Supplementary Information

Sequence Redesign of Glycosyltransferases for Enhanced Heterologous Expression and Glycosylation Efficiency in *Escherichia coli*

Shanshan Zhang[1,2], Yilei Han[3*], Yujing Ding[1,2], Xiaofan Lin[1,2], Zhipeng Yuan[1,2], Xinxiao Sun[1,4*], Qipeng Yuan[1,4], Zheng Liu[3], Yifei Zhang[1,2*]

1. State Key Laboratory of Chemical Resources Engineering, Beijing University of Chemical Technology, Beijing 100029, China
2. Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China
3. Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
4. College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China.

[*]Corresponding authors:

Yilei Han: hanyl@mail.tsinghua.edu.cn;

Xinxiao Sun: sunxx@mail.buct.edu.cn;

Yifei Zhang: yifeizhang@mail.buct.edu.cn

Table of Content

Materials

Strains, Plasmids and Materials. The genes of TOGT from *Arabidopsis thaliana* and UGT84A56 from *Aesculus chinensis* were synthesized and codon-optimized to *Escherichia coli* (*E. coli*) by Genewiz (Beijing, China). Recombination strains containing corresponding plasmids used in this study are listed in Table S6, respectively. Gene segments used in this study were generated by polymerase chain reaction (PCR) with the corresponding primers as shown in Table S7. All primers, insertions and mutations of target genes were sequenced by Genewiz (Beijing, China). Plasmids were constructed by standard enzyme digestion and ligation or homologous recombination method. DNA modifying enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific. Plasmid extraction and gel purification kits were bought from Omega Bio-tek, Inc. *E. coli* DH5α and BL21(DE3) were purchased from TransGen Biotech for cloning and protein expression. Esculetin, aesculin, and cichoriin were purchased from Macklin Biotechnology (China). All chemicals and reagents are of analytical grade.

Methods

Sequence Redesign and Filtering

Seed Sequences and Structural Modeling. This study focused on two parental enzymes: TOGT from *Nicotiana tabacum* and UGT84A56 from *Aesculus chinensis.* Both are monomeric members of the glycosyltransferase GT-B fold family. Functionally, they catalyze O-glycosylation reactions by transferring glucose from the donor UDP-glucose (UDPG) to esculetin at either the 6-OH or 7-OH positions, producing aesculin and cichoriin, respectively. The complete amino acid sequences of TOGT and UGT84A56 are provided in Supplementary Tables S5 and S8. Due to the absence of experimentally resolved crystal structures for these enzymes, their three-dimensional models were predicted using the AlphaFold2 (AF2) algorithm, resulting structures served as the basis for subsequent structure-guided protein design and mechanistic analyses.

Structure-Based Fixation. The three-dimensional structures of TOGT and UGT84A56 were

predicted using the AlphaFold2 (AF2) algorithm, as described in the publication by Jumper et al[1]. Briefly, the protein sequences were input into the AlphaFold2 pipeline via the official ColabFold interface (https://colab.research.google.com/). The sequences were provided in FASTA format, and the AF2 model was run using default parameters, including the use of MSA (multiple sequence alignment) from UniRef30 and templates from PDB. All structure predictions and analyses were conducted using Python scripts with the PyMOL (Schrödinger, New York, NY, USA) and Chimera (UCSF, San Francisco, CA, USA) software packages for visualization and further refinement.

Structural models of the parental enzymes were first obtained via AlphaFold2 prediction. To define the residues critical for substrate recognition and catalysis, molecular docking was performed with the glycosyl donor UDP-glucose and acceptor substrates (see below for docking details). Residues in direct contact with the ligands were designated as first-shell residues and fixed during sequence design to preserve enzymatic function. Residues interacting with the first-shell residues, either via sidechain packing or polar contacts, were defined as second-shell residues and similarly fixed. This structure-based strategy effectively constrained all sites within 8 Å of the ligand-binding pocket, allowing targeted diversification of peripheral sites while maintaining the catalytic core.

Sequence-Based Fixation. Evolutionary conservative analysis was performed using Consurf server (https://consurf.tau.ac.il/)[2]. The server evaluates amino acid conservation using multiple sequence alignment (MSA) and evolution-based algorithms. First, the sequence of the target protein (in FASTA format) was input to the Consurf platform, and the best evolutionary model was selected as WAG. MSA generation parameters: the number of HMMER search iterations was set to 2, and the number of sequence alignments was set to 300, with the remaining parameters set to default. Residues with a conservation score above 5 (including 5, on a scale of 1 to 9) were retained, ensuring that functionally critical regions were maintained throughout the design process.

Redesign of the Protein Sequences. ProteinMPNN was used to redesign the sequences of TOGT and UGT84A56. The code used in this work is available on the ProteinMPNN github repository (commit 0a72127, June 9, 2022). Structure-based sites and sequenced-based residues for TOGT and UGT84A56 were excluded from design. Temperature sampling parameters 0.1 were used during redesign. A model of ProteinMPNN trained with 0.2 Å noise applied to training set protein backbones was used to perform sequence generation. Sequence redesign by selection of fixed residues.

Sequence Filtering. To refine and assess the quality of the generated protein structures, several

selection criteria were employed. Initially, the global score was used to evaluate the overall quality of the generated sequence, with lower values indicating a higher likelihood of correctly folding into the target structures. The pLDDT (predicted Local Distance Difference Test) score and pTM (inter-residue predicted TM-score) score were used to evaluate the reliability of the models. A higher pLDDT value indicates greater confidence in the accuracy and stability of the predicted structure. Similarly, pTM evaluates the precision of predicted inter-chain interaction regions, with higher scores reflecting more accurate predictions of protein-protein interactions. Models with higher pLDDT and pTM scores were prioritized for subsequent analysis. To further validate the structural integrity, the predicted designs were aligned with the native protein structures using TM-align[3], and RMSD (Root Mean Square Deviation) values were calculated. Models exhibiting higher RMSD values were subjected to additional filtering steps. Finally, SoluProt[4] solubility predictions was incorporated, with only higher-scoring candidates were selected for further experimental validation. To prioritize the top-performing designed enzymes, a multi-parameter scoring scheme was applied. Initially, all candidates were ranked in ascending order based on their global score, and the top 20 designs with the lowest global score values were selected for further analysis. Four structural and biophysical parameters—predicted TM-score (pTM), predicted Local Distance Difference Test (pLDDT), root mean square deviation (RMSD), and solubility—were then subjected to min-max normalization. Positive normalization was applied to pTM, pLDDT, and solubility (i.e., higher values were considered favorable), whereas RMSD was normalized in a reverse manner, reflecting its inverse correlation with structural quality (i.e., lower values indicate better structures). A weighted scoring function was subsequently used to integrate these normalized features into a composite performance score, defined as:

$$\text{Score} = 0.1 \times pTM_{norm} + 0.2 \times pLDDT_{norm} + 0.3 \times solubility_{norm} + 0.4 \times RMSD_{inv\text{-}norm}$$

This weighting scheme emphasizes structural stability and solubility (RMSD and solubility), while still accounting for overall structural confidence metrics (pTM and pLDDT). The final composite score was used to rank and select optimized designed enzymes for experimental validation.

Experiments and Calculation

Heterogonous Expression and Purification. The recombinant plasmid pETduet1-UGTs were transformed into *E. coli* BL21(DE3) for heterologous expression. The *E. coli* cells were cultured (5 mL) overnight in Luria-Bertani (LB) medium containing 100 μg/mL ampicillin at 37°C with shaking

(220 rpm). Subsequently, 1% seeding cultures were transferred into 100 mL LB-medium containing 100 μg/mL of ampicillin and grown at 37°C and 220 rpm in 250 mL shake flasks. After the OD_{600} nm reached 0.6–0.8, the expression of recombinant UGTs were induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 16°C, 220 rpm for 18 h. The cells were harvested by centrifugation at 6,000 rpm for 2 min at 4°C and then resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM sodium chloride, 10 mM imidazole, and 10% glycerol), the cells were disrupted by a high-pressure homogenizer (Union-Biotech Co., Ltd., Shanghai, China). Crude samples (cell lysate) were centrifuged at 6,000g for 30 min. The supernatant was incubated with a Ni-NTA resin for 1 h at 4°C. The non-specifically adsorbed proteins were removed by washing with 50 mL wash buffer (50 mM Tris-HCl, 300 mM NaCl and 50 mM imidazole at pH 8.0). Then, the His-tagged protein was eluted with elution buffer (50 mM Tris-HCl, 300 mM NaCl and 300 mM imidazole at pH 8.0). The purified proteins were concentrated using a 10 kDa Amicon Ultra centrifuge tube (Millipore, Burlington, USA) and was then exchanged with HEPES buffer (10 mM, pH 8.0) by passing through the HiTrap desalting column. Before the enzyme reaction, the enzyme concentration was quantified by the BCA Protein Assay Kit.

Protein Expression Quantification. 1 mL of medium after overexpression of the recombinant strain was centrifuged at 10,000 g for 5 min and then resuspended in an appropriate amount of HEPES (10 mM, pH 7.5) buffer so that the OD_{600} values of each fraction was the same. The cell solutions were disrupted by a JY98-IIIDN ultrasonic homogenizer sound arrest (Feiqi, Nanjing, China) on ice at 120 W for 1 min to obtain the protein. The soluble protein present in the supernatant and the inclusion body present in the precipitate were separated after centrifugation at 4°C, 12,000g for 5 min. Protein fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with 12% stacking gel and 5% separating gel and stained with Coomassie blue. To detect the expression of target proteins, Western blotting was also performed. Briefly, proteins on gels not stained with Coomassie Brilliant Blue were transferred onto a PVDF membrane (Millipore, Burlington, MA, USA) using a wet transfer method at 120 V for 100min at 4°C. After transfer, the membrane was blocked with 5% non-fat dry milk (Sigma-Aldrich) in 1× TBST (Tris-buffered saline with 0.1% Tween 20) for 1 hour at room temperature to prevent non-specific binding. The membrane was incubated overnight at 4°C with primary antibodies specific to the target proteins: anti-His-tag (1:1500, Cell Signaling Technology, Danvers, MA, USA). Following primary antibody incubation, the membrane was washed three times with TBST for 8

minutes each round. The membrane was then incubated for 1 hour at room temperature with HRP-conjugated secondary antibody (1:2000, Jackson Immuno Research, West Grove, PA, USA). After washing, protein bands were visualized using an ECL chemiluminescence detection kit (Pierce, Rockford, IL, USA) according to the manufacturer's instructions. The chemiluminescent signal was detected using a ChemiDoc™ MP Imaging System (Bio-Rad, Hercules, CA, USA). Band intensities were quantified using ImageJ software (NIH, Bethesda, MD, USA).

Crude Enzyme Lysate Activity Determination. To determine the crude lysate catalytic activity of TOGT, UGT84A56 and their designs, 200 μL cell lysates heterologously expressing at 25℃ under the same conditions were incubated with 1 mM substrate in a 400 μL reaction system containing 2 mM UDPG and 50 mM Tris-HCl pH 8.0 at 30°C for 60 min. Then, the reaction was terminated with 1-fold volume of methanol. Samples were prepared by centrifugation at 12,000 g for 10 min and filtered with polyethersulfone syringe filters (0.22 μm in pore size) and analyzed by HPLC as described below. Three parallel experiments were carried out.

Kinetic Parameters Measurement. The kinetic parameters of UGTs were measured in 400 μL reaction mixtures containing 2 mM UDPG, 50 mM Tris-HCl pH 8.0, a certain concentration of enzymes (0.5-1 μM), and different concentrations of substrate (10-500 μM) at 30 °C for 30/60 min. Then, reaction termination and sample preparation and assay were consistent with the above description. The Michaelis–Menten constant (K_m) and the turnover number (k_{cat}) were determined by fitting the initial velocity to the Michaelis-Menten equation.

Melting Temperature Determination. The protein melting temperature (T_m) was assessed using Nano Differential Scanning Fluorimetry (NanoDSF) with a Prometheus NT.48 instrument (NanoTemper Technologies, Munich, Germany) following the manufacturer protocol. Briefly, purified proteins (at a concentration of 1-2 mg/mL) were dissolved in 10 mM HEPES (pH 8.0) to ensure their physiological relevance. The protein solutions were loaded into standard capillaries (NanoTemper Technologies), and the emission intensities at 330 nm and 350 nm were adjusted to ensure fluorescence readings above 2000 RFU. The samples were then heated from 20°C to 95°C at a rate of 1°C/min. The intrinsic fluorescence of tryptophan residues was measured at both excitation wavelengths of 330 nm and 350 nm. The temperature-dependent shift in the fluorescence ratio (F350/F330) was used to determine the melting temperature (T_m), which serves as an indicator of protein thermal stability. For each protein sample, the experiment was performed in triplicate to

ensure reproducibility.

Culture Media and Conditions in Feeding /De Novo Synthesis Experiments. M9Y medium (20 g/L glycerol, 5 g/L yeast extract, 4 g/L NH_4Cl, 6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 2 g/L MOPS, 0.5 g/L NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) was used in this study. For feeding experiments, 100 mg/L ampicillin was added to M9Y. In the *de novo* synthesis experiments, 10 μM $FeSO_4$, 100 mg/L ampicillin, and 50 mg/L kanamycin were additionally supplemented.

Plasmids were electroporated into the host strain, which was then incubated at 37°C for 18 hours. Single colonies were picked and grown overnight in LB medium with appropriate antibiotics. The overnight culture was sub cultured (1% inoculum) into 50 mL of the corresponding medium and grown at 37°C until OD_{600} reached ~0.6. Protein expression was induced by adding 50 μL of 0.5 mM IPTG, and the culture was further incubated at 30°C with shaking (220 rpm). Samples were collected at designated time points. OD_{600} was measured using a UV-Vis spectrophotometer, and product analysis was performed using HPLC.

Culture Media and Conditions in Fed-Batch Experiments. Seed cultures of OD_{600} about 8-10 were transferred to a 3 L bioreactor containing 1 L of medium at 10% inoculum. The bioreactor fermentation medium contained, 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L $(NH4)_2SO_4$, 15 g/L yeast extract, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 75 mg/L VB1, 100 mg/L ampicillin and 50 mg/L kanamycin, 1 mL/L TES. TES contained 0.15 g/L $CuSO_4 \cdot 5H_2O$, 1 g/L $ZnSO_4 \cdot 7H_2O$, 3 g/L $FeSO_4 \cdot 7H_2O$, 1 g/L $MnCl_2 \cdot 4H_2O$ and 0.5 g/L $CoCl_2 \cdot 6H_2O$. In this process, the initial glucose concentration was 25 g/L, and strains were cultured at 37°C. When OD_{600} of the solution in the fermenter reached 15-20, 0.5 mM IPTG was added to induce protein expression. The pH was set at 7.0 by automatically adding ammonia to the flow and the dissolved oxygen was set at 30% by linkage of speed and dissolved oxygen. The feeding solution was 500 g/L glucose and the glucose concentration was maintained below 10 g/L. Data were reported as mean values from two independent experiments.

High-Performance Liquid Chromatography (HPLC) Analysis. Target product identification and quantification were performed using high-performance liquid chromatography (HPLC) with a Shimadzu HPLC system. All samples were analyzed by reverse-phase chromatography coupled with a UV detector. The separation was achieved using a ChromCore 120 C18 (5μm, 4.6×250 mm) reverse-phase column. The mobile phase consisted of two components: solvent A, 0.2%

trifluoroacetic acid (TFA), and solvent B, methanol. The column temperature was maintained at 40°C, and the flow rate was set to 1 mL/min. A gradient elution program was employed as follows: 0–10 min, 5–35% solvent B; 10–20 min, 35–85% solvent B; 20–25 min, 85–5% solvent B. UV detection was carried out at 345 nm, where the absorption is primarily attributed to the conjugated aromatic ring systems and the phenolic hydroxyl groups (-OH) present in both compounds. A 1 mL sample was centrifuged at 12,000 rpm for 10 minutes, and the supernatant was filtered through a 0.22 μm membrane before injection into the HPLC system.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations of the native and optimized designs were performed using GROMACS 2018[5] with the CHARMM36[6] force field. The protein was solvated in a TIP3P water box with a minimal distance of 1.0 nm from the box edge to the protein, and charge neutrality was achieved by adding Na^+ and Cl^- ions. Energy minimization was carried out using steepest descent (10,000 steps) followed by conjugate gradient optimization to avoid unfavorable interactions, with a force constant of 100 kJ/mol/nm on the protein backbone. The system was then equilibrated in the NVT ensemble, gradually heated to 300 K over 100 ps using the V-rescale thermostat, and pressure was maintained at 1 bar for 1 ns with the Berendsen barostat. A 100 ns production run was performed without positional restraints. Simulations were performed with a 2 fs timestep, saving the coordinates, energy, and velocity trajectory coordinates every 0.5 ns. The root-mean-square deviation (RMSD) was calculated for protein backbone atoms with the initial structure of MD simulations as the reference. The root-mean-square fluctuation (RMSF) was also calculated for backbone atoms of the trajectories. The radius of gyration (R_g) and the solvent accessible surface area (SASA) were calculated for the whole protein.

Hydrophilicity Analysis and Protein Substrate Channel Prediction. The Protscale tool (https://web.expasy.org/protscale/) was used to assess the hydrophobicity and hydrophilicity of the target protein sequence. The amino acid sequences of the target protein TOGT, UGT84A56, and their designs were input into the Protscale platform in FASTA format. The "Amino acid scale" was set to the default Hphob./Kyte & Doolittle scale, with a sliding window size of 9 and a linear weighting model to ensure the accuracy and reproducibility of the analysis. After running the analysis, the hydrophobicity/hydrophilicity values of the protein amino acids were downloaded and subsequently visualized and analyzed using GraphPad Prism 9.0.

The Proteins*Plus*[7] - Structure-Based Modeling Support Server platform (https://proteins.plus/) was

used to predict the substrate channels of proteins. This tool uses advanced algorithms to analyze the 3D structure of proteins, identifying potential substrate channels and their probable channel sizes. All analyses are done with default settings, ensuring efficient and accurate prediction results.

Figures

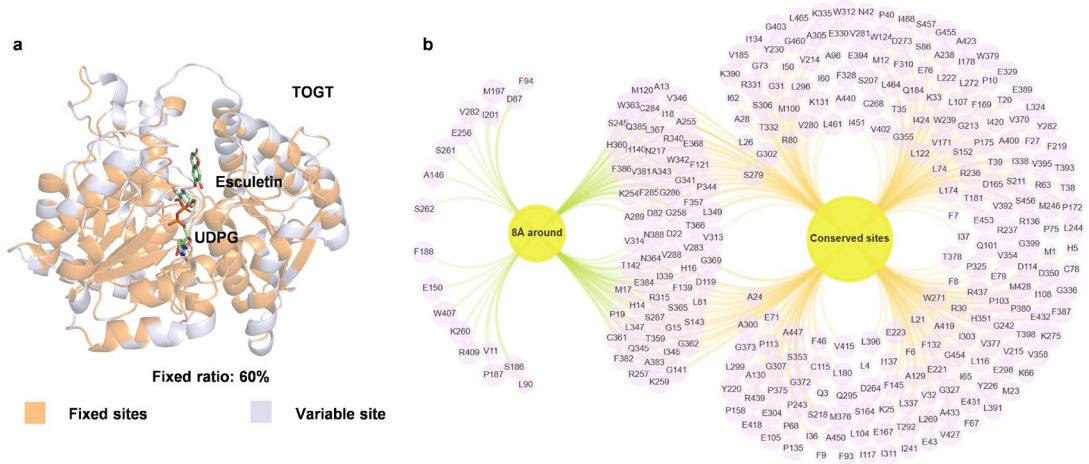


Fig. S1. Residue map of fixed sites used in TOGT sequence redesign. a, Structural overview of fixed and variable (designable) sites in TOGT during sequence redesign. The TOGT structure is shown in cartoon representation, with fixed sites (60% of the sequence) highlighted in orange and variable sites in light purple. The bound substrate esculetin and donor UDP-glucose (UDPG) are displayed in stick representation. **b,** Residues within 8 Å of the conserved sites (left) include the first- and second-shell residues surrounding the active site were fixed to maintain functional integrity. Conserved sites (right) represent evolutionarily constrained sites that were fixed during design to avoid random mutations in high-risk regions.

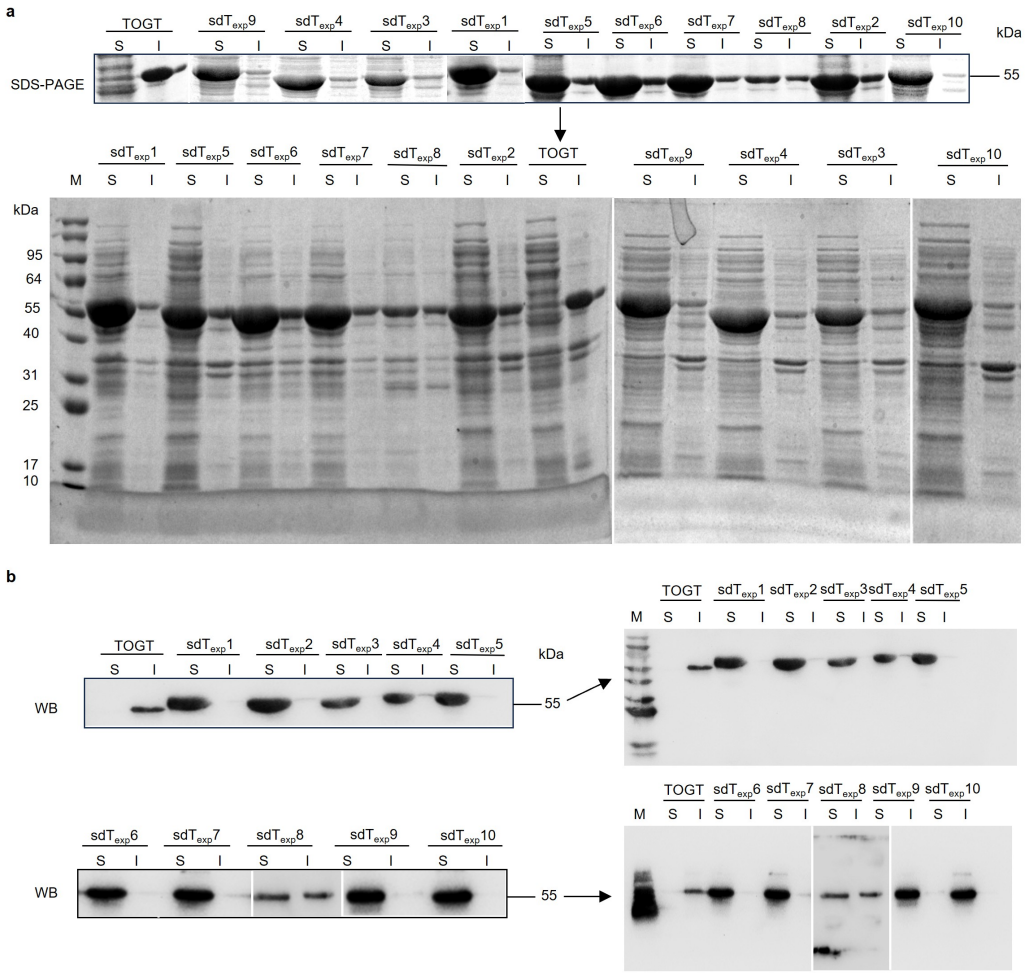


Fig. S2. Soluble expression characterization of TOGT and designs. a, SDS-PAGE analysis of the TOGT and sdT$_{exp}$ designs (sdT$_{exp}$1-sdT$_{exp}$10) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate. The arrow points to the corresponding uncropped original scanned image, where all bands are clearly visible. **b,** Western blot of the expression of TOGT and sdT$_{exp}$ designs (sdT$_{exp}$1-sdT$_{exp}$10) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate. The right-hand panel, as indicated by the arrow, presents the uncropped original western blot image.

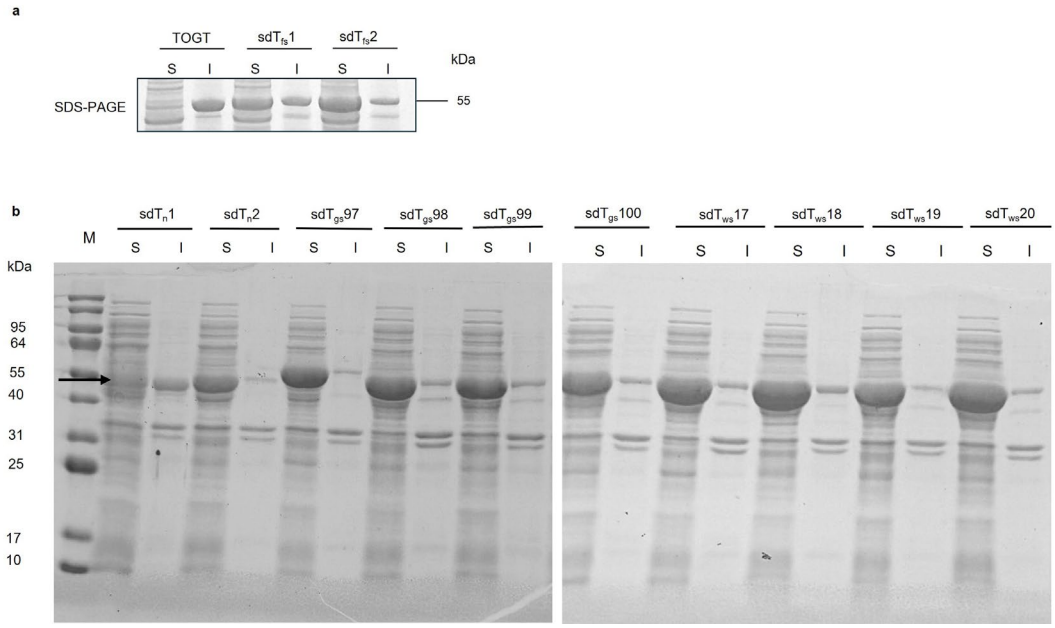


Fig. S3. Soluble expression characterization of TOGT and other four strategy designs. a, SDS-PAGE analysis of the TOGT and sdT_{fs} designs ($sdT_{fs}1$, $sdT_{fs}2$) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate. **b,** SDS-PAGE analysis of the expression of other three strategy designs (sdT_n1, sdT_n2, $sdT_{gs}98$-$sdT_{gs}100$ and $sdT_{ws}17$-$sdT_{ws}20$) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate.

