Synthetic and Systems Biotechnology 10 (2025) 185–193

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Enhancing substrate specificity of microbial transglutaminase for precise nanobody labeling

Xinglong Wang [a,b,c], Kangjie Xu [c], Haoran Fu [c], Qiming Chen [c], Beichen Zhao [e], Xinyi Zhao [c], Jingwen Zhou [b,c,d,*]

[a] School of Food Science and Technology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China
[b] Engineering Research Center of Ministry of Education on Food Synthetic Biotechnology and School of Biotechnology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China
[c] Science Center for Future Foods, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China
[d] Jiangsu Province Engineering Research Center of Food Synthetic Biotechnology, Jiangnan University, Wuxi, 214122, China
[e] Department of Chemical and Materials Engineering, The University of Auckland, Auckland, 1010, New Zealand

ARTICLE INFO

Keywords:
Enzyme-ligand docking
Virtual mutagenesis
Streptomyces mobaraenesis transglutaminase
Nanobody-drug conjugation
Allergen detection

ABSTRACT

Streptomyces mobaraenesis transglutaminase (smTG) can be used for site-specific labeling of proteins with chemical groups. Here, we explored the use of modified smTG for the biosynthesis of nanobody-fluorophore conjugates (NFC). smTG catalyzes the conjugation of acyl donors containing glutamine with lysine-containing acceptors, which can lead to non-specific cross-linking. To achieve precise site-specific labeling, we employed molecular docking and virtual mutagenesis to redesign the enzyme's substrate specificity towards the peptide GGGGQR, a non-preferred acyl donor for smTG. Starting with a thermostable and highly active smTG variant (TGm2), we identified that single mutations G250H and Y278E significantly enhanced activity against GGGGQR, increasing it by 41 % and 1.13-fold, respectively. Notably, the Y278E mutation dramatically shifted the enzyme's substrate preference, with the activity ratio against GGGGQR versus the standard substrate CBZ-Gln-Gly rising from 0.05 to 0.93. In case studies, we used nanobodies 1C12 and 7D12 as labeling targets, catalyzing their conjugation with a synthetic fluorophore via smTG variants. Nanobodies fused with GGGGQR were successfully site-specifically labeled by TGm2-Y278E, in contrast to non-specific labeling observed with other variants. These results suggest that engineering smTG for site-specific labeling is a promising approach for the biosynthesis of antibody-drug conjugates.

1. Introduction

Streptomyces mobaraenesis transglutaminase (smTG) catalyzes the formation of a covalent bond between a glutamine residue from an acyl donor and a lysine residue from an acyl acceptor [1]. Due to its catalytic crosslinking function, smTG has been widely used for site-specific labeling of protein-based drugs and in the conjugation of antibody-drug conjugates (ADCs) [2,3]. However, non-specific labeling at multiple glutamine sites within target proteins can result in an unstable labeling ratio [4,5], posing potential risks for ADCs [6]. smTG exhibits high selectivity towards acyl donors, favoring hydrophobic residues or residues with large uncharged side chains adjacent to the N-terminal of glutamine, and disfavors residues without side chains [4,7]. Therefore, engineering smTG to alter its substrate preferences can be a promising strategy for achieving a stable labeling ratio in ADCs and site-specific labeling of proteins.

To characterize the substrate preferences of smTG, a previous study using phage display revealed that smTG exhibited high affinity towards the peptide WALQRP [8]. Conversely, it showed very low activity against the peptide GQR, as identified through peptide array analysis [9]. Similarly, the peptide sequence GGGQRGG, which is akin to GQR, was also not favored by smTG, as confirmed using amplified peptides [4,7]. smTG-induced protein labeling, such as the conjugation of interferon and growth hormone with PEG, can extend the half-life of these proteins in the human body [2]. However, the uncertainty regarding the conjugation site and the polymer-PEG ratio significantly impacts its utility.

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author. Science Center for Future Foods, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China.
E-mail address: zhoujw1982@jiangnan.edu.cn (J. Zhou).

https://doi.org/10.1016/j.synbio.2024.10.003
Received 26 July 2024; Received in revised form 11 October 2024; Accepted 14 October 2024
Available online 16 October 2024

Zhao et al. [10] conducted directed evolution on smTG and identified five mutants that efficiently recognize and modify the Gln141 of human growth hormone, demonstrating the feasibility of modifying smTG substrate preferences to achieve specific labeling. Steffen et al. [11] reported that *Kutzneria albida* transglutaminase displayed significantly different substrate preferences compared to smTG, showing hyperactivity towards the peptide sequence YRYRQ. They demonstrated that fusing YRYRQ to the C-terminal of immunoglobulin G enables site-specific labeling. This inspired our study to modify smTG and use fusion tags for inducing protein site-specific labeling.

Emerging tools for protein structure modeling and virtual mutagenesis have significantly benefited the rational design process [12–14]. Designing enzyme substrate specificity is a complex task that requires not only an accurate enzyme structure but also a highly precise docking pose, since serial mutations need to be introduced into the enzyme in the docking pose to calculate binding energy changes [15]. Previous strategies have emphasized the importance of conformation sampling in enzyme-substrate docking [16,17]. Larger sampling sizes enhance the search for accurate enzyme-substrate binding poses, with the distance between catalytic residues and binding energy being crucial for determining pose accuracy [18]. According to the 'Theozyme' theory, there is an ideal distance between the catalytic residue of an enzyme and the residue being attacked on a substrate [19,20]. This distance constraint allows the catalytic reaction to occur even if the enzyme binds the substrate in different poses. The Rosetta script can apply these distance constraints to filter poses that meet the required distance criteria [21]. Therefore, combining docking with distance constraints and virtual mutagenesis provides a robust solution for computationally screening potential mutations.

In our previous study, we engineered *Escherichia coli* for the secretory expression of active smTG, finding that an N-terminal FRAPD tag remained due to protease cleavage [22]. This residual tag can significantly enhanced the thermostability of smTG. We also modified several smTG variants and discovered FRAPD-smTG-S2P–S23V–Y24N-E28T-S199A-A265P-A287P–K294L (TGm2) exhibited notably higher thermostability and specific activity than smTG [13,23]. This variant displays a theoretical pI of 7 and does not display a strong surface charge, minimizing interference with substrates containing charged residues. Since both smTG and TGm2 exhibited the same ratio of 0.05 against the peptide GGGGQR compared to the standard substrate CBZ-Gln-Gly [9], we selected TGm2 as the modification target. This study introduced a structure-based rational design method, using molecular docking and virtual mutagenesis of substrate-binding residues to select beneficial variants. Variants were chosen based on the binding delta delta Gibbs free energy (ddG), and the top 20 variants were validated experimentally. For site-specific labeling of nanobodies, we first expressed and purified 1C12 [24] for nanobody-fluorophore conjugation (NFC) assay. We then conducted a case study by using 7D12 [25] to validate the site-specific labelling capacity induced by modified smTG. We investigated the mechanism behind the increased specific activity through molecular dynamics simulation.

