFP-sTyr expression, including expression medium, Tyr addition, SO_4^{2-} addition, glycerol addition, which did not alter the expression level of sfGFP-sTyr (Supplementary Fig. 6). To examine the contribution

https://doi.org/10.1038/s41467-022-33111-4

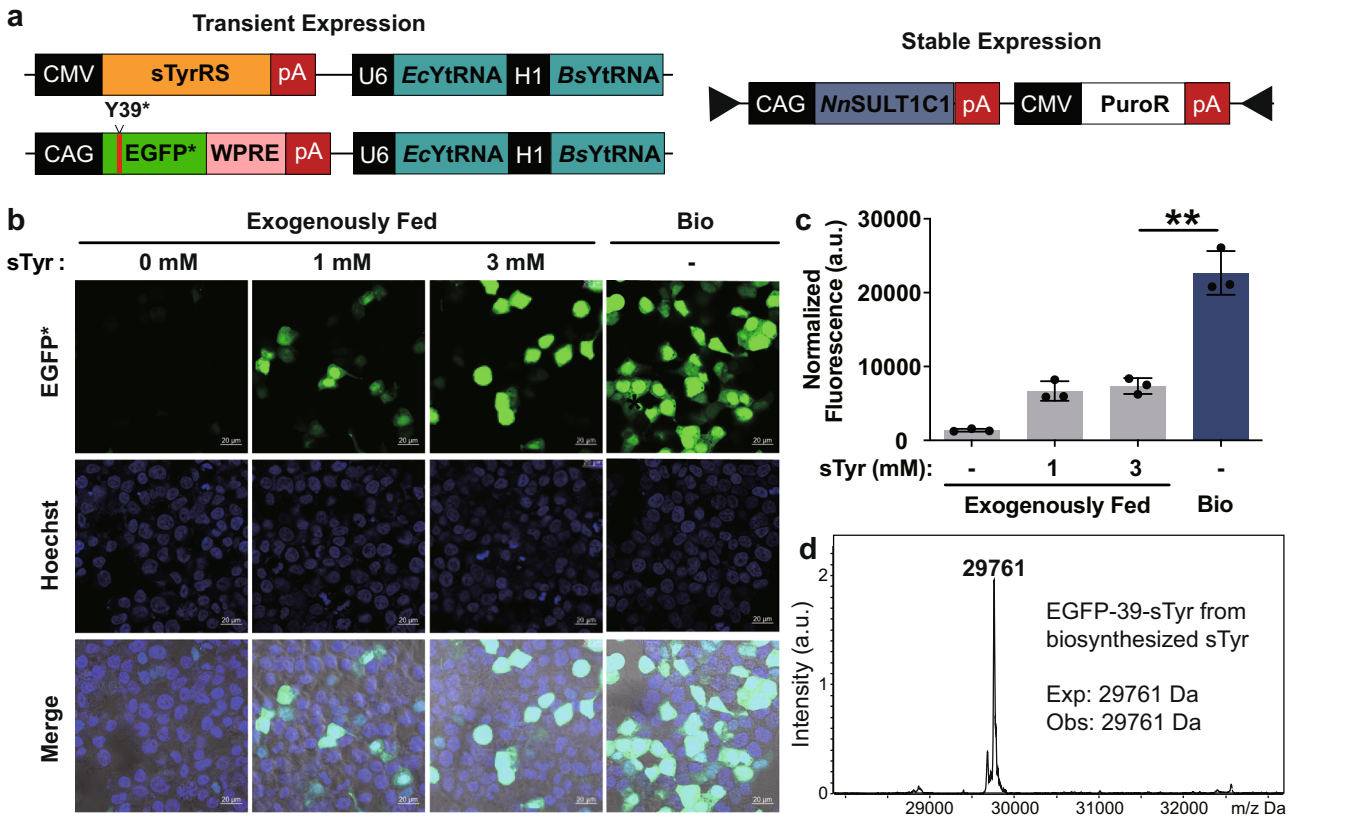


Fig. 4 | Generation of completely autonomous mammalian cells with sTyr-containing proteins. a Schematic representation of genetic circuits used for generating completely autonomous mammalian cells with sTyr-containing proteins. **b** Confocal images of HEK293T (exogenously fed) and HEK293T- *Nn*SULT1C1 (bio) cells expressing sTyrRS, $tRNA_{CUA}$ and EGFP containing an amber codon at Tyr39 position. Scale bar = 20 μm. **c** Flow cytometric analysis of EGFP expression levels of HEK293T (exogenously fed) and HEK293T- *Nn*SULT1C1 (bio) cells with sTyrRS, $tRNA_{CUA}$ and EGFP containing an amber codon at Tyr39 position. The normalized fluorescence was calculated by multiplying the geometric mean fluorescence by the percentage of EGFP-positive cells. Error bars represent standard deviations. **d** Mass spectra analysis of EGFP with sTyr (EGFP-39-sTyr) purified from HEK293T- *Nn*SULT1C1 cells. **c** Data are plotted as the mean ± standard deviation from $n = 3$ independent samples. **c** Two-sided unpaired *t*-tests were performed with $p = 0.001$. $^{**}p < 0.01$. a.u. stands for arbitrary unit.

