



Generative AI-driven enzyme engineering unlocks sustainable valorization of broccoli waste into sulforaphane

Hao Luo^{a,b}, Siyuan Guo^c, Yihong Bao^b, Shan Lu^{b,*}, Fengxia Lv^{a,*}, Ping Zhu^{a,*}

^a College of Food Science and Technology, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

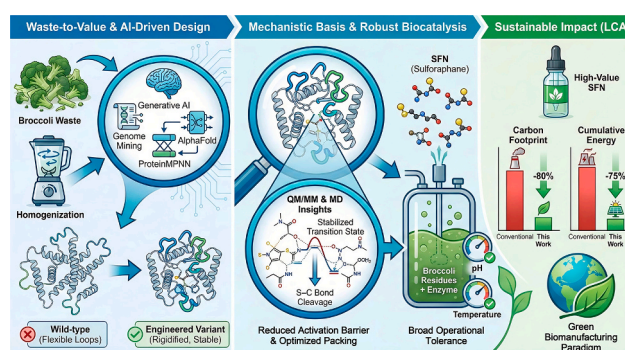
^b Key Laboratory of Saline-Alkali Vegetation Ecology Restoration of the Ministry of Education, College of Life Sciences, Northeast Forestry University, Harbin 150040, Heilongjiang, China

^c College of Connected Computing, Vanderbilt University, Nashville 37240, TN, United States of America

HIGHLIGHTS

- Generative AI pipeline engineered a robust microbial β -thioglucosidase.
- Variant achieved >9-fold catalytic efficiency improvement toward glucoraphanin.
- QM/MM and MD revealed loop rigidification lowers S—C bond cleavage barrier.
- Efficient sulforaphane recovery from solid broccoli waste via enzyme biocatalysis.
- Enzymatic route reduced carbon footprint by >70% versus chemical synthesis.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Generative protein design
Life cycle assessment
Sulforaphane
Waste valorization

ABSTRACT

Current production of sulforaphane (SFN), a health-promoting isothiocyanate, relies mainly on energy-intensive plant extraction or chemical synthesis, limiting sustainable large-scale manufacture. Here, a waste-to-value framework integrating generative artificial intelligence (AI), multi-scale modeling, and life-cycle assessment (LCA) was established to engineer a robust microbial biocatalyst for low-carbon valorization of broccoli waste. Genome mining identified a novel *Lactobacillus*-derived β -thioglucosidase as a promising scaffold. A ProteinMPNN-AlphaFold design pipeline targeting flexible loops near the active site yielded an improved variant with markedly enhanced catalytic performance and thermal robustness, with k_{cat} increasing from 98.5 ± 3.2 to $421.3 \pm 9.1 \text{ min}^{-1}$ and K_m decreasing from 4.6 ± 0.2 to $0.9 \pm 0.1 \text{ mM}$. Molecular dynamics and hybrid quantum mechanics/molecular mechanics (QM/MM) analyses indicated that loop rigidification, improved hydrophobic packing, and a reshaped electrostatic environment promoted transition-state stabilization and reduced the calculated activation barrier for S—C bond cleavage from 19.2 to $17.5 \text{ kcal mol}^{-1}$. When applied to homogenized broccoli residues in a heterogeneous solid-liquid system, the engineered enzyme showed high SFN yields and broad tolerance across pH and temperature conditions. Comparative LCA further revealed ~70% and ~60% reductions in global warming potential and cumulative energy demand, respectively, relative to conventional routes. As the first study to integrate AI-driven protein engineering with rigorous sustainability metrics for SFN

* Corresponding authors.

E-mail addresses: shanlu@nju.edu.cn (S. Lu), lufengxia@njau.edu.cn (F. Lv), pingzhu@njau.edu.cn (P. Zhu).

<https://doi.org/10.1016/j.biortech.2026.134988>

Received 1 January 2026; Received in revised form 25 May 2026; Accepted 25 May 2026

Available online 26 May 2026

0960-8524/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

production, this work offers a scalable paradigm for the green biomanufacturing of high-value natural products from agricultural residues.

1. Introduction

The transition toward low-carbon biomanufacturing is driving the search for sustainable routes to produce high-value bioactive compounds. Sulforaphane (SFN), a potent isothiocyanate with significant chemopreventive and metabolic regulatory properties (Mokhtari et al., 2018; Shawky et al., 2025), represents a critical target for this transition. Despite its growing demand in pharmaceutical and nutraceutical sectors, industrial SFN supply is currently constrained by inefficient production methods. Conventional extraction from plant biomass is hindered by the variable activity of endogenous enzymes, significant mass transfer limitations, and the generation of substantial organic waste (Asif Ali et al., 2023; Sharma et al., 2021). Alternatively, chemical synthesis, while scalable, suffers from poor atom economy and relies on hazardous reagents, creating environmental burdens that contradict the principles of green chemistry. Therefore, establishing a robust, enzymatically driven process for SFN production is essential to overcome these bottlenecks and achieve scalable, sustainable manufacturing.

Enzymatic biotransformation provides a green and selective route for SFN biosynthesis, but its overall efficiency strongly depends on both substrate accessibility and enzyme catalytic performance. Cruciferous agricultural residues, particularly broccoli waste, offer an abundant and renewable reservoir of glucoraphanin, the stable precursor of SFN (Fahey et al., 2015; James et al., 2012). Valorizing these residues aligns with the goals of the circular bioeconomy; however, efficiently unlocking this potential remains a challenge (Ye et al., 2022; Wu et al., 2021). The bioconversion of glucoraphanin into SFN is catalyzed by myrosinase (β -thioglucosidase, EC 3.2.1.147), which hydrolyzes the thioglucosidic bond to release an aglycone intermediate that spontaneously rearranges into SFN (Fahey et al., 2015; Galadova et al., 2022). Plant-derived myrosinases are traditionally regarded as benchmark biocatalysts for this process (Okunade et al., 2018; Sivapalan et al., 2018); however, their practical utilities are often constrained by narrow pH and temperature tolerance, limited turnover efficiency, and poor adaptability in heterogeneous reaction systems. Therefore, the discovery and engineering of robust and efficient myrosinases are essential for establishing a scalable and sustainable route to green SFN biosynthesis directly from native glucoraphanin-rich biomass.

Microbial β -thioglucosidases have recently emerged as promising alternatives to plant enzymes, offering advantages such as higher expression yields and potentially superior stability (Wang et al., 2022; Ye et al., 2022; Youseif et al., 2022). Novel enzymes identified from *Shewanella* and *Lactobacillus* species have demonstrated encouraging hydrolytic capabilities (Li et al., 2022; Ye et al., 2022). Lactiplantibacillus plantarum strains have been identified as natural producers of functional β -thioglucosidases capable of hydrolyzing glucoraphanin to yield sulforaphane and erucin under mild, near-neutral conditions (Lai et al., 2009; Liang et al., 2025; Plaza-Vinuesa et al., 2022). Compared with plant-derived counterparts, microbial β -thioglucosidases generally exhibit a broader temperature operating window while maintaining catalytic activity and long-term stability, making them particularly advantageous for continuous bioreactor operation and biotransformation processes. However, naturally occurring microbial enzymes often lack the specific catalytic efficiency and kinetic robustness required for economically viable processes, particularly when dealing with complex solid-liquid heterogeneous substrates. Furthermore, the molecular mechanisms governing their substrate specificity and transition-state stabilization remain understudied, impeding rational improvement. To bridge the gap between natural enzyme discovery and industrial application, a paradigm shift involving data-driven design, multi-scale mechanistic analysis, and sustainability-oriented evaluation is

required. Recent advances in generative AI have further highlighted the potential of data-driven enzyme design for sustainable biocatalysis (Wen et al., 2025). A key limitation of current SFN production technologies is the lack of robust and engineerable biocatalysts for efficient conversion of heterogeneous broccoli-waste substrates.