2. Materials and methods

2.1. Materials

The plasmid pETDUET/*pelB-TrxA-proH-TGm2/TAMEP* for active secretory expressing TGm2 was constructed in our previous study [22], where proH and TAMEP are the pro-region from *Streptomyces hygroscopicus* transglutaminase and transglutaminase activating metalloprotease [22,26]. The gene encoding for pelB-proH-smTG (smTG GenBank accession number: EU301664) was synthesized and cloned into pETDUET/*pelB-TrxA-proH-TGm2/TAMEP* to replace the original expression cassette of pelB-TrxA-proH-TGm2 conducted by GENEWIZ (Suzhou, China). The gene encoding for 1C12 (GenBank accession number: CAA2601632.1) and 7D12 (GenBank accession number: 4KRL_B), and the fusion tag GGGGQR were synthesized by GENEWIZ, and cloned into pET-22b obtaining pET-22b/1C12-GGGGQR and pET-22b/7D12-GGGGQR (the expression cassette was shown in Table S1). To enable the exposure of His-tag and GGGGQR tag for purification and site-specific labelling purpose, we added a short GSS linker prior to His-tag and GGGGQR tag, respectively.

2.2. DNA manipulation

Polymerase chain reaction (PCR) was conducted to introduce point mutations into TGm2 and smTG using pETDUET/*pelB-TrxA-proH-TGm2/TAMEP* and pETDUET/*pelB-proH-TGm2/TAMEP* as template, respectively. The gene encoding for the fusion tag GGGGQR was removed from pET-22b/1C12-GGGGQR by PCR. The primers used for PCR were listed in Table S2, and the achieved linearized PCR products were circularized using one-step cloning (ClonExpress II One Step Cloning Kit, Vazyme) or Blunting Kination Ligation Kit (TaKaRa, Dalian, China) before competent cells transformation. The.

2.3. Protein expression and purification

All plasmids were chemical transformed into *E. coli* BL21 (DE3), and single colony was cultured using LB medium under 37 °C for 10 h in a shaking incubator. The initial culture was transferred into TB medium by a percentage of 2 %, the induction was carried out by supplemented IPTG to 0.1 mM while the cell density (OD_{600}) reach 1.0, the cells were continuously cultivated under 20 °C for 40 h. Cells were spun down by centrifugation, and the supernatant was collected for purification. Affinity purification was carried out using HisTrap column (GE healthcare). The column was pre-equivalent by a solution contains 50 mM Tris-HCl, 100 mM NaCl and 15 mM imidazole, pH 7.8. After sample injection, the column was washed using the same solution as pre-equivalent, and proteins fused with His-tag was eluted with solution contains 50 mM Tris-HCl, 100 mM NaCl and 150 mM imidazole. Size exclusion chromatography (SEC) was carried out using HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare, New York, USA) to desalt protein solution. smTG variants were finally eluted in a buffer of 50 mM Tris-HCl, pH 8.0, and nanobody was eluted using PBS, pH 7.5. The concentration of purified proteins were determined using Bradford assay according manufacturer instructions (Beyotime, Shanghai, China).

2.4. Transglutaminase activity assay

To measure transglutaminase activity against CBZ-Gln-Gly, the substrate solution was as follows: Tris-HCl 200 mM, 100 mM hydroxylamine, 10 mM GSH, 30 mM CBZ-Gln-Gly, pH 6.0. To measure the activity against peptide GGGGQR, the substrate solution was as follows, Tris-HCl 200 mM, 100 mM hydroxylamine, 10 mM glutathione reduced (GSH), 30 mM GGGGQR, 20 % (v/v) DMSO, pH 6.0. To measure the activity, protein samples and substrate solution were pre-warmed under 37 °C for 5 min. The reaction lasts for 10 min under 37 °C and terminated by adding 60 μL termination solution. The termination solution obtained by mixing same volume of 3 M HCl, 12 % trichloroacetic acid and 5 % $FeCl_3{\cdot}H_2O$ (in 0.1 M HCl) [27]. The reaction solution was centrifuged and measure the extinction citation under 525 nm wavelength using microtiter plate reader (FluoStar, BMG Labtech GmbH, Offenburg). The reference sample was made by adding 60 μL of sample protein to 60 μL termination solution and reacted at 37 °C for 10 min, followed by adding 150 μL substrate solution. The reaction produced 1 μmol L-glutamic acid γ-monohydroxamater per min was defined as 1 U of transglutaminase activity [13].

2.5. smTG variants catalyzed NFC

The purified nanobody was prepared at a concentration of 2 mg/mL in PBS (pH 7.5) with the addition of 10 mM GSH. The fluorophore

biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl) was synthesized by Sangon Biotech (Shanghai, China) and dissolved in water at a concentration of 50 mg/mL. For the catalytic reaction, 1 mL of the nanobody solution was combined with 200 μL of the fluorophore solution, followed by the addition of 100 μL of smTG variants at a concentration of 1 mg/mL. The reaction mixture was incubated at 37 °C for 1 h to complete the catalysis. After incubation, the samples were treated with 8 M urea at a 1:3 ratio, followed by centrifugation. The supernatant was then collected for loading onto size-exclusion chromatography (SEC).

2.6. Thermostability analysis

The smTG variants were adjusted to a concentration of 1 mg/mL, and their melting temperature (T_m) was determined using a differential scanning calorimeter (Nano-DSC, TA Instruments, New Castle, USA). The system was maintained at a pressure of 3 atm, and the temperature was increased from 40 °C to 90 °C at a rate of 1 °C per minute. The enthalpy change data collected during the scan were used to calculate the T_m.

2.7. Building rosetta scripts for molecular docking and virtual mutagenesis

The Rosetta script for molecular docking were built using CreateDistanceConstraint mover, HighResDocker mover, InterfaceAnalyzerMover [28], MinMover and FastRelax [29] mover, and AtomicDistance filter to enable all exported docking poses satisfied our preset distance restraint (Table S3). The enzyme monomer structure file was written with its substrate to obtain a combined file, followed using the molecular docking script to search the accurate docking poses (Table S3). Discovery studio Libdock were also carried out to align with our built Rosetta dock script, and the parameters were shown as Table S4.

The virtual mutagenesis script was built using SavePoseMover, InterfaceDdGMover, FastDesign mover, FastRelax mover, MinMover, and ReportToDB mover (Table S3). The docking pose was used as the input binding structure, and its binding score was calculated before mutation. After mutations was introduced into enzyme, the binding structure was refined using MinMover and FastRelax mover, and the obtained binding score was used to calculate the binding ddg (Table S5).

2.8. Molecular dynamics simulation

The structure of FRAPD-TGm2 was modeled by AlphaFold-2 and refined using Rosetta relax [13]. The TGm2 variants carried single mutations were modeled by Rosetta remodel using FRAPD-TGm2 as template [30]. Peptide GGGGQR was built using Rosetta BuildPeptide, and molecular docking was carried out using our self-wrote Rosetta script. Molecular dynamics (MD) simulation on enzyme-substrate complex was performed by Gromacs-2020 (Uppsala University, Uppsala, Sweden) that the complex structures were embedded with FF14sb force field, and solvated in SPC/E water in a cubic box with 1.2 nm to the horizon [31,32]. The system was neutralized and ion charged by Na^+ and Cl^-. The solvated structures were energy minimized through steepest descent step that both short-range van der Waals and short-range electrostatic interactions were truncated at 14 Å. Isochoric-isothermal ensemble (NVT) and isothermal-isovolumetric ensemble (NPT) were carried out under 300 K for 100 and 200 ps, respectively. The simulation was carried out under 330 K for 100 ns.