of the biosynthetic pathway to intracellular sTyr concentration, we measured the intracellular sTyr concentrations in cells when sTyr was either biosynthesized or delivered via exogenous feeding. To our delight, the cellular concentration of sTyr in cells endowed with the sTyr biosynthetic pathway is 756.3 μM, which is 28-fold higher than that from cells exogenously fed with 1 mM sTyr and higher even than in cells fed with 27 mM sTyr (Fig. 3e). Consistent with these intracellular levels of sTyr, endogenous biosynthesis of sTyr results in much higher sfGFP-sTyr expression than that produced via exogenous feeding (Fig. 3d, e). To further investigate the efficiency and specificity of incorporation of biosynthesized sTyr in these autonomous *E. coli* cells, sfGFP-sTyr proteins derived from exogenously fed sTyr and from biosynthesized sTyr were purified by Ni^{2+}-NTA affinity chromatography and characterized by SDS-PAGE and ESI-MS. Intact sfGFP was only expressed after exogenous sTyr feeding or after induction of sTyr biosynthesis. The yield of sfGFP-sTyr derived from biosynthetic sTyr is 5.67 mg/L sfGFP-sTyr under the optimal condition, compared with 1.5 mg/L sfGFP-sTyr produced by feeding with 1 mM exogenous sTyr (Fig. 3f). The mass of sfGFP-sTyr produced from biosynthetic sTyr was 27, 674 Da, which is in good agreement with the calculated mass. (Fig. 3g, h). To test the activity of *Nn*SULT1C1 in vitro, its kinetics values were measured. It exhibits a K_m, V_{max}, and K_{cat} of 0.60 μM, 85.76 nmol/min/mg and 3.08 min^{-1}, respectively (Supplementary Fig. 7). Its catalytic efficiency ($V_{max}/K_m = 85.36\ s^{-1}\ mM^{-1}$) is comparable with the activity of human SULT1C1 reported previously[69,70].

Biosynthesis and genetic incorporation of sTyr in mammalian cells

Post-translational tyrosine sulfation occurs exclusively in eukaryotes. Although this modification has been estimated to occur on 1% of all tyrosine residues in eukaryotic proteomes, its functional significance is not well understood[41,71,72]. One approach to determine the biological importance of protein tyrosine sulfation is to express sulfated protein in living cells in a site-specific and homogeneous fashion, a goal that is difficult to achieve by chemical synthesis or recombinant expression. Genetic code expansion based on *E. coli*-derived tyrosyl-tRNA synthetase (*Ec*TyrRS)/tRNA has been proven to overcome these challenges by site-specifically incorporating sTyr in proteins in mammalian cells[73,74]. To promote the efficient expression of mammalian proteins sulfated on specific tyrosines, we have generated mammalian cells equipped with both sTyr biosynthetic and translational machinery. To generate mammalian cells capable of biosynthesizing sTyr, we used piggybac system to stably integrate *Nn*SULT1C1 into the genome of HEK293T cells, yielding the HEK293T- *Nn*SULT1C1 cell line (Fig. 4a)[75]. The *Ec*TyrRS/tRNA pair was used to construct pAcBac2.tR4-sTyrRS/EGFP*, containing EGFP with a stop codon at position 39 as well as two copies of *E. coli* and *Bacillus stearothermophilus* $tRNA_{CUA}^{Tyr}$ (Fig. 4a)[76]. To evaluate the function of *Nn*SULT1C1 in mammalian cells, pAcBac2.tR4-sTyrRS/EGFP* was transfected into HEK293T and HEK293T-*Nn*SULT1C1 cells, which were then incubated in the presence or absence of exogenous sTyr. The expression of EGFP was monitored by confocal microscopy 2 days after transfection. As expected, the addition of 1 mM

sTyr to HEK293T cells resulted in moderate expression of full-length EGFP, while minimal EGFP fluorescence was observed in the absence of sTyr addition (Fig. 4b). Gratifyingly, higher expression of EGFP was observed in HEK293T-*Nn*SULT1C1 cells without exogenous sTyr addition than that seen in HEK293T cells fed with 3 mM sTyr. In addition to confocal imaging, flow cytometry was used to quantify expression levels of EGFP in cells fed with exogenous sTyr and in cells biosynthesizing sTyr. As shown in Fig. 4c and Supplementary Fig. 8, significantly higher EGFP fluorescence was observed in HEK293T-*Nn*SULT1C1 cells endowed with sTyr biosynthetic capability than in HEK293T cells fed with 3 mM sTyr. As direct evidence of sTyr biosynthesis in mammalian cells, cellular sTyr concentration in HEK293T-*Nn*SULT1C1 is more than that in HEK293T cells fed with 3 mM sTyr (Supplementary Fig. 9). The fidelity of site-specific incorporation of sTyr was evaluated by mass spectral analysis of purified sTyr-containing EGFP proteins. The observed mass was 29, 761 Da, consistent with the calculated mass of EGFP with sTyr at position 39 and observed mass of EGFP39sTyr purified from HEK293T with external sTyr addition (Fig. 4d and Supplementary Fig. 10). These results demonstrate that the generation of mammalian cells autonomously able to biosynthesize sTyr and incorporate it into proteins significantly enhances the expression level of sTyr-containing protein in mammalian cells.