In this work, an integrated “molecule-to-process” framework is presented that combines generative artificial intelligence (AI), multi-scale modeling, and life cycle assessment (LCA) to engineer a superior biocatalyst for sustainable SFN production. The novelty of this study lies in addressing this bottleneck through AI-guided enzyme design integrated with mechanistic modeling and sustainability assessment, rather than through process optimization alone. First, a novel *Lactobacillus*-derived β -thioglucosidase scaffold was identified via genome mining. To overcome the limitations of natural variants, a deep learning-based design pipeline (combining ProteinMPNN and AlphaFold) was orchestrated to evolve the enzyme for enhanced catalytic efficiency and thermal robustness. Beyond performance metrics, hybrid quantum mechanics/molecular mechanics (QM/MM) simulations was employed to decode the atomistic basis of the reaction, revealing how loop dynamics and electrostatic preorganization lower the activation barrier. Finally, the practical viability of the engineered enzyme was demonstrated through the efficient conversion of broccoli waste in a heterogeneous system, quantified by an LCA that confirms a drastic reduction in carbon footprint compared to traditional routes. This study establishes a holistic, AI-enabled strategy for converting agricultural residues into high-value products, offering a blueprint for next-generation green biomanufacturing.

2. Materials and methods

2.1. Sequence mining and phylogenetic analysis

Putative β -thioglucosidase (LA) homologs were retrieved from the NCBI non-redundant (nr) and MGnify databases using the myrosinase crystal structure from *Sinapis alba* (Protein Data Bank (PDB) ID: 1MYR) as the query sequence. Homologous protein sequences were identified through BLASTp searches with an E-value threshold of $<1e-10$. Redundant sequences were removed using CD-HIT with a 90% identity cutoff. Multiple sequence alignment was performed with MAFFT v7.505 using the L-INS-i algorithm. A maximum-likelihood phylogenetic tree was constructed in MEGA X based on the LG + G + I substitution model with 1000 bootstrap replicates to assess node reliability. The resulting tree was visualized and annotated using iTOL v6, and sequences were classified into bacterial, fungal, and plant clades to support comparative evolutionary interpretation.

2.2. Structural modeling and conservation analysis

The 3D structure of the *Lactobacillus acidophilus*-derived β -thioglucosidase (LA) was predicted using AlphaFold3 with default template-free modeling settings (Jumper et al., 2021). Structural comparison with the canonical myrosinase from *Sinapis alba* (PDB ID: 1MYR) was performed in PyMOL v2.5, and backbone structural similarity was quantified using root-mean-square deviation (RMSD) calculated by least-squares fitting. Residue-level evolutionary conservation was estimated using ConSurf (Ashkenazy et al., 2016). Homologous sequences were retrieved using HMMER v3.3, generating an initial dataset of 16,645 sequences. Redundancy reduction was conducted with CD-HIT at a 70% sequence identity threshold, resulting in a non-redundant dataset of 4,734 unique sequences. The top 300 representative homologs were aligned, and conservation scores (range 0–9) were computed under a

Bayesian inference framework.

2.3. ProteinMPNN-based sequence design and rational mutagenesis

Residue conservation and local structural confidence within the active-pocket-associated regions (residues 162–177, 372–384, and 418–444) were used as design constraints based on ConSurf analysis and AlphaFold3-derived confidence metrics. Sequence redesign was carried out using the publicly available ProteinMPNN framework with the official developer-released implementation (Dauparas et al., 2022). The official repository, Colab notebook, helper scripts for residue-level constraint specification, and residue-constraint examples used in this study are available at: <https://github.com/dauparas/ProteinMPNN>; https://colab.research.google.com/github/dauparas/ProteinMPNN/blob/main/colab_notebooks/quickdemo.ipynb; https://github.com/dauparas/ProteinMPNN/tree/main/helper_scripts; https://github.com/dauparas/ProteinMPNN/blob/main/examples/submit_example_4.sh; https://github.com/dauparas/ProteinMPNN/blob/main/examples/submit_example_4_non_fixed.sh. More than 800 candidate β -thioglucosidase variants were generated under two strategies: (i) fixing approximately 50% of highly conserved residues, and (ii) restricting mutational exploration to catalytic-pocket residues. All candidates were evaluated using AlphaFold3, and variants with mean predicted Local Distance Difference Test (pLDDT) > 94.7 and backbone RMSD < 0.8 Å were retained. The top 10 variants were selected for expression and biochemical validation.

2.4. Protein expression and purification

The target gene sequences were synthesized by Sangon Biotech (Shanghai, China) and amplified using the AG12202 DNA polymerase (Accurate Biotechnology [Hunan] Co., Ltd., Changsha, China) prior to cloning into the pET-28a(+) expression vector containing an N-terminal His-tag. Expression was performed in *E. coli* BL21(DE3) at 20 °C for 12 h following induction with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cells were lysed by sonication, and proteins were purified using Ni²⁺-NTA affinity chromatography (Qiagen). Purified proteins were dialyzed in 20 mM phosphate buffer (pH 6.0) and analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (>95% purity). Protein concentration was measured using the Bradford assay.

2.5. Enzyme kinetics and stability measurements

Steady-state kinetics were performed using 0.05–20 mM glucoraphanin in 50 mM phosphate buffer (pH 6.0) at 40 °C. Reactions were quenched with 10% methanol, and sulforaphane (SFN) formation was quantified using high-performance liquid chromatography with diode-array detection (HPLC–DAD) (Agilent 1260, 254 nm). Kinetic constants (Michaelis constant, K_m and turnover number, k_{cat}) were derived by nonlinear regression in GraphPad Prism 9. Changes in folding free energy ($\Delta\Delta G$) values for single-point and designed mutations were calculated using FoldX v5.0. The input structure was first preprocessed using the RepairPDB function to correct steric clashes and optimize side-chain geometry. Mutational free energy changes were obtained using the BuildModel command, with five independent replicates performed for each mutation to ensure reproducibility. The final $\Delta\Delta G$ values represent the mean of all replicates, where positive $\Delta\Delta G$ values indicate destabilizing effects and negative values indicate increased stability relative to the wild-type protein (Schymkowitz et al., 2005). All measurements were conducted in triplicate, and data are presented as mean $\pm$ standard deviation (SD).

2.6. Molecular dynamics (MD) simulations

All-atom molecular dynamics simulations were performed for both the wild-type (WT) and the Design 7 mutant–substrate complexes using

GROMACS 2022 and the AMBER99SB-ILDN (Lindorff-Larsen et al., 2010) force field. Each system was solvated in a TIP3P explicit water model within a dodecahedral box with a minimum distance of 10 Å from any protein atom to the box boundary and neutralized with Na⁺/Cl[−] counterions. After energy minimization, systems underwent equilibration consisting of 100 ps constant number, volume, and temperature (NVT) and 100 ps constant number, pressure, and temperature (NPT) phases with position restraints on the protein backbone. Production simulations were run for 100 ns at 310 K using the Nosé–Hoover thermostat and the Parrinello–Rahman barostat. Trajectories were analyzed using standard GROMACS modules to compute RMSD, RMSF, solvent-accessible surface area (SASA), and radius of gyration (Rg). Free energy landscapes (FELs) were constructed based on principal component analysis (PCA) of C α fluctuations, and the resulting distributions were projected using the first two principal components with gmx sham.

2.7. QM/MM reaction mechanism calculations

Quantum mechanics/molecular mechanics (QM/MM) calculations were performed using the ONIOM (Our own *N*-layered Integrated molecular Orbital and Molecular mechanics) approach as implemented in Gaussian 16 (Revision C.01) (Frisch et al., 2016). The two-layer ONIOM (QM:MM) scheme was applied, partitioning the enzyme–substrate complex into a high-accuracy QM layer and a surrounding MM layer.

The QM layer comprised the glucoraphanin substrate and the four key catalytic residues directly involved in the S–C bond cleavage reaction: the catalytic nucleophile (Glu170), the general acid/base catalyst (Glu377), and two residues contributing to substrate binding and transition-state stabilization (Trp422 and Lys436). The QM region was treated at the B3LYP/6-31G(d) level of theory (Becke, 1993; Lee et al., 1988), which has been validated for the mechanistic study of glycoside hydrolases and thioglucosidases (Jafari et al., 2021; Jafari et al., 2024). The remainder of the protein and explicit water molecules within a 25 Å sphere of the active site were treated at the MM level using the AMBER force field as implemented in Gaussian. Covalent bonds crossing the QM/MM boundary were capped with hydrogen link atoms, which were inserted along the original bond vector to maintain the valency of the QM fragment.