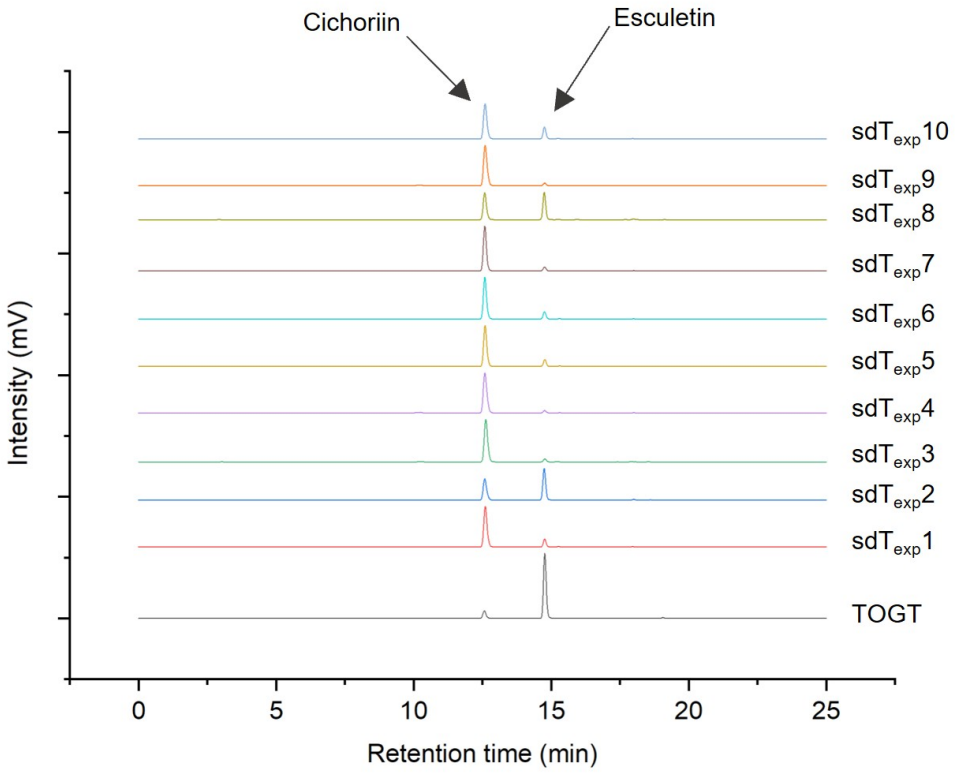
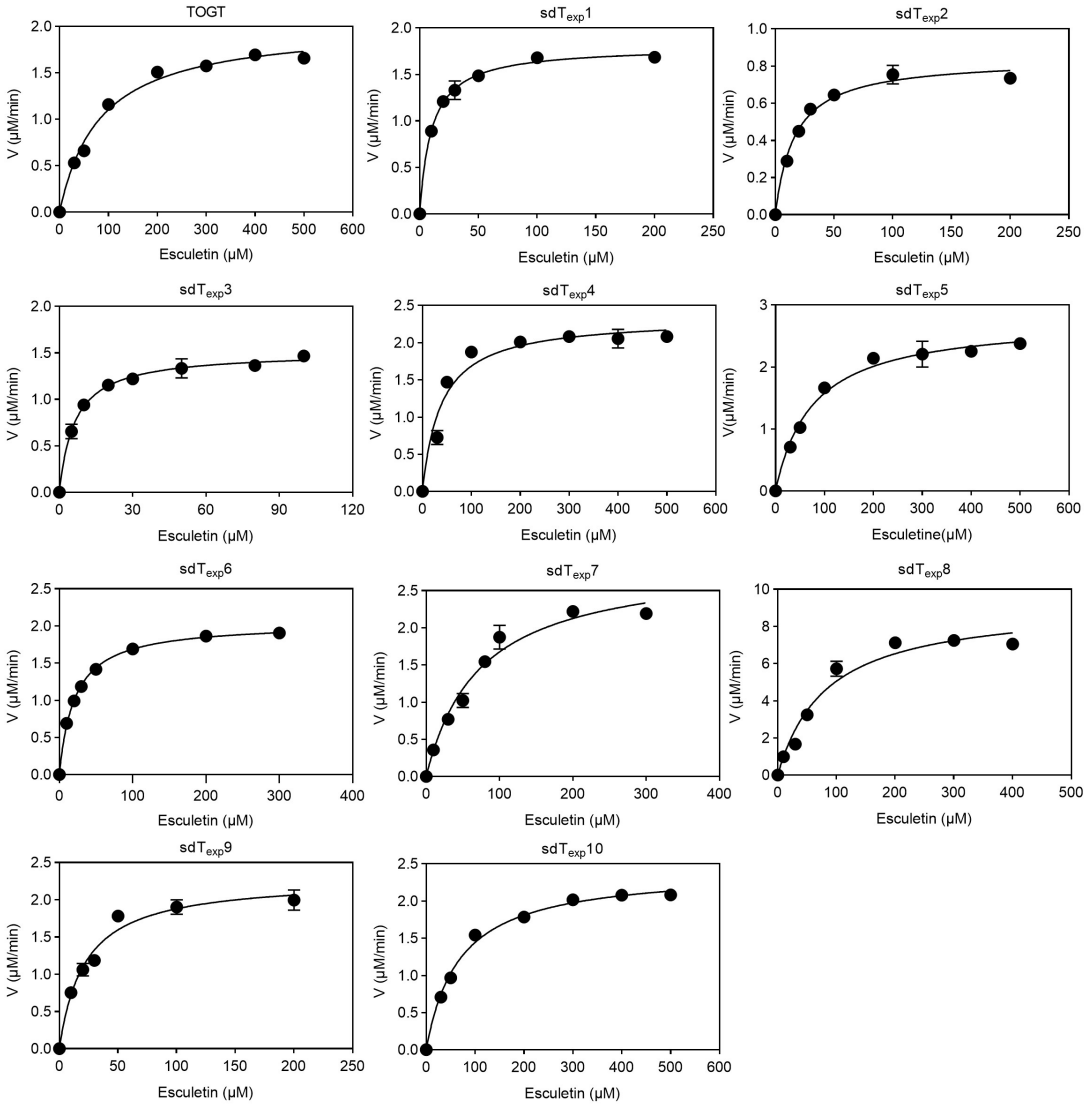
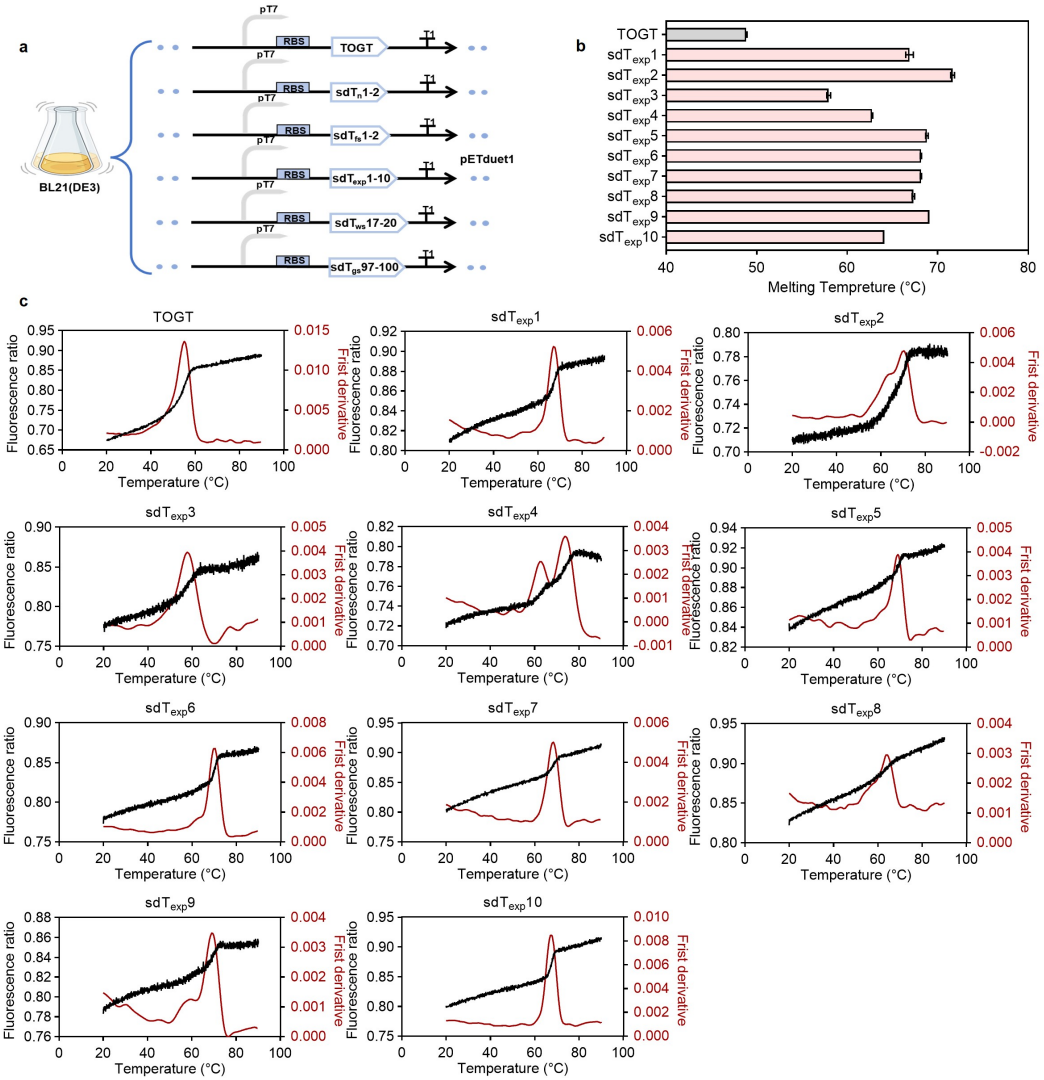


Fig. S4. Substrate and product profiling of TOGT and its designs. The chromatogram shows the retention times of the native TOGT and its sdT$_{exp}$ designs (sdT$_{exp}$1-sdT$_{exp}$10) in substrate addition assays. Both product (cichoriin) and substrate (esculetin) are indicated by arrows. No additional peaks are observed for the designs, which show a single peak corresponding to the product with the same retention time as the native TOGT. Each trace corresponds to a different design. Retention times are plotted on the x-axis, and peak height represents intensity (mV).



286

287 **Fig. S5. Michaelis-Menten curves of TOGT and sdT$_{exp}$ designs.**



288

289 **Fig. S6. Expression constructs and thermal stability of TOGT its designs. a,** Schematic illustration
290 of the expression constructs for native TOGT and designs, cloned into the pETDuet-1 vector under the
291 control of the T7 promoter, and expressed in *E. coli* BL21(DE3). **b, c,** Melting temperatures (T_m) of
292 TOGT and its sdT_{exp} designs measured by Nano Differential Scanning Fluorimetry (Nano DSF). All
293 designs (pink) exhibited improved thermostability relative to native TOGT (grey). The fluorescence ratio
294 (black line) and its first derivative (red line) are plotted as a function of temperature for TOGT and
295 designed enzymes. The peak in the first derivative curve indicates the apparent T_m of each protein. The
296 data suggest that designs exhibit increased thermal stability compared to the native TOGT.

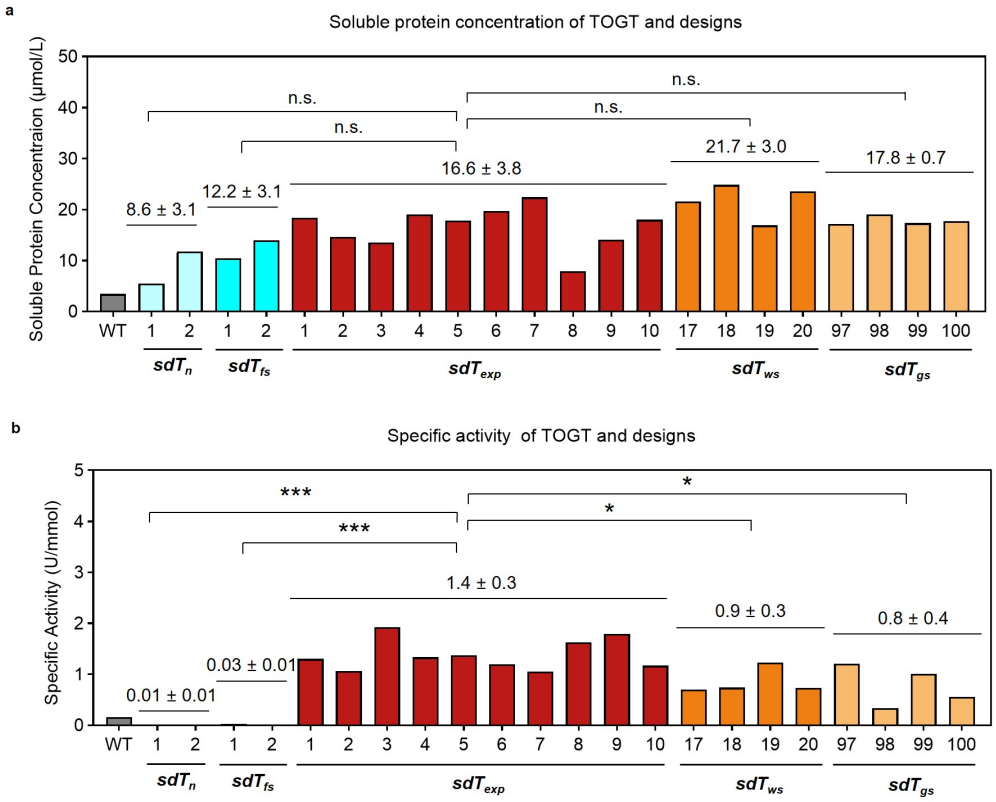


Fig. S7. Soluble expression levels and specific activity of TOGT and its designs. a, Soluble protein concentrations of TOGT designs expressed in *E. coli.* Protein levels were quantified from the equal volume and are shown as mean ± s.d. Statistical significance was assessed using Welch's t-test (one-tailed, unequal variance). **b,** Specific activities of native TOGT (WT) and designs. Designed groups include sdT_n (fixing no residues, using only ProteinMPNN for sequence regeneration without sequence filtering criteria), sdT_{fs} (fixing only residues in the first shell and conserved residues with sequence filtering criteria), sdT_{exp} (simultaneously fixing residues in the first shell and second shell as well as conserved residues with sequence filtering criteria), sdT_{ws} (bottom four designs in weight score within the sdT_{exp} group) and sdT_{gs} (bottom four designs in global score within the sdT_{exp} group). The specific activity of the sdT_{exp} group was significantly higher than that of the other groups. Significance levels are indicated as follows: $P \leqslant 0.001$ is denoted by ***, $P \leqslant 0.01$ is denoted by **, $P \leqslant 0.05$ is denoted by *, and $P > 0.05$ is denoted by n.s. (no significant).

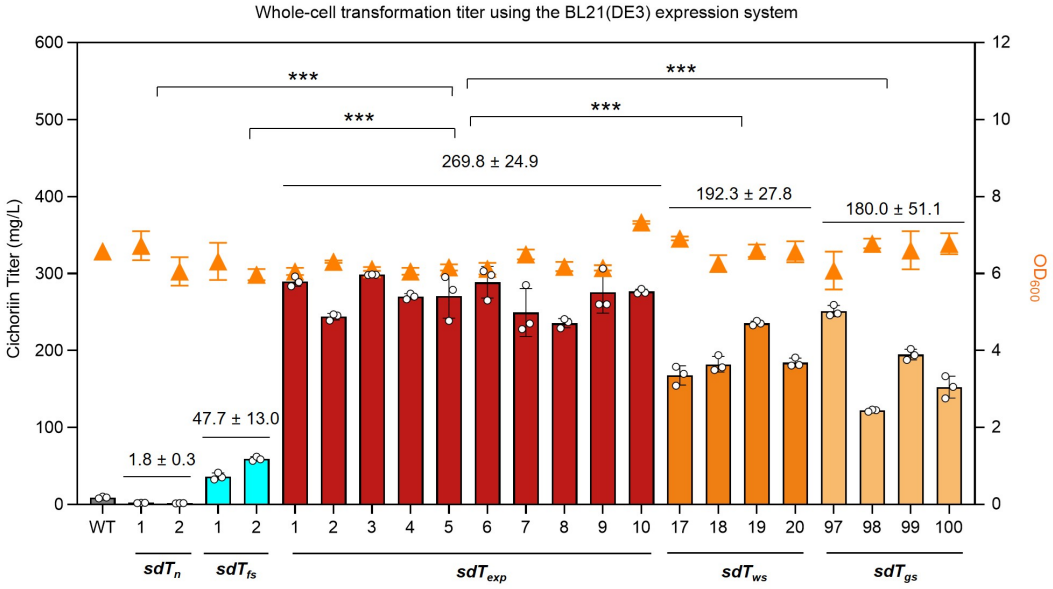


Fig. S8. Whole-cell biotransformation using *E. coli* BL21(DE3) expressing TOGT and the designs with esculetin as the substrate. Cichoriin Titers (bars, left axis) and corresponding cell densities (triangles, right axis) of native TOGT (grey) and designs were determined after whole-cell biotransformation. Designs include sdT_n designs, sdT_{fs} designs, sdT_{exp} designs, sdT_{ws} designs, and sdT_{gs} designs. Bars represent mean ± s.d. of three biological replicates. Statistical significance was assessed using Welch's t-test (one-tailed, unequal variance). n.s., $p > 0.05$, no significant; *, $p \leqslant 0.05$; **, $p \leqslant 0.01$; ***, $p \leqslant 0.001$. sdT_{exp} designs showed significantly enhanced titers.

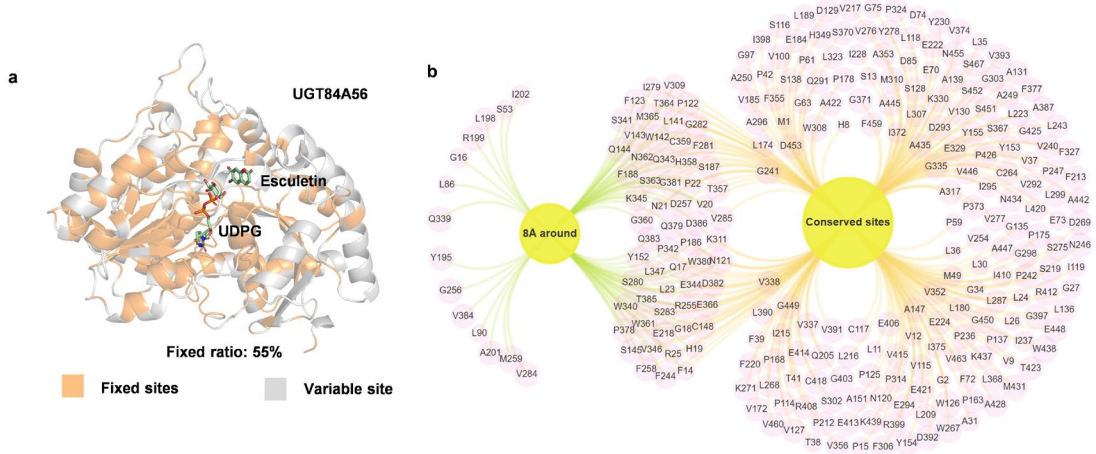


Fig. S9. Residue map of fixed sites used in UGT84A56 sequence redesign. a, Structural overview of fixed and variable (designable) sites in UGT84A56 during sequence redesign. The UGT84A56 structure is shown in cartoon representation, with fixed sites (65% of the sequence) highlighted in orange and variable sites in gray. The bound substrate esculetin and donor UDP-glucose (UDPG) are displayed in stick representation. **b,** Residues within 8 Å of the conserved sites (left) include the first- and second-shell residues surrounding the active site were fixed to maintain functional integrity. Conserved sites (right) represent evolutionarily constrained sites that were fixed during design to avoid random mutations in high-risk regions.

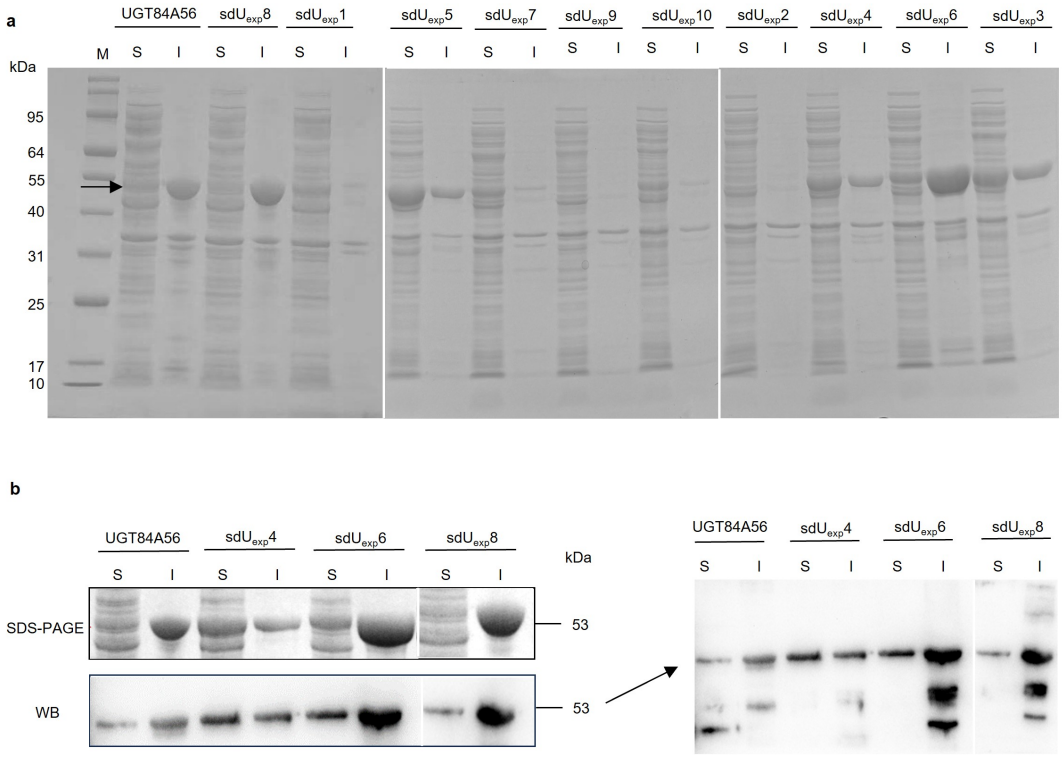


Fig. S10. Soluble expression characterization of UGT84A56 and its designs. a, SDS-PAGE analysis of UGT84A56 and its sdU$_{exp}$ designs (sdU$_{exp}$1-sdU$_{exp}$10) by *E. coli* BL21 (DE3) cells. The position of the target band is indicated by an arrow. **b,** SDS-PAGE and Western blot analysis of selected sdU$_{exp}$ designs (sdU$_{exp}$4, sdU$_{exp}$6, and sdU$_{exp}$8) with higher crude lysate activity than the native enzyme. Enhanced soluble expression of the designs is evident in both SDS-PAGE and WB analyses. "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate. The right-hand panel, as indicated by the arrow, presents the uncropped original western blot image.

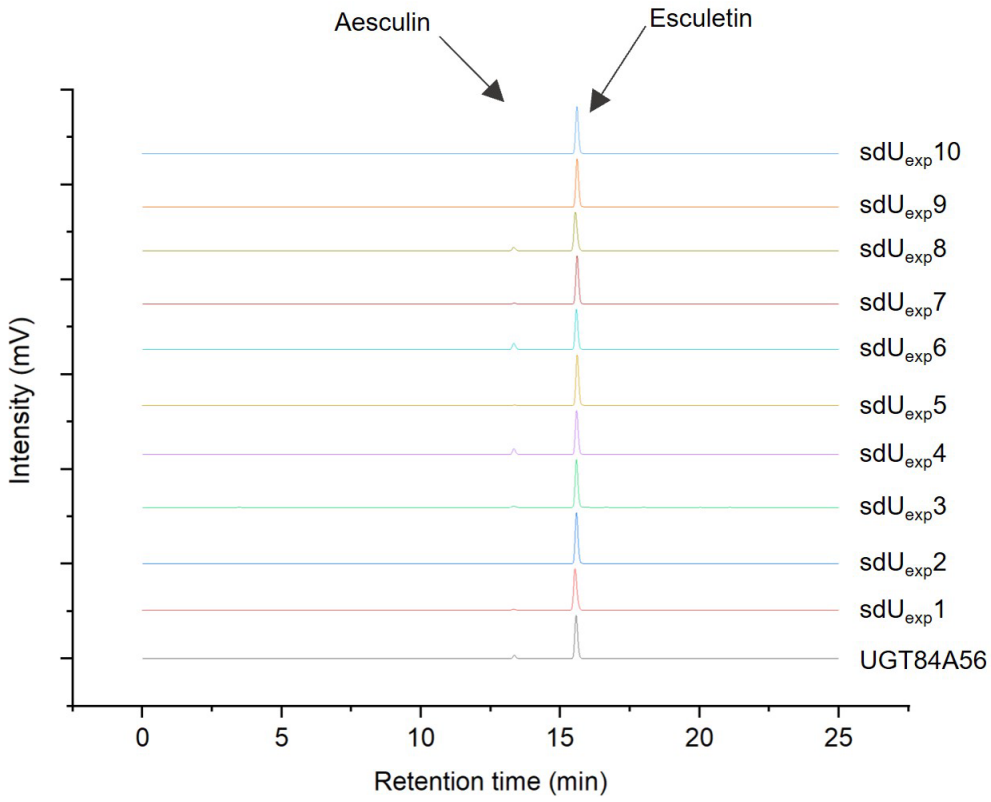
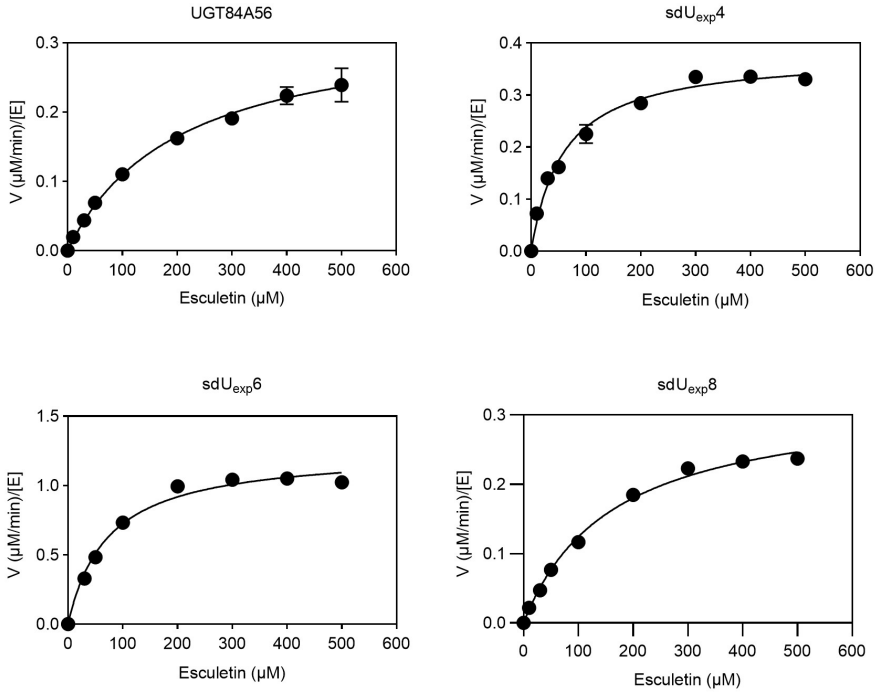


Fig. S11. Substrate and product profiling of UGT84A56 and its designs. The chromatogram shows the retention times of the native UGT84A56 and its sdU$_{exp}$ designs (sdU$_{exp}$1-sdU$_{exp}$10) in substrate addition assays. Both product (aesculin) and substrate (esculetin) are indicated by arrows. No additional peaks are observed for the designs, which show a single peak corresponding to the product with the same retention time as the native UGT84A56. Each trace corresponds to a different design. Retention times are plotted on the x-axis, and peak height represents intensity (mV).