Using completely autonomous sTyr biosynthetic cells to synthesize potent thrombin inhibitors with site-specific sulfation

Thrombin inhibitors represent an important class of anticoagulants used to prevent blood clotting. In addition, several thrombin inhibitors from hematophagous organisms have been shown to facilitate the acquisition and digestion of bloodmeal[77–79]. Recent studies have reported that post-translational sulfation of these proteins has a dramatic effect on their inhibitory activity[31,80]. For example, tyrosine sulfation of hirudin increases its affinity for thrombin by more than 10-fold[80,81]. Tyrosine sulfation of madanin-1 and chimadanin significantly increases their affinities for thrombin by promoting strong electrostatic interactions with positively-charged residues (Fig. 5a). Current methods for studying these site-specifically sulfated thrombin inhibitors rely heavily on solid-phase peptide synthesis and subsequent chemical ligation, processes that are time-consuming and may result in sub-optimal protein folding[31,82,83]. To explore the generation sTyr-containing thrombin inhibitors using cells endowed with autonomous sTyr biosynthetic machinery, we chose both madanin-1 and chimadanin identified in the salivary gland of *haemaphysalis longicornis* (Fig. 5b)[31,84]. As shown in Fig. 5a, sulfation of madanin-1 converts Tyr32 and Tyr35 to negative residues, thus enhancing madanin-1's direct electrostatic interaction with the ε-amino groups of K236 and K240 located within the exosite II site of thrombin. To express the site-specifically sulfated thrombin inhibitors, we constructed plasmids encoding the thrombin inhibitor and substituted with amber codons at either or both of the indicated Tyr sites. sTyr-containing inhibitors were expressed by transforming ΔcysH BW25113 cells with pEvol-*Nn*SULT1C1-cysDNCQ, pUltra-sTyr, and a plasmid encoding the thrombin inhibitor. In parallel, we utilized the ΔcysH BW25113 cells lacking the sTyr biosynthetic systems but exogenously fed with 3 mM sTyr. The site-specific sulfation of madanin-1 and chimadanin was further validated using SDS-PAGE and ESI-MS analysis (Fig. 5c and Supplementary Figs. 11–13).

To test the thrombin inhibiting activity of the wildtype inhibitors and their sTyr-containing mutants, we performed chromogenic thrombin amidolytic activity assays in the presence of a range of concentrations of each inhibitor. Compared with wildtype madanin-1 ($Ki = 16.0 \pm 0.9$ nM), incorporation of a single sTyr at either Tyr32 ($Ki = 1.3 \pm 0.1$ nM) or Tyr35 ($Ki = 6.1 \pm 0.6$ nM) position significantly enhanced its inhibition of thrombin (Fig. 5d and Supplementary Fig. 14). To our delight, madanin-1 mutants sulfated at both Tyr32 and Tyr35 exhibited the highest potency ($Ki = 0.5 \pm 0.1$ nM) against thrombin activity (Fig. 5d and Supplementary Fig. 14). Following a similar trend, incorporating a single biosynthesized sTyr at either Tyr28 or Tyr31 of chimadanin yields more potent inhibition of thrombin activity ($Ki = 0.6 \pm 0.1$ nM and 1.5 ± 0.1 nM, respectively) than achieved with wildtype chimadanin ($Ki = 12.9 \pm 0.1$ nM, Fig. 5e and Supplementary Fig. 14). Double sulfation of chimadanin at both Tyr28 and Tyr31 further improved its *Ki* to 0.1 nM, consistent with the madanin-1 study (Fig. 5e and Supplementary Fig. 14). Furthermore, sTyr-containing thrombin inhibitors prepared using cells with completely autonomous sTyr biosynthetic machinery are more potent than chemically synthesized ones[31]. This may due to the fact that co-translational folding is more efficient than that achieved via chemical synhesis. These data demonstrate the advantages of producing therapeutic proteins with site-specific sTyr modifications using completely autonomous cells with the ability to biosynthesize and genetically encode the sTyr.

Discussion

In this research, we have generated completely autonomous bacterial and mammalian cells endowed with machinery for both sTyr biosynthesis and site-specific incorporation into proteins. *Nn*SULT1C1-mediated biosynthesis of sTyr from tyrosine and PAPS was discovered using a SSN, and the unique specificity of *Nn*SULT1C1 for tyrosine was systematically explored using both bioinformatic and computational methods. Use of *Nn*SULT1C1 and other optimized components allowed us to engineer both bacterial and mammalian cells capable of autonomously biosynthesizing sTyr and genetically incorporating it into proteins. The resulting cells produce site-specifically sulfated proteins at higher yields than cells exogenously fed with 3–27 mM sTyr. The value of these completely autonomous cells was further demonstrated via their use in the preparation of therapeutic sTyr-containing proteins with enhanced efficacy.

More than 300 ncAAs have been genetically incorporated into proteins in a site-specific manner, providing powerful tools for investigating protein structures and functions[1–3,6,85–93]. To date, utilizing these ncAAs in the context of Genetic Code Expansion has required both exogenous feeding and good membrane permeability of chemically-synthesized ncAAs. Cell membranes are poorly permeable to ncAAs with charged or polar structures. Thus, intracellular biosynthesis of these ncAAs is likely to significantly expand the utility of Genetic Code Expansion technology. Attempts to engineer cells for autonomous ncAA biosynthesis without external addition of precursors have frequently been hindered by the scarcity of verified biosynthetic pathways for producing ncAAs at high concentrations. For this reason, biosynthetic pathways for *p*AF, *p*Thr, 5HTP, and DOPA are the only ones that have been applied to bacterial cells for intracellular ncAA biosynthesis from simple carbon sources[12,19–22]. We expect that the combination of bioinformatics and ncAA screening methods reported in this work can be a powerful strategy for enlarging the repertoire of biosynthesized ncAA for Genetic Code Expansion. Our study further reports the construction of a completely autonomous mammalian cell line capable of biosynthesizing sTyr and incorporating it into proteins in response to the amber codon. The creation of additional mammalian cells with the endogenous ability to biosynthesize ncAAs and use them for protein synthesis will expand the preparation of therapeutic proteins, as well as allow application of the Genetic Code Expansion technology at the level of whole organisms.