Protonation states of the active-site residues were assigned based on the established double-displacement (retaining) mechanism of β -thioglucosidases (Jafari et al., 2021): Glu170 was set in its deprotonated (carboxylate) form to serve as the nucleophile, whereas Glu377 was set in its protonated (carboxylic acid) form to act as the general acid/base catalyst, consistent with the reactant state geometry. All other titratable residues were set to their standard protonation states at pH 6.0.

Starting geometries for the WT and Design 7 systems were derived from the representative snapshots at the production trajectory of the respective 100 ns MD simulations, selecting the lowest-energy frame within the dominant conformational basin of the free energy landscape. Geometric optimizations were performed without constraints at the B3LYP/6-31G(d) level until the maximum gradient fell below 4.5×10^{-4} . The reaction coordinate for S–C bond cleavage was defined as the distance between the anomeric carbon (C1) of glucoraphanin and the sulfur atom of the aglycone moiety. Energy profiles were constructed by scanning this coordinate from the reactant state (RS) to the product state (PS) in 0.1 Å increments, with full geometry optimization at each point. The transition state (TS) was identified as the energy maximum along the profile and confirmed by the absence of imaginary frequencies orthogonal to the reaction coordinate. Activation barriers (E_a) for both WT and Design 7 were reported as the energy difference between the TS and the RS. Representative RS, TS, and PS structures were visualized and analyzed using PyMOL v2.5.

2.8. Enzymatic transformation of broccoli waste

Broccoli tissue waste (outer leaves and stems) was washed, dried,

and ground to <100 µm. Biotransformations were performed with 5 mg·kg⁻¹ enzyme dosage in 50 mM phosphate buffer (pH 6.0) at 40 °C for 8 h under 150 rpm shaking. Control reactions without enzyme and with heat-inactivated enzyme were conducted in parallel under the same conditions. SFN was extracted using dichloromethane and quantified via HPLC. Reaction kinetics were analyzed under varying parameters: time (0–12 h), pH (2–10), temperature (20–80 °C), and particle size (50–300 µm). Experiments were conducted in triplicate.

2.9. Life cycle assessment (LCA) and mass–energy balance

A gate-to-gate life cycle assessment (LCA) was conducted using SimaPro v9.4, following ISO 14040 and ISO 14044 standards. The functional unit (FU) was defined as 1 kg of sulforaphane (SFN) produced. All foreground inventory data were adjusted to this functional unit and modeled using the ecoinvent 3.9 database. In the base-case scenario, broccoli waste was treated as a zero-burden co-product. Environmental impact assessment was performed using the ReCiPe 2016 midpoint (H) method. Key indicators included global warming potential (GWP) and cumulative energy demand (CED) (Pereira et al., 2024). These two indicators were selected as the primary metrics because the present comparison focused on the carbon and energy burdens of three laboratory-scale SFN production routes.

Comparative modeling was carried out for three production routes: (i) enzymatic biotransformation, (ii) plant-derived extraction, and (iii) chemical synthesis. Process parameters and background data for extraction and synthetic routes were obtained from Becker et al. and Aldaghi et al. (Aldaghi et al., 2023; Becker et al., 2023), respectively. A detailed life cycle inventory, including foreground and background flows and the specific ecoinvent processes used where applicable, was compiled for the comparative assessment. A scenario-based sensitivity analysis was performed on the GWP indicator to evaluate the influence of energy source and SFN yield variability. CED, which is dominated by mass-balance-driven inventory flows and scales approximately linearly with yield, was reported as point estimates in the base case. Material and energy flows were visualized using Sankey v5.0 to illustrate mass balance and hotspot contributions along the production chain.

2.10. Statistical analysis

All experimental data were collected in biological triplicates (n = 3) unless otherwise stated. Two-group comparisons were analyzed using unpaired two-tailed Student's *t*-tests, whereas multi-condition experiments were analyzed using one-way or two-way analysis of variance (ANOVA) with Tukey's multiple-comparison test, as appropriate. Differences were considered significant at *p* < 0.05. Data are presented as mean ± standard deviation (SD). Kinetic parameter regression and curve fitting were performed using GraphPad Prism 9 and OriginPro 2023.

Kinetic data describing monotonic time-dependent increases in response were fitted to a one-phase association model using GraphPad Prism (v10). This nonlinear regression model assumes a single exponential process that approaches equilibrium, and is defined as:

$$Y = Y_0 + (\text{Plateau} - Y_0)(1 - e^{-kX})$$

where *Y* represents the measured response, *Y*₀ is the initial value at *X* = 0, Plateau corresponds to the asymptotic maximum response, *k* is the first-order rate constant, and *X* represents time. Model fitting was performed using least-squares minimization, and goodness-of-fit was assessed using *R*², residual distribution analysis, and Akaike Information Criterion (AIC) comparisons.

Peak-shaped response curves, such as temperature- or pH-dependent activity profiles, were fitted using a Gaussian model in GraphPad Prism (v10). The Gaussian function describes a symmetric bell-shaped response centered at the optimal condition and is expressed as:

$$Y = A \cdot \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right)$$

where *A* represents the peak amplitude, *μ* denotes the position of the maximum (optimal condition), *σ* corresponds to the standard deviation reflecting curve width, and *X* represents the experimental variable. Model parameters were estimated using nonlinear least-squares regression, and model suitability was evaluated through residual analysis and AIC-based model selection.

3. Results and discussion

3.1. Discovery of a novel *Lactobacillus*-derived β-thioglucosidase for bioprocess intensification

While plant-derived myrosinases serve as the natural catalysts for glucosinolate hydrolysis (Atwell et al., 2015), their industrial utility is often compromised by poor stability and limited expression yields. Similarly, research into microbial counterparts remains sparse, with few characterized enzymes capable of efficient SFN production (Dmytriv et al., 2025). To overcome these limitations and expand the biocatalytic repertoire, microbial genomic and metagenomic databases was mined for uncharacterized β-thioglucosidase homologs. A candidate enzyme from *Lactobacillus acidophilus* (designated LA) was identified through sequence–function screening. Phylogenetic analysis revealed that LA forms a distinct clade within the *Lactobacillus* family, clearly separated from both plant-derived myrosinases and previously reported bacterial myrosinase-like systems (e.g., from *Lactiplantibacillus plantarum* WCFS1) (Fig. 1A) (Liang et al., 2025; Plaza-Vinuesa et al., 2022). This evolutionary divergence suggests a unique catalytic architecture distinct from known scaffolds.

Multiple sequence alignment (Fig. 1C) further demonstrated that the overall sequence identity between LA and representative plant myrosinases (*Arabidopsis thaliana* TGG1 and *Brassica napus* Myr) was below 30%, indicating substantial divergence at the primary structure level. Nevertheless, several conserved catalytic motifs characteristic of myrosinase-type β-thioglucosidases were retained, including the Glu-Gln-Glu catalytic triad and the glycone-binding pocket residues (Asn-His-Trp). These features suggest that LA preserves the essential catalytic machinery for S–C bond cleavage, despite overall sequence variability. The low global homology but partial motif conservation implies a case of functional convergence rather than strict homology, a phenomenon previously observed in microbial glycosidases evolving from general β-glucosidases toward thioglucosidase specificity in the gut microbiota context (Lozupone et al., 2008; Munoz-Munoz et al., 2017).

To assess structural conservation, a homology model of LA was generated and superimposed on the crystal structure of *Sinapis alba* myrosinase (PDB ID: 1MYR) (Fig. 1B). The overall fold of LA displayed the canonical (β/α)-barrel architecture typical of glycoside hydrolase family 1, confirming its catalytic scaffold is evolutionarily conserved. However, significant deviations were observed in the loops surrounding the substrate-binding pocket, particularly in the 270–300 and 440–470 regions, which correspond to variable domains in microbial glucosidases associated with substrate recognition and microenvironment flexibility (Liu et al., 2019; Rogowski et al., 2014). The altered conformation of these loops in LA may provide a broader substrate channel and facilitate interaction with bulky glucoraphanin-like substrates.