346

347 **Fig. S12. Michaelis-Menten curves of UGT84A56, sdU$_{exp}$ designs (sdU$_{exp}$4, sdU$_{exp}$6 and sdU$_{exp}$8).**

348

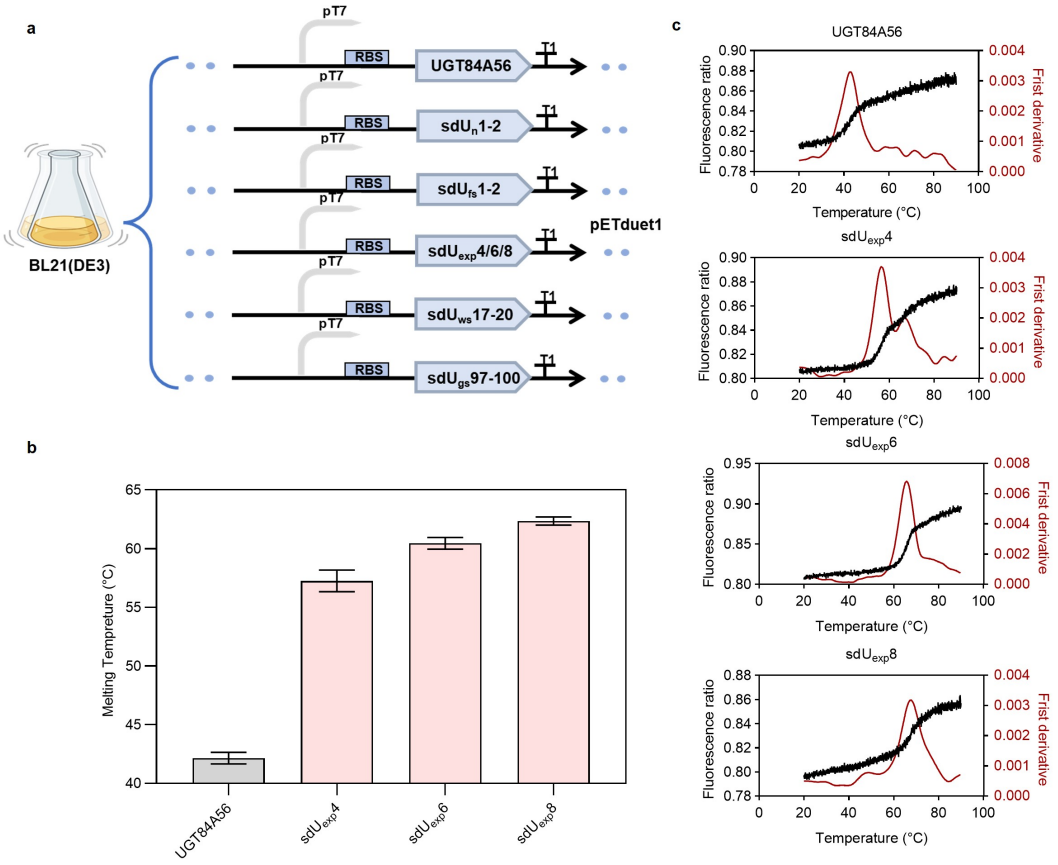


Fig. S13. Expression constructs and thermal stability of UGT84A56 and its designs. a, Schematic illustration of the expression constructs for native UGT84A56 and designs cloned into the pETDuet-1 vector under the control of the T7 promoter, and expressed in *E. coli* BL21(DE3). **b, c,** Melting temperatures (T_m) of UGT84A56 and its designs measured by Nano DSF. All designs (pink) exhibited improved thermostability relative to native UGT84A56 (grey). The fluorescence ratio (black line) and its first derivative (red line) are plotted as a function of temperature for UGT84A56 and three designs. The peak in the first derivative curve indicates the apparent T_m of each protein. The data suggest that sdU$_{exp}$4, sdU$_{exp}$6, and sdU$_{exp}$8 exhibit increased thermal stability compared to the native UGT84A56.

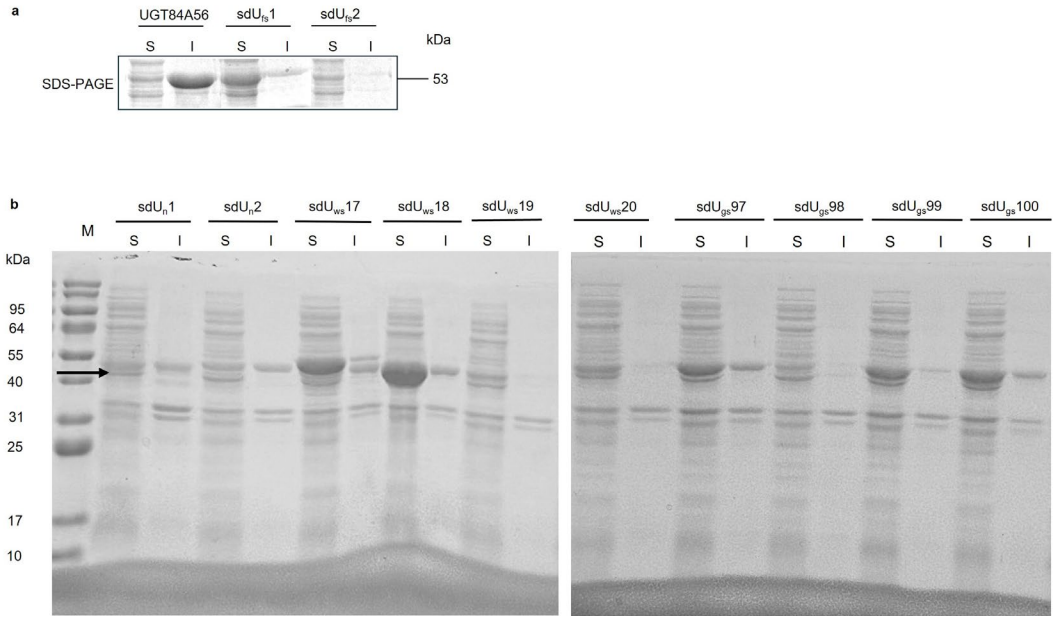


Fig. S14. Soluble expression characterization of UGT84A56 and other four strategy designs. a, SDS-PAGE analysis of the UGT84A56 and sdU_{fs} designs ($sdU_{fs}1$ and $sdU_{fs}2$) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate. **b,** SDS-PAGE analysis of the expression of other three strategy designs (sdU_p1, sdU_p2, $sdU_{ws}17$-$sdU_{ws}20$ and $sdU_{gs}97$-$sdU_{gs}100$) by *E. coli* BL21 (DE3) cells, "S" represents the soluble fraction of the cell lysate, "I" represents the precipitates of the cell lysate.

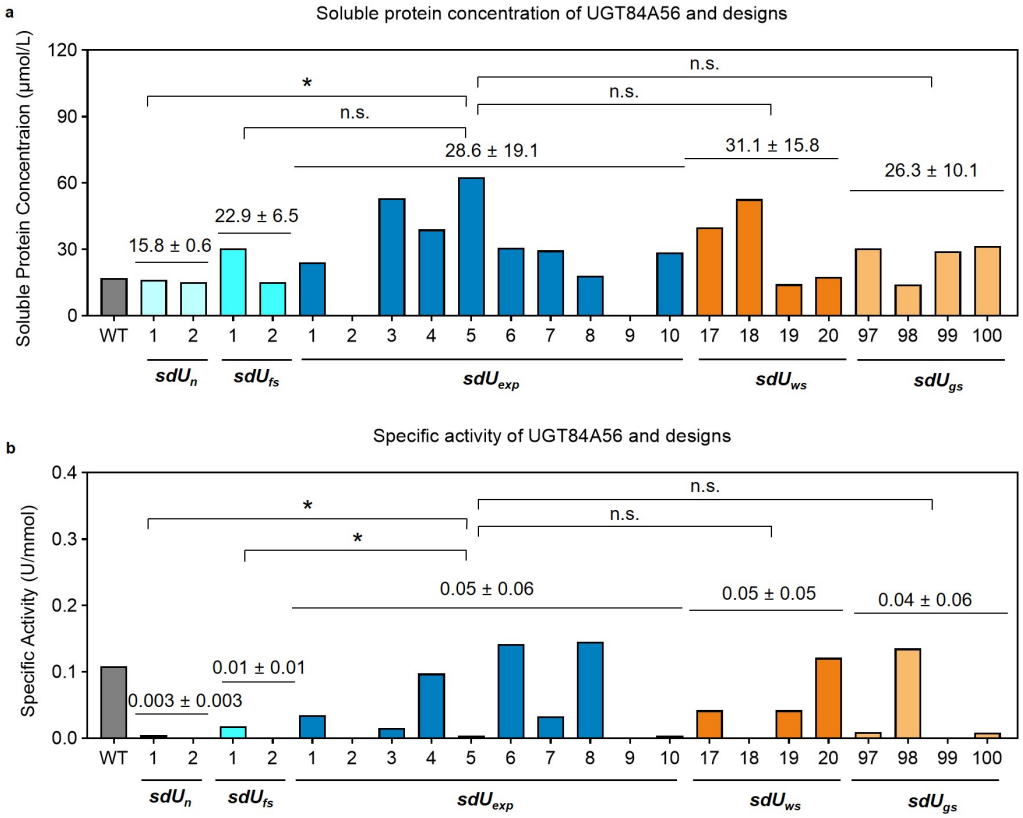
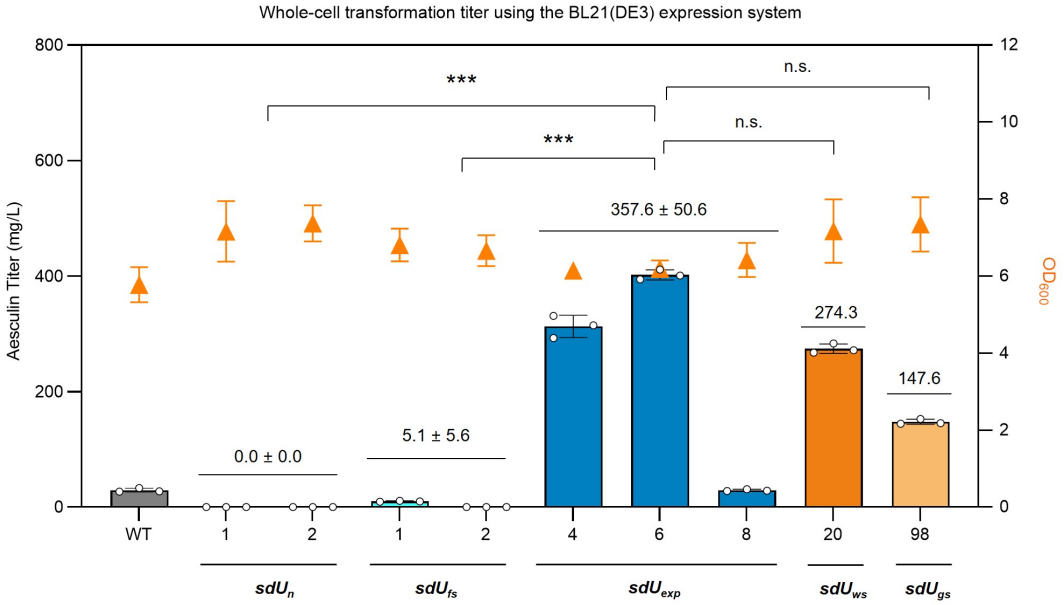


Fig. S15. Soluble expression levels and specific activity and of UGT84A56 its designs. a, Soluble protein concentrations of UGT84A56 designs expressed in *E. coli.* Protein levels were quantified from the equal volume and are shown as mean ± s.d. sdU_{exp} exhibited comparable soluble protein expression to sdU_{gs} and sdU_{ws}, but exceeded that of sdU_{n}. However, two enzymes from the designed groups failed to express, suggesting that the design model still requires further optimization to reliably generate soluble enzymes. **b,** Specific activities of native UGT84A56 (WT) and designs. Designed groups include sdU_{n} (fixing no residues, using only ProteinMPNN for sequence regeneration without sequence filtering criteria), sdU_{fs} (fixing only residues in the first shell and conserved residues with sequence filtering criteria), sdU_{exp} (simultaneously fixing residues in the first shell and second shell as well as conserved residues with sequence filtering criteria), sdU_{ws} (bottom four designs in weight score within the sdU_{exp} group) and sdU_{gs} (bottom four designs in global score within the sdU_{exp} group). The specific activity of the sdU_{exp} group had no significant to sdU_{ws} and sdU_{gs} groups. Statistical significance was assessed using Welch's t-test (one-tailed, unequal variance). Significance levels are indicated as follows: $P \leqslant 0.001$ is denoted by ***, $P \leqslant 0.01$ is denoted by **, $P \leqslant 0.05$ is denoted by *, and $P > 0.05$ is denoted by n.s. (no significant).



384

385 **Fig. S16. Whole-cell biotransformation using *E. coli* BL21(DE3) expressing UGT84A56 and the**
386 **designs with esculetin as the substrate.** Aesculin Titers (bars, left axis) and corresponding cell densities
387 (triangles, right axis) of native UGT84A56 (grey) and designs were determined after whole-cell
388 biotransformation. Designs include sdU_n designs, sdU_{fs} designs, sdU_{exp} designs, sdU_{ws} designs, and sdU_{gs}
389 designs. Bars represent mean ± s.d. of three biological replicates. Statistical significance was assessed
390 using Welch's t-test (one-tailed, unequal variance). n.s., $p > 0.05$, no significant; *, $p \leqslant 0.05$; **, $p \leqslant$
391 0.01; ***, $p \leqslant 0.001$. sdU_{exp} designs showed significantly enhanced titers to sdU_n designs and sdU_{fs}
392 designs.

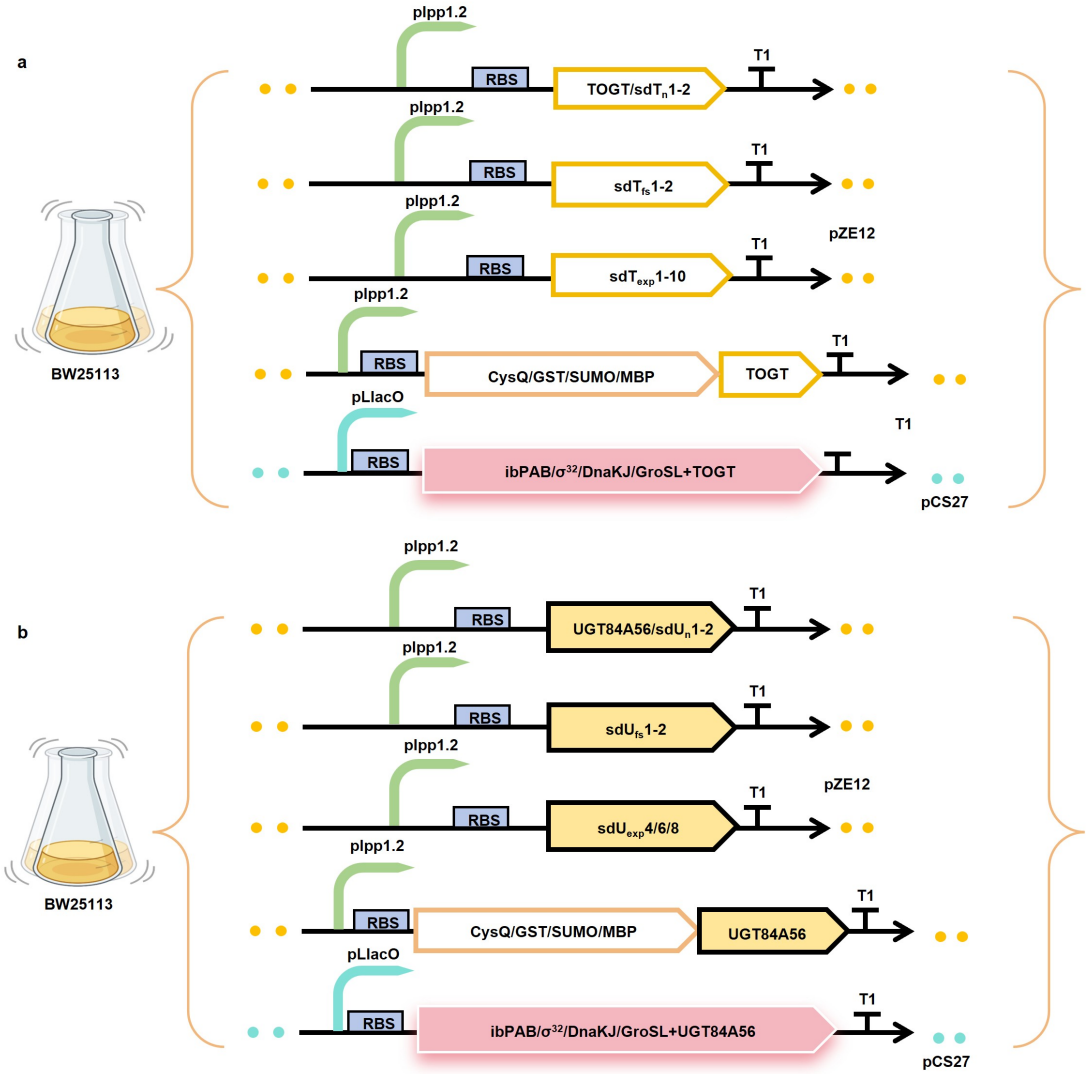
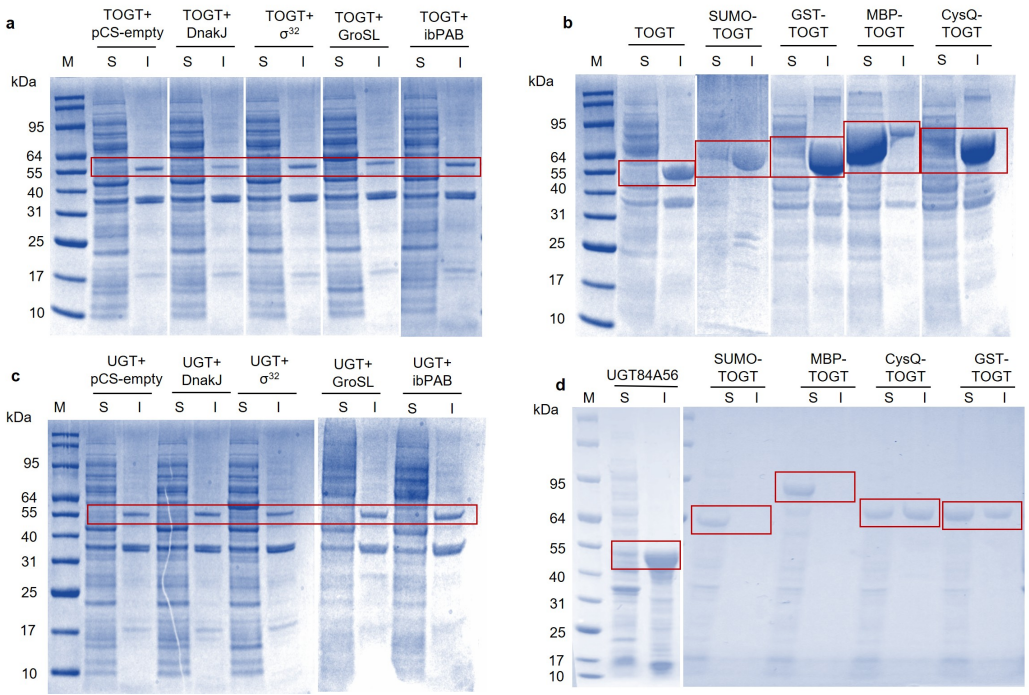


Fig. S17. a, Schematic diagram of plasmid constructs used for the expression of TOGT and its designs in *E. coli* strain BW25113. The plasmid pZE12 and pCS27 are used for the expression system. The plasmids include promoters (plpp1.2 or pLlacO), ribosome binding sites (RBS), and terminator (T1). Constructs include TOGT and its designs. Co-expression with fusion-tags (CysQ, GST, SUMO, MBP) and chaperones (ibPAB, σ^{32}, DnaKJ, GroSL) were also shown. **b,** Schematic diagram of plasmid constructs used for the expression of UGT84A56 and its designs in *E. coli* strain BW25113. The plasmid pZE12 and pCS27 are used for the expression system. The plasmids include promoters (plpp1.2 or pLlacO), ribosome binding sites (RBS), and terminator (T1). Constructs include UGT84A56 and its designs. Co-expression with fusion-tags (CysQ, GST, SUMO, MBP) and chaperones (ibPAB, σ^{32}, DnaKJ, GroSL) were also shown.



404

405 **Fig. S18. a,** SDS-PAGE analysis of TOGT co-expressed with different molecular chaperones (DnaKJ,
406 σ^{32}, GroSL, and ibPAB) in soluble (S) and insoluble (I) fractions. The red box highlights the bands
407 corresponding to TOGT. **b,** SDS-PAGE analysis of TOGT fused with various fusion-tags (SUMO, GST,
408 MBP, and CysQ) in soluble (S) and insoluble (I) fractions. The red box highlights the bands
409 corresponding to the fusion proteins. **c,** SDS-PAGE analysis of UGT84A56 co-expressed with different
410 molecular chaperones (DnaKJ, σ^{32}, GroSL, and ibPAB) in soluble (S) and insoluble (I) fractions. The
411 red box highlights the bands corresponding to TOGT. **d,** SDS-PAGE analysis of UGT84A56 fused with
412 various fusion-tags (SUMO, MBP, CysQ, and GST) in soluble (S) and insoluble (I) fractions. The red
413 box highlights the bands corresponding to the fusion proteins.

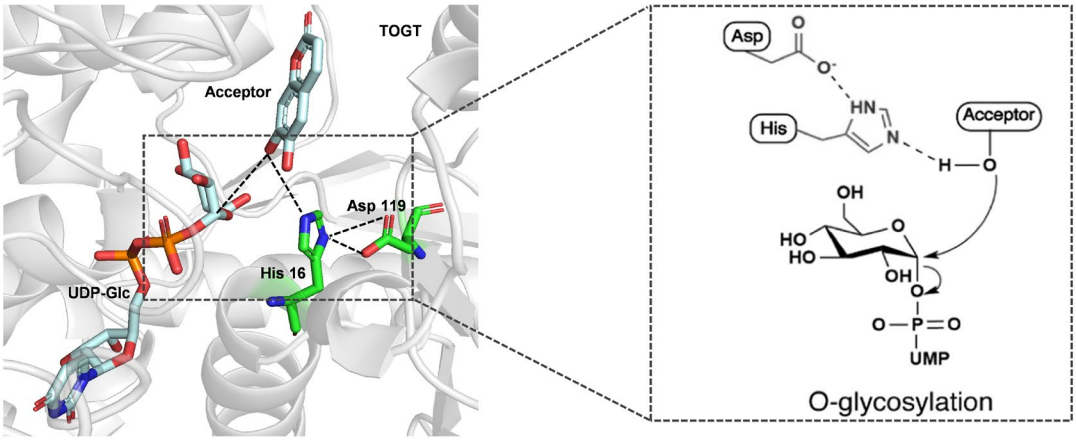
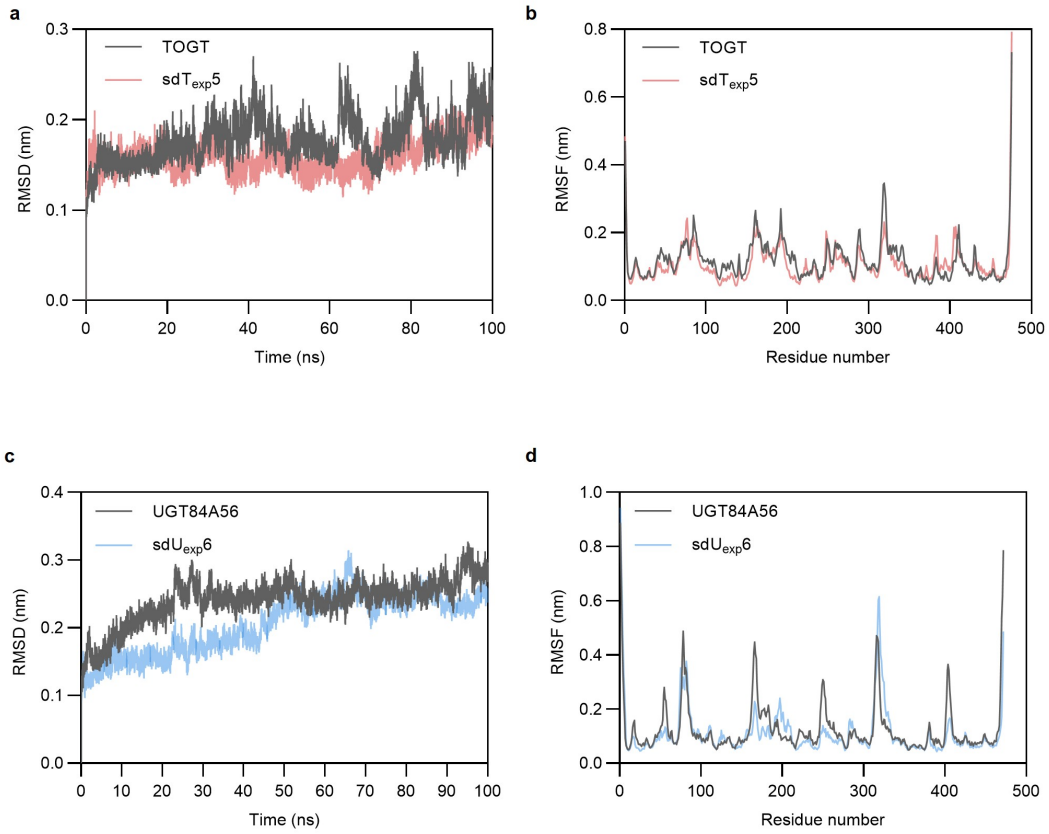


Fig. S19. Proposed catalytic mechanisms of TOGT in O-glycosylation. The Histidine-Aspartate dyad deprotonates the 7-hydroxyl group of the acceptor esculetin molecules, and the deprotonated acceptor develops nucleophilic properties, which lead to a nucleophilic attack on the anomeric carbon of the Glc portion replacing the UDP portion and ultimately leading to the formation of the product cichoriin.