Methods

Sequence similarity network (SSN)

The SSN was generated by inputting the amino acid sequence of *Rn*SULT1A1 as query sequence at https://efi.igb.illinois.edu/efi-est/. The UniProt database was selected and the *e* value was set as 5. The resulting network was finalized by setting the alignment score threshold as 110 to generate edges representing pairwise sequence

https://doi.org/10.1038/s41467-022-33111-4

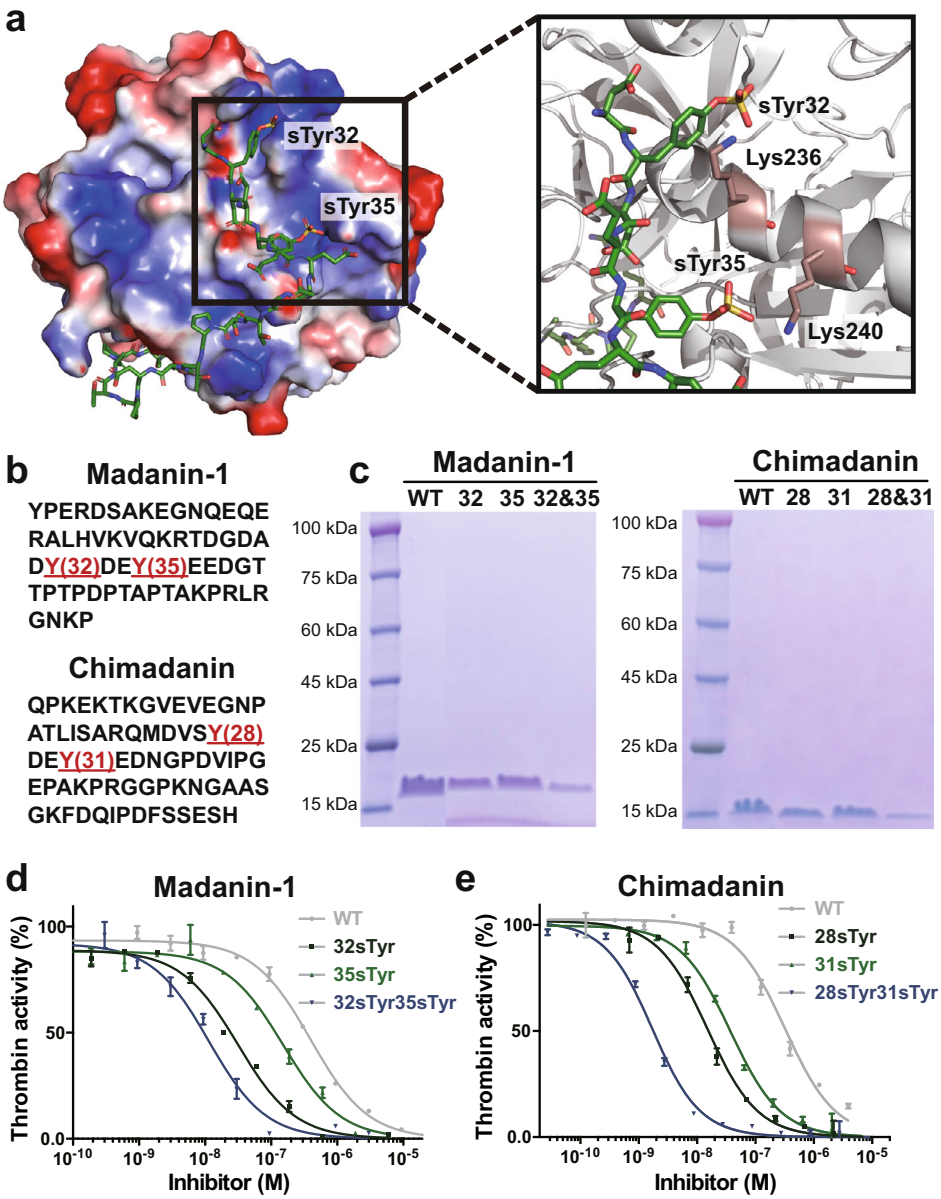


Fig. 5 | Production of thrombin inhibitors with site-specific sTyr insertion using completely autonomous *E. coli*. **a** Madanin-1 with sulfation at Tyr32 and Tyr35 positions binds to exosite II site in thrombin, analyzed from PDB: 5L6N (https://www.rcsb.org/structure/5l6n). Surface representation of positive electrostatic potential in blue and negative electrostatic potential in red. **b** Amino acid sequences of madanin-1 and chimadanin. Sulfation sites are shown in red. **c** SDS-PAGE analysis of thrombin inhibitors with site-specific sTyr insertion expressed in completely autonomous *E. coli*. **d**, **e** Inhibition of thrombin activity by madanin-1 and chimadanin proteins. **d**, **e** Data were plot as mean ± standard error from $n = 3$ independent samples and fitted into Morrison model.

similarities. The representative node network with %ID of 80% was downloaded in the format of xgmml and visualized within Cytoscape.