Collectively, these results establish that the *Lactobacillus acidophilus*-derived β-thioglucosidase represents a novel microbial myrosinase with structural conservation of the catalytic core but distinct sequence and loop-region adaptations. This unique combination of conserved enzymatic machinery and divergent architecture suggests potential for operational substrate accessibility and stability compared to plant-derived myrosinases, which are often limited by narrow pH and temperature windows (Ye et al., 2022; Youseif et al., 2022). Thus, LA

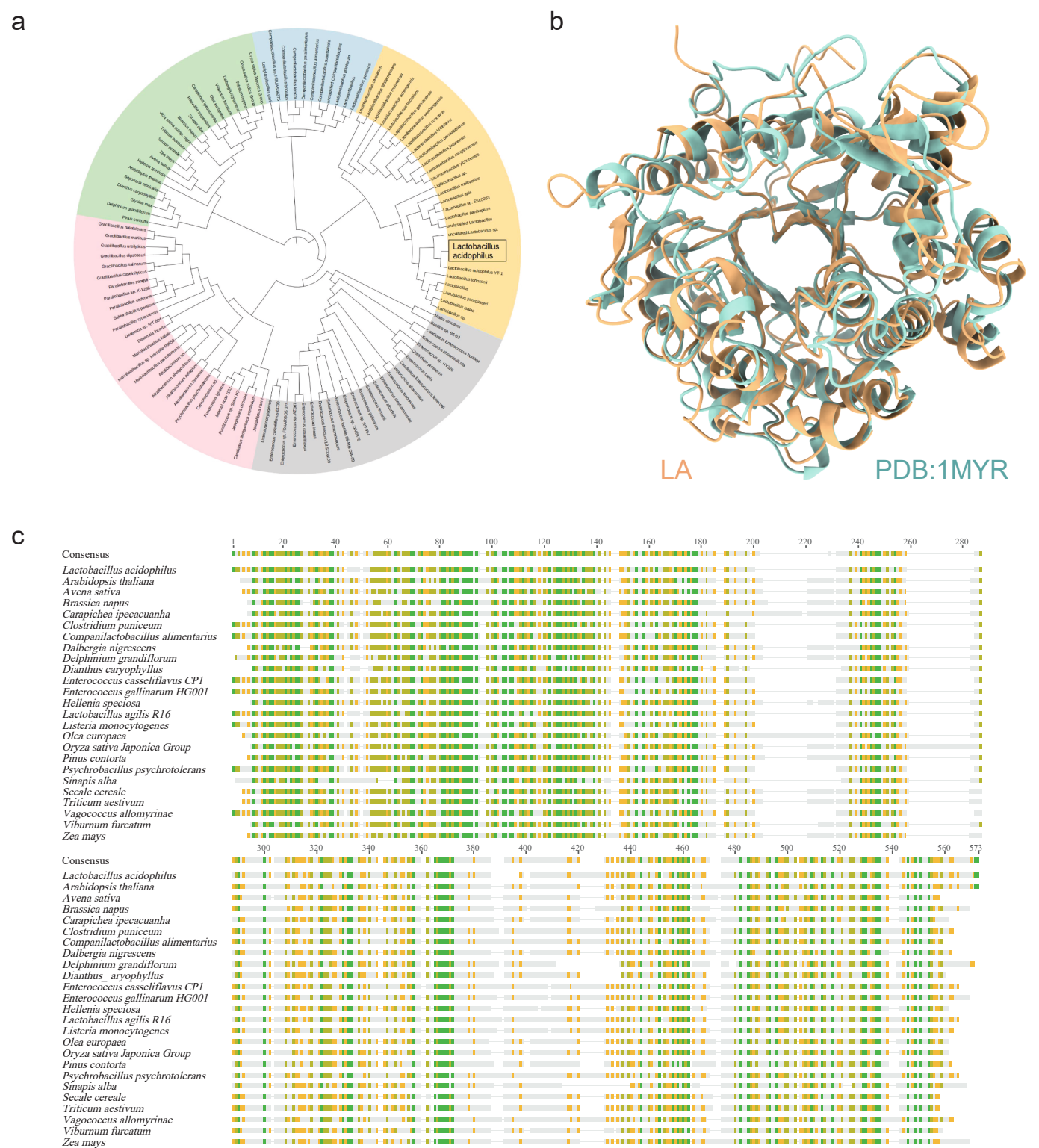


Fig. 1. Discovery and structural characterization of a novel *Lactobacillus acidophilus*-derived β -thioglucosidase. a, Phylogenetic tree analysis of β -thioglucosidases and myrosinase-like enzymes from microbial and plant sources. b, Comparative structural modeling and superposition between LA and Sinapis alba myrosinase (PDB ID: 1MYR). c, Multiple sequence alignment between LA and representative myrosinases from plants and bacteria.

represents a novel catalytic scaffold that combines a conserved reaction center with evolutionary adaptations favorable for bioprocess applications.

3.2. Generative AI-guided design enables catalytic and stability enhancement

To tailor LA for the harsh conditions of industrial processing, a generative design pipeline combining ProteinMPNN and AlphaFold was employed (Fig. 2A). ProteinMPNN, a graph neural network trained on