3. Results

3.1. Developing rosetta script for molecular docking and virtual mutagenesis

To evaluate the performance of our custom-built Rosetta docking script, we compared its results with those obtained using Discovery Studio Libdock [33]. According to the "Theozyme" theory, the catalytic residue and the atom being attacked must be in close proximity to facilitate catalysis, such as a nucleophilic attack [19,20]. We measured the distance between the SG atom in Cys64 of smTG and the C-beta atom in the OAS of the co-crystallized acyl donor, showing it to be 5 Å (Fig. 1). Due to differences between the residues OAS and Gln, using the same distance for the SG atom in Cys64 of smTG and the C-beta atom in Gln of GGGGQR was not feasible, as it could result in strong repulsion forces. Therefore, we applied a more tolerant distance constraint, filtering out any docking poses of TGm2 with GGGGQR where this distance exceeded 6.1 Å.

Using our Rosetta docking script, we generated 100 docking poses, automatically filtering out those that did not meet the distance constraints, ultimately retaining 22 binding complexes (Fig. 1). Discovery Studio Libdock exported 28 docking poses, with Pose1, Pose2, and Pose4 satisfying the distance constraint (Fig. 1). The overall energy of the docking complex was used to represent binding affinity and the accuracy of the binding pose. Over 95 % of the docking complexes achieved using our script displayed lower overall energy than those obtained by Libdock (Fig. 1), indicating a fully refined binding complex with correct enzyme-substrate interaction. The complex structure with the lowest score showed a distance of 4.1 Å between the SG and C-beta atoms (Fig. 1). This structure was further used as input for virtual mutagenesis and ddG calculation.

Previous strategies have highlighted that substrate preference modification can be achieved by introducing novel bonds between enzymes and substrates through mutagenesis [34,35] Therefore, we used energy terms to evaluate local energy changes upon mutation, representing local interaction networks. We designed a Rosetta script to track energy changes upon single mutations by integrating repack and minimization modules. Additionally, InterfaceDdGMover was used to calculate ddG after mutations. Substrate binding site prediction revealed 17 key residues critically affecting acyl donor binding in the catalytic pocket of smTG, including residues 62–65, 75, 252–256, 276–280, and 285 [36]. Considering that neighboring residue mutations can result in the backbone movement of selected residues, we further included 2 residues flanking the predicted binding residues, extending our virtual scan library to residues 60–67, 73–77, 250–258, and 274–287 (Fig. 2). *In silico* saturation mutagenesis on TGm2 was performed using our Rosetta energy scan script, selecting the top 20 single mutations with the lowest binding ddG for experimental validation (Fig. 2).

3.2. Characterizing substrate specific activity of TGm2 variants

The plasmids carrying single mutations on TGm2 were transformed into *E. coli.* The recombinant smTG variants were purified and initially characterized their activity against CBZ-Gln-Gly and the peptide GGGGQR. We measured the specific activity of the secretory expressed active smTG variants, finding that only V252Q, Y278E, and G250H with detectable activity against both substrates. These results suggested that mutations within the catalytic pocket of TGm2 may disrupt enzyme folding, leading to undetected activity in the extracellular fraction. TGm2 variants with single mutations V252Q, Y278E, and G250H were purified from the extracellular fractions (Fig. 3a). To verify the significance by engineering the substrate binding pocket to modify enzyme preferences, we also purified smTG variants including smTG-S23V–Y24N–K294L (TGm) and smTG-S2P–S23V–Y24N–S199A-K294L (TGm1). These two variants have previously been shown to exhibit higher specific activity against CBZ-Gln-Gly and greater thermostability compared to smTG [13] (Fig. 3a–S1).

We compared the specific activity of smTG variants, showing that except for TGm2-Y278E and TGm2-G250H, all other variants displayed an activity ratio for peptide GGGGQR to CBZ-Gln-Gly of approximately 0.05 (Table 1), despite significant differences in specific activity against CBZ-Gln-Gly. TGm2 exhibited the highest activity against CBZ-Gln-Gly

	Pose_id	total_score
Discovery Studio	Pose1	-957.426
	Pose2	-972.789
	Pose4	-952.014
Rosetta dock script	1_0005	-990.596
	1_0003	-1007.58
	1_0019	-998.938
	1_0023	-1012.66
	1_0027	-1006.07
	1_0028	-999.239
	1_0034	-1008.96
	1_0036	-999.877
	1_0040	-997.886
	1_0043	-1005.98
	1_0044	-991.424
	1_0049	-995.29
	1_0051	-994.762
	1_0056	-986.175
	1_0061	-1003.88
	1_0067	-998.3
	1_0073	-912.432
	1_0074	-986.308
	1_0075	-1013.28
	1_0086	-991.015
	1_0091	-1021.28
	1_0093	-986.669

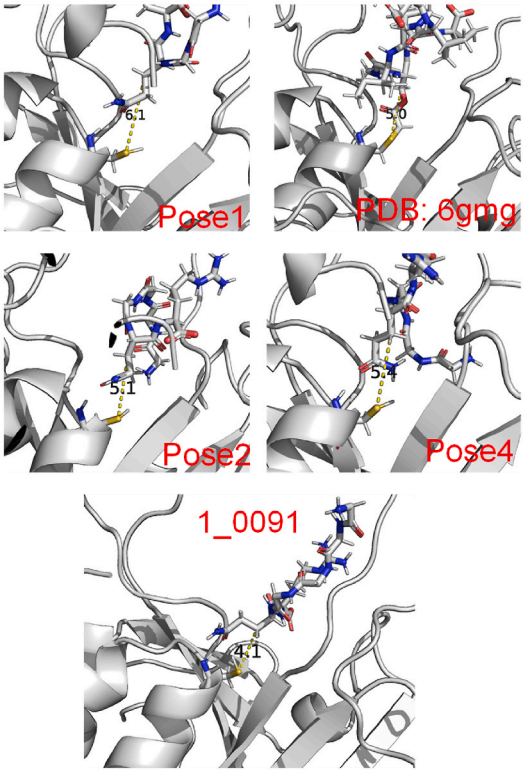

Fig. 1. Analysis of docking poses achieved using Discovery Studio Libdock and Rosetta script
The docking poses were achieved by restricting the distance between the SG atom in Cys64 of TGm2 and the C-beta atom in Gln of GGGGQR do not exceed 6.1 Å. The left table showed the energy of docking complex that satisfied the distance constraints, and the right figures showed the docking complex achieved by crystallization (PDB: 6gmg), Discovery Studio Libdock (Pose1, Pose 2, Pose 4), and our created Rosetta script (1_0091).