Optimized expression of sfGFP-sTyr from sTyr biosynthesis

ΔcysH BW25113 cells, transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG, and pEvol-*Nn*SULT1C1-cysDNCQ, were grown in Luria-Bertani (LB) medium at 37 °C. 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. After 6 h induction, the cells were diluted five times to OD 0.6. Expression of reporter sfGFP and sTyrRS were induced with 1 mM IPTG. Additional *l*-arabinose was also added to maintain its final concentration of 15 mg/L. The control cells transformed with pUltra-sTyrRS, pET22b-T5-sfGFP151TAG and pEvol-empty were grown under the same condition with an indicated concentration of sTyr. After growth at 30 °C for 18 h, cells were harvested by centrifugation at 4750 × *g* for 10 min and used for GFP fluorescence and cell optical density measurements. 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.

Predicting the structure of *Nn*SULT1C1 by AlphaFold

The structure of *Nn*SULT1C1 was predicted by AlphaFold2 using GitHub AlphaFold code 2.0. The database, including reduced BFD, PDB70, MGnify, and Uniclust30, was used to filter structural templates. All other settings were set as default. Based on pLDDT, the top structure was output and used in this study.

Protein-ligand docking

The protein-ligand docking process was performed by Glide v8.1 using Schrödinger software package v2018.4. Glide uses the OPLS3 force field to evaluate the docking procedure. OPLS3 is an enhanced version of the OPLS_2005 all-atom force field to provide a larger coverage of

organic functionality. Four protein structures, including 2zvq, 2a3r, 2gwh, and the predicted structure of *Nn*SULT1C1, are taken into consideration for docking. The PAPS-binding site for the predicted *Nn*SULT1C1 structure is inferred by aligning with the structure of 2a3r. For other structures, we used the original PAP sites in reported co-crystal structures to install PAPS. A short run of protein-ligand energy minimization was performed to remove the steric clashes for each of the complexes. The docking box was inferred from the position of dopamine in 2a3r. The RMSD is set to 0.5 to sample the distinct conformations. All parameters are set to default SP mode in the Glide software. The number of maximum output poses for each docking protein was set to 200 and the top 50 poses ranked by Emodel score were picked out. The best docking pose for each complex was compared using the Glide docking score, an empirical scoring function that approximates the ligand binding free energy in the unit of kcal/mol.

Characterization of HEK293T-*Nn*SULT1C1 with confocal microscopy and flow cytometry

To generate HEK293T-*Nn*SULT1C1, HEK293T were transfected with PB-*Nn*SULT1C1 (100 ng) and Piggybac transposases plasmids (20 ng) with Polyjet In Vitro DNA Transfection Reagent (SignaGen Laboratories). 1 μg/mL puromycin was added to culture medium from Day 2 to Day 7 for selecting cells with genomic integration of *Nn*SULT1C1. The puromycin concentration was raised to 3 μg/mL from Day 8 and maintained in the future. HEK293T and HEK293T-*Nn*SULT1C1 cells were cultured in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37 °C and 5% CO_2. 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 the indicated concentration of sTyr. Mediums were changed 12–16 h after transfection. After 48 h of the transfection, cells were used for confocal microscopy where nucleus staining was performed by incubating cells with Hoechst 33342 (Life Technologies). After being washed with PBS (pH 7.4) for three times, cells were imaged with Zeiss LSM710 confocal microscopy. The rest of cells were used for flow cytometry analysis with Sony SA3800 Flow Cytometer where a total of 20,000 cells were analyzed for each sample. Data were processed with FlowJo. Reported data are the average measurement of three independent samples prepared at the same time with the standard deviation.

Thrombin activity assay

N-(*p*-Tosyl)-GPR-*p*NA acetate (Cayman Chemicals) was used as a chromogenic substrate to test the amidolytic activity of human α-thrombin (Haematologic Technologies). Purified chi and mad inhibitors were buffer-exchanged to the assay buffer (pH 8) containing 50 mM Tris-HCl, 50 mM NaCl using PD-10 columns. Inhibition assays were performed in the assay buffer with 0.14 nM human α-thrombin, 100 μM substrate, and varying concentrations of inhibitors. The activity of thrombin was monitored by absorption at 405 nm. Inhibition constants (K_i) were determined based on a Morrison equation within GraphPad Prism. Three independent samples were prepared for each group.

Statistics and reproducibility

All statictics analysis were performed using GraphPad Prism. Similar results were obtained from three independent experiments.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated in this study are included in the paper and supplementary information. Plasmids for pEvol-*Nn*SULT1C1-cysDNCQ, pET22b-T5-chi28TAG, pET22b-T5-chi31TAG, pET22b-T5-chi28T-AG31TAG, pET22b-T5-mad32TAG, pET22b-T5-mad35TAG, pET22b-T5-mad32TAG35TAG, as well as other essential constructs developed by this work, are available on Addgene via https://www.addgene.org/Han_Xiao/. Source data are provided with this paper.