420

421 **Fig. S20. Molecular dynamics simulation analysis of native and designed glycosyltransferases. a, b,**
422 Structural dynamics of TOGT and its designed enzyme sdT$_{exp}$5: **a,** Time-dependent root-mean-square
423 deviation (RMSD) profiles over 100 ns indicate that sdT$_{exp}$5 exhibits enhanced structural stability
424 compared to native TOGT. **b,** Per-residue root-mean-square fluctuation (RMSF) of TOGT and sdT$_{exp}$5.
425 **c, d,** Structural dynamics of UGT84A56 and its designed enzyme sdU$_{exp}$6: **c,** RMSD trajectories show
426 that sdU$_{exp}$6 maintains lower overall conformational deviation throughout the simulation. **d,** RMSF
427 profiles of UGT84A56 and sdU$_{exp}$6.

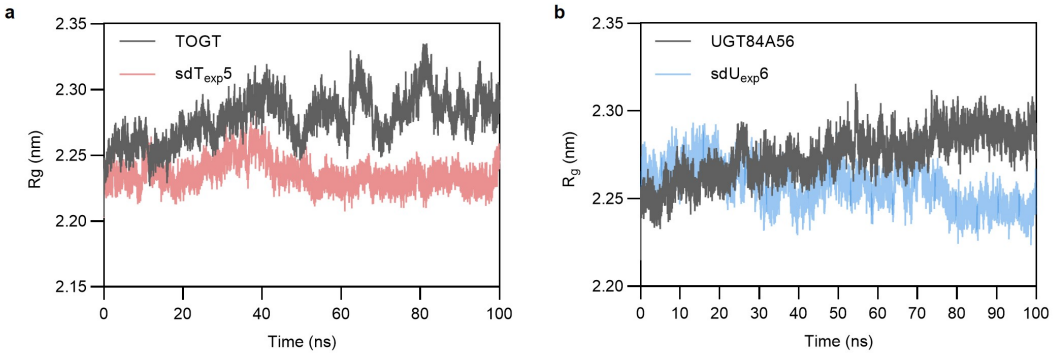
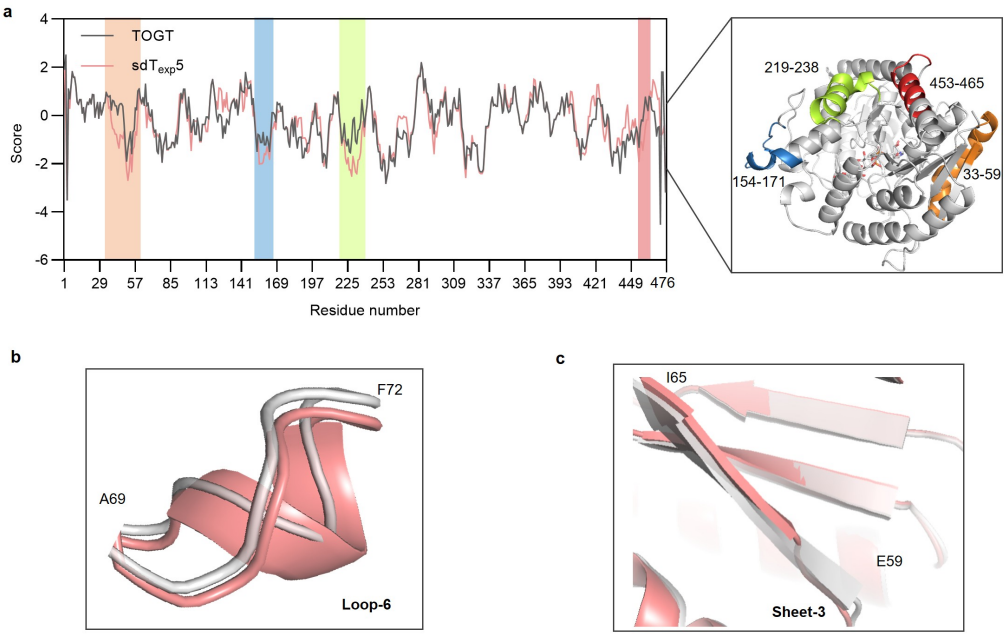


Fig. S21. Time evolution of the radius of gyration (R_g) for native and engineered glycosyltransferases over 100 ns molecular dynamics simulations. a, R_g profiles of TOGT and its designed enzyme sdT$_{exp}$5. The sdT$_{exp}$5 maintains a consistently lower and more stable R_g throughout the simulation, indicating enhanced compactness and structural integrity relative to native TOGT. **b,** R_g profiles of UGT84A56 and its designed enzyme sdU$_{exp}$6. Similar to sdT$_{exp}$5, sdU$_{exp}$6 exhibits reduced R_g values with lower fluctuation amplitudes compared to UGT84A56, suggesting improved conformational stability and rigidity.



436

437 **Fig. S22. Hydrophobicity analysis and structural comparison between native TOGT and its**
438 **designed enzyme sdT$_{exp}$5. a,** Hydrophobicity profiles of TOGT and sdT$_{exp}$5 plotted as a function of
439 residue number. Lower scores indicate increased hydrophilicity. Several regions exhibiting enhanced
440 hydrophilicity in sdT$_{exp}$5 (highlighted with colored bars) correspond to surface-exposed segments
441 (residues 26-58, 145-162, 218-235, and 453-465), as mapped onto the 3D structure (right panel). These
442 shifts align with the rational design strategy of selectively modifying non-conserved outer-shell residues
443 to improve protein solubility. **b, c,** Structural overlays highlighting local secondary structure differences
444 between TOGT (gray) and sdT$_{exp}$5 (pink). **b,** Loop-6 region exhibits a slight conformational
445 rearrangement in sdT$_{exp}$5 which may lead to increased rigidity. **c,** β-sheet-3 region shows local structural
446 deviation, potentially contributing to altered hydrophobicity.

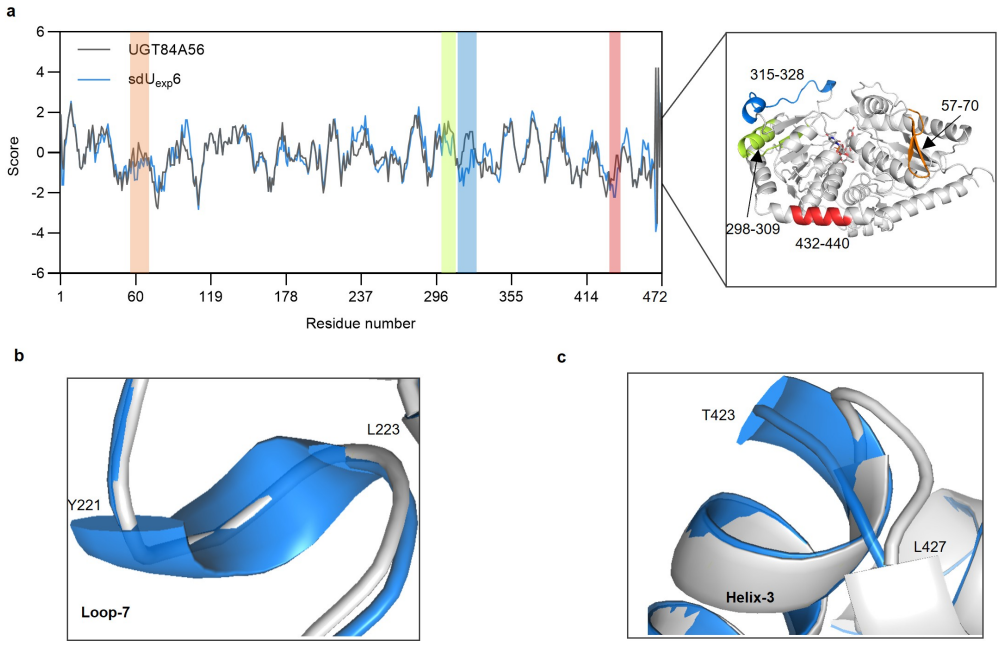


Fig. S23. Hydrophobicity analysis and structural comparison between native UGT84a56 and its designed enzyme sdU$_{exp}$6. a, Hydrophobicity profiles of UGT84A56 and sdU$_{exp}$6 plotted as a function of residue number. Lower scores indicate increased hydrophilicity. Several regions exhibiting enhanced hydrophilicity in sdU$_{exp}$6 (highlighted with colored bars) correspond to surface-exposed segments (residues 57-70, 298-309, 315-328 and 432-440), as mapped onto the 3D structure (right panel). These also shifts align with the rational design strategy of selectively modifying non-conserved outer-shell residues to improve protein solubility. **b, c,** Structural overlays highlighting local secondary structure differences between UGT84A56 (gray) and sdU$_{exp}$6 (blue). **b,** Loop-7 region exhibits a slight conformational rearrangement in sdU$_{exp}$6 which may lead to increased rigidity. **c,** Superimposed structures reveal localized conformational alterations in the Helix-3 region. Relative to the native, the sdU$_{exp}$6 design displays an extended backbone conformation and a reshaped helix-loop junction, potentially contributing to modified surface topology and enhanced structural rigidity.

Tables

Table S1: Detailed information of the TOGT designs.

The molecular weight, amino acid length, RMSD compared to the retained template structures by VMD 1.9.3, pLDDT and pTM-score given by AlphaFold 2, the Global-Score given by ProteinMPNN, sequence identity between the designs and TOGT was obtained by NCBI Blast alignment. Normalized weighted scoring was performed on designs. The four evaluated parameters—predicted Template Modeling score (pTM), predicted Local Distance Difference Test (pLDDT), solubility, and root-mean-square deviation (RMSD)—were assigned weights of 0.1, 0.2, 0.3, and 0.4, respectively, to reflect their relative contributions to overall designs performance.

Protein	Molecular Weight (kDa)	Length (aa)	pTM	RMSD (Å)	pLDDT	Solubility	Per. Ident (%)	Weighted Score
TOGT	53.65	476	-	-	-	-	-	-
$sdT_{exp}1$	53.95	476	0.935	0.49	95.4	0.404	76.26	0.815
$sdT_{exp}2$	53.80	476	0.931	0.46	94.9	0.424	76.89	0.794
$sdT_{exp}3$	53.47	476	0.933	0.46	94.9	0.403	77.73	0.721
$sdT_{exp}4$	53.89	476	0.933	0.50	95.4	0.385	76.05	0.713
$sdT_{exp}5$	53.66	476	0.929	0.42	94.9	0.369	76.05	0.600
$sdT_{exp}6$	53.45	476	0.939	0.68	95.1	0.383	76.89	0.476
$sdT_{exp}7$	54.02	476	0.932	0.50	94.8	0.352	77.10	0.438
$sdT_{exp}8$	53.92	476	0.932	0.62	94.8	0.384	77.31	0.436
$sdT_{exp}9$	54.01	476	0.939	0.76	94.8	0.395	76.68	0.370
$sdT_{exp}10$	53.99	476	0.932	0.62	94.9	0.354	76.47	0.335
$sdT_{exp}11$	53.55	476	0.926	0.62	94.8	0.354	76.05	0.276
$sdT_{exp}12$	53.63	476	0.933	0.63	94.5	0.351	76.89	0.229
$sdT_{exp}13$	53.68	476	0.927	0.69	94.9	0.354	76.47	0.226
$sdT_{exp}14$	53.98	476	0.929	0.65	94.6	0.353	77.94	0.212
$sdT_{exp}15$	53.83	476	0.931	0.78	95.2	0.352	77.31	0.210
$sdT_{exp}16$	53.89	476	0.934	0.66	94.5	0.351	76.05	0.202
$sdT_{ws}17$	53.80	476	0.932	0.71	94.6	0.353	76.89	0.164
$sdT_{ws}18$	54.00	476	0.929	0.75	94.8	0.351	76.47	0.138
$sdT_{ws}19$	53.46	476	0.923	0.72	94.7	0.352	75.63	0.115
$sdT_{ws}20$	53.50	476	0.923	0.72	94.7	0.352	76.47	0.115

Table S2: Detailed information of the UGT84A56 designs.

The molecular weight, amino acid length, RMSD compared to the retained template structures by VMD 1.9.3, pLDDT and pTM-score given by AlphaFold 2, the Global-Score given by ProteinMPNN, sequence identity between the designs and TOGT was obtained by NCBI Blast alignment. Normalized weighted scoring was performed on designs. The four evaluated parameters—pTM, pLDDT, solubility and RMSD—were assigned weights of 0.1, 0.2, 0.3 and 0.4, respectively, to reflect their relative contributions to overall designs performance.

Protein	Molecular Weight (kDa)	Length (aa)	pTM	RMSD (Å)	pLDDT	Solubility	Per. Ident (%)	Weighted Score
UGT84A56	52.85	472	-	-	-	-	-	-
$sdU_{exp}1$	52.42	472	0.943	0.48	96.0	0.589	77.75	0.880195
$sdU_{exp}2$	52.43	472	0.944	0.45	96.1	0.425	76.27	0.793277
$sdU_{exp}3$	52.46	472	0.942	0.47	96.1	0.424	76.91	0.716692
$sdU_{exp}4$	52.53	472	0.943	0.50	96.1	0.456	77.75	0.680405
$sdU_{exp}5$	52.29	472	0.941	0.51	95.6	0.520	77.33	0.589324
$sdU_{exp}6$	52.86	472	0.939	0.52	95.6	0.548	77.12	0.577865
$sdU_{exp}7$	52.26	472	0.939	0.51	95.4	0.518	77.54	0.518621
$sdU_{exp}8$	52.44	472	0.938	0.50	95.4	0.490	77.12	0.502807
$sdU_{exp}9$	52.10	472	0.943	0.50	95.9	0.353	76.48	0.500573
$sdU_{exp}10$	52.43	472	0.936	0.58	95.6	0.513	76.06	0.335046
$sdU_{exp}11$	52.30	472	0.942	0.58	95.8	0.423	76.69	0.326146
$sdU_{exp}12$	52.69	472	0.934	0.59	95.3	0.424	78.18	0.301108
$sdU_{exp}13$	52.62	472	0.936	0.56	95.6	0.420	77.33	0.274962
$sdU_{exp}14$	52.44	472	0.936	0.59	95.6	0.403	76.48	0.167819
$sdU_{exp}15$	52.63	472	0.938	0.58	96.1	0.353	77.12	0.151547
$sdU_{exp}16$	52.11	472	0.935	0.58	95.4	0.413	76.91	0.149905
$sdU_{ws}17$	52.85	472	0.939	0.58	95.4	0.423	77.54	0.145302
$sdU_{ws}18$	52.33	472	0.935	0.59	95.3	0.352	77.33	0.138025
$sdU_{ws}19$	52.46	472	0.933	0.61	95.1	0.418	77.75	0.084454
$sdU_{ws}20$	52.75	472	0.936	0.59	95.5	0.351	76.06	0.077273

Table S3: Detailed information of the TOGT and design enzyme sdT$_{exp}$5.

The depth and Surface-Volume-Ratio of the active pocket, and calculated volume are summarized below. The calculations for the pocket channel characteristics indicated that the designs have smaller Surface-Volume-Ratio, suggesting more compact pockets with higher specificity, thus minimizing promiscuity.

Protein	Surface-Volume-Ratio	Volume (Å^3) *	Depth (Å) *
TOGT	0.98	1131	29.00
sdT$_{exp}$5	0.89	1107	28.63

* The data were calculated by Protein plus.

Table S4: Detailed information of the UGT84A56 and design enzyme $sdU_{exp}6$.

The depth and Surface-Volume-Ratio of the active pocket, and calculated volume are summarized below. The calculations for the pocket channel characteristics indicated that the designs have smaller Surface-Volume-Ratio, suggesting more compact pockets with higher specificity, thus minimizing promiscuity.

Protein	Surface-Volume-Ratio	Volume (Å^3) *	Depth (Å) *
UGT84A56	1.01	1291	28.18
$sdU_{exp}6$	1.04	1200	30.63

* The data were calculated by Protein plus.

Table S5. Amino acid sequences and nucleotide sequences (codon optimized for *E. coli* expression) of the native TOGT and designer enzymes sdT$_{exp}$1-sdT$_{exp}$10. Amino acid sequences in black words, nucleotide sequences in blue words.

Native TOGT

MGQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPLNEFVFSKAIQRNKHLGIEIEIRLIKFPAVENGLP
EECERLDQIPSDEKLPNFFKAVAMMQEPLEQLIEECRPDCLISDMFLPWTTDTAAKFNIPRIVFHGTSFFALCVE
NSVRLNKPFKNVSSDSETFVVPDLPHEIKLTRTQVSPFERSGEETAMTRMIKTVRESDSKSYGVVFNSFYELETD
YVEHYTKVLGRRAWAIGPLSMCNRDIEDKAERGKKSSIDKHECLKWLDSKKPSSVVYVCFGSVANFTASQLHELA
MGIEASGQEFIWVVRTELDNEDWLPEGFEERTKEKGLIIRGWAPQVLILDHESVGAFVTHCGWNSTLEGVSGGVP
MVTWPVFAEQFFNEKLVTEVLKTGAGVGSIQWKRSASEGVKREAIAKAIKRVMVSEEADGFRNRAKAYKEMARKA
IEEGGSSYTGLTTLLEDISTYSSTGH

ATGGGTCAGCTGCATTTTTTCTTTTTCCCGGTGATGGCACACGGTCATATGATTCCGACCCTGGATATGGCAAAA
CTGTTTGCAAGCCGTGGTGTTAAAGCAACCATTATTACCACCCCGCTGAATGAATTTGTTTTCAGCAAAGCAATT
CAGCGCAATAAGCACCTGGGTATCGAAATCGAAATCCGTCTGATCAAATTTCCGGCCGTTGAAAATGGTCTGCCG
GAAGAATGTGAACGTCTGGATCAGATTCCGAGCGATGAAAAACTGCCGAATTTTTTTAAAGCCGTGGCCATGATG
CAGGAGCCGCTGGAACAGTTAATTGAAGAATGTCGTCCGGATTGTCTGATTAGCGATATGTTTCTGCCGTGGACC
ACCGATACCGCAGCAAAATTTAATATTCCGCGTATTGTGTTTCACGGCACCAGCTTTTTTGCACTGTGTGTTGAA
AATAGCGTGCGCCTGAATAAACCGTTTAAGAACGTGAGCAGCGATAGCGAAACCTTTGTGGTTCCGGATCTGCCG
CATGAAATTAAACTGACCCGTACCCAGGTTAGCCCGTTTGAACGTAGCGGTGAAGAAACCGCAATGACCCGTATG
ATTAAAACCGTTCGTGAAAGCGATAGCAAAAGCTATGGTGTTGTTTTTAACAGCTTTTACGAGCTGGAAACCGAC
TACGTTGAGCATTATACCAAAGTTCTGGGTCGTCGTGCATGGGCAATTGGTCCGTTAAGCATGTGTAATCGTGAT
ATTGAAGACAAGGCCGAACGCGGCAAAAAGAGCAGCATTGATAAACATGAATGCCTGAAATGGCTGGACAGCAAA
AAACCGAGCAGCGTTGTTTATGTTTGTTTCGGCAGCGTTGCCAACTTCACCGCAAGCCAGTTACATGAACTGGCA
ATGGGTATTGAAGCAAGCGGTCAGGAATTTATTTGGGTTGTTCGTACCGAACTGGATAATGAAGATTGGCTGCCG
GAAGGTTTTGAAGAACGTACCAAAGAAAAAGGCCTGATTATCCGTGGTTGGGCACCGCAGGTTCTGATTCTGGAT
CATGAAAGCGTTGGTGCATTTGTTACCCATTGTGGTTGGAATAGCACCCTGGAAGGTGTTAGCGGTGGTGTTCCG
ATGGTTACCTGGCCGGTTTTTGCAGAACAGTTTTTTAATGAGAAGCTGGTGACCGAGGTGCTGAAGACCGGTGCA
GGTGTTGGTAGCATTCAGTGGAAACGTAGCGCAAGCGAAGGTGTTAAACGTGAAGCAATTGCAAAAGCCATTAAG
CGCGTGATGGTTAGCGAAGAAGCCGACGGTTTTCGTAATCGTGCAAAAGCATATAAAGAGATGGCCCGTAAAGCC
ATTGAGGAGGGTGGTAGCAGCTATACCGGTCTGACCACCCTGCTGGAAGATATTAGCACCTATAGCAGCACCGGT
CATTAA

Amino acid sequences and nucleotide sequences of designer enzymes

> sdT$_{exp}$1

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEDVFSKEIERWKPRGIKIEIRLIKFPAKEFGLP
EECERLDQLPSDADLPKFFAALKAMQEPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIVFHGTSFFARVVE
YSVRKHKPWKDVKSDTEKFVVPDLPHKIELTKTQVSPFIRSEEKTPMSELIEEMLESDKKSWGVVFNSFYELEKD
YFDHYTNVLKRRAWAIGPLSMCNTTLEDKAERGKKSSIDPAECLAWLDSKAPDSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESESWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATVWKRNDADGVSREAIAAAIRRVMTSEEAPAYRERARELAEKAKKA
IEEGGSSDKGLQELLEDIRAFRLANH
ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCGCACGGTCACATGATCCCGACCTTGGACATGGCTAAA
CTGTTCGCGAGCCGTGGTGTTAAAGCGACTATCATCACCACTCCAGGTAACGAAGACGTGTTCTCCAAAGAAATC

GAACGTTGGAAACCGCGTGGTATCAAGATCGAAATCCGTCTGATCAAATTTCCAGCGAAAGAGTTCGGCCTGCCG
GAAGAATGCGAACGTCTGGACCAGCTTCCGTCTGACGCTGACCTGCCGAAATTCTTCGCTGCTCTGAAAGCAATG
CAAGAACCGCTGGAAGAACTCATCGCGGAAGAGAAACCGGACTGCCTGATCTCTGATATGTTTCTGCCGTGGACC
GTTGACGTAGCGGCTAAATTCAACATTCCACGTATCGTGTTCCACGGCACCTCTTTCTTCGCTCGTGTTGTTGAA
TACTCTGTACGTAAACACAAACCGTGGAAAGACGTTAAATCCGACACCGAGAAATTCGTTGTTCCGGATCTGCCG
CACAAGATCGAACTGACCAAGACTCAGGTGTCTCCGTTCATCCGTTCTGAAGAGAAGACTCCGATGTCCGAACTG
ATCGAAGAAATGCTGGAATCTGACAAGAAATCTTGGGGTGTGGTGTTCAACTCTTTCTACGAACTGGAGAAAGAC
TACTTCGATCACTACACCAACGTTCTGAAACGTCGTGCGTGGGCTATCGGTCCACTGTCTATGTGCAACACCACC
TTGGAAGACAAAGCGGAACGTGGTAAGAAATCCAGCATCGATCCAGCAGAATGCCTGGCTTGGCTGGATTCTAAA
GCGCCGGATTCCGTTGTTTACGTTTGCTTCGGTTCTGTTGCGCGTCTGACTCCGGAACAGCTGCGTGAACTGGCA
TTGGGTATCGAAGCGAGCGGTCAAGACTTCATCTGGGTTGTTCGTCCGGACTACGAATCTGAATCTTGGCTGCCG
GAAGGCTTCGAAGAACGTACTAAAGACAAAGGTCTGATCATCCGTGGTTGGGCTCCACAGGTTCTGATCTTGGAT
CATCCGTCTGTTGGTGCGTTCGTTACTCACTGTGGCTGGAACTCTACTCTGGAAGGTGTTGCAGGTGGTGTTCCA
ATGGTAACCTGGCCGGTATTCGCGGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTGAAGACCGGTGCT
AGCGTAGGTGCGACCGTGTGGAAACGTAACGACGCGGATGGTGTTTCTCGTGAAGCGATCGCAGCAGCGATCCGT
CGTGTTATGACCTCTGAAGAAGCTCCGGCGTACCGTGAACGTGCGCGTGAACTGGCTGAGAAAGCGAAGAAAGCA
ATCGAAGAAGGTGGTTCCAGCGATAAAGGTCTGCAAGAACTGCTGGAAGATATCCGTGCGTTCCGTCTGGCTAAC
CACTAA