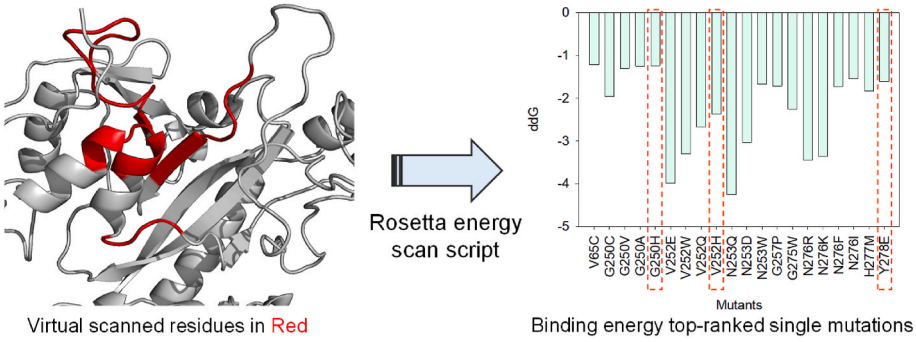

Fig. 2. Prediction of binding ddG using Rosetta script
The left figure showed the residues for virtual mutagenesis (red color), and the right figure showed the binding ddG of single mutations achieved by our created Rosetta script. Noted that the right figure only provided the top 20 ranked single mutations based on the ddG result.

at 68 U/mg, while its activity against peptide GGGGQR was 3.42 U/mg (Table 1). These results confirmed that engineering the substrate binding pocket can enhance its activity towards a specific target. The single mutation Y278E showed much lower activity against CBZ-Gln-Gly at 5.18 U/mg, but its activity against peptide GGGGQR was 4.81 U/mg, which is 41 % higher than TGm2 (Fig. 3b). Notably, the activity ratio for peptide GGGGQR to CBZ-Gln-Gly for TGm2-Y278E reached 93 %, the highest among all variants. To assess the impact of the Y278E mutation, we introduced it into smTG and measured its specific activity. The smTG-Y278E variant showed significantly enhanced activity against GGGGQR, reaching 2.33 U/mg, but reduced activity against CBZ-Gln-Gly at 2.19 U/mg. The activity ratio for GGGGQR versus CBZ-Gln-Gly was 0.94 (Table 1), similar to the trend of TGm2-Y278E, confirming its effect on altering smTG substrate preferences. However, the T_m of TGm2-Y278E was 66.2 °C, 1.6 °C lower than TGm2, suggesting a destabilized effect of Y278E mutation (Fig. S2).

Among these variants, TGm2-G250H exhibited the highest activity against peptide GGGGQR at 7.3 U/mg, which is 1.13-fold higher than TGm2 (Fig. 3b). However, this increase in specific activity against GGGGQR resulted in a 28 % reduction in activity against CBZ-Gln-Gly (Table 1). The activity ratio of TGm2-G250H for GGGGQR versus CBZ-Gln-Gly was 0.17, indicating a continued preference for CBZ-Gln-Gly as the substrate (Fig. 3b).

3.3. *smTG variants induced site-specific labeling*

Nanobody 1C12 naturally contains seven Gln residues. To verify the site-specific labeling ability of modified smTG variants, we generated 1C12 fused with the GGGGQR tag (Fig. 4a). The nanobodies were purified from *E. coli* (Fig. S3), and labeling was performed using the

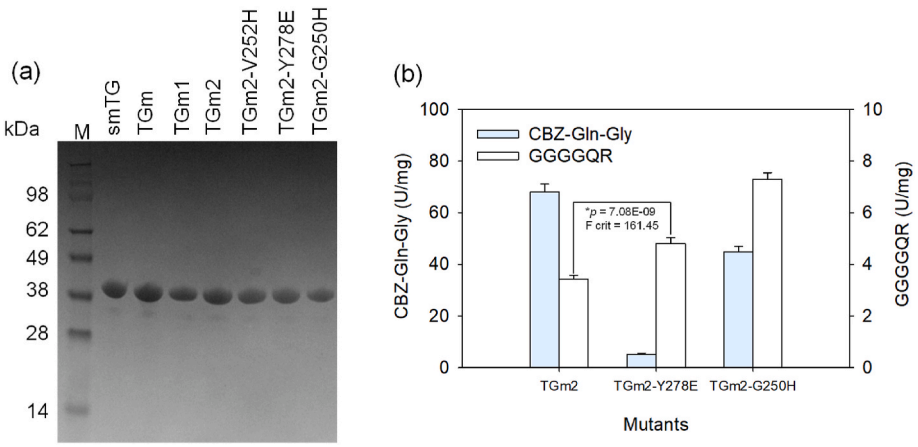

Fig. 3. Substrate preferences validation of single mutations (a) Visualization of purified smTG variants on SDS-PAGE; (b) Specific activity towards substrate CBZ-Gln-Gly and peptide GGGGQR of smTG variants. Statistically significant differences analysis carried out using Excel Variance Analysis that the obtained *p* value comparison between the catalytic activity against GGGGQR of TGm2 and TGm2-Y278E was indicated with an asterisk.

Table 1
Activity against CBZ-Gln-Gly and peptide GGGGQR of smTG variants.

Mutations	Activity (CBZ-Gln-Gly, U/mg)	Activity (GGGGQR, U/mg)	Ratio (GGGGQR vs CBZ-Gln-Gly)
smTG	24.92	1.22	0.05
TGm	36.29	1.93	0.05
TGm1	49.12	2.46	0.05
TGm2	68	3.42	0.05
TGm2-V252H	49.02	2.77	0.06
TGm2-Y278E	5.18	4.81	0.93
TGm2-G250H	44.90	7.30	0.17
smTG-Y278E	2.33	2.19	0.94

*TGm: smTG-S23V–Y24N–K294L.
*TGm1: smTG-S2P–S23V–Y24N–S199A-K294L.
*TGm2: smTG-S2P–S23V–Y24N-E28T-S199A-A265P-A287P–K294L.

synthetic biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl). Initially, we performed non-specific labeling based on 1C12 and fluorophore, evaluating the efficiency of smTG, TGm2, TGm2-G250H, and TGm2-Y278E. After treatment, the samples were fully denatured and subjected to SEC analysis. The negative control without smTG variants displayed two peaks at 200–250 mL, corresponding to the elution of 1C12 (without the fusion tag) and biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl) (Fig. 4b). In contrast, samples treated with smTG, TGm2, and TGm2-G250H showed multiple peaks, while TGm2-Y278E treatment resulted in nonobvious peak before the two primary peaks at 200–250 mL (Fig. 4b). Next, we prepared 1C12-GGGGQR for validating the site-specific labeling efficiency of smTG variants. SEC analysis revealed multiple peaks (peak 1–4) following treatment with smTG, TGm2, and TGm2-G250H, while TGm2-Y278E resulted in an obvious single peak (peak 5) preceding the double peaks at 200–250 mL (Fig. 4c).

To confirm that TGm2-Y278E specifically labels the fusion tag GGGGQR, we analyzed the eluted samples using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). For 1C12-GGGGQR treated with TGm2, peak 1–4 were identified that corresponded to the molecular weight of 1C12-GGGGQR labeled at 4–7 sites with biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl) (Fig. S4). Peak 5 from TGm2-Y278E treatment had a molecular weight matching a single site labeling (Fig. S4). To further confirm that TGm2-Y278E induces site-specific labeling of the GGGGQR tag, we cleaved the sample from peak 5 using trypsin. One of the resulting fragments had a molecular weight consistent with the total weight of FWGQGTLVTVSSGGGGQR and biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl) (Fig. S5). These results indicate that TGm2-Y278E catalyzes a single site-specific labeling event, compared to the multiple labeling sites observed with other smTG variants.