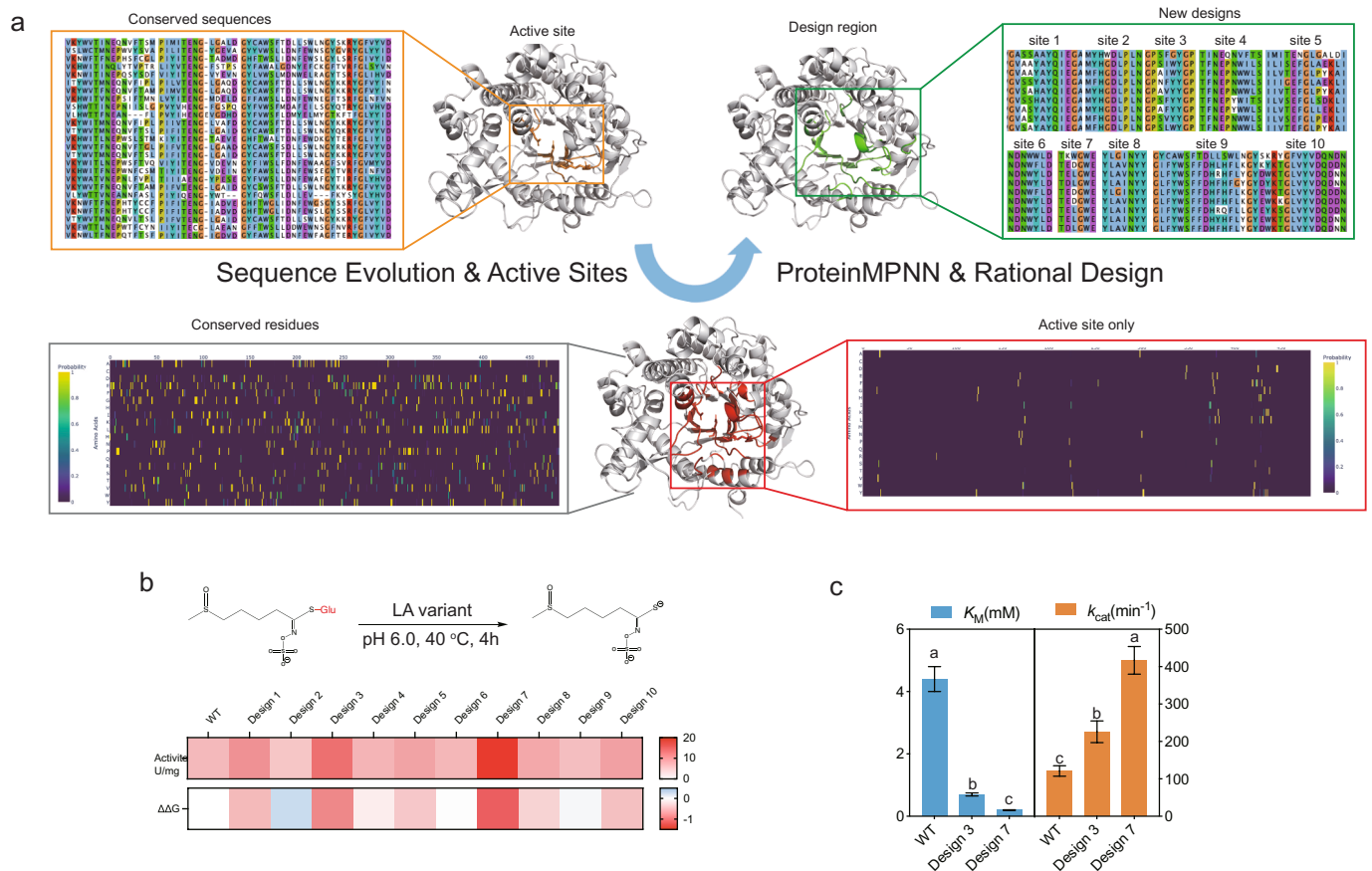


Fig. 2. Design schemes for the optimization of *Lactobacillus*-derived β -thioglucosidase. a, Integrated workflow for AI-assisted and structure-guided design of *Lactobacillus acidophilus* β -thioglucosidase (LA). Sequence conservation analysis identified flexible and catalytically relevant residues near the active site. ProteinMPNN sampling and rational mutagenesis generated 10 mutation regions with locally optimized sequence probabilities and structural compatibility. Heatmaps depict amino acid sampling probabilities for conserved sequence (left) and the active site (right), highlighting targeted design regions. b, Screening of 10 designed variants for catalytic activity and folding stability. c, Steady-state kinetic parameters (K_M and k_{cat}) for WT and representative variants (Designs 3 and 7).

millions of protein backbones, enables the efficient generation of sequence variants that remain compatible with three-dimensional structural constraints (Dauparas et al., 2022; Sumida et al., 2024). Based on multiple sequence alignment and conservation analysis, three flexible regions surrounding the active pocket (residues 162–177, 372–384, and 418–444) were identified (Fig. 2A, left). These regions exhibited low evolutionary conservation but spatial proximity to the catalytic cavity, making them ideal for adaptive mutagenesis. The residue probability heatmap revealed high tolerance in these loop regions, consistent with natural variability observed among microbial β -thioglucosidases.

To balance catalytic fidelity and adaptability, a dual-layer constraint strategy was introduced by fixing the top 50% conserved residues and all residues within 10 Å of the catalytic center (Ribeiro et al., 2020; Wang et al., 2012). Using this setup, ProteinMPNN generated over 600 sequences under conserved constraints and more than 200 under active-site constraints (Fig. 2A, middle). All designed variants were evaluated by AlphaFold 3 yielding high structural confidence (pLDDT > 94.7) and low deviation from the wild-type enzyme (RMSD < 0.8 Å).

Among all the variants from 10 mutation sites (Fig. 2A, right), 10 designs exhibited the most favorable probability distributions for local backbone compatibility (Fig. 2B). Notably, Design 7 emerged as the most promising variant, featuring Gly226Leu, Gly228Leu, and Lys436Trp substitutions. Structural model revealed that these mutations reinforced the active site loop, which could lower entropic penalties and accelerate glycosidic bond cleavage (Chong et al., 2024; Johnson & Holyoak, 2010; Pegan et al., 2013). Experimental validation confirmed

the computationally predicted improvements in catalytic performance and stability (Fig. 2B–C). Among the ten designed variants, Design 7 exhibited the highest catalytic activity and lowest folding free energy change (Fig. 2B). Kinetic assays further revealed that Design 7 achieved a k_{cat} of $421.3 \pm 9.1 \text{ min}^{-1}$, compared with $98.5 \pm 3.2 \text{ min}^{-1}$ for the wild-type enzyme, while its K_M decreased from $4.6 \pm 0.2 \text{ mM}$ to $0.9 \pm 0.1 \text{ mM}$ (Fig. 2C). This indicates improved catalytic efficiency (k_{cat}/K_M), reflecting improved substrate affinity and turnover rate. These results suggest that the present AI-guided redesign strategy improves catalytic performance beyond the conventional optimization of broccoli processing or exogenous myrosinase supplementation, thereby extending SFN-related research from process-level enhancement to enzyme-level engineering (Wu et al., 2021; Zhang et al., 2022). Together, these findings demonstrate that integrating generative AI with physics-based constraints can effectively unlock “hidden” enzyme performance, yielding a robust biocatalyst suitable for scalable manufacturing.

3.3. Structural dynamics of enzymatic performance leap

To uncover the molecular mechanism driving the enzymatic performance leap, comparative molecular dynamics (MD) simulations were performed (Fig. 3a–f). The root-mean-square deviation (RMSD) analysis revealed that the mutant maintained higher conformational stability, equilibrating around 25 ns compared to around 70 ns for WT (Fig. 3a). This reduced fluctuation indicates overall structural rigidification induced by the introduced mutations. Similarly, root-mean-square fluctuation (RMSF) profiles exhibited lower flexibility for the mutant,