> sdT$_{exp}$2

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEWVFSKEIERWKKDGIDIEIRLIKFPAKEFGLP
EECERLDDVPSDEYLPKFFAALEAMQAPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIVFHGTSPFALVVE
RSVRKHKPWKDVKSDTEEFVVPDLPHEIKLTRTQVSPFLRSGEKTPMSELIERMRASDEKSWGVVFNSFYELEKD
YFDHYTNTLKRRAWAIGPLSMCNKTLKDKAERGKKSSIDPEECLAWLDSKAPGSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESEDWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATVWKRENADGVSREAIAAAIRRVMTSEEAPAYRERAKALAEAAKKA
IEEGGSSDKGLEDLLADIREFRLANH
ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCTCACGGTCACATGATCCCAACTCTGGACATGGCGAAA
CTGTTCGCTTCTCGTGGTGTTAAAGCGACTATCATCACCACTCCGGGTAACGAATGGGTGTTCTCCAAAGAGATC
GAACGTTGGAAGAAAGACGGTATCGACATCGAAATCCGTCTGATCAAATTCCCGGCTAAAGAATTTGGTCTGCCG
GAAGAATGCGAACGTCTGGATGACGTTCCGTCTGACGAATACCTGCCGAAATTCTTCGCTGCACTGGAAGCGATG
CAGGCTCCATTGGAAGAACTGATCGCAGAAGAGAAACCAGACTGCCTGATCTCTGACATGTTTCTGCCGTGGACC
GTTGACGTTGCAGCGAAATTCAACATTCCGCGTATCGTGTTCCACGGCACCAGTCCGTTCGCTTTGGTTGTTGAA
CGTTCTGTTCGTAAACACAAACCGTGGAAAGACGTTAAATCCGACACCGAAGAGTTCGTTGTTCCGGATCTGCCG
CACGAAATCAAACTGACTCGTACTCAGGTTTCTCCGTTTCTGCGTTCTGGTGAGAAGACTCCGATGTCTGAACTT
ATCGAACGTATGCGTGCTTCTGACGAGAAATCTTGGGGTGTTGTCTTCAACTCTTTCTACGAACTGGAGAAAGAC
TACTTCGACCACTACACCAACACCTTGAAACGTCGTGCTTGGGCTATCGGTCCACTGTCTATGTGCAACAAGACT
CTGAAAGACAAAGCGGAACGTGGTAAGAAATCTTCCATCGATCCAGAAGAATGCCTTGCTTGGCTGGACTCTAAA
GCACCGGGTTCTGTGGTTTACGTGTGCTTCGGTTCTGTAGCTCGTCTGACTCCAGAACAGCTGCGTGAACTGGCT
CTGGGTATCGAAGCATCTGGTCAAGACTTCATCTGGGTTGTTCGTCCAGACTACGAATCTGAAGACTGGCTGCCA
GAAGGTTTCGAAGAACGTACCAAAGATAAAGGTCTGATCATCCGTGGTTGGGCACCACAGGTTCTGATCCTGGAC
CATCCGTCCGTTGGTGCTTTCGTTACTCACTGCGGTTGGAACTCTACTCTGGAAGGTGTTGCTGGTGGTGTTCCG
ATGGTGACTTGGCCAGTGTTCGCAGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTGAAGACCGGTGCA
TCTGTTGGTGCGACCGTTTGGAAACGTGAGAACGCTGACGGTGTTAGCCGTGAAGCTATCGCTGCTGCAATCCGT

CGTGTTATGACCTCTGAAGAAGCACCGGCATACCGTGAACGTGCGAAAGCACTGGCTGAAGCAGCAAAGAAAGCG
ATCGAAGAAGGTGGTTCTTCTGACAAAGGTCTGGAAGACTTGCTGGCAGACATCCGTGAGTTCCGTCTTGCGAAC
CACTAA

> sdT$_{exp}$3

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEPVFSKEIERWKKDGIEIEIRLIKFPAKEFGLP
EECERLDQVPSDDQLPNFFAALAAMQKPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIIFHGTSLFALVVE
ESVRKHKPYLSVKSDTEEFVVPDLPHEIKLTRTQVSPFLRSGKETPMTKLIKAMRESDKKSWGVVFNSFYELEKD
YYNHYTEKLKRRAWAIGPLSMCNKTLEDKAERGKKSSIDPEECLAWLDSKAPGSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESEDWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATKWRRYDADGVSREAIAAAIRTVMTSEEAAGYRERAKELAEKAKKA
IEEGGSSDNGLKALLADIAAFRAAGH
ATGAAACAGCTGCACTTCTTCTTCTTTCCAGTAATGGCGCACGGTCACATGATCCCGACCTTGGACATGGCGAAA
CTGTTCGCATCTCGTGGTGTTAAAGCGACCATCATCACCACTCCGGGTAACGAACCGGTGTTCTCCAAAGAAATC
GAACGTTGGAAGAAAGACGGTATCGAAATCGAAATCCGTCTGATCAAATTTCCAGCGAAAGAATTTGGTCTGCCG
GAAGAATGCGAACGTCTGGATCAGGTGCCATCCGACGATCAGCTTCCGAACTTCTTCGCGGCACTGGCTGCGATG
CAGAAACCGCTGGAAGAACTGATCGCTGAAGAGAAACCGGACTGTCTGATCTCTGACATGTTTCTGCCATGGACC
GTTGATGTTGCTGCGAAATTCAACATCCCGCGTATCATCTTCCACGGCACCTCTCTGTTCGCTCTGGTTGTTGAA
GAATCTGTTCGTAAACACAAACCGTACCTGTCTGTTAAGAGCGACACCGAAGAGTTCGTTGTTCCGGATCTGCCA
CACGAAATCAAACTGACTCGTACTCAGGTTAGCCCGTTCTTGCGTTCTGGTAAAGAAACTCCGATGACCAAACTG
ATCAAAGCGATGCGTGAATCCGACAAGAAGTCTTGGGGTGTTGTCTTCAACAGCTTCTACGAACTGGAGAAAGAC
TACTACAACCACTACACCGAGAAACTGAAACGTCGTGCATGGGCTATCGGTCCGCTGTCCATGTGCAACAAGACC
TTGGAAGACAAAGCGGAACGTGGTAAGAAATCCAGCATCGATCCAGAAGAATGCTTGGCTTGGCTCGACTCTAAA
GCGCCGGGTTCTGTGGTTTACGTTTGCTTCGGTTCTGTTGCGCGTCTGACTCCGGAACAGCTGCGTGAACTGGCA
CTGGGTATCGAAGCGTCTGGTCAAGACTTCATCTGGGTTGTACGTCCGGACTACGAATCTGAAGACTGGCTGCCG
GAAGGTTTCGAAGAACGTACCAAAGACAAAGGCCTGATCATCCGTGGTTGGGCTCCGCAGGTACTGATTCTGGAT
CATCCATCTGTGGGTGCGTTCGTTACTCACTGCGGTTGGAACTCCACTCTGGAAGGTGTTGCTGGTGGTGTTCCG
ATGGTGACCTGGCCAGTATTCGCGGAACAGTTCTTCAACGAGAAACTGGTTACTGAAGTTCTGAAGACCGGCGCA
TCTGTAGGTGCGACCAAATGGCGTCGTTACGACGCTGATGGTGTTTCTCGTGAAGCGATCGCTGCTGCAATCCGT
ACTGTTATGACCTCTGAAGAAGCTGCTGGTTACCGTGAACGTGCGAAAGAACTGGCAGAGAAAGCGAAGAAAGCT
ATCGAAGAAGGTGGTAGCTCTGATAACGGTCTGAAAGCACTGCTGGCGGATATCGCGGCTTTCCGTGCGGCTGGT
CACTAA

> sdT$_{exp}$4

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEDVFSKEIERWKERGIDIEIRLIKFPAKEFGLP
EECERLDDIPSDEYLPKFFAALAAMQAPLEALIAEEKPDCLISDMFLPWTTDLAAKFNIPRIVFHGTSLFARVVE
YSVRKHKPFLEVKSDEEEFVVPDLPHKITLTRTQVSPFLRSGEKTPMSELIEAMYESDEKSWGVVFNSFYELEPE
YYKHYTEVLKRRAWAIGPLSMCNKTLKDKAERGKKSSIDPEECLAWLDSKAPDSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYPDESWLPEGFEERTKEKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGARVWRRENADGVSREAIAAAIRTVMTSEEAPAYRARARELAEKAKRA
IEEGGSSDLGLEELLKDIAEYRLANH
ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCACACGGTCACATGATCCCGACTCTGGACATGGCGAAA
CTGTTCGCTTCTCGTGGTGTTAAAGCGACCATCATCACTACTCCGGGTAACGAAGATGTCTTCAGCAAAGAAATC
GAACGTTGGAAAGAACGTGGTATCGACATCGAAATCCGTCTGATCAAATTCCCGGCTAAAGAATTTGGCCTGCCG
GAAGAATGTGAACGTTTGGACGATATTCCATCTGACGAATACCTGCCGAAATTCTTCGCAGCGCTGGCAGCTATG

CAGGCACCACTGGAAGCGCTGATCGCTGAAGAGAAACCGGATTGCCTGATCTCCGATATGTTTCTGCCATGGACC
ACCGATCTGGCTGCGAAATTCAACATTCCACGTATCGTGTTCCACGGTACTTCTCTGTTCGCTCGTGTGGTTGAA
TACTCCGTTCGTAAACACAAACCGTTCTTGGAAGTTAAATCCGACGAAGAAGAGTTCGTGGTGCCGGATCTGCCA
CACAAGATCACTCTGACTCGTACTCAGGTTTCTCCGTTTCTGCGTTCTGGTGAGAAGACTCCGATGTCCGAACTG
ATCGAAGCAATGTACGAATCTGACGAGAAATCTTGGGGCGTAGTCTTCAACAGCTTCTACGAACTGGAACCGGAA
TACTACAAACATTACACCGAAGTTCTGAAACGTCGTGCTTGGGCAATCGGTCCACTGAGCATGTGCAACAAGACT
CTGAAAGATAAAGCGGAACGTGGTAAGAAATCTTCCATCGATCCAGAAGAATGCCTGGCATGGCTGGATTCCAAA
GCACCGGATTCCGTGGTGTACGTATGCTTCGGTTCCGTTGCGCGTCTGACTCCGGAACAGCTGCGCGAACTGGCG
CTGGGTATCGAAGCATCCGGTCAAGACTTCATCTGGGTTGTTCGTCCGGACTATCCGGACGAATCTTGGCTGCCG
GAAGGTTTCGAAGAACGTACCAAAGAGAAAGGTCTGATCATCCGTGGTTGGGCACCGCAGGTTCTGATTCTGGAC
CATCCATCTGTTGGTGCATTCGTTACTCATTGCGGTTGGAACTCTACTCTGGAAGGTGTTGCAGGTGGTGTTCCG
ATGGTGACCTGGCCGGTATTCGCGGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTACTGAAGACCGGTGCT
TCCGTAGGTGCTCGTGTTTGGCGTCGTGAGAACGCTGACGGTGTTTCTCGTGAAGCGATCGCAGCGGCGATCCGT
ACCGTTATGACCTCTGAAGAAGCGCCGGCATATCGTGCTCGTGCACGTGAACTGGCGGAGAAAGCGAAACGTGCA
ATCGAAGAAGGTGGTTCTTCCGATCTGGGCCTGGAAGAACTGCTGAAAGACATCGCAGAATACCGTCTGGCGAAC
CATTAA

> sdT$_{exp}$5

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEDVFSKEIKRWEKRGIKIDIRLIKFPAKEFGLP
EECERLDDVPSDEDLPKFFAALRAMQEPLEKLIAEEKPDCLISDMFLPWTVDVAAKFNIPRIIFHGTSLFARVVE
ISVRKHQPWKTVKSDTEEFVVPDLPHEIKLTKTQVSPFLRSGEETPMSKLIAEMYESDEKSWGVVFNSFYELEPD
YYKHYTEKLKRRAWAIGPLSMCNKTLEDKAERGKKSSIDPEECLAWLDSKEPGSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESEDWLPEGFEERTREKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATVWKREDADGVSREAIAAAIRRVMTSEEAPAYRERAKALAKKAKAA
IEEGGSSDKGLDDLLADIAAFRAAGH

ATGAAACAGCTGCACTTCTTCTTCTTCCCGGTTATGGCTCATGGCCACATGATCCCGACTCTGGACATGGCGAAA
CTGTTCGCTTCTCGTGGTGTTAAAGCGACCATCATCACCACTCCGGGTAACGAAGACGTGTTCTCTAAAGAGATC
AAACGTTGGGAGAAACGTGGTATCAAGATCGACATCCGTCTGATCAAATTTCCAGCGAAAGAATTTGGTCTGCCA
GAAGAATGCGAACGTCTGGACGACGTTCCGTCTGACGAAGACCTGCCAAAGTTCTTCGCGGCTCTGCGTGCAATG
CAAGAACCGCTGGAAAAACTGATCGCGGAAGAGAAACCGGACTGCCTGATCTCTGACATGTTTCTGCCATGGACC
GTTGACGTTGCAGCGAAATTCAACATTCCGCGTATCATCTTCCACGGCACCAGCCTGTTCGCACGTGTTGTTGAA
ATCTCTGTTCGTAAACACCAACCGTGGAAGACTGTTAAATCTGACACCGAAGAATTTGTTGTACCGGACCTGCCA
CACGAAATCAAACTGACCAAGACTCAGGTTTCTCCATTTCTGCGTTCTGGTGAAGAAACTCCAATGAGCAAACTG
ATCGCTGAAATGTACGAATCCGACGAGAAATCTTGGGGTGTTGTATTCAACAGCTTCTACGAACTGGAACCGGAC
TACTACAAACACTACACCGAGAAACTGAAACGTCGTGCTTGGGCTATCGGTCCGCTTTCCATGTGCAACAAAACC
TTGGAAGACAAAGCAGAACGTGGTAAGAAATCTTCCATCGATCCGGAAGAATGCCTGGCATGGCTGGACTCTAAA
GAACCAGGTTCTGTTGTTTACGTTTGCTTCGGTTCTGTTGCTCGTCTGACTCCGGAACAGCTGCGTGAACTGGCT
CTGGGTATCGAAGCGTCTGGTCAAGACTTCATCTGGGTTGTTCGTCCGGATTACGAATCTGAAGACTGGTTGCCG
GAAGGTTTCGAAGAACGTACTCGTGAGAAAGGTCTGATCATCCGTGGTTGGGCACCACAGGTACTGATCTTGGAC
CATCCGTCCGTTGGTGCGTTCGTTACTCACTGCGGCTGGAACTCTACTCTGGAAGGTGTTGCGGGTGGTGTTCCG
ATGGTTACTTGGCCGGTATTCGCGGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTGAAGACCGGTGCA
TCTGTTGGTGCTACCGTTTGGAAACGTGAAGACGCAGATGGTGTTTCTCGTGAAGCGATCGCGGCAGCGATCCGT
CGTGTTATGACCTCCGAAGAAGCGCCAGCTTACCGTGAACGTGCTAAAGCGCTGGCTAAGAAAGCGAAAGCGGCT
ATCGAAGAAGGTGGTTCTTCTGACAAAGGCCTGGACGATCTGCTGGCGGATATCGCAGCTTTCCGTGCAGCGGGT
CATTAA

> **sdT$_{exp}$6**

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEDVFSKEIERWKERGIKIDIRLIKFPAKEHGLP
EECERLDQLPSDEDLPKFFAALAAMQAPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIIFHGTSLFARVVE
ESVIKYKPWKNVKSDTEEFVVPDLPHEIKLTKTQVSPFIRSEEETPMSALIKAMRESDKKSWGVVFNSFYELEKD
YYDHYTKVLKRRAWAIGPLSMCNKTLEDKAERGKKSSIDPEECLAWLDSKAPGSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYPSEDWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLQTGASVGATVWKRRNADGVSREAIAAAIRRVMTSEEAPAYRARAKELAAKAKKA
IEEGGSSDNGLEALLADIAAFRAANH

ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCTCACGGTCACATGATTCCAACTCTGGACATGGCGAAA
CTGTTCGCAAGCCGTGGTGTTAAAGCGACCATCATCACCACTCCGGGTAACGAAGACGTGTTCTCTAAAGAAATC
GAACGTTGGAAAGAACGTGGTATCAAGATCGACATCCGTCTGATCAAATTCCCGGCTAAAGAACACGGTTTGCCG
GAAGAATGCGAACGTCTGGATCAGCTGCCATCTGATGAAGACCTGCCGAAATTCTTCGCGGCTCTGGCGGCGATG
CAGGCGCCGTTGGAAGAACTGATCGCTGAAGAGAAACCGGACTGCCTGATCTCCGACATGTTCTTGCCATGGACC
GTTGACGTTGCGGCTAAATTCAACATCCCGCGTATCATCTTCCATGGCACCTCTCTGTTCGCTCGTGTTGTTGAA
GAATCTGTTATCAAATACAAACCGTGGAAGAACGTTAAATCCGACACCGAAGAATTTGTTGTTCCGGACCTGCCG
CACGAAATCAAACTGACCAAGACTCAGGTTTCTCCGTTCATCCGTTCTGAAGAAGAAACTCCGATGTCTGCGCTG
ATCAAAGCGATGCGTGAATCTGACAAGAAGTCTTGGGGTGTTGTGTTCAACAGCTTCTACGAACTGGAGAAAGAT
TACTACGACCACTACACCAAAGTTCTGAAACGTCGTGCGTGGGCTATCGGTCCGCTGTCCATGTGCAACAAGACT
CTGGAAGACAAAGCGGAACGTGGTAAGAAATCTTCTATCGATCCAGAAGAATGCCTGGCTTGGCTGGACTCTAAA
GCTCCAGGTTCCGTGGTTTACGTTTGCTTCGGTTCTGTTGCGCGTCTGACTCCGGAACAGTTGCGTGAACTGGCT
CTGGGTATCGAGGCTTCTGGCCAAGACTTCATCTGGGTTGTTCGTCCGGACTATCCGTCCGAAGACTGGCTGCCA
GAAGGTTTCGAAGAACGTACCAAAGACAAAGGTCTGATCATCCGTGGTTGGGCACCACAGGTACTGATCTTGGAT
CATCCGTCTGTAGGTGCGTTCGTTACTCACTGCGGTTGGAACTCTACCTTGGAAGGTGTAGCAGGTGGTGTTCCA
ATGGTTACTTGGCCGGTATTCGCTGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTCCAGACCGGTGCA
TCTGTTGGTGCTACCGTTTGGAAACGTCGTAACGCAGACGGTGTTTCTCGTGAAGCAATCGCTGCTGCTATCCGT
CGTGTTATGACCTCCGAAGAAGCGCCGGCTTACCGTGCACGTGCCAAAGAACTGGCTGCTAAAGCGAAGAAAGCG
ATCGAAGAAGGTGGTTCTTCCGACAACGGTCTGGAAGCGTTGCTGGCTGATATCGCAGCATTCCGTGCTGCTAAC
CACTAA

> **sdT$_{exp}$7**

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEEVFSKEIERWKKRGINIEIRLIKFPAKEFGLP
EECERLDDIPSDEYLPKFFKALEAMQEPLEELIAEEKPDCLISDMFLPWTVDLAAKFNIPRIIFHGTSLFARVVE
YSVRKHQPWKTVKSDTEEFVVPDLPHEIKLTRTQVSPFLRSGEKTPMSELIEKMYASDEKSWGVVFNSFYELEKD
YYDYYTNTLKRRAWAIGPLSMCNTTLEDKAERGKKSSIDPEECLAWLDSKAPDSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESESWLPEGFEERTKEKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGSTVWKRFDADGVSREAIAAAIRRVMTGEEAPGFRARARELAEKARKA
IEEGGSSDNGLKELLADIAAYRESNH

ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCGCACGGTCACATGATCCCGACCCTGGACATGGCGAAA
CTGTTCGCTTCTCGTGGTGTTAAAGCGACCATCATCACCACTCCGGGTAACGAAGAAGTGTTCTCCAAAGAAATC
GAACGTTGGAAGAAACGTGGCATCAACATCGAAATCCGTCTGATCAAATTTCCGGCGAAAGAGTTCGGCTTGCCG
GAAGAATGCGAACGTCTGGATGACATTCCGTCTGACGAATACCTGCCGAAATTCTTCAAAGCGCTTGAAGCGATG
CAAGAACCACTGGAAGAACTGATCGCGGAAGAGAAACCGGACTGCCTGATCTCCGATATGTTCTTGCCGTGGACC
GTTGACCTGGCGGCGAAATTCAACATTCCGCGTATCATCTTCCACGGTACTTCTCTGTTCGCACGTGTGGTGGAA
TACTCTGTTCGTAAACACCAGCCATGGAAAACCGTGAAATCCGACACCGAAGAATTTGTTGTTCCGGATCTGCCG

CACGAAATCAAACTGACTCGTACTCAGGTTTCTCCGTTCTTGCGTTCTGGTGAGAAGACTCCGATGTCCGAACTG
ATCGAGAAGATGTACGCGTCCGACGAGAAGAGCTGGGGCGTTGTGTTCAACAGCTTCTACGAACTGGAGAAAGAC
TACTACGACTACTACACCAACACTCTGAAACGTCGTGCATGGGCTATCGGTCCGCTGTCTATGTGCAACACCACT
CTTGAAGATAAAGCTGAACGTGGTAAGAAATCTAGCATCGATCCGGAAGAGTGCCTGGCATGGCTTGACTCTAAA
GCGCCAGATTCCGTTGTTTACGTGTGCTTCGGTTCTGTTGCTCGTCTGACTCCGGAACAGCTGCGCGAACTGGCG
CTGGGTATCGAGGCTTCTGGTCAAGACTTCATCTGGGTTGTTCGTCCGGATTACGAATCCGAATCTTGGCTGCCG
GAAGGTTTCGAAGAACGTACCAAAGAGAAAGGTCTGATCATCCGTGGTTGGGCACCACAGGTTCTGATTCTGGAC
CATCCGTCTGTTGGCGCTTTCGTTACTCACTGCGGTTGGAACTCTACTCTGGAAGGTGTTGCTGGTGGTGTACCG
ATGGTTACCTGGCCAGTGTTCGCTGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTGAAGACCGGTGCG
TCTGTTGGTTCCACCGTTTGGAAACGTTTCGACGCTGACGGTGTGTCTCGTGAAGCTATCGCTGCGGCAATTCGT
CGTGTTATGACTGGCGAAGAAGCACCGGGTTTCCGTGCACGTGCTCGTGAACTGGCGGAGAAAGCTCGTAAAGCG
ATCGAAGAAGGTGGTAGCTCCGACAACGGTCTGAAAGAACTGCTGGCAGACATCGCGGCATACCGTGAATCCAAC
CACTAA

> sdT$_{exp}$8

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEPVFSKEIERWKEEGINIEIRLIKFPAKEHGLP
EECERLDQVPSDEYLPKFFAALESMQAPLEELIAEEKPDCLISDMFLPWTTDIAAKFNIPRIVFHGTSFFALVVE
YSVRKYKPWKKVKSDTEEFVVPDLPHEIRLTRTQVSPFITSEEETPMSKRIKAMYESDEKSWGVVFNSFYELEKD
YFDHYTKTLKRRAWAIGPLSMCNKTLKDKAERGKKSSIDPEECLAWLDSKAPNSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESEDWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATKWKRRDADGVSREAIAAAIRRVMTSEEAPAYRARAKELAEKAKKA
IEEGGSSDNGLEELLADIAAFRAANH

ATGAAACAGCTGCACTTCTTCTTCTTCCCGGTTATGGCTCACGGCCACATGATTCCGACTCTGGACATGGCTAAA
CTGTTCGCGAGCCGTGGCGTAAAGGCGACTATCATCACCACTCCGGGCAACGAACCAGTGTTCTCTAAAGAAATC
GAACGTTGGAAAGAAGAAGGTATCAACATCGAAATCCGTCTGATCAAATTCCCGGCCAAAGAACATGGCCTGCCA
GAAGAATGTGAACGTCTGGATCAGGTTCCGTCTGACGAGTATCTGCCAAAGTTCTTCGCGGCACTGGAAAGCATG
CAGGCTCCACTGGAAGAATTGATCGCGGAAGAGAAACCGGACTGCCTGATCTCCGACATGTTTCTGCCGTGGACC
ACTGACATCGCGGCGAAATTCAACATTCCACGTATCGTCTTCCATGGCACCTCTTTCTTTGCACTGGTTGTGGAA
TACAGCGTGCGTAAATACAAACCGTGGAAGAAAGTTAAATCCGACACCGAAGAATTTGTTGTTCCGGACCTGCCG
CACGAAATCCGCCTGACTCGTACTCAGGTTAGTCCATTCATCACCTCTGAAGAAGAAACTCCAATGTCCAAACGT
ATCAAAGCTATGTACGAATCCGACGAGAAATCTTGGGGTGTTGTGTTCAACAGCTTCTACGAACTGGAGAAAGAC
TACTTCGATCACTACACCAAGACTCTGAAACGTCGTGCTTGGGCTATCGGTCCGCTGTCTATGTGCAACAAGACT
CTGAAAGACAAAGCGGAACGTGGTAAAAAATCCAGCATCGATCCGGAAGAATGCCTGGCATGGCTGGATAGCAAA
GCACCGAACAGCGTTGTTTACGTTTGCTTCGGCTCCGTTGCTCGTCTGACTCCGGAACAGCTGCGTGAACTGGCA
CTGGGTATCGAAGCCTCTGGTCAAGACTTCATCTGGGTTGTTCGTCCAGATTACGAATCTGAAGACTGGCTGCCG
GAAGGTTTCGAAGAACGTACCAAAGACAAAGGTCTGATCATCCGTGGTTGGGCTCCGCAGGTGCTGATTCTGGAC
CATCCGTCTGTTGGTGCGTTCGTGACTCACTGCGGTTGGAACTCTACTCTGGAAGGTGTAGCAGGTGGTGTTCCA
ATGGTTACCTGGCCAGTATTCGCAGAACAGTTCTTCAACGAGAAACTGGTGACCGAAGTTCTGAAGACCGGTGCG
TCTGTAGGTGCGACCAAATGGAAACGTCGTGACGCTGACGGTGTGTCTCGTGAAGCGATCGCTGCTGCTATCCGT
CGTGTAATGACCTCCGAAGAAGCGCCGGCGTACCGTGCTCGTGCGAAAGAACTGGCGGAGAAAGCGAAGAAAGCG
ATCGAAGAAGGTGGTTCTTCCGATAACGGTCTGGAAGAACTGCTGGCAGATATCGCTGCTTTCCGTGCTGCTAAC
CACTAA