To validate this approach, we conducted a case study using nanobody 7D12 fused with the GGGGQR tag (Fig. S3). 7D12 contains six native Gln residues (Fig. 4d). When mixed with fluorophore and treated with TGm2 and TGm2-Y278E, SEC analysis revealed three peaks prior to the final two peaks at 200–250 mL (Fig. 4e). The molecular weights of these peaks (peak 1–3) corresponded to labeling at 3–5 sites by biotin-PEG(10000)-$(Lys)_3$-Lys(Dansyl) (Fig. S6). In contrast, treatment with TGm2-Y278E again produced a single peak (peak 4) preceding the final two, with the molecular weight consistent with 7D12-GGGGQR labeled by a single fluorophore molecule (Fig. 4e–S6). These results demonstrate that TGm2-Y278E consistently induces single-site labeling for both 1C12-GGGGQR and 7D12-GGGGQR.

3.4. Molecular dynamics simulation

To investigate the mechanism underlying the enhanced substrate preferences, MD simulations were conducted based on the docking complexes of TGm2 vs GGGGQR, TGm2-G250H vs GGGGQR, and TGm2-Y278E vs GGGGQR. RMSD analysis revealed that all enzymes exhibited stabilization after 5 ns of simulation, with TGm2-G250H showing the highest fluctuation [13] (Fig. 5a). Notably, peptide GGGGQR showed the least fluctuation when bound to TGm2-Y278E (Fig. 5b), indicating a tight binding. RMSF analysis indicated that while modifications in substrate preferences did not significantly alter the flexibility of substrate-binding loops (Fig. 5c), certain regions (residues 95–102, 181–185, 207–210, 265–270, 295–297, and the N-terminal FRAPDDP region) showed increased flexibility in TGm2-G250H and TGm2-Y278E (Fig. 5c), though these were outside the catalytic center (Fig. S7). Thus, key binding residues within the catalytic pocket likely play a crucial role in substrate binding and orientation, potentially influenced by overall structural flexibility.

Analysis of the distance between the SG atom of Cys64 in TGm2 variants and the C-beta atom of the Gln residue in GGGGQR throughout the simulation revealed that TGm2-Y278E vs GGGGQR maintained the shortest distance consistently below 6 Å (Fig. 5d), indicating favorable exposure for nucleophilic attack. In contrast, TGm2-G250H vs GGGGQR and TGm2 vs GGGGQR maintained distances around 11 Å and 21 Å, respectively, suggesting less optimal exposure (Fig. 5d).

Inter-molecule interaction analysis [37] indicated that TGm2-G250H vs GGGGQR exhibited the highest binding affinity in the last 10 ns of simulation, while TGm2-Y278E vs GGGGQR showed the lowest

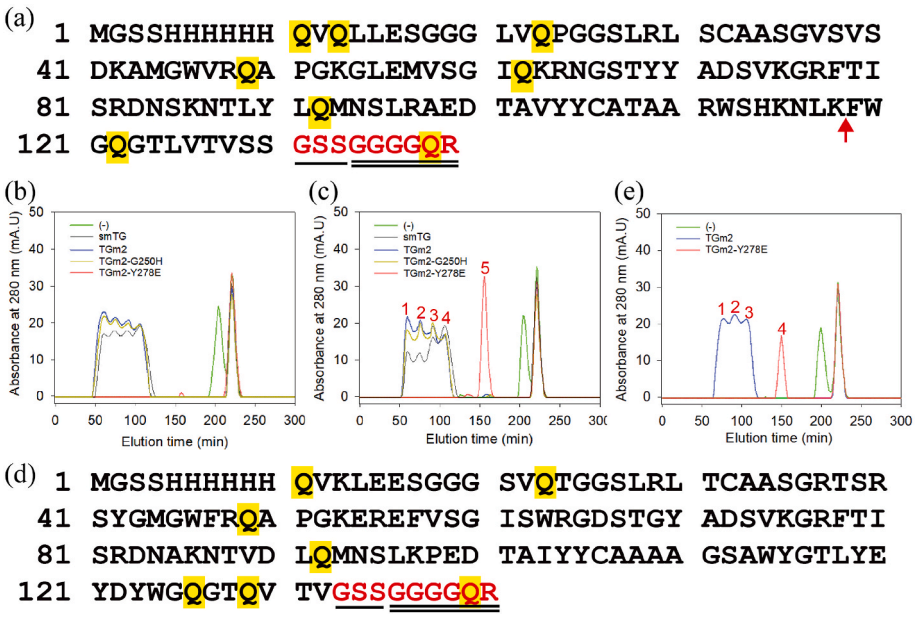

Fig. 4. Characterizing site-specific labelling induced by TGm2 variants
The sequence of nanobody 1C12-GGGGQR (a) and 7D12-GGGGQR (d), the Gln residues were highlighted. The GSS linker (single underline) and GGGGQR tag (double underline) were colored in red. A red arrow was used to indicate the trypsin cleavage site; SEC analysis was conducted on crosslinking induced by smTG variants for (b) 1C12, (c) 1C12-GGGGQR, and (e) 7D12-GGGGQR, with peaks in (c) and (e) numbered. The eluted samples were used for MALDI-TOF MS analysis (shown in Figs. S4, S5, S6). The SEC elution flow rate was set at 1 mL/min.

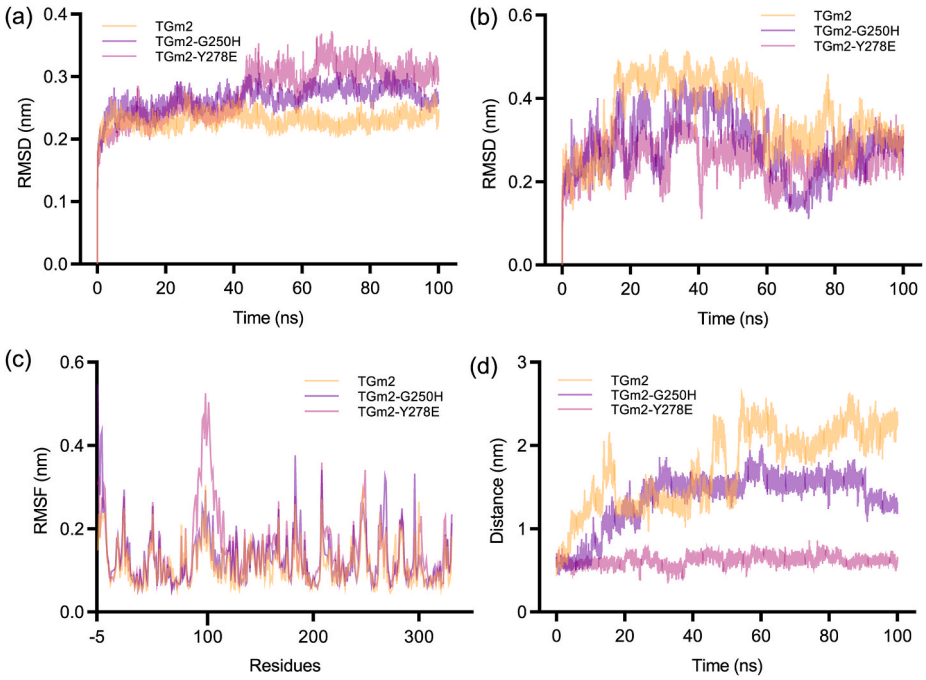

Fig. 5. Molecular dynamics simulation analysis (a) RMSD analysis for enzyme; (b) RMSD analysis for substrate peptide GGGGQR; (c) RMSF analysis; (d) Distance analysis between the SG atom of TGm2 variants' Cys64 and the C-beta atom of Gln residue in GGGGQR. Molecular dynamics was carried out using Gromacs-2020.