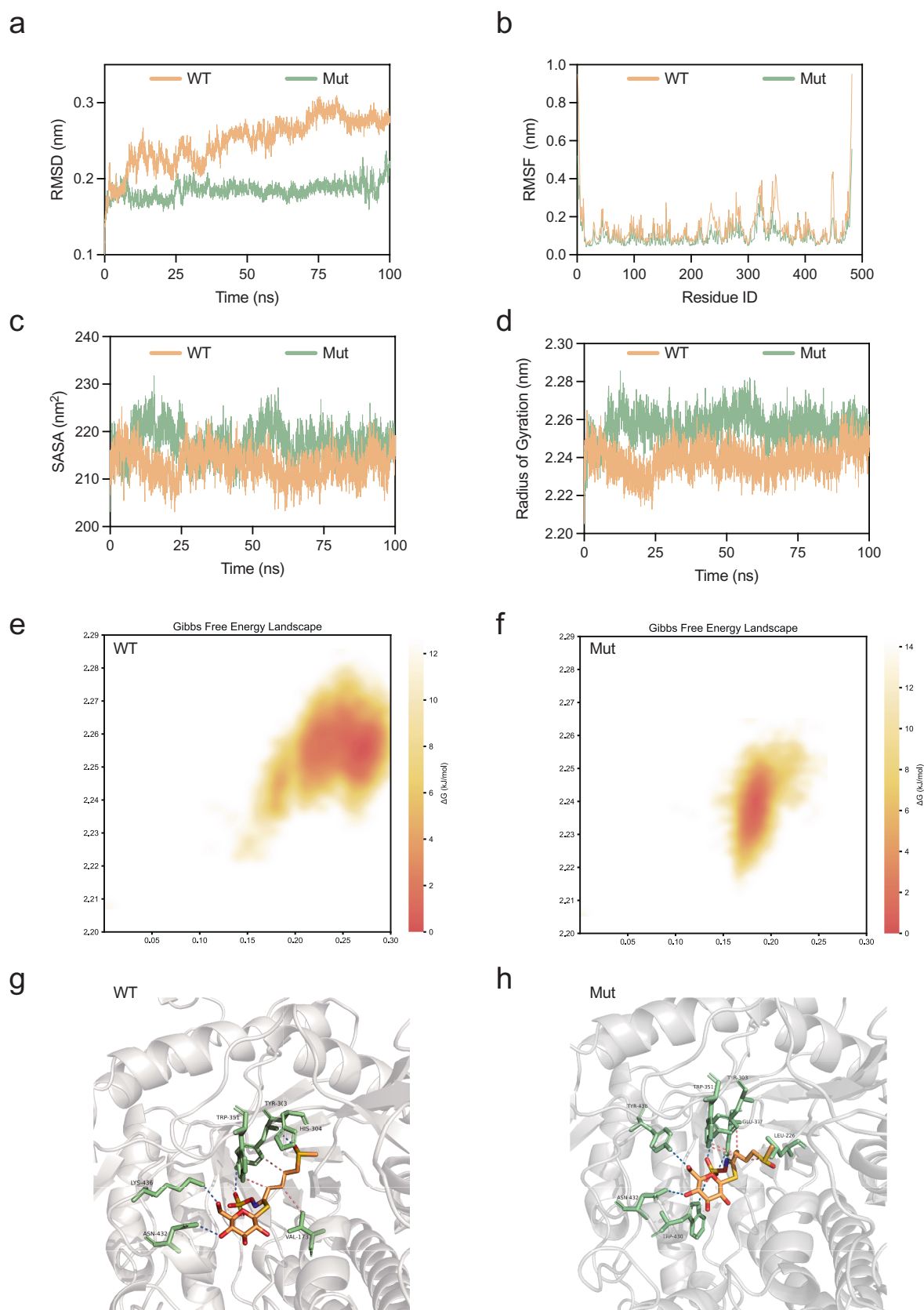


Fig. 3. Molecular dynamics analysis and substrate binding of the engineered LA variant. a, Root-mean-square deviation (RMSD) plots of Design 7 mutant and wild-type (WT) enzyme over 100 ns MD simulation. b, Root-mean-square fluctuation (RMSF) profiles of Design 7 mutant and wild-type (WT) enzyme. c, Solvent-accessible surface area (SASA) analysis of Design 7 mutant and wild-type (WT) enzyme. d, Radius of gyration (Rg) plots of Design 7 mutant and wild-type (WT) enzyme. e, f, Gibbs free energy landscapes of Design 7 mutant and wild-type (WT) enzyme. g, h, Representative binding conformations of glucoraphanin within the WT and Design 7 catalytic pockets. Representative hydrogen-bonding and hydrophobic interactions are indicated by blue and red dashed lines, respectively.

particularly in loop regions near residues 420–460 that form the catalytic cleft (Fig. 3b). The dampened local motions suggest strengthened packing interactions around the active site, stabilizing substrate positioning and catalytic geometry. This phenomenon agrees with reports that loop stabilization can reduce conformational entropy and enhance transition-state preorganization in glycosidases (Quirke & Crich, 2021; Walker et al., 2024).

In addition, the solvent-accessible surface area (SASA) and radius of gyration (R_g) analyses confirmed that the mutant enzyme retained a compact yet accessible catalytic architecture (Fig. 3c–d). The slightly increased SASA ($\sim 222 \text{ nm}^2$ vs. 214 nm^2 in WT) implies improved solvent exposure of the active pocket, facilitating glucoraphanin binding, while the reduced R_g fluctuation reflects enhanced structural compactness. These traits align with the features of thermostable β -glycosidases (Chen et al., 2021; Mohsin et al., 2019). The Gibbs free energy landscapes further supported this observation (Fig. 3e–f). The WT enzyme displayed a broad, shallow basin, while the mutant exhibited a deeper, narrower energy minimum, indicative of improved conformational stability and catalytic efficiency.

To further probe enzyme–substrate interactions, the active-site pocket of the WT and Design 7 mutant was analyzed (Fig. 3g–h). In the WT complex, glucoraphanin was mainly surrounded by Tyr303, His304, Trp351, Lys436, Asn432, and Val173, whereas in Design 7 the substrate was contacted by Tyr303, Trp351, Glu377, Asn432, Trp430, Tyr438, and Leu226. Compared with WT, Design 7 formed additional

hydrogen bonds and hydrophobic contacts, resulting in a denser and more cohesive interaction network. These strengthened interactions reduced substrate mobility and oriented the glucosinolate moiety more favorably toward the catalytic Glu residue, thereby facilitating S–C bond cleavage and transition-state stabilization. Collectively, the structural comparison and MD analysis suggest that the AI-guided mutations enhanced local substrate packing and catalytic preorganization in Design 7, consistent with the experimentally observed increase in k_{cat} and decrease in K_m .

3.4. QM/MM elucidation of transition-state stabilization

To gain atomic-level insights into the catalytic mechanism responsible for the enhanced activity of the engineered LA mutant, QM/MM (quantum mechanics/molecular mechanics) simulations were performed to characterize the reaction coordinate of glucoraphanin hydrolysis. The catalytic process was partitioned into three major states: reactant state (RS), transition state (TS), and product state (PS) (Fig. 4a). In the WT enzyme, the reaction barrier was calculated to be $19.2 \text{ kcal mol}^{-1}$, which is consistent with reported values for classical plant myrosinases ($19\text{--}21 \text{ kcal mol}^{-1}$ (Botti et al., 1995; Jafari et al., 2024; Jafari et al., 2021)). In contrast, the mutant (Design 7) displayed a significantly reduced activation barrier of $17.5 \text{ kcal mol}^{-1}$, accompanied by a more stable product state ($\Delta E = -12.1 \text{ kcal mol}^{-1}$). This about $1.7 \text{ kcal mol}^{-1}$ reduction in the barrier aligns with the experimentally