References

1. Wang, L., Xie, J. & Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
2. Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nat. Chem. Biol.* **3**, 29–35 (2007).
3. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
4. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
5. Dien, V. T., Morris, S. E., Karadeema, R. J. & Romesberg, F. E. Expansion of the genetic code via expansion of the genetic alphabet. *Curr. Opin. Chem. Biol.* **46**, 196–202 (2018).
6. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
7. Agostini, F. et al. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* **56**, 9680–9703 (2017).
8. Manandhar, M., Chun, E. & Romesberg, F. E. Genetic Code Expansion: Inception, Development, Commercialization. *J. Am. Chem. Soc.* **143**, 4859–4878 (2021).
9. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292**, 498–500 (2001).
10. Furter, R. Expansion of the genetic code: Site-directed p-fluorophenylalanine incorporation in Escherichia coli. *Protein Sci.* **7**, 419–426 (1998).
11. Giese, C. et al. Intracellular uptake and inhibitory activity of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem* **3**, 1449–1456 (2008).
12. Zhang, M. S. et al. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* **14**, 729–736 (2017).
13. Luo, X. et al. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* **13**, 845–849 (2017).
14. Hoppmann, C. et al. Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* **13**, 842–844 (2017).
15. Bundy, B. C. & Swartz, J. R. Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein–Protein Click Conjugation. *Bioconjug. Chem.* **21**, 255–263 (2010).
16. Ko, W., Kumar, R., Kim, S. & Lee, H. S. Construction of Bacterial Cells with an Active Transport System for Unnatural Amino Acids. *ACS Synth. Biol.* **8**, 1195–1203 (2019).
17. Burkovski, A. & Krämer, R. Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Appl. Microbiol. Biotechnol.* **58**, 265–274 (2002).
18. Palacín, M., Estévez, R., Bertran, J. & Zorzano, A. Molecular Biology of Mammalian Plasma Membrane Amino Acid Transporters. *Physiol. Rev.* **78**, 969–1054 (1998).
19. Mehl, R. A. et al. Generation of a Bacterium with a 21 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
20. Chen, Y. et al. A noncanonical amino acid-based relay system for site-specific protein labeling. *Chem. Commun.* **54**, 7187–7190 (2018).
21. Rogerson, D. T. et al. Efficient genetic encoding of phosphoserine and its non-hydrolyzable analog. *Nat. Chem. Biol.* **11**, 496–503 (2015).
22. Chen, Y. et al. Creation of Bacterial Cells with 5-Hydroxytryptophan as a 21st Amino Acid Building Block. *Chem.* **6**, 2717–2727 (2020).