(Fig. 6a). Analysis of enzyme residue contributions revealed that key binding residues for TGm2 vs GGGGQR were located outside the acyl donor binding area (Fig. 6b). Cluster analysis highlighted that the top three clusters, which represented the majority of structures, showed the substrate positioned away from the catalytic center for TGm2 (Fig. S8). In contrast, for TGm2-G250H vs GGGGQR and TGm2-Y278E vs GGGGQR, the top three clusters depicted the substrate around the catalytic center (Fig. S8). Residues F251, V252, F254, N276, H277, and L285 were identified as making significant contributions to substrate binding for TGm2-G250H and TGm2-Y278E (Fig. 6c–e).

4. Discussion

In this study, we developed a Rosetta script for high-resolution molecular docking and virtual mutagenesis. Using TGm2 and its unfavorable substrate peptide GGGGQR as targets, we compared our docking

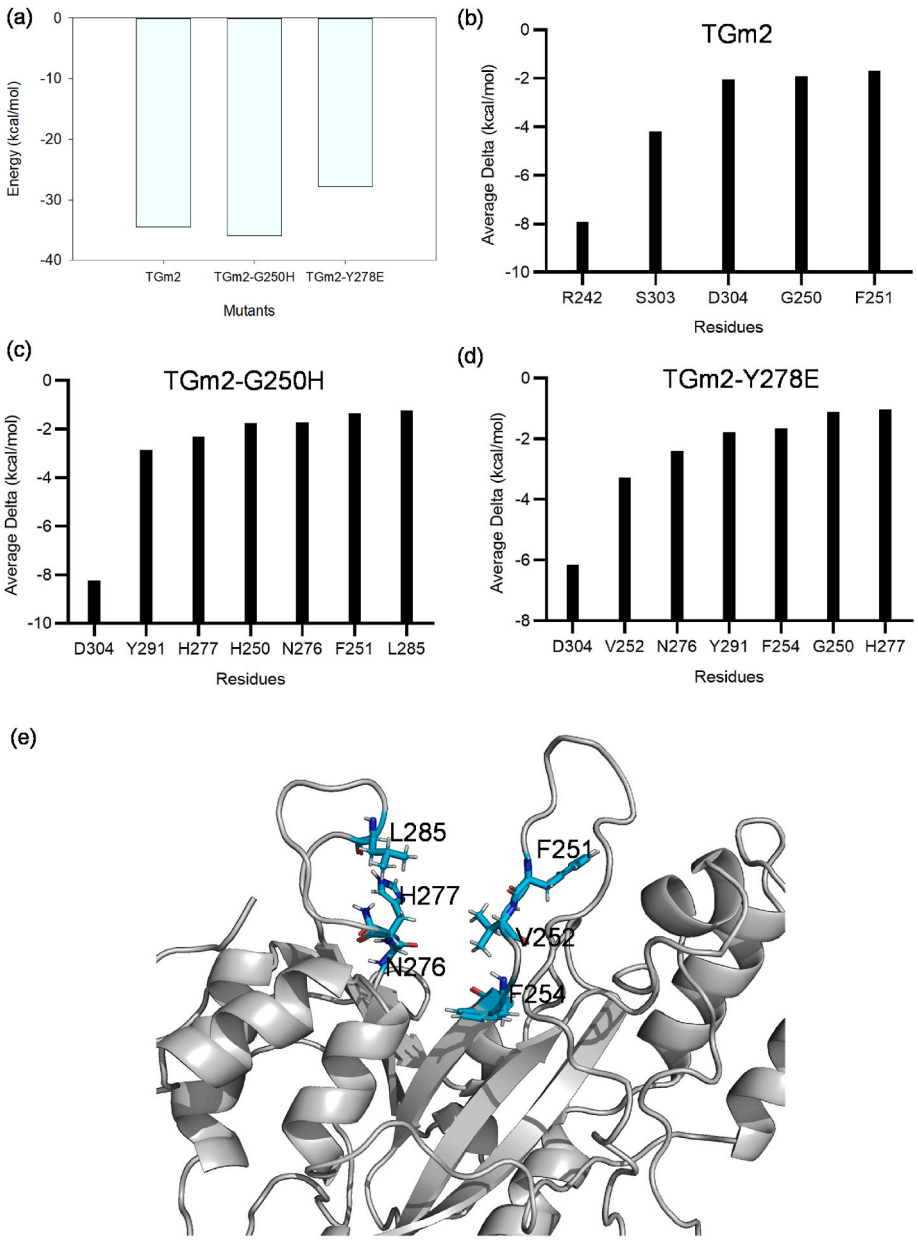

Fig. 6. Enzyme-substrate interaction analysis (a) Overall binding affinity between enzyme and substrate; (b) (c) (d) Residue decomposition for enzyme-substrate binding; (e) Visualization of key residues for enzyme-substrate binding. Binding affinity calculated using gmx_MMPBSA [52] of TGm2 variants with GGGGQR. The end 10 ns of the simulation was used for analysis.

script's performance with Discovery Studio Libdock. Our script produced reliable and refined TGm2 vs GGGGQR complex structures. Based on these docking poses, we conducted virtual mutagenesis on the catalytic pocket residues and experimentally validated the top 20 single mutations. We identified two variants with significantly increased activity against GGGGQR, dramatically altering their substrate preference. These engineered TGm2 variants were used to catalyze NFC, with TGm2-Y278E demonstrating high regioselectivity in labeling nanobodies fused with GGGGQR. MD simulations highlighted the importance of engineering catalytic pocket residues to modify substrate preferences. This work provides a strategy for rationally engineering enzyme substrate specificity.

smTG-induced protein site-labeling was used to engineer antibody-drug and antibody-small molecule conjugations for disease treatment or immunoassays [24,38,39]. Antibody-drug conjugates (ADCs) can be generated through chemical synthesis or enzyme catalysis using phosphopantetheinyl transferase, protein farnesyltransferase, or formylglycine-generating enzyme [40–44]. Chemical synthesis enables a consistent antibody-drug ratio [40,41], but enzyme catalysis is cleaner with fewer unwanted byproducts and reduced energy consumption [45]. However, enzyme substrate preferences can affect the antibody-drug ratio, influencing ADC efficiency or causing toxicity [46]. Thus, achieving steady, site-specific labeling by enzymes remains challenging. Here, we aim to modify the substrate preferences of smTG by engineering the catalytic pocket. The single mutation Y278E enhances the specific activity against GGGGQR for both smTG and TGm2, increasing their preference for GGGGQR over CBZ-Gln-Gly. In contrast, other variants like TGm and TGm1, despite having significantly higher activity against CBZ-Gln-Gly, maintained an activity ratio of 0.05 for GGGGQR to CBZ-Gln-Gly. This confirms the effectiveness of rationally designing the substrate binding pocket to enhance specific activity.