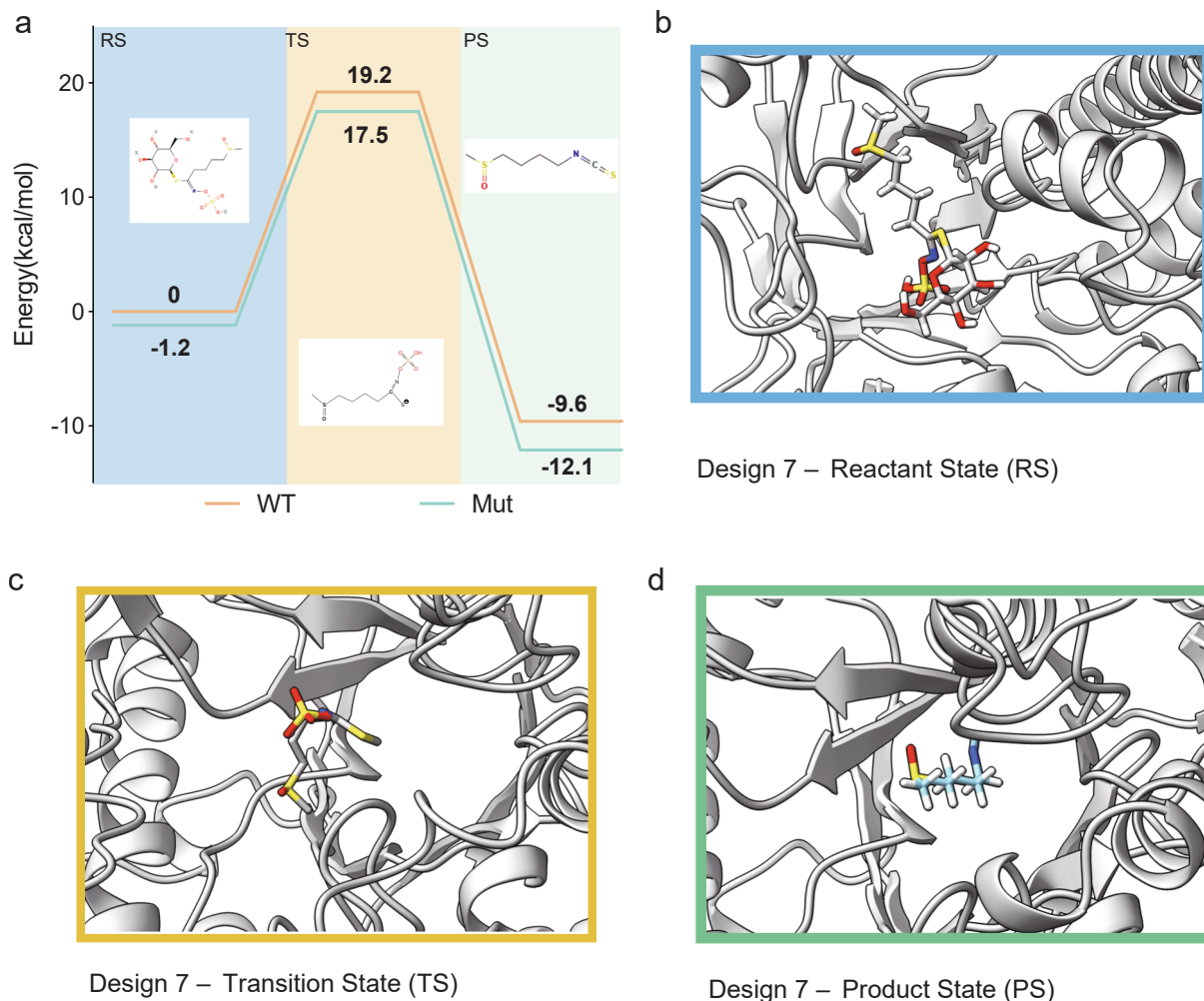


Fig. 4. QM/MM analysis of the reaction energy landscape for glucoraphanin hydrolysis catalyzed by LA variants. a, Potential energy profile comparing the WT (orange) and Design 7 mutant (green) along the reaction coordinate. Insets show representative reactant (RS), transition (TS), and product (PS) states. b,c,d. Representative QM/MM snapshots of the catalytic pocket at the RS (blue frame), TS (orange frame), and PS (green frame), respectively, for the Design 7 system.

observed enhancement in catalytic efficiency (k_{cat}/K_m), and is therefore interpreted as mechanistic support for improved transition-state stabilization rather than as a stand-alone quantitative measure. Fig. 4b-d show the representative QM/MM-optimized RS, TS, and PS structures of the Design 7 system. Structural inspection of these QM/MM-optimized states revealed that the thioglucosidic S—C bond cleavage in Design 7 was associated with improved positioning of neighboring residues, particularly Tyr303 and Glu377, leading to slightly stronger electrostatic complementarity around the catalytic center. Comparable electrostatic fine-tuning effects have been reported to accelerate transition-state formation in engineered β -glucosidases (Namchuk & Withers, 1995; Vega et al., 2025).

Overall, the QM/MM results indicate that Design 7 enhances glucoraphanin hydrolysis through a combination of optimized electrostatic

stabilization and improved substrate orientation. The QM/MM results therefore support the experimental kinetic data by showing that the structural rearrangements in Design 7 were associated with a lower calculated activation barrier for S—C bond cleavage.

3.5. High-efficiency valorization of broccoli waste in heterogeneous systems

Moving from molecular design to process application, the enzyme's performance in converting homogenized broccoli waste, a complex, heterogeneous substrate was evaluated. The ground and sieved residues exhibited fine particulate structures (Fig. 5a), enhancing substrate accessibility for enzymatic hydrolysis. Time-course analysis showed that the mutant enzyme displayed substantially higher SFN release rates than

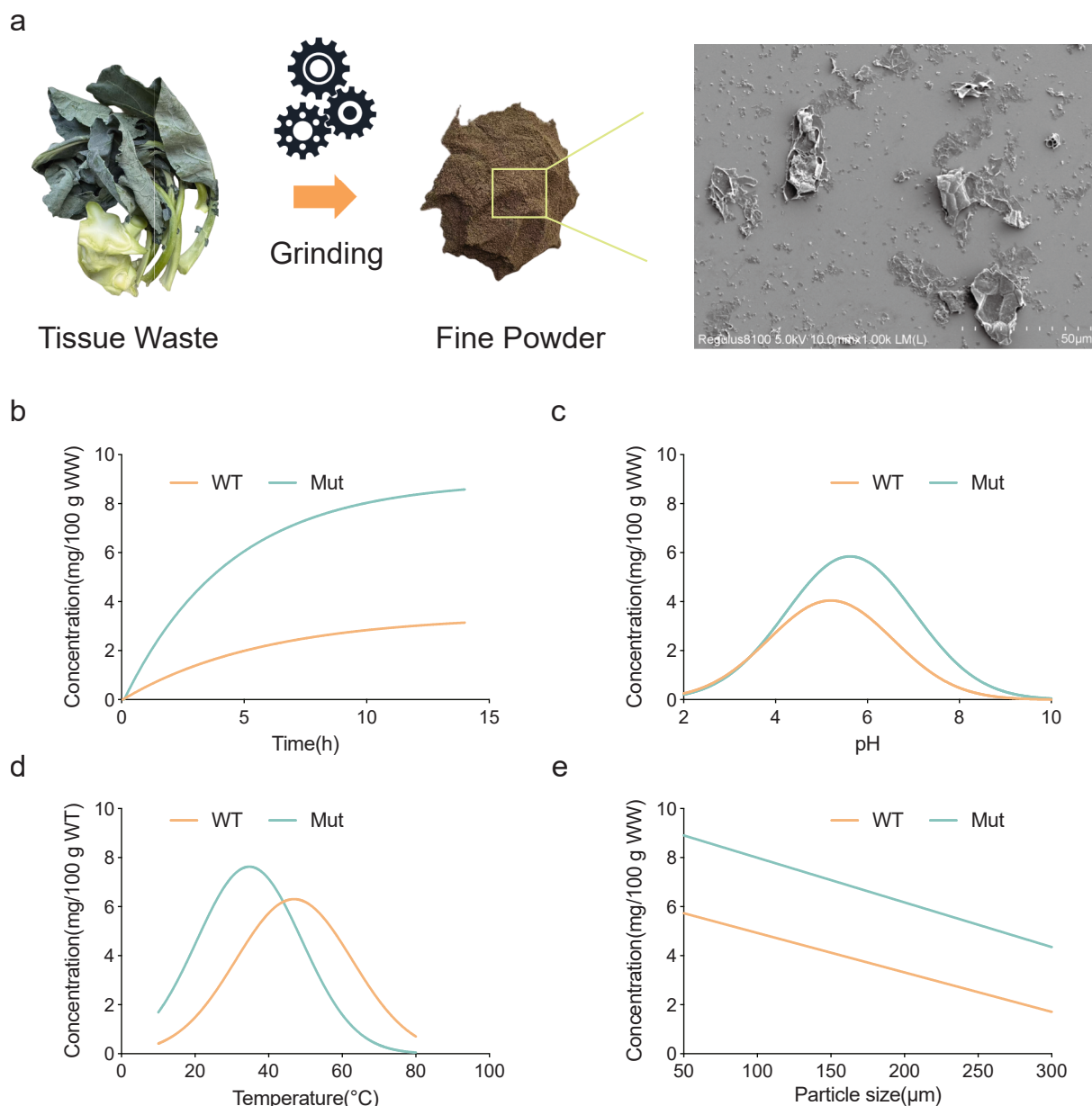


Fig. 5. Enzymatic transformation of sulforaphane from broccoli waste. a, Preparation of the broccoli tissue substrate. Broccoli waste was ground into a fine powder to increase surface area and substrate accessibility, as confirmed by scanning electron microscopy (SEM, right). b, Time-course profiles of sulforaphane (SFN) formation catalyzed by wild-type (WT) and Design 7 mutant using homogenized broccoli tissue. c, pH-dependent activity curves of SFN formation catalyzed by WT and Design 7 mutant. d, Temperature-dependent activity profiles of SFN formation catalyzed by WT and Design 7 mutant. e, Effect of plant tissue particle size of SFN formation catalyzed by WT and Design 7 mutant. Data are shown as mean $\pm$ SD from three independent replicates. Curves in b–e were fitted to the experimental data; statistical analysis for multi-condition experiments is described in Section 2.10.

the WT under identical conditions (Fig. 5b). After 12 h of reaction, SFN yield reached 8.7 mg per 100 g wet weight (WW) for the mutant, corresponding to an increase in catalytic conversion. This observation is consistent with the growing view that broccoli by-products are an underutilized yet bioactive-rich resource, and further highlights the value of integrating biocatalyst engineering with residue valorization and sustainability assessment (Wang et al., 2025; Sharma et al., 2021). This improvement aligns with previous findings that engineered or microbial myrosinases exhibit enhanced activity in solid-state matrices due to improved substrate diffusion and reduced product inhibition (Hao et al., 2024).

pH optimization experiments indicated that both enzymes exhibited bell-shaped activity profiles centered around weakly acidic conditions (Fig. 5c). However, the mutant displayed a broader and slightly shifted optimum from pH 5.5 to 6.5, whereas WT activity dropped sharply beyond pH 6. These results demonstrate improved catalytic robustness under near-neutral conditions, which are typical in plant tissue matrices. Similar pH broadening has been observed in glycosidases with optimized electrostatic surface distributions (Xia et al., 2016; Zhang et al., 2020). Temperature-dependent assays (Fig. 5d) revealed that the mutant retained higher activity across a wider thermal range (20–55 °C) with an optimum around 40 °C, while WT peaked near 50 °C. This shift toward moderate temperatures may reflect improved protein packing and reduced unfolding rate, consistent with molecular dynamics predictions of greater conformational rigidity in Design 7 and with previous studies linking conformational dynamics to enzyme stability and catalytic performance (Chong et al., 2024; Walker et al., 2024).