> sdT$_{exp}$9

MKQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEWVFSKEIERWKERGIKIEIRLIKFPAKEHGLP
EECERLDDVPSDEDLPKFFAALESMQAPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIIFHGTSFFALVVE
YSVRKHKPWKEVKSDTEEFVVPDLPHEIKLTRTQVSPFFRSEEETPMSKRIKAMYESDEKSWGVVFNSFYELEKD
YYDYYTKKLKRRAWAIGPLSMCNKTLEDKAERGKKSSIDPAECLAWLDSKAPESVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESESWLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATVWKRRNADGVSREAIAAAIRRVMTSEEAPAYRARAKELAAKAKKA
IEEGGSSDKGLEELLADIAEFRAANH
ATGAAACAGCTGCACTTCTTCTTCTTTCCGGTTATGGCTCACGGTCACATGATCCCGACTCTGGATATGGCTAAA
CTGTTCGCTAGCCGTGGTGTTAAAGCGACTATCATCACCACTCCAGGTAACGAATGGGTGTTCTCCAAAGAAATC
GAACGTTGGAAAGAACGTGGTATCAAGATCGAAATCCGTCTGATCAAGTTTCCAGCGAAAGAACACGGTTTGCCG
GAAGAATGCGAACGTCTGGATGACGTACCGTCTGACGAAGACTTGCCGAAATTCTTCGCAGCTCTGGAATCCATG
CAGGCTCCGCTGGAAGAACTGATCGCGGAAGAGAAACCGGATTGCCTGATCTCTGACATGTTTCTGCCGTGGACC
GTTGACGTTGCGGCTAAATTCAACATTCCGCGTATCATCTTCCACGGTACTTCTTTCTTCGCTCTGGTGGTTGAA
TACTCTGTTCGTAAACACAAACCGTGGAAAGAAGTTAAATCCGACACCGAAGAGTTCGTAGTACCGGATCTTCCG
CACGAAATCAAACTGACTCGTACTCAGGTTTCTCCGTTCTTCCGTTCTGAAGAAGAAACTCCGATGTCTAAGCGT
ATCAAAGCGATGTACGAATCCGACGAGAAGAGCTGGGGTGTAGTGTTCAACTCTTTCTACGAACTGGAGAAAGAC
TACTACGACTACTACACCAAGAAACTGAAACGTCGTGCATGGGCTATCGGTCCGCTGTCCATGTGCAACAAGACC
CTGGAAGACAAAGCTGAACGTGGTAAGAAATCTTCTATCGATCCGGCTGAATGCTTGGCGTGGCTGGACTCCAAA
GCGCCAGAATCTGTTGTGTACGTTTGCTTCGGTAGCGTTGCGCGTCTGACTCCGGAACAGCTGCGTGAACTGGCA
CTGGGTATCGAAGCGTCTGGTCAAGACTTCATCTGGGTTGTTCGTCCGGACTACGAATCTGAATCTTGGCTGCCA
GAAGGTTTCGAAGAACGTACCAAAGACAAAGGCCTGATCATCCGTGGTTGGGCACCACAGGTTCTGATCTTGGAT
CACCCGTCTGTAGGCGCATTCGTCACTCATTGCGGTTGGAACTCTACTCTGGAAGGTGTTGCTGGTGGCGTACCA
ATGGTTACCTGGCCGGTATTCGCTGAACAGTTCTTCAACGAGAAACTGGTTACCGAAGTTCTGAAGACCGGTGCA
TCTGTAGGTGCGACTGTTTGGAAACGTCGTAACGCTGATGGTGTTTCTCGTGAAGCAATCGCAGCTGCAATCCGT
CGTGTTATGACCTCTGAAGAAGCACCGGCGTATCGTGCGCGTGCTAAAGAACTGGCGGCGAAAGCTAAGAAAGCG
ATCGAAGAAGGTGGTTCTTCCGACAAAGGTCTGGAAGAGCTGCTGGCAGACATCGCTGAGTTCCGTGCTGCGAAC
CACTAA

> sdT$_{exp}$10

MRQLHFFFFPVMAHGHMIPTLDMAKLFASRGVKATIITTPGNEWVFRKEIERWKERGINIEIRLIKFPAKEHGLP
EECERLDDIPSDEDLPKFFAALESMQAPLEELIAEEKPDCLISDMFLPWTVDVAAKFNIPRIIFHGTSFFARVVE
YSVIKHKPWKKVKSDTEEFVVPDLPHEIKLTRTQVSPFITSGKETPMSKLIKKMRESDEKSWGVVFNSFYELEKD
YFDHYTKTLKRRAWAIGPLSMCNKTLEDKAERGKKSSIDPEECLAWLDSKAPNSVVYVCFGSVARLTPEQLRELA
LGIEASGQDFIWVVRPDYESESYLPEGFEERTKDKGLIIRGWAPQVLILDHPSVGAFVTHCGWNSTLEGVAGGVP
MVTWPVFAEQFFNEKLVTEVLKTGASVGATKWKRNDADGVSREAIAAAIRRVMTSEEAPAYRARAKELAEKAKKA
IEEGGSSDNGLKALLDDIRAFRLANH

ATGCGTCAGCTGCACTTCTTCTTCTTTCCGGTAATGGCTCACGGTCACATGATCCCGACTCTGGACATGGCGAAA
CTGTTCGCTAGCCGTGGCGTTAAAGCGACCATCATCACCACTCCAGGTAACGAATGGGTGTTCCGTAAAGAAATC
GAACGTTGGAAAGAACGTGGTATCAACATCGAAATCCGTCTGATCAAATTTCCGGCTAAAGAACACGGTTTGCCA
GAAGAATGCGAACGTCTGGACGACATTCCGTCTGATGAGGATCTGCCGAAATTCTTCGCTGCTCTGGAATCTATG
CAGGCACCGCTGGAAGAACTGATCGCAGAAGAGAAACCGGACTGCCTGATCTCCGACATGTTCTTGCCATGGACT
GTTGACGTTGCGGCGAAATTCAACATTCCACGTATCATCTTCCATGGCACCTCTTTCTTCGCGCGTGTTGTGGAA
TACAGCGTTATCAAACACAAACCGTGGAAGAAAGTTAAATCCGACACCGAAGAGTTCGTTGTTCCGGATCTTCCG
CACGAAATCAAACTGACTCGTACTCAGGTTTCTCCGTTCATCACCTCTGGCAAAGAAACTCCGATGTCTAAACTG

```
ATCAAGAAGATGCGTGAATCTGACGAGAAGAGCTGGGGTGTTGTGTTCAACTCTTTCTACGAACTGGAGAAAGAC
TACTTCGACCACTACACCAAGACTCTGAAACGTCGTGCTTGGGCGATCGGTCCGCTGTCTATGTGCAACAAGACT
CTGGAAGACAAAGCGGAACGTGGCAAGAAATCTTCTATCGATCCGGAAGAATGCCTGGCTTGGCTGGACTCCAAA
GCTCCGAACTCTGTTGTTTACGTTTGCTTCGGTAGCGTAGCACGTCTGACTCCGGAACAGCTGCGTGAACTGGCT
CTGGGTATCGAGGCTTCTGGTCAAGACTTCATCTGGGTTGTTCGTCCGGACTACGAATCCGAATCTTACTTGCCG
GAAGGTTTCGAAGAACGTACCAAAGATAAAGGCCTGATCATCCGTGGTTGGGCACCACAGGTTCTGATCTTGGAC
CATCCATCTGTTGGTGCGTTCGTTACTCACTGCGGTTGGAACTCTACTCTGGAAGGTGTTGCTGGTGGTGTACCG
ATGGTTACCTGGCCAGTGTTCGCTGAACAGTTCTTCAACGAGAAACTGGTTACTGAAGTTCTGAAGACCGGCGCT
TCCGTTGGTGCTACCAAATGGAAACGTAACGACGCAGACGGTGTAAGCCGTGAAGCAATCGCGGCTGCTATCCGT
CGTGTGATGACCTCTGAAGAAGCGCCGGCTTACCGTGCACGTGCGAAAGAACTGGCTGAGAAAGCGAAGAAAGCA
ATCGAAGAAGGTGGTAGCTCTGACAACGGTCTGAAAGCGCTGCTGGATGACATCCGTGCATTCCGTCTGGCTAAC
CACTAA
```

Table S6: Strains and plasmids used in this study.

Strains	Description*	Source
E. coli BW25113	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Lab storage
E. coli DH5α	*recA1 endA1gyrA96thi-1hsdR17supE44relA1lac*	TransGen
E. coli BL21 (DE3)	F^- *ompT hsdSB* ($r_B^-m_B^-$) *gal dcm* (DE3)	TransGen
BW1	BW25113 Δ*pykA*Δ*pykF* ::10xPtrc *AroG* ::50xPtrc *TyrA*	Lab storage
BL-T0	BL21 (DE3), pET-T0	This study
BL-T1	BL21 (DE3), pET-T1	This study
BL-T2	BL21 (DE3), pET-T2	This study
BL-T3	BL21 (DE3), pET-T3	This study
BL-T4	BL21 (DE3), pET-T4	This study
BL-T5	BL21 (DE3), pET-T5	This study
BL-T6	BL21 (DE3), pET-T6	This study
BL-T7	BL21 (DE3), pET-T7	This study
BL-T8	BL21 (DE3), pET-T8	This study
BL-T9	BL21 (DE3), pET-T9	This study
BL-T10	BL21 (DE3), pET-T10	This study
BL-dT1	BL21 (DE3), pET-dT1	This study
BL-dT2	BL21 (DE3), pET-dT2	This study
BL-T-p1	BL21 (DE3), pET-T-p1	This study
BL-T-p2	BL21 (DE3), pET-T-p2	This study
BL-T-ws1	BL21 (DE3), pET-T-ws1	This study
BL-T-ws2	BL21 (DE3), pET-T-ws2	This study
BL-T-ws3	BL21 (DE3), pET-T-ws3	This study
BL-T-ws4	BL21 (DE3), pET-T-ws4	This study
BL-T-gs1	BL21 (DE3), pET-T-gs1	This study
BL-T-gs2	BL21 (DE3), pET-T-gs2	This study
BL-T-gs3	BL21 (DE3), pET-T-gs3	This study
BL-T-gs4	BL21 (DE3), pET-T-gs4	This study
BL-U0	BL21 (DE3), pET-U0	This study
BL-U1	BL21 (DE3), pET-U1	This study
BL-U2	BL21 (DE3), pET-U2	This study
BL-U3	BL21 (DE3), pET-U3	This study
BL-U4	BL21 (DE3), pET-U4	This study

BL-U5	BL21 (DE3), pET-U5	This study
BL-U6	BL21 (DE3), pET-U6	This study
BL-U7	BL21 (DE3), pET-U7	This study
BL-U8	BL21 (DE3), pET-U8	This study
BL-U9	BL21 (DE3), pET-U9	This study
BL-U10	BL21 (DE3), pET-U10	This study
BL-dU1	BL21 (DE3), pET-dU1	This study
BL-dU2	BL21 (DE3), pET-dU2	This study
BL-U-p1	BL21 (DE3), pET-U-p1	This study
BL-U-p2	BL21 (DE3), pET-U-p2	This study
BL-U-ws1	BL21 (DE3), pET-U-ws1	This study
BL-U-ws2	BL21 (DE3), pET-U-ws2	This study
BL-U-ws3	BL21 (DE3), pET-U-ws3	This study
BL-U-ws4	BL21 (DE3), pET-U-ws4	This study
BL-U-gs1	BL21 (DE3), pET-U-gs1	This study
BL-U-gs2	BL21 (DE3), pET-U-gs2	This study
BL-U-gs3	BL21 (DE3), pET-U-gs3	This study
BL-U-gs4	BL21 (DE3), pET-U-gs4	This study
BL-CT	BL21 (DE3), pET-CT	This study
BL-GT	BL21 (DE3), pET-GT	This study
BL-ST	BL21 (DE3), pET-ST	This study
BL-MT	BL21 (DE3), pET-MT	This study
BW-T0	BW25113, pZE-T0	This study
BW-T1	BW25113, pZE-T1	This study
BW-T2	BW25113, pZE-T2	This study
BW-T3	BW25113, pZE-T3	This study
BW-T4	BW25113, pZE-T4	This study
BW-T5	BW25113, pZE-T5	This study
BW-T6	BW25113, pZE-T6	This study
BW-T7	BW25113, pZE-T7	This study
BW-T8	BW25113, pZE-T8	This study
BW-T9	BW25113, pZE-T9	This study
BW-T10	BW25113, pZE-T10	This study
BW-dT1	BW25113, pZE-dT1	This study
BW-dT2	BW25113, pZE-dT2	This study
BW-T-p1	BW25113, pZE-T-p1	This study

BW-T-p2	BW25113, pZE-T-p2	This study
BW-U0	BW25113, pZE-U0	This study
BW-U4	BW25113, pZE-U4	This study
BW-U6	BW25113, pZE-U6	This study
BW-U8	BW25113, pZE-U8	This study
BW-dU1	BW25113, pZE-dU1	This study
BW-dU2	BW25113, pZE-dU2	This study
BW-U-p1	BW25113, pZE-U-p1	This study
BW-U-p2	BW25113, pZE-U-p2	This study
BW-CT	BW25113, pZE12-CT	This study
BW-GT	BW25113, pZE12-GT	This study
BW-ST	BW25113, pZE12-ST	This study
BW-MT	BW25113, pZE12-MT	This study
BW-18	BW-T0, pCS27-ibPAB	This study
BW-19	BW-T0, pCS27-σ^{32}	This study
BW-20	BW-T0, pCS27-DnakJ	This study
BW-21	BW-T0, pCS27-GroSL	This study
BW-22	BW-T0, pCS27-empty	This study
BW-23	BW-U0, pCS27-ibPAB	This study
BW-24	BW-U0, pCS27-σ^{32}	This study
BW-25	BW-U0, pCS27-DnakJ	This study
BW-26	BW-U0, pCS27-GroSL	This study
BW-26	BW-U0, pCS27-empty	This study
BW	BW1, pCS27-TFH4	This study
BW-TOGT	BW, pZE-D5CT	This study
BW-sdT5	BW, pZE-D5CsdT5	This study
BW-sdT6	BW, pZE-D5CsdT6	This study
BW-UGT84A56	BW, pZE-D5CU	This study
BW-sdU4	BW, pZE-D5CsdU4	This study
BW-sdU6	BW, pZE-D5CsdU6	This study

Plasmids	**Description***	**Source**
pZE12	Plpp1.2, colE ori, luc, AmpR	Lab storage
pCS27	$P_L$lacO1, p15A ori, KanR	Lab storage
pETDuet-1	PT7, pBR322 ori, AmpR	Lab storage
pZE-xc	pZE containing Plpp1.0-AtCOSY	Lab storage
pZE-T0	pZE containing Plpp1.2-TOGT	This study

pZE-T1	pZE containing Plpp1.2-sdT1	This study
pZE-T2	pZE containing Plpp1.2-sdT2	This study
pZE-T3	pZE containing Plpp1.2-sdT3	This study
pZE-T4	pZE containing Plpp1.2-sdT4	This study
pZE-T5	pZE containing Plpp1.2-sdT5	This study
pZE-T6	pZE containing Plpp1.2-sdT6	This study
pZE-T7	pZE containing Plpp1.2-sdT7	This study
pZE-T8	pZE containing Plpp1.2-sdT8	This study
pZE-T9	pZE containing Plpp1.2-sdT9	This study
pZE-T10	pZE containing Plpp1.2-sdT10	This study
pZE-dT1	pZE containing Plpp1.2-dT1	This study
pZE-dT2	pZE containing Plpp1.2-dT2	This study
pZE-T-p1	pZE containing Plpp1.2-T-p1	This study
pZE-T-p2	pZE containing Plpp1.2-T-p2	This study
pZE-U0	pZE containing Plpp1.2-UGT84A56	This study
pZE-U4	pZE containing Plpp1.2-sdU4	This study
pZE-U6	pZE containing Plpp1.2-sdU6	This study
pZE-U8	pZE containing Plpp1.2-sdU8	This study
pZE-dU1	pZE containing Plpp1.2-dU1	This study
pZE-dU2	pZE containing Plpp1.2-dU2	This study
pZE-U-p1	pZE containing Plpp1.2-U-p1	This study
pZE-U-p2	pZE containing Plpp1.2-U-p2	This study
pZE-D5CT	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-TOGT	This study
pZE-D5CT1	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-sdT5	This study
pZE-D5CT2	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-sdT6	This study
pZE-D5CU	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-UGT84A56	This study
pZE-D5CU1	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-sdU4	This study
pZE-D5CU2	pZE containing Plpp1.2-D5-AtCOSY-Plpp1.2-sdU6	This study
pZE-CT	pZE containing Plpp1.2-CysQ-TOGT	This study
pZE-GT	pZE containing Plpp1.2-GST-TOGT	This study
pZE-ST	pZE containing Plpp1.2-SUMO-TOGT	This study
pZE-MT	pZE containing Plpp1.2-MBP-TOGT	This study
pCS-TFH4	pCS27 containing $P_L$lacO1-*Rg*TAL-Fre*Kp*HpaBC-$P_L$lacO1-*At*4CL	Lab storage
pCS-ibPAB	pCS27 containing ibPA-ibPB	This study
pCS-σ^{32}	pCS27 containing σ^{32}	This study
pCS-DnakJ	pCS27 containing Dnak-DnaJ	This study

pCS-GroSL	pETDuet-1, GroS-GroL	This study
pET-D5	pETDuet-1, *At*F6'H* (D102E/ E163K/E190K)	Lab storage
pET-T0	pETDuet-1, TOGT	This study
pET-T1	pETDuet-1, sdT1	This study
pET-T2	pETDuet-1, sdT2	This study
pET-T3	pETDuet-1, sdT3	This study
pET-T4	pETDuet-1, sdT4	This study
pET-T5	pETDuet-1, sdT5	This study
pET-T6	pETDuet-1, sdT6	This study
pET-T7	pETDuet-1, sdT7	This study
pET-T8	pETDuet-1, sdT8	This study
pET-T9	pETDuet-1, sdT9	This study
pET-T10	pETDuet-1, sdT10	This study
pET-dT1	pETDuet-1, dT1	This study
pET-dT2	pETDuet-1, dT2	This study
pET-T-p1	pETDuet-1, T-p1	This study
pET-T-p2	pETDuet-1, T-p2	This study
pET-T-ws1	pETDuet-1, T-ws1	This study
pET-T-ws2	pETDuet-1, T-ws2	This study
pET-T-ws3	pETDuet-1, T-ws3	This study
pET-T-ws4	pETDuet-1, T-ws4	This study
pET-T-gs1	pETDuet-1, T-gs1	This study
pET-T-gs2	pETDuet-1, T-gs2	This study
pET-T-gs3	pETDuet-1, T-gs3	This study
pET-T-gs4	pETDuet-1, T-gs4	This study
pET-U0	pETDuet-1, UGT84A56	This study
pET-U1	pETDuet-1, sdU1	This study
pET-U2	pETDuet-1, sdU2	This study
pET-U3	pETDuet-1, sdU3	This study
pET-U4	pETDuet-1, sdU4	This study
pET-U5	pETDuet-1, sdU5	This study
pET-U6	pETDuet-1, sdU6	This study
pET-U7	pETDuet-1, sdU7	This study
pET-U8	pETDuet-1, sdU8	This study
pET-U9	pETDuet-1, sdU9	This study
pET-U10	pETDuet-1, sdU10	This study

pET-dU1	pETDuet-1, dU1	This study
pET-dU2	pETDuet-1, dU2	This study
pET-U-p1	pETDuet-1, U-p1	This study
pET-U-p2	pETDuet-1, U-p2	This study
pET-U-ws1	pETDuet-1, U-ws1	This study
pET-U-ws2	pETDuet-1, U-ws2	This study
pET-U-ws3	pETDuet-1, U-ws3	This study
pET-U-ws4	pETDuet-1, U-ws4	This study
pET-U-gs1	pETDuet-1, U-gs1	This study
pET-U-gs2	pETDuet-1, U-gs2	This study
pET-U-gs3	pETDuet-1, U-gs3	This study
pET-U-gs4	pETDuet-1, U-gs4	This study
pET-CT	pETDuet-1, CysQ-TOGT	This study
pET-GT	pETDuet-1, GST-TOGT	This study
pET-ST	pETDuet-1, SUMO-TOGT	This study
pET-MT	pETDuet-1, MBP-TOGT	This study

801

802 **Table S7: Primers used in this study.**

Primer	Sequence (5’ to 3’)
TOGT-pZEHR-R	ctgcatgcgctagcaagcttttaatgaccggtgctgctataggtg
SUMO-togt-pZEHR-F	agagaacagattggtggtggtcagctgcatttttttcttttcccg
SUMO-togt-pZEHR-R	gaaaaaatgcagctgaccaccaccaatctgttctctgtgagc
MBP-togt-pZEHR-F	taaagaggagaaaggtaccatgaaaatcgaagaaggtaaactggtaatctggat
GST-togt-pZEHR-F	tcattaaagaggagaaaggtaccatgaaattgttctacaaaccgggtgcc
CysQ-togt-pZEHR-F	taaagaggagaaaggtaccatgttagatcaagtatgccagcttgcac
GST-TOGT-F	gtcagcggaaggcttaaagggtcagctgcatttttttcttttcccgg
TOGT-R	ttaccagactcgagggtaccttaatgaccggtgctgctataggtgct
BamHI-GST-F	gggaaaggatccgaaattgttctacaaaccggg
GST-TOGT-R	gcagctgacccttaagccttccgctgacagcg
pET-TOGT-HR-F	ctatagcagcaccggtcattaaggtaccctcgagtctggtaaagaaaccg
SUMO-TOGT-F	cagagaacagattggtggtggtcagctgcatttttttcttttcccgg
SUMO-plasmid-R	aagaaaaaatgcagctgaccaccaccaatctgttctctgtgagcctc
MBP-TOGT-F	ccctgaaagacgcgcagactggtcagctgcatttttttcttttcccgg
MBP-plasmid-R	gaaaaagaaaaaatgcagctgaccagtctgcgcgtctttcagggcttc
CysQ-F	gggaaaggatccgttagatcaagtatgccagcttgcacgg
CysQ-R	aagaaaaaatgcagctgaccgtaaatagacactctgaaccccggattc
CysQ-TOGT-F	ccggggttcagagtgtctatttacggtcagctgcatttttttcttttcccgg
BamHI-MBP-F	gggaaaggtaccgaaaatcgaagaaggtaaactggtaa
MBP-UGT84A56-R	ccaggctctggctacccatagtctgcgcgtctttcagg
UGT84A56-MBP-F	cgcgcagactatgggtagccagagcctggtt
UGT84A56-KpnI-R	gggaaaggtaccttaacaggtaacttccacgctacgg

BamHI-SUMO-F	gggaaaggatccgtcggactcagaagtcaatcaagaag
UGT84A56-SUMO-F	agattggtggtatgggtagccagagcctggtt
SUMO-UGT84A56-R	Aaccaggctctggctacccataccaccaatctgttctctgtgagc
BamHI-CysQ-F	gggaaaggatccgttagatcaagtatgccagcttgcacgg
UGT84A56-CysQ-F	agagtgtctatttacatgggtagccagagcctggttc
CYSQ-UGT84A56-R	Gaaccaggctctggctacccatgtaaatagacactctgaaccccggattc
UGT84A56-GST-F	agcggaaggcttaaagatgggtagccagagcctggtt
GST-UGT84A56-R	aaccaggctctggctacccatctttaagccttccgctgacagc
BamHI-GST-F	gggaaaggatccgaaattgttctacaaaccggg
KpnI-sdU4-F	gggaaaggtaccatggcaccgccaccac
XbaI-sdU4-R	gggaaatctagattatgccgccagttccgcaga
KpnI-sdU6-F	gggaaaggtaccatgccaccggaaccactg
XbaI-sdU6-R	gggaaatctagattatgccgccagttctttggatttcg
KpnI-sdU8-F	gggaaaggtaccatgccggcaccgccgctggttcacg
XbaI-sdU8-R	gggaaatctagattatgccgccagttcagcagaac
KpnI-UGT84A56-F1	gggaaaggtaccatgggtagccagagcctgg
XbaI-UGT84A56-R	gggaaatctagattaacaggtaacttccacgctacgg
denovo-tongyong-F	gggaaaactagtatcaaaaaaatattctcaacataaaaaactttgtgtaatacttgtaacg
denovo-tongyong-R	gggaaagagctcacaacagataaaacgaaaggcccagtc
KpnI-sdT1-4/6/7/10-F	gggaaaggtaccatgaaacagctgcacttcttcttctttc
XbaI-sdT1-R	gggaaatctagatttagtggttcgcagcacggaac
KpnI-sdT2-F	gggaaaggtaccatgaaacagctgcacttcttcttctttc
XbaI-sdT2-R	gggaaatctagatttagtggttagccagacggaacg
XbaI-sdT3-R	gggaaatctagatttaattgttcgccagacggtattct

XbaI-sdT4-R	gggaaatctagatttagtgaccagccgcacg
KpnI-sdT5/8-F	gggaaaggtaccatgaaacagctgcacttcttcttcttc
XbaI-sdT5-R	gggaaatctagatttaattacccgctgcacgg
XbaI-sdT6-R	gggaaatctagatttagtggttggattcacggtatgcc
XbaI-sdT7-R	gggaaatctagatttagtggttagcagcacggaatg
XbaI-sdT8-R	gggaaatctagatttagtggttagcagcacggaaag
KpnI-sdT9-F	gggaaaggtaccatgcgtcagctgcacttcttc
XbaI-sdT9-R	gggaaatctagatttagtggttagccagacggaatgc
XbaI-sdT10-R	gggaaatctagatttagtggttcgcaagacggaact
KpnI-TOGT-F	gggaaaggtaccatgggtcagctgcatttttttctttttcc
XbaI-TOGT-R	gggaaatctagattaatgaccggtgctgctataggtg

803

Table S8. Amino acid sequences and nucleotide sequences (codon optimized for *E. coli* expression) of the native UGT84A56 and designer enzymes sdU$_{exp}$1-sdU$_{exp}$10. Amino acid sequences in black, nucleotide sequences in blue.