Engineering substrate preferences was hard to achieve, it relied on the researcher's experience and the support of acknowledging the key binding residues [35,47]. Rational design enzyme substrate specificity

focuses on enhancing the binding affinity between enzyme and substrate [13,48,49], since easy to capture substrates may provide more possibilities for catalysis. Generally, analyzing the interaction network between enzyme and substrate is important to find out the key binding residues [50,51]. In this study, we designed molecular docking and virtual mutagenesis methods based on Rosetta script for achieving an accurate enzyme-substrate pose and predicting mutations that can enhance the enzyme-substrate binding affinity [21]. Our result showed that three out of the 20 predicted variants could be purified after fermentation, and two displayed higher activity against its not favored substrate peptide GGGGQR. The single mutation Y278E changed the substrate preferences from 0.05 to 0.93, obviously affected the regioselectivity of smTG [4,7]. However, most variants were not stable and not able to be purified, highlighting that mutating residues within the catalytic center may negatively influence the folding of smTG.

Even though we have used our developed Rosetta scripts for engineering the substrate preference of smTG, but the success rate of 10 % was relative low. Thus, the developed scripts can be further improved, requiring more experimental data to optimize the protocol. Meanwhile, we showed the engineered smTG can be utilized to generate NFC, its further utility for generating ADC, and critical assessment of the generated products needed to be performed to ensure the feasibility.

CRediT authorship contribution statement

Xinglong Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kangjie Xu:** Methodology, Investigation, Data curation. **Haoran Fu:** Methodology, Investigation. **Qiming Chen:** Methodology, Investigation. **Beichen Zhao:** Software, Data curation. **Xinyi Zhao:** Formal analysis, Data curation. **Jingwen Zhou:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Data availability statement

Supplementary Figs. S1–8, Tables S1–S5 are included in the supplementary material.

Declaration of competing interest

The authors declare no competing financial interest.

Acknowledgements

This study was funded by the Natural Science Foundation of Jiangsu Province (BK20202002), the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (Grant No. SN-ZJU-SIAS-0013), China Postdoctoral Science Foundation (2023M741403), Jiangsu Funding Program for Excellent Postdoctoral Talent (2023ZB037), and the National First-class Discipline Program of Light Industry Technology and Engineering (QGJC20230102).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2024.10.003.

References

[1] Kashiwagi T, Yokoyama K-i, Ishikawa K, Ono K, Ejima D, Matsui H, Suzuki E-i. Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. J Biol Chem 2002;277(46):44252–60.
[2] Spolaore B, Raboni S, Satwekar AA, Grigoletto A, Mero A, Montagner IM, Rosato A, Pasut G, Fontana A. Site-specific transglutaminase-mediated conjugation of interferon α-2b at glutamine or lysine residues. Bioconjugate Chem 2016;27(11):2695–706.
[3] Strop P. Versatility of microbial transglutaminase. Bioconjugate Chem 2014;25(5):855–62.
[4] Fiebig D, Schmelz S, Zindel S, Ehret V, Beck J, Ebenig A, Ehret M, Fröls S, Pfeifer F, Kolmar H, Fuchsbauer H-L, Scrima A. Structure of the dispase autolysis-inducing protein from *Streptomyces mobaraensis* and glutamine cross-linking sites for transglutaminase. J Biol Chem 2016;291(39):20417–26.
[5] Juettner NE, Schmelz S, Bogen JP, Happel D, Fessner W-D, Pfeifer F, Fuchsbauer H-L, Scrima A. Illuminating structure and acyl donor sites of a physiological transglutaminase substrate from *Streptomyces mobaraensis*. Protein Sci 2018;27(5):910–22.
[6] Ricart AD. Antibody-drug conjugates of calicheamicin derivative: gemtuzumab ozogamicin and inotuzumab ozogamicin. Clin Cancer Res 2011;17(20):6417–27.
[7] Ohtsuka T, Ota M, Nio N, Motoki M. Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors. Biosci Biotechnol Biochem 2000;64(12):2608–13.
[8] Sugimura Y, Yokoyama K, Nio N, Maki M, Hitomi K. Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. Arch Biochem Biophys 2008;477(2):379–83.
[9] Malešević M, Migge A, Hertel TC, Pietzsch M. A fluorescence-based array screen for transglutaminase substrates. Chembiochem 2015;16(8):1169–74.
[10] Zhao X, Shaw AC, Wang J, Chang C-C, Deng J, Su J. A novel high-throughput screening method for microbial transglutaminases with high specificity toward Gln141 of human growth hormone. J Biomol Screen 2010;15(2):206–12.
[11] Steffen W, Ko FC, Patel J, Lyamichev V, Albert TJ, Benz J, Rudolph MG, Bergmann F, Streidl T, Kratzsch P, Boenitz-Dulat M, Oelschlaegel T, Schraeml M. Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins. J Biol Chem 2017;292(38):15622–35.
[12] Jumper J, Evans R, Pritzel A, Green T, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 2021:1–11.
[13] Wang X, Du J, Zhao B, Wang H, Rao S, Du G, Zhou J, Chen J, Liu S. Significantly improving the thermostability and catalytic efficiency of *Streptomyces mobaraenesis* transglutaminase through combined rational design. J Agric Food Chem 2021;69(50):15268–78.
[14] Wang X, Xu K, Tan Y, Liu S, Zhou J. Possibilities of using de novo design for generating diverse functional food enzymes. Int J Mol Sci 2023;24(4):3827.
[15] Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 2011;79(3):830–8.
[16] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model 2021;61(8):3891–8.
[17] Guterres H, Im W. Improving protein-ligand docking results with high-throughput molecular dynamics simulations. J Chem Inf Model 2020;60(4):2189–98.
[18] Cui J, Feng Y, Yang T, Wang X, Tang H. Computer-aided designing peptide inhibitors of human hematopoietic prostaglandin D2 synthase combined molecular docking and molecular dynamics simulation. Molecules 2023;28(15):5933.
[19] Tantillo DJ, Jiangang C, Houk KN. Theozymes and compuzymes: theoretical models for biological catalysis. Curr Opin Chem Biol 1998;2(6):743–50.
[20] Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. *De novo* enzyme design using Rosetta3. PLoS One 2011;6(5):e19230.
[21] Fleishman SJ, Andrew LF, Corn JE, Eva-Maria S, Khare SD, Nobuyasu K, Justin A, Paul M, Florian R, Gordon L. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One 2011;6(6):e20161.
[22] Wang X, Zhao B, Du J, Xu Y, Zhu X, Zhou J, Rao S, Du G, Chen J, Liu S. Active secretion of a thermostable transglutaminase variant in *Escherichia coli*. Microb. cell fact. 2022;21(1):74.
[23] Mu D, Lu J, Shu C, Li H, Li X, Cai J, Luo S, Yang P, Jiang S, Zheng Z. Improvement of the activity and thermostability of microbial transglutaminase by multiple-site mutagenesis. Biosci Biotechnol Biochem 2018;82(1):106–9.
[24] Shen C, Jin J, Huang Z, Meng M, Lin M, Hu X, Zhu Q, Xu C, Chen W, Lin J, Zhang X, Liu Y, Liu X. Screening and identification of anti-idiotypic nanobody capable of broad-spectrum recognition of the toxin binding region of Lepidopteran cadherins and mimicking domain II of Cry2Aa toxin. J Agric Food Chem 2024;72(3):1582–91.
[25] Schmitz Karl R, Bagchi A, Roovers Rob C, en Henegouwen van Bergen, Paul MP, Ferguson Kathryn M. Structural evaluation of EGFR inhibition mechanisms for nanobodies/VHH domains. Structure 2013;21(7):1214–24.
[26] Liu Y, Lin S, Zhang X, Liu X, Wang J, Lu F. A novel approach for improving the yield of *Bacillus subtilis* transglutaminase in heterologous strains. J Ind Microbiol Biotechnol 2014;41(8):1227–35.
[27] Folk JE, Cole PW. Transglutaminase: mechanistic features of the active site as determined by kinetic and inhibitor studies. Biochim Biophys Acta 1966;122(2):244–64.
[28] Stranges PB, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. Protein Sci 2013;22(1):74–82.
[29] Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D. Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci 2014;23(1):47–55.
[30] Huang Po-Ssu, Yih-En Ban, Richter Andrew, Florian Andre, Ingemar Vernon. RosettaRemodel: a generalized framework for flexible backbone protein design. PLoS One 2011;6. 0024109.