23. Völler, J.-S. & Budisa, N. Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* **48**, 1–7 (2017).
24. Wang, Y. et al. Expanding the Structural Diversity of Protein Building Blocks with Noncanonical Amino Acids Biosynthesized from Aromatic Thiols. *Angew. Chem. Int. Ed.* **60**, 10040–10048 (2021).
25. Exner, M. P. et al. Design of S-Allylcysteine in Situ Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant. *ChemBioChem* **18**, 85–90 (2017).
26. Chen, Y. et al. Biosynthesis and Genetic Incorporation of 3,4-Dihydroxy-L-Phenylalanine into Proteins in Escherichia coli. *J. Mol. Biol.* **434**, 167412–167421 (2022).
27. Veldkamp, C. T. et al. Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12. *Sci. Signal.* **1**, ra4–ra4 (2008).
28. Ludeman, J. P. & Stone, M. J. The structural role of receptor tyrosine sulfation in chemokine recognition. *Br. J. Pharmacol.* **171**, 1167–1179 (2014).
29. Farzan, M. et al. Tyrosine Sulfation of the Amino Terminus of CCR5 Facilitates HIV-1 Entry. *Cell* **96**, 667–676 (1999).
30. Choe, H. et al. Tyrosine Sulfation of Human Antibodies Contributes to Recognition of the CCR5 Binding Region of HIV-1 gp120. *Cell* **114**, 161–170 (2003).
31. Thompson, R. E. et al. Tyrosine sulfation modulates activity of tick-derived thrombin inhibitors. *Nat. Chem.* **9**, 909–917 (2017).
32. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1. *Cell* **103**, 467–479 (2000).
33. Westmuckett, A. D., Thacker, K. M. & Moore, K. L. Tyrosine Sulfation of Native Mouse Psgl-1 Is Required for Optimal Leukocyte Rolling on P-Selectin In Vivo. *PLOS ONE* **6**, e20406 (2011).
34. Lee, S.-W. et al. A Type I–Secreted, Sulfated Peptide Triggers XA21-Mediated Innate Immunity. *Science* **326**, 850–853 (2009).
35. Li, X., Hitomi, J. & Liu, C. C. Characterization of a Sulfated Anti-HIV Antibody Using an Expanded Genetic Code. *Biochemistry* **57**, 2903–2907 (2018).
36. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24**, 1436–1440 (2006).
37. Liu, C. C., Cellitti, S. E., Geierstanger, B. H. & Schultz, P. G. Efficient expression of tyrosine-sulfated proteins in E. coli using an expanded genetic code. *Nat. Protoc.* **4**, 1784–1789 (2009).
38. Liu, C. C. et al. Protein evolution with an expanded genetic code. *Proc. Natl Acad. Sci. USA* **105**, 17688–17693 (2008).
39. Liu, C. C., Choe, H., Farzan, M., Smider, V. V. & Schultz, P. G. Mutagenesis and Evolution of Sulfated Antibodies Using an Expanded Genetic Code. *Biochemistry* **48**, 8891–8898 (2009).
40. Schwessinger, B. et al. A second-generation expression system for tyrosine-sulfated proteins and its application in crop protection. *Integr. Biol.* **8**, 542–545 (2016).
41. Yang, Y.-S. et al. Tyrosine Sulfation as a Protein Post-Translational Modification. *Molecules* **20**, 2138–2164 (2015).
42. Gamage, N. et al. Human Sulfotransferases and Their Role in Chemical Metabolism. *Toxicol. Sci.* **90**, 5–22 (2006).
43. Allali-Hassani, A. et al. Structural and Chemical Profiling of the Human Cytosolic Sulfotransferases. *PLOS Biol.* **5**, e97 (2007).
44. Suiko, M., Kurogi, K., Hashiguchi, T., Sakakibara, Y. & Liu, M.-C. Updated perspectives on the cytosolic sulfotransferases (SULTs) and SULT-mediated sulfation. *Biosci. Biotechnol. Biochem.* **81**, 63–72 (2017).
45. Jendresen, C. B. & Nielsen, A. T. Production of zosteric acid and other sulfated phenolic biochemicals in microbial cell factories. *Nat. Commun.* **10**, 1–10 (2019).
46. Chen, Y. et al. Addition of Isocyanide-Containing Amino Acids to the Genetic Code for Protein Labeling and Activation. *ACS Chem. Biol.* **14**, 2793–2799 (2019).
47. Copp, J. N., Akiva, E., Babbitt, P. C. & Tokuriki, N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry* **57**, 4651–4662 (2018).
48. Gerlt, J. A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
49. Blanchard, R. L., Freimuth, R. R., Buck, J., Weinshilboum, R. M. & Coughtrie, M. W. H. A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics* **14**, 199–211 (2004).
50. Schlee, D. Review of Numerical Taxonomy. The Principles and Practice of Numerical Classification. *Syst. Zool.* **24**, 263–268 (1975).
51. Varin, L., Marsolais, F., Richard, M. & Rouleau, M. Biochemistry and molecular biology of plant sulfotransferases. *FASEB J.* **11**, 517–525 (1997).
52. Hirschmann, F., Krause, F. & Papenbrock, J. The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* **5**, 556 (2014).
53. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
54. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
55. Jin, S. et al. Protein Structure Prediction in CASP13 Using AWSEM-Suite. *J. Chem. Theory Comput.* **16**, 3977–3988 (2020).
56. Jin, S. et al. Molecular-replacement phasing using predicted protein structures from AWSEM-Suite. *IUCrJ* **7**, 1168–1178 (2020).
57. Berger, I., Guttman, C., Amar, D., Zarivach, R. & Aharoni, A. The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. *PLOS ONE* **6**, e26794 (2011).
58. Lu, J.-H. et al. Crystal structure of human sulfotransferase SULT1A3 in complex with dopamine and 3′-phosphoadenosine 5′-phosphate. *Biochem. Biophys. Res. Commun.* **335**, 417–423 (2005).
59. Bidwell, L. M. et al. Crystal structure of human catecholamine sulfotransferase. *J. Mol. Biol.* **293**, 521–530 (1999).
60. Friesner, R. A. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
61. Banks, J. L. et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780 (2005).
62. Zhu, S., Wu, J., Du, G., Zhou, J. & Chen, J. Efficient synthesis of eriodictyol from L-tyrosine in Escherichia coli. *Appl. Environ. Microbiol.* **80**, 3072–3080 (2014).
63. Wei, T., Cheng, B.-Y. & Liu, J.-Z. Genome engineering Escherichia coli for L-DOPA overproduction from glucose. *Sci. Rep.* **6**, 30080 (2016).
64. Bang, H. B., Lee, Y. H., Kim, S. C., Sung, C. K. & Jeong, K. J. Metabolic engineering of Escherichia coli for the production of cinnamaldehyde. *Microb. Cell Fact.* **15**, 16 (2016).
65. Chu, L. L. et al. Metabolic Engineering of Escherichia coli for Enhanced Production of Naringenin 7-Sulfate and Its Biological Activities. *Front. Microbiol.* **9**, 1671 (2018).
66. Badri, A., Williams, A., Xia, K., Linhardt, R. J. & Koffas, M. A. G. Increased 3′-Phosphoadenosine-5′-phosphosulfate Levels in Engineered Escherichia coli Cell Lysate Facilitate the In Vitro Synthesis of Chondroitin Sulfate A. *Biotechnol. J.* **14**, e1800436 (2019).
67. Neuwald, A. F. et al. cysQ, a gene needed for cysteine synthesis in Escherichia coli K-12 only during aerobic growth. *J. Bacteriol.* **174**, 415–425 (1992).
68. Spiegelberg, B. D., Xiong, J.-P., Smith, J. J., Gu, R. F. & York, J. D. Cloning and Characterization of a Mammalian Lithium-sensitive Bisphosphate 3′-Nucleotidase Inhibited by Inositol 1,4-Bisphosphate*. *J. Biol. Chem.* **274**, 13619–13628 (1999).