Native UGT84A56

MGSQSLVHVFLVSFPGQGHVNPLLRLGIRLASKGLLVTFSTPESIGKEMRKASNITDEPVPIGDGYIRFEFFEDG
WDENEPRLLDLDQYLPQLELVGKEVIPRMIRKHAEEGRPVSCLINNPFIPWVSDVAESLGLPSAMLWVQSCACFT
AYYYYYHGMVPFPSEKDPEIDVQLPCMPLLKYDEVPSFLYPTTPYPCLRRAILGQYNKLDKPFCILVESFQELEH
DIIEYMSKISPIKTVGPLFMNPKAANSAVRGDFMKADNCIEWLDSKPPSSVVYISFGSVVYLKQEQVDEIAYGLL
NSGVSFLWVMKPLPKDAGLELVLPEGFLEKVGDNGKVVQWSPQEKVLAHPSVACFVTHCGWNSTMESLSSGIPVI
AFPQWGDQVTDAVYLVDVFKTGIRMCRGEAEDRIIPREEVEKCLLEATTGPKAENMKQNALKWKAAAEEAVAEGG
SSDKNIQDFVDEVKRRSVEVTC

ATGGATGTAGCCAGAGCCTGGTTCATGTTTTTCTGGTTAGCTTTCCGGGTCAGGGTCATGTTAATCCGCTGCTGC
GTCTGGGTATTCGTCTGGCAAGTAAAGGTCTGCTGGTTACCTTTAGCACCCCGGAAAGCATTGGTAAAGAAATGC
GTAAAGCAAGCAATATCACCGACGAACCGGTTCCGATTGGTGATGGTTATATTCGTTTTGAGTTTTTCGAGGACG
GCTGGGACGAAAACGAACCGCGTCTGCTGGATCTGGATCAGTATCTGCCGCAGCTGGAACTGGTTGGTAAAGAAG
TTATTCCGCGTATGATTCGCAAACACGCAGAAGAAGGTCGTCCGGTTAGCTGTCTGATTAATAATCCGTTTATTC
CGTGGGTGAGCGATGTTGCAGAAAGCCTGGGTCTGCCGAGCGCAATGTTATGGGTTCAAAGCTGTGCATGTTTTA
CCGCATATTATTACTACTACCACGGCATGGTTCCGTTTCCGAGCGAAAAAGATCCGGAAATTGATGTTCAGCTGC
CGTGTATGCCGCTGCTGAAATATGATGAAGTTCCGAGCTTTCTGTATCCGACCACCCCGTATCCGTGTCTGCGTC
GTGCAATTCTGGGTCAGTATAATAAACTGGATAAGCCGTTTTCGATCCTGGTGGAGAGCTTTCAGGAACTGGAAC
ATGATATTATTGAGTACATGAGCAAGATCAGCCCGATCAAGACCGTGGGTCCGCTGTTTATGAATCCGAAAGCAG
CAAATAGCGCAGTTCGTGGTGATTTTATGAAAGCCGATAATTGCATCGAGTGGCTGGATAGCAAACCGCCGAGCA
GCGTTGTTTATATTAGCTTTGGTAGCGTTGTGTACCTGAAGCAGGAACAGGTTGATGAAATTGCCTATGGCCTGC
TGAATAGCGGTGTTAGCTTTCTGTGGGTTATGAAACCGCTGCCGAAAGATGCAGGTCTGGAACTGGTTCTGCCGG
AAGGTTTTCTGGAAAAAGTTGGTGATAATGGCAAGGTGGTTCAGTGGAGCCCGCAGGAAAAAGTTCTGGCACATC
CGAGCGTTGCATGTTTTGTTACCCATTGTGGTTGGAATAGCACCATGGAAAGCCTGAGCAGCGGTATTCCGGTTA
TTGCATTTCCGCAGTGGGGTGATCAGGTTACCGATGCAGTTTATCTGGTTGATGTTTTTAAGACCGGCATTCGCA
TGTGTCGTGGTGAAGCAGAAGATCGTATTATTCCGCGTGAAGAAGTTGAAAAATGCCTGCTGGAAGCAACCACCG
GTCCGAAAGCAGAAAATATGAAACAGAATGCCCTGAAGTGGAAGGCAGCAGCAGAAGAAGCAGTTGCAGAAGGTG
GTAGCAGCGATAAAAATATTTCAGGATTTTGTGGACGAGGTGAAGCGCCGTAGCGTGGAAGTTACCTGTTAA

Amino acid sequences and nucleotide sequences of designer enzymes

> sdU$_{exp}$1

MPAPPLVHVLLVSFPGQGHVNPLLRLGLHLAARGLLVTFTTPAWVGEKMRKASGVTTEPYPVGAGEIRFEFFEDG
WAPDEPRLRDLDVYLPQLERVGRETLPALIRAHAEAGRPVSCLINNPFIPWVSDVAESLGLPSAMLWVQSPACLL
AYYYYYHNLVPFPSEENPDIDVQLPKLPLLKWDEVPSFLHPTSPYPALRRAILGQYARLDKPFCILVESFYELEK
DVIEYLSKILPIKCVGPLFLNPRAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGKDFLWVMKPLPKDAGEELTLPEGFLEKVGDRGLVVQWSPQEKVLQHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAYLLVDVFKVGIRMCRGEAENRIIPREEVARCIKEATEGPLAEEMKANAAKWKEKALAAVAEGG
SSDKNIQAFVDEVRARSEELAA

ATGCCGGCACCACCACTGGTTCACGTACTGCTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTCTGCATCTGGCAGCACGTGGTCTGCTGGTTACCTTCACCACTCCGGCTTGGGTTGGTGAGAAGATGCGT
AAAGCGTCTGGTGTTACCACCGAACCGTACCCGGTAGGTGCGGGTGAAATCCGTTTCGAATTTTTCGAAGATGGT
TGGGCACCGGACGAACCACGTCTGCGTGATCTGGACGTGTACCTGCCTCAGCTGGAACGTGTTGGTCGTGAAACC
TTGCCAGCTCTGATCCGTGCACATGCTGAAGCTGGTCGTCCGGTTAGCTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCGGAATCTCTCGGTCTGCCGTCTGCGATGCTGTGGGTTCAGTCTCCGGCATGCTTGCTG
GCGTACTACTATTACTACCACAACCTGGTTCCGTTTCCGTCTGAAGAGAACCCGGACATCGACGTTCAGCTGCCT
AAACTGCCGCTGCTGAAATGGGACGAAGTTCCGTCTTTTCTGCATCCGACCAGTCCGTATCCGGCTCTGCGTCGT
GCTATTCTGGGTCAGTACGCACGTCTGGACAAACCGTTCTGCATTCTGGTTGAGTCTTTCTACGAACTGGAGAAA
GACGTGATCGAATACCTGTCTAAGATTCTGCCGATCAAATGCGTTGGTCCGCTGTTTCTGAATCCGCGTGCAGCG
GACTCTCCAGTACGTGGTGACTTCATGCCGGCAGACGAATGCCTGCCATGGCTGGACACCAAACCGCCGAACTCT
GTGGTTTACATCAGCTTCGGTTCTGTTGTTTACCTGCCACAAGAACAGGTTGACGAAATCGCGGAAGGTCTGCTG
CTGTCTGGTAAAGACTTTCTGTGGGTTATGAAACCGCTGCCGAAAGATGCAGGTGAAGAACTGACTCTGCCAGAA
GGTTTCTTGGAGAAAGTTGGTGATCGTGGTCTGGTTGTTCAGTGGAGTCCGCAAGAGAAAGTTCTCCAGCATCCG
GCAGTTGCAGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAATCTCTGGCATCTGGTATTCCGGTTATC
TGCTTTCCGCAGTGGGGTGATCAGGTTACCGACGCTTACCTGCTGGTTGACGTGTTCAAAGTTGGTATCCGTATG
TGCCGTGGTGAAGCTGAGAACCGTATCATTCCGCGTGAAGAAGTTGCGCGTTGCATCAAAGAAGCGACTGAAGGT
CCGCTGGCTGAAGAAATGAAAGCGAACGCGGCGAAATGGAAAGAGAAAGCTCTGGCAGCTGTTGCGGAAGGTGGT
AGCTCTGACAAGAACATCCAGGCGTTCGTTGACGAAGTTCGTGCTCGTTCTGAAGAACTGGCGGCTTAA

> sdU$_{exp}$2

MAAEPLVHVLLVSFPGQGHVNPLLRLGLHLASLGLLVTFTTPSWYGEKMRKASGVTTEPYPVGEGEIRFEFFEDG
WAPDEPRRRDLDVYLPQLEKVGSVTIPALVRRHAEEGRPVSCLINNPFIPWVSDVAASLGLPSAMLWVQSPACLL
QYYYYHHKLVPFPSEANPKIDVQLPKLPLLKWDEVPSFLHPTSPYPALRRAILGQYDRLDKPFCILVESFYELEK
DVIEYLSKILPIKCVGPLFLNPRAADSPVRGDFMPADECLAWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGEDFLWVMKPLPKDAGKELTLPEGFLEKVGDRGLVVQWSPQEKVLQHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAKLLVDVFKVGIRMCRGEAENRVIPREEVARCIREATHGPLAAERKANAAKWAAKAKAAVAEGG
SSDKNLQEFVDEVRARSKELAA

ATGGCGGCTGAACCGCTGGTTCACGTTCTTCTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTCTGCACCTGGCGTCTCTGGGTCTGCTGGTTACCTTCACCACTCCGTCTTGGTACGGTGAGAAGATGCGT
AAGGCTTCTGGTGTTACCACCGAACCGTATCCGGTTGGTGAGGGTGAAATCCGTTTCGAGTTCTTCGAAGACGGT
TGGGCTCCGGATGAACCGCGTCGTCGTGACCTGGACGTTTACCTGCCACAGCTGGAAAAGGTTGGTTCTGTTACC
ATTCCGGCGCTGGTTCGTCGTCATGCGGAAGAAGGTCGTCCGGTTTCTTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCAGCGAGCCTGGGTCTGCCGTCTGCGATGCTGTGGGTTCAGTCTCCGGCTTGCCTGCTT
CAGTACTACTACTACCACCACAAACTGGTTCCGTTTCCGTCTGAAGCGAATCCGAAGATCGACGTTCAGCTGCCG
AAACTGCCGCTGCTGAAATGGGACGAAGTTCCGAGCTTTCTGCACCCAACCTCTCCGTACCCGGCTCTGCGTCGT
GCGATTCTGGGTCAGTACGACCGTCTGGACAAACCGTTCTGCATCTTGGTGGAATCTTTCTACGAACTGGAGAAA
GACGTTATCGAATACCTGTCTAAGATTCTGCCGATCAAATGCGTTGGTCCGCTGTTTCTGAATCCGCGTGCAGCT
GACTCTCCGGTTCGTGGTGACTTCATGCCAGCAGACGAATGCCTGGCTTGGCTGGACACCAAACCGCCGAACTCT
GTTGTTTACATCTCTTTCGGTTCTGTTGTGTACCTGCCGCAAGAACAGGTTGACGAAATCGCGGAAGGTCTGCTG
CTGTCTGGTGAAGACTTTCTGTGGGTTATGAAACCACTGCCGAAAGACGCGGGTAAAGAACTGACCTTGCCGGAA
GGTTTCTTGGAGAAAGTTGGTGACCGTGGTCTGGTTGTTCAGTGGAGTCCGCAAGAGAAAGTTCTCCAGCATCCG
GCGGTAGCTGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAATCTCTGGCGTCTGGTATTCCGGTTATC
TGCTTTCCGCAGTGGGGTGACCAGGTTACCGACGCGAAACTGCTGGTTGACGTGTTCAAAGTTGGTATCCGTATG

TGCCGTGGTGAAGCTGAGAACCGTGTTATTCCGCGTGAAGAAGTTGCTCGTTGCATCCGTGAAGCGACTCACGGT
CCACTGGCTGCGGAACGTAAAGCGAACGCTGCGAAATGGGCGGCGAAAGCGAAAGCAGCAGTTGCGGAAGGTGGT
TCTTCTGACAAGAACCTGCAAGAATTTGTTGACGAAGTTCGTGCACGTTCCAAAGAACTGGCGGCATAA

> sdU$_{exp}$3

MAPAPLVHVLLVSFPGQGHVNPLLRLGLHLASLGLLVTFTTPSWVGDKIRKASGVTSEPYPVGKGELRFEFFEDG
WAPDEPRRRDLDVYLPQLEAVGSRELPRLIRRQAEAGRPVSCLINNPFIPWVSDVAESLGLPSAMLWVQSPACLL
AYYYYYHNLVPFPSEENPDIDVQLPKLPLLKPDEVPSFLHPTSPYPALRRAILGQYARLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFINPKAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGKDFLWVMKPRPKDAGEELKLPEGFLEKVGDRGLVVQWSPQEKVLRHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGEAENRIIPREEVARCIKEATEGPLAEEMKANAAKWKELAEKAVAEGG
SSDKNLQAFVDEVRARSAELAA

ATGGCTCCAGCACCGCTGGTTCACGTTCTGCTGGTTTCTTTCCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTTTGCACCTGGCTTCTCTGGGTCTGCTGGTTACCTTCACCACTCCGTCTTGGGTTGGTGACAAGATCCGT
AAAGCATCTGGTGTTACCTCTGAACCGTACCCAGTTGGTAAAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT
TGGGCACCGGATGAACCACGTCGTCGTGATCTGGACGTTTACCTGCCGCAGCTTGAAGCGGTTGGTAGCCGTGAA
CTGCCACGTCTGATCCGTCGTCAAGCAGAAGCAGGTCGTCCGGTTTCTTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCTGAAAGCCTGGGTCTGCCATCTGCGATGCTGTGGGTTCAGAGTCCGGCTTGCTTGCTG
GCGTACTACTATTACTACCACAACCTGGTTCCGTTTCCGTCTGAAGAGAATCCGGACATCGACGTTCAGCTGCCG
AAATTGCCGCTGCTGAAACCAGATGAAGTTCCGTCTTTTCTGCATCCGACTTCTCCGTATCCGGCACTTCGTCGT
GCAATCTTGGGTCAGTACGCACGTCTGGACAAACCGTTCTGCATTCTGGTTGAGTCTTTCTACGAACTGGAGAAA
GAAGTGATCGAATACCTGTCCAAGATTCTGCCGATCAAATGCGTTGGTCCGCTGTTCATCAATCCGAAAGCTGCG
GATTCTCCGGTACGTGGTGACTTCATGCCAGCTGACGAATGCCTGCCATGGCTGGACACCAAACCGCCGAACTCT
GTGGTTTACATCAGCTTCGGTTCTGTTGTGTACCTTCCGCAGGAACAGGTTGACGAAATCGCGGAAGGTCTGCTG
CTGTCTGGTAAAGACTTTCTGTGGGTTATGAAACCGCGTCCGAAAGATGCTGGTGAAGAACTGAAACTGCCGGAA
GGTTTCTTGGAGAAAGTTGGTGATCGTGGTCTGGTTGTTCAGTGGTCCCCACAAGAGAAAGTTCTGCGTCATCCA
GCAGTTGCAGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAATCTCTGGCATCTGGTATTCCGGTTATC
TGCTTTCCGCAATGGGGTGATCAGGTTACCGATGCACACCTGCTGGTTGACGTGTTCAAGGTTGGTATCCGTATG
TGTCGTGGTGAAGCAGAGAACCGTATCATCCCACGTGAAGAAGTTGCTCGTTGCATCAAAGAAGCGACTGAAGGT
CCACTGGCAGAAGAAATGAAAGCTAACGCTGCGAAATGGAAAGAACTGGCGGAGAAAGCTGTTGCAGAAGGTGGT
TCTTCTGACAAGAACCTTCAGGCTTTCGTTGACGAAGTTCGTGCACGTTCCGCAGAACTGGCTGCTTAA

> sdU$_{exp}$4

MAPPPLVHVFLVSFPGQGHVNPLLRLGLHLASRGLLVTFSTPSWVGDKMRKASGVTTEPYPVGAGELRFEFFEDG
WAPDEPRRSDLDVYLPQLEEVGKKVLPEIIRRHAEEGRPVSCLINNPFIPWVSDVAEELGLPSAMLWVQSPACLL
AYYYYYHKLVPFPSEENPDIDVQLPKLPLLKPDEVPSFLHPTSPYPALRRAILGQYSRLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFLNPRAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGEDFLWVMKPLPKDAGKELTLPEGFLEKVGDRGLVVQWSPQEKVLAHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGERENRIIPREEVARCILEATHGPKAEELKANAAKWKEAAAKAVAEGG
SSDKNLQEFVDFVRAKSAELAA

ATGGCACCGCCACCACTGGTTCACGTGTTCTTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTCTGCACCTGGCATCTCGTGGTCTTCTGGTTACCTTCTCTACTCCAAGCTGGGTTGGTGACAAGATGCGT
AAAGCGTCTGGTGTTACCACCGAACCGTATCCAGTTGGTGCAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT

TGGGCTCCGGACGAACCGCGTCGTTCTGACCTGGACGTGTACCTGCCACAGCTGGAAGAAGTGGGTAAGAAAGTG
CTGCCGGAAATCATCCGTCGTCACGCAGAAGAAGGTCGTCCGGTTTCTTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCAGAAGAACTGGGTCTGCCGTCTGCGATGCTGTGGGTTCAGAGTCCGGCATGCCTGCTG
GCGTACTACTACTATTACCACAAACTGGTTCCGTTTCCATCTGAAGAGAACCCGGACATCGACGTTCAGCTGCCG
AAACTGCCGCTGCTGAAACCAGACGAAGTTCCGTCTTTTCTGCATCCGACCTCTCCGTATCCGGCTCTGCGTCGT
GCGATCTTGGGTCAGTACTCTCGTCTGGACAAACCGTTCTGCATTCTGGTTGAATCTTTCTACGAGCTGGAGAAA
GAAGTGATCGAATACCTGTCTAAGATTCTGCCGATCAAATGCGTAGGTCCGCTGTTTCTGAATCCGCGTGCTGCT
GACTCTCCAGTTCGTGGTGACTTCATGCCGGCGGATGAATGCCTGCCGTGGCTGGACACCAAACCGCCGAACTCT
GTGGTTTACATCTCTTTCGGTTCTGTTGTTTACCTGCCGCAAGAACAGGTTGACGAAATCGCAGAAGGTCTGCTG
CTGTCTGGTGAAGACTTTCTGTGGGTTATGAAACCACTGCCGAAAGACGCGGGTAAAGAACTGACTCTGCCGGAA
GGCTTTCTGGAGAAAGTTGGTGATCGTGGTCTGGTTGTTCAGTGGTCCCCGCAAGAGAAAGTTCTGGCTCATCCA
GCTGTTGCTGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAAAGCCTGGCGTCTGGTATTCCGGTTATC
TGCTTCCCACAGTGGGGTGATCAGGTTACCGACGCGCATCTGCTGGTTGACGTGTTCAAGGTTGGTATCCGTATG
TGCCGTGGTGAACGTGAGAACCGCATCATTCCGCGTGAAGAAGTTGCGCGTTGTATTCTGGAAGCGACTCACGGT
CCGAAAGCGGAAGAACTGAAAGCGAACGCGGCGAAATGGAAAGAAGCTGCGGCTAAAGCGGTTGCTGAAGGTGGT
TCTTCTGACAAGAACCTGCAAGAGTTCGTTGACTTCGTTCGTGCGAAATCTGCGGAACTGGCGGCATAA

> sdU$_{exp}$5

MAAEPLVHVLLVSFPGQGHVNPLLRLGIHLASRGLLVTFTTPAWVGDKIRKASGVTTEPVPVGKGELRFEFFEDG
WAPDEPRLRDLDVYLPQLEAVGRRVIPELVRRHAEEGRPVSCLINNPFIPWVSDVAAELGLPSAMLWVQSPACLL
AYYYYYHKLVPFPSEENPDIDVQLPKLPLLKPDEVPSFLHPTSPYPALRRAILGQYANLDKPFCILVESFYELEK
EVIEHLSKILPIKCVGPLFLNPEAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGADFLWVMKPLPKDAGEELTLPEGFLEKVGDRGLVVQWSPQEKVLAHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGERENRIIPREEVARCIKEATEGPLAAEMKANAAKWKALAQKAVAEGG
SSDKNIQAFVDEVRARSKELAA

ATGGCAGCAGAACCGCTGGTACACGTTCTGCTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTTCTGCGT
CTGGGTATCCACCTGGCTTCTCGTGGTCTGCTGGTTACCTTCACCACTCCAGCATGGGTTGGTGACAAGATCCGT
AAAGCGTCTGGTGTTACCACTGAACCAGTACCGGTTGGTAAAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT
TGGGCTCCGGATGAACCACGTCTGCGTGATCTGGACGTTTACCTGCCACAGCTGGAAGCTGTTGGTCGTCGTGTT
ATTCCGGAACTGGTTCGTCGTCACGCTGAAGAAGGTCGTCCAGTTTCTTGCCTGATCAACAACCCGTTCATCCCA
TGGGTTTCTGACGTTGCTGCTGAACTGGGTCTGCCATCTGCGATGCTGTGGGTTCAGTCTCCAGCTTGCCTGCTG
GCGTACTACTATTACTACCACAAACTGGTTCCATTTCCGTCTGAAGAGAATCCAGACATCGACGTTCAGCTGCCG
AAACTGCCACTGCTGAAACCGGACGAAGTTCCGAGCTTTCTGCATCCGACCTCTCCGTATCCAGCTCTGCGTCGT
GCTATTCTGGGTCAGTACGCAAACCTGGACAAACCGTTCTGCATTCTGGTTGAGAGCTTCTACGAACTGGAGAAA
GAAGTTATCGAACACCTGTCCAAGATTCTGCCAATCAAGTGCGTTGGTCCACTGTTTCTGAATCCGGAAGCGGCG
GATAGTCCGGTTCGTGGTGACTTCATGCCGGCAGATGAATGCCTGCCATGGCTGGACACCAAACCGCCGAACTCT
GTTGTGTACATCAGCTTCGGTTCTGTTGTTTACCTTCCGCAAGAACAGGTTGACGAAATCGCGGAAGGTTTGCTG
CTGTCTGGTGCGGACTTTCTGTGGGTTATGAAACCACTGCCGAAAGATGCTGGTGAAGAACTGACTCTGCCAGAA
GGTTTCTTGGAGAAAGTTGGTGACCGTGGTTTGGTTGTTCAGTGGAGCCCGCAAGAGAAAGTTCTGGCTCATCCA
GCAGTTGCTGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAATCTCTGGCTTCTGGTATCCCAGTTATC
TGCTTTCCGCAGTGGGGTGATCAGGTTACCGATGCTCACCTGCTGGTTGACGTGTTCAAAGTTGGTATCCGTATG
TGCCGTGGTGAACGTGAGAACCGTATCATTCCGCGTGAAGAAGTTGCACGTTGCATCAAAGAAGCGACTGAAGGT
CCACTGGCAGCGGAAATGAAGGCAAACGCAGCGAAATGGAAAGCTCTGGCACAGAAAGCAGTGGCTGAAGGTGGT
TCTTCTGACAAGAACATCCAGGCGTTCGTTGACGAAGTTCGTGCACGTTCCAAAGAACTGGCAGCATAA

> **sdU$_{exp}$6**

MPPEPLVHVLLVSFPGQGHVNPLLRLGLRLAERGLLVTFTTPAWMGEKMRKASGVTEEPYPVGKGEIRFEFFEDG
WAPDEPRLRDLDVYLPQLEKVGSVVIPELVRREAERGRPVSCLINNPFIPWVSDVAESLGLPSAMLWVQSPACLL
AYYYYYHKLVPFPSEENPKIDVQLPKLPLLKWDEVPSFLHPTSPYPALRRAILGQYSRLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFINPRAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVFLPQEQVDEIAEGLL
LSGKDFLWVMKPLPKDAGEELKLPEGFLEKVGDRGLVVQWSPQEKVLKHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAKLLVDVFKVGIRMCRGEAENRIIPREEVARCILEATEGPLAEEMKKNAAKWKEKALAAVAEGG
SSDKNIQEFVDLVRAKSKELAA

ATGCCACCGGAACCACTGGTTCACGTTCTGCTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTTGGTCTGCGTCTGGCGGAACGTGGTCTGCTGGTGACCTTCACTACTCCGGCATGGATGGGTGAGAAGATGCGT
AAAGCGTCTGGTGTTACCGAAGAACCATATCCGGTTGGTAAAGGTGAAATCCGTTTCGAGTTCTTCGAAGACGGT
TGGGCACCAGATGAACCGCGTTTGCGTGACCTGGACGTTTACCTGCCGCAGCTGGAGAAGGTTGGTTCTGTGGTT
ATTCCAGAACTGGTACGTCGTGAAGCTGAACGTGGTCGTCCGGTTTCTTGCCTGATCAACAATCCGTTCATCCCG
TGGGTTTCTGACGTTGCTGAAAGCCTCGGTCTGCCATCTGCGATGCTGTGGGTTCAGAGTCCAGCGTGCCTGCTG
GCGTACTACTACTATTACCACAAACTGGTTCCGTTCCCATCTGAAGAGAATCCGAAGATCGATGTTCAGCTGCCA
AAGCTGCCGCTGCTGAAATGGGACGAAGTTCCGAGCTTTCTGCATCCGACCTCTCCGTATCCGGCTCTTCGTCGT
GCTATTCTGGGTCAGTACTCTCGTCTGGACAAACCGTTCTGCATTCTGGTTGAATCTTTCTACGAACTGGAGAAA
GAAGTTATCGAATACCTGTCCAAAATCTTGCCGATCAAATGCGTTGGTCCACTGTTCATCAATCCGCGTGCAGCT
GATTCTCCGGTTCGTGGTGACTTCATGCCAGCTGATGAATGCCTGCCATGGCTGGACACCAAACCGCCGAACTCT
GTTGTGTACATCAGCTTCGGTTCCGTTGTGTTTCTGCCGCAAGAACAGGTTGACGAAATCGCTGAAGGTCTGCTG
CTGTCTGGTAAAGACTTTCTGTGGGTTATGAAACCACTGCCGAAAGATGCAGGTGAAGAACTGAAACTGCCGGAA
GGCTTTCTGGAGAAAGTTGGTGACCGTGGTCTGGTTGTTCAGTGGAGTCCGCAAGAGAAAGTGCTGAAACATCCG
GCTGTAGCTGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAAAGCCTGGCGTCTGGTATCCCGGTTATC
TGCTTTCCGCAGTGGGGTGATCAGGTTACCGACGCGAAACTGCTGGTTGACGTGTTCAAAGTGGGTATCCGTATG
TGCCGTGGTGAAGCGGAGAACCGTATCATCCCGCGTGAAGAAGTTGCGCGTTGTATTCTGGAAGCAACCGAAGGT
CCACTGGCTGAAGAAATGAAGAAGAACGCGGCTAAATGGAAAGAGAAAGCACTGGCTGCTGTTGCAGAAGGTGGT
TCTTCTGACAAGAACATCCAAGAGTTCGTTGACCTGGTTCGTGCGAAATCCAAAGAACTGGCGGCATAA