[31] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–219–25.
[32] Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 2011;40(7):843.
[33] Alam S, Khan F. Virtual screening, docking, ADMET and system pharmacology studies on Garcinia caged xanthone derivatives for anticancer activity. Sci Rep 2018;8(1):5524.
[34] Acevedo-Rocha CG, Gamble CG, Lonsdale R, Li A, Nett N, Hoebenreich S, Lingnau JB, Wirtz C, Farés C, Hinrichs H, Deege A, Mulholland AJ, Nov Y, Leys D, McLean KJ, Munro AW, Reetz MT. P450-Catalyzed regio- and diastereoselective steroid hydroxylation: efficient directed evolution enabled by mutability landscaping. ACS Catal 2018;8:3395–410. https://doi.org/10.1021/acscatal.8b00389.
[35] Grobe S, Badenhorst CPS, Bayer T, Hamnevik E, Wu S, Grathwol CW, Link A, Koban S, Brundiek H, Großjohann B, Bornscheuer UT. Engineering regioselectivity of a P450 monooxygenase enables the synthesis of ursodeoxycholic acid via 7β-hydroxylation of lithocholic acid. Angew Chem Int Ed 2021;60(2):753–7.
[36] Wang X, Xu K, Zeng X, Linghu K, Zhao B, Yu S, Wang K, Yu S, Zhao X, Zeng W, Wang K, Zhou J. Machine learning-assisted substrate binding pocket engineering based on structural information. Briefings Bioinf 2024;25(5):bbae381.
[37] Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with gromacs. J Chem Theor Comput 2021;17(10):6281–91.
[38] Zhou L, Li Y, Yang X, Gu H, Duan Y, Fu H, Wang A, Liu K, Gao Y, Song B, Li Y, Jiang Y, Zhang J, Wang C, Wang M, Li Z, Xu Y, Wang C, Wang Y. Effect of prior anticoagulation therapy on stroke severity and in-hospital outcomes in patients with acute ischemic stroke and atrial fibrillation. Int J Cardiol 2023;385:62-70.
[39] El Alaoui M, Sivado E, Jallas A-C, Mebarki L, Dyson MR, Perrez F, Valsesia-Wittmann S, El Alaoui S. Antibody and antibody fragments site-specific conjugation using new Q-tag substrate of bacterial transglutaminase. Cell Death Dis 2024;10(1):79.
[40] Sun MMC, Beam KS, Cerveny CG, Hamblett KJ, Blackmore RS. Reduction-alkylation strategies for the modification of specific monoclonal antibody bisulfides. Bioconjugate Chem 2005;16(5):1282–90.
[41] Mcdonagh CF, Eileen T, Lori W, Webster JB, Alley SC, Kristine K, Jamie A, Ivan S, Hamblett KJ, Francisco JA. Engineered antibody-drug conjugates with defined sites and stoichiometries of drug attachment. Protein Eng Des Sel 2006;19(7):299.
[42] Junutula JR, Bhakta S, Raab H, Ervin KE, Eigenbrot C, Vandlen R, Scheller RH, Lowman HB. Rapid identification of reactive cysteine residues for site-specific labeling of antibody-Fabs. J Immunol Methods 2008;332(1–2):41–52.
[43] Dozier JK, Khatwani SL, Wollack JW, Wang YC, Schmidt-Dannert C, Distefano MD. Engineering protein farnesyltransferase for enzymatic protein labeling applications. Bioconjugate Chem 2014;25(7):1203–12.
[44] York D, Baker J, Holder PG, Jones LC, Drake PM, Barfield RM, Bleck GT, Rabuka D. Generating aldehyde-tagged antibodies with high titers and high formylglycine yields by supplementing culture media with copper(II). BMC Biotechnol 2016;16(1):23.
[45] Agarwal PK. A biophysical perspective on enzyme catalysis. Biochemist 2019;58(6):438–49.
[46] Ricart AD. Antibody-drug conjugates of calicheamicin derivative: gemtuzumab ozogamicin and inotuzumab ozogamicin. Clin Cancer Res 2011;17(20):6417–27.
[47] Polic V, Auclair K. Controlling substrate specificity and product regio- and stereo-selectivities of P450 enzymes without mutagenesis. Bioorg Med Chem 2014;22(20):5547–54.
[48] Akbarzadeh A, Pourzardosht N, Dehnavi E, Ranaei Siadat SO, Zamani MR, Motallebi M, Nikzad Jamnani F, Aghaeepoor M, Barshan Tashnizi M. Disulfide bonds elimination of endoglucanase II from *Trichoderma reesei* by site-directed mutagenesis to improve enzyme activity and thermal stability: an experimental and theoretical approach. Int J Biol Macromol 2018:1201572–80.
[49] Wang W, Su S, Wang S, Ye L, Yu H. Significantly improved catalytic efficiency of caffeic acid O-methyltransferase towards N-acetylserotonin by strengthening its interactions with the unnatural substrate's terminal structure. Enzym Microb Technol 2019:1251–5.
[50] Aggarwal R, Gupta A, Chelur V, Jawahar CV, Priyakumar UD. DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. J Chem Inf Model 2021;62(21):5069–79.
[51] Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Improving detection of protein-ligand binding sites with 3D segmentation. Sci Rep 2020;10(1):5035.
[52] Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with gromacs. J Chem Theor Comput 2021;17(10):6281–91.