69. Lu, L.-Y., Hsu, Y.-C. & Yang, Y.-S. Spectrofluorometric assay for monoamine-preferring phenol sulfotransferase (SULT1A3). *Anal. Biochem.* **404**, 241–243 (2010).
70. Dajani, R. et al. Kinetic Properties of Human Dopamine Sulfotransferase (SULT1A3) Expressed in Prokaryotic and Eukaryotic Systems: Comparison with the Recombinant Enzyme Purified fromEscherichia coli. *Protein Expr. Purif.* **16**, 11–18 (1999).
71. Moore, K. L. Protein tyrosine sulfation: a critical posttranslation modification in plants and animals. *Proc. Natl Acad. Sci. USA* **106**, 14741–14742 (2009).
72. Seibert, C. & Sakmar, T. P. Toward a framework for sulfoproteomics: synthesis and characterization of sulfotyrosine-containing peptides. *Biopolymers* **90**, 459–477 (2008).
73. He, X. et al. Functional genetic encoding of sulfotyrosine in mammalian cells. *Nat. Commun.* **11**, 4820 (2020).
74. Italia, J. S. et al. Genetically encoded protein sulfation in mammalian cells. *Nat. Chem. Biol.* **16**, 379–382 (2020).
75. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl Acad. Sci.* **108**, 1531–1536 (2011).
76. Chatterjee, A., Xiao, H., Bollong, M., Ai, H.-W. & Schultz, P. G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl Acad. Sci. USA* **110**, 11803–11808 (2013).
77. Tanaka-Azevedo, A.M., Morais-Zani, K., Torquato, R.J.S. & Tanaka, A.S. Thrombin Inhibitors from Different Animals. *J. Biomed. Biotechnol.* **2010**, 641025 (2010).
78. Koh, C. Y. & Kini, R. M. Molecular diversity of anticoagulants from haematophagous animals. *Thromb. Haemost.* **102**, 437–453 (2009).
79. Kazimírová, M. & Štibrániová, I. Tick salivary compounds: their role in modulation of host defences and pathogen transmission. *Front. Cell. Infect. Microbiol.* **3**, 43 (2013).
80. Hsieh, Y. S. Y., Wijeyewickrema, L. C., Wilkinson, B. L., Pike, R. N. & Payne, R. J. Total synthesis of homogeneous variants of hirudin P6: a post-translationally modified anti-thrombotic leech-derived protein. *Angew. Chem. Int. Ed. Engl.* **53**, 3947–3951 (2014).
81. Corral-Rodríguez, M. A., Macedo-Ribeiro, S., Pereira, P. J. B. & Fuentes-Prior, P. Leech-derived thrombin inhibitors: from structures to mechanisms to clinical applications. *J. Med. Chem.* **53**, 3847–3861 (2010).
82. Watson, E. E. et al. Mosquito-Derived Anophelin Sulfoproteins Are Potent Antithrombotics. *ACS Cent. Sci.* **4**, 468–476 (2018).
83. Watson, E. E. et al. Rapid assembly and profiling of an anticoagulant sulfoprotein library. *Proc. Natl Acad. Sci.* **116**, 13873–13878 (2019).
84. Nakajima, C. et al. A Novel Gene Encoding a Thrombin Inhibitory Protein in a cDNA Library from *Haemaphysalis longicornis* Salivary Gland. *J. Vet. Med. Sci.* **68**, 447–452 (2006).
85. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).
86. Xiao, H. & Schultz, P.G. At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harb. Perspect. Biol.* **8**, a023945 https://doi.org/10.1101/cshperspect.a023945 (2016).
87. Iannuzzelli, J. A. & Fasan, R. Expanded toolbox for directing the biosynthesis of macrocyclic peptides in bacterial cells. *Chem. Sci.* **11**, 6202–6208 (2020).
88. Owens, A. E., Iannuzzelli, J. A., Gu, Y. & Fasan, R. MOrPH-PhD: An Integrated Phage Display Platform for the Discovery of Functional Genetically Encoded Peptide Macrocycles. *ACS Cent. Sci.* **6**, 368–381 (2020).
89. Huang, Y. & Liu, T. Therapeutic applications of genetic code expansion. *Synth. Syst. Biotechnol.* **3**, 150–158 (2018).
90. Chen, C. et al. Genetic-code-expanded cell-based therapy for treating diabetes in mice. *Nat. Chem. Biol.* **18**, 47–55 (2022).
91. Zhang, S. & Ai, H. A general strategy to red-shift green fluorescent protein-based biosensors. *Nat. Chem. Biol.* **16**, 1434–1439 (2020).
92. Chen, Z., Ren, W., Wright, Q. E. & Ai, H. Genetically Encoded Fluorescent Probe for the Selective Detection of Peroxynitrite. *J. Am. Chem. Soc.* **135**, 14940–14943 (2013).
93. Guo, J. & Niu, W. Genetic Code Expansion Through Quadruplet Codon Decoding. *J. Mol. Biol.* **434**, 167346-167357 (2022).

Acknowledgements

We thank Dr. Xiao Laboratory members for insightful comments. This work was supported by the Cancer Prevention Research Institute of Texas (CPRIT RR170014 to H.X.), NIH (R35-GM133706, R21-CA255894, and R01-AI165079 to H.X.), the Robert A. Welch Foundation (C-1970 to H.X.), US Department of Defense (W81XWH-21-1-0789 to H.X.), the John S. Dunn Foundation Collaborative Research Award (to H.X.), and the Hamill Innovation Award (to H.X.), Center for Theroretical Biological Physics (NSF grant PHY-2019745 to P.G.W.) and D. R. Bullard Welch Chair at Rice University (Grant C-0016 to P.G.W). H.X. is a Cancer Prevention & Research Institute of Texas (CPRIT) scholar in cancer research.

Author contributions

Y.C., P.W., and H.X. designed the project. Y.C., M.Z., A.C. and Y.W. constructed plasmids. Y.C. and K.W. performed the SSN analysis. S.J. conducted structure prediction and docking experiments. Y.C. and Y.H. expressed and purified proteins. Y.C., S.W., and Z.T. carried out confocal microscopy. All other experiments were performed by Y.C. and H.X. Y.C., S.J., P.W., and H.X. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-33111-4.

Correspondence and requests for materials should be addressed to Han Xiao.

Peer review information *Nature Communications* thanks Nediljko Budisa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.