> **sdU$_{exp}$7**

MAPAPLVHVLLVSFPGQGHVNPLLRLGIHLAERGLLVTFTTPAWVGDKIRKASGVTTEPVPVGEGELRFEFFEDG
WAPDEPRRRDLDVYLPQLEKVGSVELPKLIRRLAEEGRPVSCLINNPFIPWVSDVAASLGLPSAMLWVQSPACLL
AYYYYYHKLVPFPSEENPDIDVQLPKLPLLKPDEVPSFLHPTSPYPALRRAILGQYANLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFINPRAADSPVRGDFMKADECLPWLDTKPPNSVVYISFGSVVFLPQEQVDEIAEGLL
LSGADFLWVMKPLPKDAGEELKLPEGFLEKVGDRGLVVQWSPQEKVLAHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGERENRIIPREEVARCIKEATQGPKAEEMKANAAKWKAAAAKAVAEGG
SSDKNIQEFVDEVRARSAELAA

ATGGCTCCAGCTCCACTGGTTCACGTTCTGCTGGTTTCTTTCCCAGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTATCCATCTGGCAGAACGTGGTCTTCTGGTTACCTTCACCACTCCAGCTTGGGTAGGTGACAAGATCCGT
AAAGCGTCTGGTGTTACCACCGAACCAGTTCCAGTTGGTGAAGGTGAACTGCGTTTCGAATTTTTCGAAGACGGT
TGGGCTCCAGATGAACCGCGTCGTCGTGATCTGGACGTTTACCTGCCACAGCTTGAGAAGGTTGGTTCTGTTGAA
CTGCCGAAATTGATCCGTCGTCTTGCGGAAGAAGGTCGTCCGGTTTCTTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCTGCTTCTCTGGGTCTGCCATCTGCGATGCTGTGGGTTCAGTCTCCGGCATGTCTGCTG
GCGTATTACTACTACTACCACAAACTGGTTCCGTTTCCGTCTGAAGAGAATCCGGATATCGACGTTCAGCTGCCG
AAACTGCCGCTGCTTAAACCGGATGAAGTTCCGAGCTTTCTGCATCCGACCTCTCCGTATCCGGCTCTGCGTCGT

GCAATTCTGGGTCAGTACGCGAACCTGGACAAACCGTTCTGCATTCTGGTCGAATCTTTCTACGAACTGGAGAAA
GAAGTGATCGAATACCTGTCCAAAATCTTGCCGATCAAATGCGTAGGTCCGCTGTTCATCAATCCGCGTGCAGCT
GACAGTCCGGTTCGTGGTGACTTCATGAAAGCGGACGAATGCCTGCCGTGGCTGGACACCAAGCCACCGAACTCT
GTTGTGTACATCAGCTTCGGTTCCGTTGTGTTTCTGCCGCAAGAACAGGTTGACGAAATCGCGGAAGGTCTGCTG
CTGTCTGGTGCTGACTTTCTGTGGGTGATGAAACCGCTGCCGAAAGACGCTGGTGAAGAACTGAAACTGCCAGAA
GGTTTCTTGGAGAAAGTTGGTGATCGTGGTCTGGTTGTTCAGTGGTCCCCGCAAGAGAAAGTTCTGGCTCATCCA
GCTGTAGCAGCATTCGTTACTCACTGCGGTTGGAACTCTACCATGGAAAGCCTGGCTTCTGGTATTCCGGTTATC
TGCTTTCCGCAGTGGGGTGATCAGGTTACCGATGCTCACCTTCTGGTTGACGTGTTCAAAGTTGGTATCCGTATG
TGCCGTGGTGAACGTGAGAACCGTATCATTCCACGTGAAGAAGTTGCTCGTTGCATCAAAGAAGCGACTCAGGGT
CCAAAAGCGGAAGAAATGAAGGCTAACGCTGCGAAATGGAAAGCTGCTGCTGCTAAAGCGGTAGCTGAAGGTGGT
TCTTCTGACAAGAACATCCAAGAGTTCGTTGACGAAGTTCGTGCACGTTCCGCGGAACTGGCAGCATAA

> sdU$_{exp}$8

MPAPPLVHVFLVSFPGQGHVNPLLRLGIHLAERGLLVTFSTPKWVGDKIRKASGVTTEPVPVGAGELRFEFFEDG
WAPDEPRRRDLDVYLPQLEEVGKKVLPAMIRAAAEAGRPVSCLINNPFIPWVSDVAAELGLPSAMLWVQSAACLL
IYYYYYHKLVPFPSEENPKIDVQLPKLPLLKWDEVPSFLHPTSPYPALRRAILGQYSRLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFINPRAADSPVRGDFMPADECLPWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGQYFLWVMKPRPKDAGEELKLPEGFLEKVGDKGLVVQWSPQEKVLKHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGELENRIIPREEVARCILEATTGPKAEEMKANAAKWAALAQKAVAEGG
SSDTNLQAFVDEVRARSAELAA

ATGCCGGCACCGCCGCTGGTTCACGTGTTCTTGGTTTCTTTCCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTATCCACCTGGCTGAACGTGGTCTGCTGGTTACCTTCTCTACTCCGAAATGGGTTGGTGACAAGATCCGT
AAAGCTAGCGGTGTTACCACCGAACCAGTTCCGGTTGGTGCAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT
TGGGCACCGGACGAACCGCGTCGTCGTGACCTGGACGTGTACCTGCCACAGCTGGAAGAAGTTGGTAAGAAAGTT
CTGCCGGCGATGATCCGTGCAGCGGCTGAAGCTGGTCGTCCAGTTAGCTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCTGCGGAACTGGGTCTGCCGTCTGCGATGCTGTGGGTTCAGTCTGCTGCATGCCTGCTG
ATCTACTACTATTACTACCACAAACTGGTTCCGTTTCCGTCTGAAGAGAATCCGAAGATCGACGTTCAGCTGCCG
AAACTGCCGCTGCTGAAATGGGATGAAGTTCCATCTTTCTTGCATCCGACCTCTCCGTATCCGGCTCTGCGTCGT
GCGATCTTGGGTCAGTACTCTCGTCTGGACAAACCGTTCTGCATCTTGGTGGAATCTTTCTACGAACTGGAGAAA
GAAGTTATCGAATACCTGTCTAAGATTCTGCCGATCAAATGCGTTGGTCCACTGTTCATCAATCCACGTGCTGCG
GATTCTCCGGTTCGTGGTGACTTCATGCCGGCAGACGAATGCCTGCCGTGGCTGGACACTAAACCGCCGAACTCC
GTTGTTTACATCTCTTTCGGTTCTGTTGTTTACCTGCCGCAAGAACAGGTTGACGAAATCGCGGAAGGTCTGCTG
CTGTCTGGTCAGTACTTTCTGTGGGTTATGAAACCGCGTCCGAAAGACGCGGGTGAAGAACTGAAACTGCCGGAA
GGTTTCTTGGAGAAAGTTGGCGACAAAGGTCTGGTTGTTCAGTGGAGTCCGCAAGAGAAAGTGCTGAAACATCCA
GCAGTAGCAGCGTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAAAGCCTGGCGTCTGGTATTCCGGTTATC
TGCTTCCCACAGTGGGGTGACCAGGTTACTGACGCGCACCTGCTGGTTGACGTGTTCAAAGTTGGTATCCGTATG
TGCCGTGGTGAACTGGAGAACCGTATCATTCCGCGTGAAGAAGTTGCACGTTGCATTCTGGAAGCGACCACTGGT
CCGAAAGCGGAAGAAATGAAAGCGAACGCGGCTAAATGGGCTGCTCTGGCGCAGAAAGCTGTTGCGGAAGGTGGT
TCTTCTGACACCAACCTTCAGGCGTTCGTTGACGAAGTTCGTGCTCGTTCTGCTGAACTGGCGGCATAA

> sdU$_{exp}$9

MAAPPLVHVHLVSFPGQGHVNPLLRLGLHLAELGLLVTFTTPAWVGEKMRKASGVTTEPYPVGKGELRFEFFEDG
WAPDEPRRRDLDVYLPQLEAVGAKVLPAQIRAAAERGRPVSCLINNPFIPWVSDVAAALGLPSAMLWVQSPACLL
AYYYYYHKLVPFPSEENPDIDVQLPKLPLLKPDEVPSFLHPTSPYPALRRAILGQYARLDKPFCILVESFYELEK

EVIEYLSKILPIKCVGPLFLNPRAADSPVRGDFMPADECMAWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGKYFLWVMKPRPKDAGEELKLPEGFLEKVGDKGLVVQWSPQEKVLAHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAYLLVDVFKVGIRMCRGEAENRIIPREEVARCILEATEGPKAEAMKANAAKWAAAAKAAVAEGG
SSDTNLQEFVDLVRARSAELAA

ATGGCTGCTCCGCCATTGGTTCACGTTCACCTGGTTAGCTTTCCGGGTCAGGGTCACGTTAATCCGCTGCTGCGT
CTGGGTCTGCATCTGGCAGAACTGGGCCTGCTGGTTACCTTCACCACTCCGGCTTGGGTTGGTGAGAAGATGCGT
AAAGCGTCTGGTGTTACCACCGAACCATACCCAGTTGGTAAAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT
TGGGCTCCGGATGAACCGCGTCGTCGTGATCTGGACGTTTACCTGCCACAGCTGGAAGCAGTAGGTGCGAAAGTG
TTGCCAGCTCAGATCCGTGCTGCTGCTGAACGTGGTCGTCCAGTTAGCTGCCTGATCAACAATCCGTTCATCCCG
TGGGTTTCTGACGTTGCTGCTGCACTGGGTCTGCCATCTGCGATGCTGTGGGTTCAGTCTCCAGCTTGTCTGCTG
GCGTACTACTACTACTACCACAAACTGGTTCCGTTCCCATCTGAAGAGAATCCGGATATCGACGTTCAGCTGCCG
AAACTGCCACTGCTGAAACCGGACGAAGTTCCGTCTTTCTTGCATCCGACCTCTCCATATCCGGCTCTGCGTCGT
GCAATCTTGGGTCAGTACGCACGTCTGGACAAACCGTTCTGCATCTTGGTTGAGTCTTTCTACGAACTGGAGAAA
GAAGTGATCGAATACCTGTCCAAAATCTTGCCGATCAAATGCGTTGGTCCACTGTTTCTGAATCCGCGTGCGGCA
GACAGTCCGGTTCGTGGTGACTTCATGCCAGCGGACGAATGCATGGCGTGGCTGGACACCAAACCGCCGAACTCT
GTTGTGTACATCAGCTTCGGTTCTGTTGTTTACCTGCCGCAAGAACAGGTGGATGAAATCGCTGAAGGTCTGCTG
CTGTCTGGTAAGTACTTTCTGTGGGTTATGAAACCGCGTCCGAAAGACGCTGGTGAAGAACTGAAACTGCCGGAA
GGCTTTCTGGAGAAAGTTGGTGACAAAGGTCTGGTTGTTCAGTGGAGTCCGCAAGAGAAAGTTCTGGCACATCCG
GCTGTAGCGGCTTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAAAGCCTGGCATCTGGTATTCCGGTTATC
TGCTTTCCACAGTGGGGTGACCAGGTTACCGACGCGTACCTGCTGGTAGACGTGTTCAAAGTTGGTATCCGTATG
TGCCGTGGTGAAGCTGAGAACCGTATCATTCCGCGTGAAGAAGTTGCGCGTTGCATTCTGGAAGCGACTGAAGGT
CCGAAAGCAGAAGCGATGAAAGCGAACGCGGCGAAATGGGCGGCAGCAGCGAAAGCAGCGGTAGCTGAAGGTGGT
TCTTCTGACACCAACCTGCAAGAGTTCGTTGACCTGGTTCGTGCTCGTTCCGCAGAACTGGCTGCTTAA

> sdU$_{exp}$10

MAPAPLVHVLLVSFPGQGHVNPLLRLGIHLASRGLLVTFTTPSWVGEKIRKASGVTTEPYPVGKGELRFEFFEDG
WAADEPRLRDLDVYLPQLEKVGRVTIPALVRRLAEEGRPVSCLINNPFIPWVSDVAAELGLPSAMLWVQSPACLL
AYYYYYHGLVPFPSEENPKIDVQLPKLPLLKWDEVPSFLHPTSPYPALRRAILGQYAKLDKPFCILVESFYELEK
EVIEYLSKILPIKCVGPLFLNPEAADSPVRGDFMPADECMAWLDTKPPNSVVYISFGSVVYLPQEQVDEIAEGLL
LSGKDFLWVMKPLPKDAGEQLTLPEGFLEKVGDRGLVVQWSPQEKVLRHPAVAAFVTHCGWNSTMESLASGIPVI
CFPQWGDQVTDAHLLVDVFKVGIRMCRGEFENRIIPREEVARCIKEATEGPLAAEMKANAAKWAELAKKAVAEGG
SSDKNIQEFVDLVRAKSAELAA

ATGGCACCGGCACCGTTGGTTCACGTTCTGCTGGTTTCTTTCCCGGGTCAAGGTCACGTTAACCCACTGCTGCGT
CTGGGTATCCACCTGGCTTCTCGTGGTCTGCTGGTTACCTTCACCACTCCAAGCTGGGTTGGTGAGAAGATCCGT
AAAGCGTCTGGTGTTACCACCGAACCGTATCCGGTTGGTAAAGGTGAACTGCGTTTCGAGTTCTTCGAAGACGGT
TGGGCAGCTGATGAACCGCGTCTGCGTGATCTGGACGTTTACCTGCCGCAGCTGGAGAAAGTTGGTCGTGTTACC
ATTCCGGCTCTGGTTCGTCGTCTGGCTGAAGAAGGTCGTCCGGTTAGCTGCCTGATCAACAATCCGTTCATTCCG
TGGGTTTCTGACGTTGCGGCTGAACTGGGTCTGCCGTCTGCTATGCTGTGGGTTCAGTCTCCGGCTTGCTTGCTG
GCGTACTACTACTACTACCACGGTCTGGTTCCGTTTCCGTCTGAAGAGAATCCGAAGATCGACGTTCAGCTGCCG
AAACTGCCGCTGCTGAAATGGGACGAAGTTCCGAGCTTTCTGCATCCGACTTCTCCGTATCCGGCACTGCGTCGT
GCAATTCTGGGTCAGTACGCGAAACTGGACAAACCGTTCTGCATTCTGGTTGAATCTTTCTACGAACTGGAGAAA
GAAGTTATCGAATACCTGTCTAAGATTCTGCCGATCAAATGCGTAGGTCCACTGTTTCTGAATCCGGAAGCGGCT
GATTCTCCGGTACGTGGTGACTTCATGCCGGCTGACGAATGCATGGCTTGGCTGGACACCAAACCGCCAAACTCT
GTTGTTTACATCAGCTTCGGTTCTGTTGTTTACCTGCCGCAAGAACAGGTTGACGAAATCGCTGAAGGTCTGCTG

1099 CTGTCTGGTAAAGACTTTCTGTGGGTTATGAAACCGCTGCCGAAAGACGCTGGTGAACAGCTGACCTTGCCGGAA
1100 GGTTTCTTGGAGAAAGTTGGTGACCGTGGTCTGGTTGTTCAGTGGTCCCCGCAAGAGAAAGTTCTGCGTCATCCA
1101 GCAGTTGCTGCTTTCGTTACTCACTGCGGTTGGAACTCTACCATGGAATCTCTGGCTTCTGGTATTCCGGTTATC
1102 TGCTTTCCGCAGTGGGGTGATCAGGTTACCGATGCTCACCTGCTGGTTGACGTGTTCAAAGTGGGTATCCGTATG
1103 TGTCGTGGTGAGTTCGAGAACCGTATCATTCCGCGTGAAGAAGTTGCGCGTTGCATCAAAGAAGCGACCGAAGGT
1104 CCGCTGGCTGCTGAAATGAAAGCTAACGCTGCAAAGTGGGCTGAACTGGCGAAGAAAGCTGTTGCAGAAGGTGGT
1105 TCTTCTGACAAGAACATCCAAGAGTTCGTTGACCTGGTACGTGCGAAATCTGCTGAACTGGCGGCTTAA

Table S9: Molecular chaperones and fusion tags used for enzyme expression in this study. * Represents the stop codon.

> ibPAB (ibPA-ibPB)

MRNFDLSPLYRSAIGFDRLFNHLENNQSQSNGGYPPYNVELVDENHYRIAIAVAGFAESELEITAQDNLLVVKGA
HADEQKERTYLYQGIAERNFERKFQLAENIHVRGANLVNGLLYIDLERVIPEAKKPRRIEIN*MRNFDLSPLMRQ
WIGFDKLANALQNAGESQSFPPYNIEKSDDNHYRITLALAGFRQEDLEIQLEGTRLSVKGTPEQPKEEKKWLHQG
LMNQPFSLSFTLAENMEVSGATFVNGLLHIDLIRNEPEPIAAQRIAISERPALNS*

>σ^{32}

MTDKMQSLALAPVGNLDSYIRAANAWPMLSADEERALAEKLHYHGDLEAAKTLILSHLRFVVHIARNYAGYGLPQ
ADLIQEGNIGLMKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRNWRIVKVATTKAQRKLFFNLRKTKQRLGWFN
QDEVEMVARELGVTSKDVREMESRMAAQDMTFDLSSDDDSDSQPMAPVLYLQDKSSNFADGIEDDNWEEQAANRL
TDAMQGLDERSQDIIRARWLDEDNKSTLQELADRYGVSAERVRQLEKNAMKKLRAAIEA*

> DnaKJ (DnaK-DnaJ)

MGKIIGIDLGTTNSCVAIMDGTTPRVLENAEGDRTTPSIIAYTQDGETLVGQPAKRQAVTNPQNTLFAIKRLIGR
RFQDEEVQRDVSIMPFKIIAADNGDAWVEVKGQKMAPPQISAEVLKKMKKTAEDYLGEPVTEAVITVPAYFNDAQ
RQATKDAGRIAGLEVKRIINEPTAAALAYGLDKGTGNRTIAVYDLGGGTFDISIIEIDEVDGEKTFEVLATNGDT
HLGGEDFDSRLINYLVEEFKKDQGIDLRNDPLAMQRLKEAAEKAKIELSSAQQTDVNLPYITADATGPKHMNIKV
TRAKLESLVEDLVNRSIEPLKVALQDAGLSVSDIDDVILVGGQTRMPMVQKKVAEFFGKEPRKDVNPDEAVAIGA
AVQGGVLTGDVKDVLLLDVTPLSLGIETMGGVMTTLIAKNTTIPTKHSQVFSTAEDNQSAVTIHVLQGERKRAAD
NKSLGQFNLDGINPAPRGMPQIEVTFDIDADGILHVSAKDKNSGKEQKITIKASSGLNEDEIQKMVRDAEANAEA
DRKFEELVQTRNQGDHLLHSTRKQVEEAGDKLPADDKTAIESALTALETALKGEDKAAIEAKMQELAQVSQKLME
IAQQQHAQQQTAGADASANNAKDDDVVDAEFEEVKDKK*MAKQDYYEILGVSKTAEEREIRKAYKRLAMKYHPDR
NQGDKEAEAKFKEIKEAYEVLTDSQKRAAYDQYGHAAFEQGGMGGGGFGGGADFSDIFGDVFGDIFGGGRGRQRA
ARGADLRYNMELTLEEAVRGVTKEIRIPTLEECDVCHGSGAKPGTQPQTCPTCHGSGQVQMRQGFFAVQQTCPHC
QGRGTLIKDPCNKCHGHGRVERSKTLSVKIPAGVDTGDRIRLAGEGEAGEHGAPAGDLYVQVQVKQHPIFEREGN
NLYCEVPINFAMAALGGEIEVPTLDGRVKLKVPGETQTGKLFRMRGKGVKSVRGGAQGDLLCRVVVETPVGLNER
QKQLLQELQESFGGPTGEHNSPRSKSFFDGVKKFFDDLTR*

> GroSL (GroS-GroL)

MNIRPLHDRVIVKRKEVETKSAGGIVLTGSAAAKSTRGEVLAVGNGRILENGEVKPLDVKVGDIVIFNDGYGVKS
EKIDNEEVLIMSESDILAIVEA*MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKD
GVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGIDKAVTAA
VEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDGTGLQDELDVVEGMQFDRGYLSP
YFINKPETGAVELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVK
APGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQ
IEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLA
DLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNAATEEYGNMIDMGILDPTKVTRS
ALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGGMGGMGGMM*

> CysQ

MLDQVCQLARNAGDAIMQVYDGTKPMDVVSKADNSPVTAADIAAHTVIMDGLRTLTPDVPVLSEEDPPGWEVRQH
WQRYWLVDPLDGTKEFIKRNGEFTVNIALIDHGKPILGVVYAPVMNVMYSAAEGKAWKEECGVRKQIQVRDARPP

LVVISRSHADAELKEYLQQLGEHQTTSIGSSLKFCLVAEGQAQLYPRFGPTNIWDTAAGHAVAAAAGAHVHDWQG
KPLDYTPRESFLNPGFRVSIY

> **GST**

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLTEGVAIMQYLAD
SVPDRQLLAPVNSISRYKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQLEKKLQYVNEALKDEHWICGQR
FTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQDALSAEGLK

> **SUMO**

MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQAD
QTPEDLDMEDNDIIEAHREQIGG

> **MBP**

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQT

Notes

The following command was used to perform sequence redesign with ProteinMPNN.

```
python $MPNN_PATH/protein_mpnn_run.py \
--jsonl_path ../parsed_pdbs_bb.jsonl \
--chain_id_jsonl ../assigned_chains.jsonl \
--fixed_positions_jsonl ../masked_pos.jsonl \
--out_folder $MPNN_OUTDIR \
--num_seq_per_target 100 \
--sampling_temp “0.1” \
--batch_size 1 \
--omit_AAs=’C’
```

Where ../assigned_chains.jsonl contains the parsed PDB chain information: {“TOGT”: [[“A”]] } /{“UGT84A56”: [[“A”]] }

This script generates 100 sequences for TOGT and 100 sequences for UGT84A56 in Second-design process with 1st and 2nd shell residues fixation.

This script generates 20 sequences for TOGT and 20 sequences for UGT84A56 in First-design process with only 1st shell residues and conserved residues fixation.

Sites fixed during the second-redesign of TOGT sequence are as follows:

[1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 42, 43, 46, 50, 60, 62, 63, 65, 66, 67, 68, 71, 73, 74, 75, 76, 78, 79, 80, 81, 82, 86, 87, 90, 93, 94, 96, 100, 101, 103, 104, 105, 107, 108, 113, 114, 115, 116, 117, 119, 120, 121, 124, 129, 130, 131, 132, 134, 135, 136, 137, 139, 140, 141, 142, 143, 145, 146, 150, 152, 158, 164, 165, 167, 169, 171, 172, 174, 175, 178, 180, 181, 184, 186, 187, 188, 197, 201, 207, 211, 213, 214, 215, 217, 218, 219, 221, 222, 223, 226, 230, 236, 237, 238, 239, 241, 242, 243, 244, 245, 246, 254, 255, 256, 257, 258, 259, 260, 261, 262, 264, 268, 269, 271, 272, 273, 275, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 295, 296, 298, 299, 300, 302, 303, 304, 305, 306, 307, 310, 311, 312, 313, 314, 315, 324, 325, 327, 328, 329, 330, 331, 332, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 353, 354, 355, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 372, 373, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 398, 399, 400, 402, 403, 407, 409, 415, 418,

419, 420, 423, 424, 427, 428, 431, 432, 433, 437, 439, 440, 447, 450, 451, 453, 454, 455, 456, 457, 460, 461, 464, 465, 468].

Sites fixed during the second-redesign of UGT84A56 sequence are as follows:

[1, 2, 8, 9, 11, 12，13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 31, 34, 35, 36, 37, 38, 39, 41, 42, 49, 53, 59, 61, 63, 70, 72, 73, 74, 75, 85, 86, 90, 97, 100, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 125, 126, 127, 128, 129, 130, 131, 135, 136, 137, 138, 139, 141, 142, 143, 144, 145, 147, 148, 151, 152, 153, 154, 155, 163, 168, 172, 174, 175, 178, 180, 184, 185, 186, 187, 188, 189, 195, 198, 199, 201, 202, 205, 209, 212, 213, 215, 216, 217, 218, 219, 220, 222, 223, 224, 228, 230, 236, 237, 240, 241, 242, 243, 244, 246, 247, 249, 250, 254, 255, 256, 257, 258, 259, 264, 267, 268, 269, 271, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 287, 291, 292，293, 294, 295, 296, 298, 299, 302, 303, 306, 307, 308, 309, 310, 311, 314, 317, 323, 324, 327, 329, 330, 335, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 349, 352, 353, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 370, 371, 372, 373, 374, 375, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 390, 391, 392, 393, 397, 398, 399, 403, 406, 408, 410, 412, 413, 414, 415, 418, 420, 421, 422, 423, 425, 426, 428, 431, 434, 435, 437, 438, 439, 442, 445, 446, 447, 448, 449, 450, 451, 452, 453, 455, 459, 460, 463, 467].

References

[1] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583–589 (2021).

[2] Goldenberg, O., Erez, E., Nimrod, G. & Ben-Tal, N. The ConSurf-DB: Pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* **37**, D323–D327 (2008).

[3] Zhang, Y., & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

[4] Hon, J. et al. SoluProt: Prediction of soluble protein expression in *Escherichia coli*. *Bioinformatics.* **37**, 23–28 (2021).

[5] Lemkul, J. A. From proteins to perturbed Hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package. *Living J. Comput. Mol. Sci.* **1**, 5068 (2019).

[6] Huang, J. et al. Biomolecular simulations: Theory and applications. *Biophys. J.* **112**, 175a–176a (2017).

[7] Schöning-Stierand, K. et al. ProteinsPlus: interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Res.* **50**, W611–W615 (2022).