The influence of particle size on SFN release was further assessed (Fig. 5e). Both enzymes exhibited decreasing activity with increasing tissue particle size, due to diffusion and accessibility limitations (Li et al., 2020). Nevertheless, the mutant consistently outperformed WT across all particle ranges, achieving 9.0 mg SFN / 100 g WW at 50 μ m. Collectively, these results demonstrate that the AI-powered LA mutant possesses superior catalytic turnover, broader operational stability, and improved substrate penetration in complex biomass systems, thereby alleviating the physical barriers imposed by plant tissue recalcitrance (Li et al., 2020; Rogowski et al., 2014).

3.6. Life cycle assessment and sustainability metrics

To quantify the environmental benefits of the present bio-route, a comprehensive mass-energy balance and LCA was conducted (Fig. 6a–b). The mass and energy flow distribution (Fig. 6a) was established for a biotransformation unit processing 100 kg·h⁻¹ of broccoli waste with an enzyme dosage of 5 mg·kg⁻¹. The enzymatic reaction consumed approximately 150 kJ·h⁻¹ of energy (mainly for mixing and

temperature control), producing 0.009 kg·h⁻¹ of SFN and 99.49 kg·h⁻¹ of residual biomass, which can be composted to generate a -20 kJ·h⁻¹ energy credit. This integrated conversion–recycling model demonstrates nearly zero waste generation and a mass conversion efficiency of ~0.009%, consistent with typical biocatalytic transformation efficiencies in lignocellulosic or glucosinolate valorization systems (Picart et al., 2015).

The LCA analysis compared three SFN production routes: enzymatic biotransformation by LA mutant (this work), plant extraction, and chemical synthesis (Fig. 6b). The results were expressed in terms of global warming potential (GWP, kg CO₂-eq·g⁻¹ SFN) and cumulative energy demand (CED, MJ·g⁻¹ SFN). The enzymatic route exhibited a GWP of ~8.3 kg CO₂-eq·g⁻¹ SFN and a CED of ~20 MJ·g⁻¹ SFN, representing ~70% and ~60% reductions, respectively, compared to plant extraction and >80% reduction relative to chemical synthesis. These results indicate lower carbon and energy burdens for the enzymatic route under the present system boundary, although other environmental midpoint categories were not assessed in this study. These improvements primarily stem from mild reaction conditions, low solvent usage, and valorization of waste biomass as compost, reducing both energy demand and emissions (Piao et al., 2026). Such LCA profiles are in line with recent analyses of enzyme-based biotransformations showing significant carbon footprint reduction compared to petrochemical synthesis routes (Goldaraz-Salamero et al., 2025).

Overall, these results confirm that the AI-guided engineering of LA can address a catalytic bottleneck while supporting a promising low-carbon route for the production of high-value phytochemicals. The present study has several limitations, including its laboratory-scale scope, the lack of systematic evaluation of enzyme lifetime under repeated operation, the variability of broccoli-waste feedstocks, uncertainties associated with the current modeling assumptions and system boundaries, and the limited environmental scope of the current LCA. It is explicitly acknowledged that the present LCA relied on scenario-based sensitivity analysis applied primarily to the GWP indicator, rather than full probabilistic uncertainty quantification across all impact categories. The reported scenario ranges therefore provide a first-order indication of the sensitivity of GWP to energy-source and yield variability, but do not constitute statistical uncertainty intervals for either GWP or CED. A comprehensive uncertainty assessment with parameter probability distributions for both indicators is identified as a priority for future work, particularly when pilot- or industrial-scale inventory data become available. Nevertheless, the results support the practical relevance of integrating enzyme engineering with residue valorization for lower-carbon biomanufacturing. From a process perspective, further translation of this system will depend on enzyme production cost, the feasibility of enzyme reuse or immobilization,

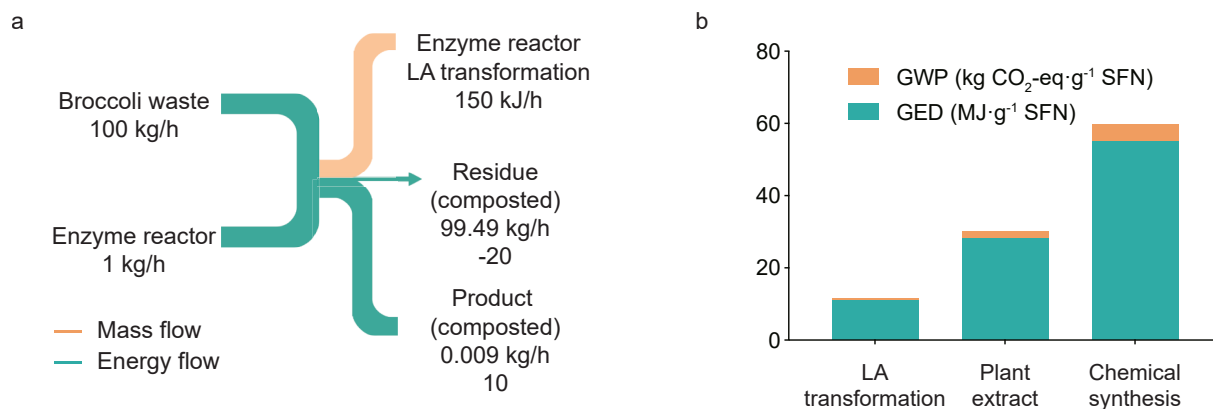


Fig. 6. Sustainability evaluation of the enzymatic biotransformation process. a, Mass energy flow Sankey diagram for the enzymatic conversion of broccoli waste into sulforaphane (SFN) using the engineered *Lactobacillus* β -thioglucosidase (LA). b, Life cycle assessment (LCA) comparison of three SFN production routes: LA enzymatic transformation, plant extraction, and chemical synthesis, based on global warming potential (GWP) and cumulative energy demand (CED).

suitable reactor configurations for heterogeneous biomass conversion, and efficient downstream SFN recovery. These aspects will also benefit from future AI-enabled process development and optimization at the industrial level (Christopher Selvam et al., 2026). Future work should include broader environmental assessment, process optimization, and scale-up evaluation.

4. Conclusions

In this study, an integrated framework merging generative AI-driven enzyme engineering, multi-scale mechanistic modeling, and life cycle assessment (LCA) was successfully established to unlock the sustainable valorization of broccoli waste into high-value sulforaphane (SFN). Addressing the limitations of traditional extraction and synthesis, a phylogenetically distinct *Lactobacillus*-derived β -thioglucosidase via genome mining was first identified. By orchestrating a ProteinMPNN-AlphaFold design pipeline with dual-layer structural constraints, a robust variant that overcame the catalytic and thermal bottlenecks inherent was engineered to the wild-type scaffold. The atomistic basis for this performance leap was rigorously elucidated through hybrid QM/MM and MD simulations. It was demonstrated that the introduced mutations induce loop rigidification and optimize hydrophobic packing, which in turn reshapes the electrostatic environment to stabilize the transition state, lowering the activation barrier for S—C bond cleavage. This “molecule-to-mechanism” insight provides a rational basis for the engineered enzyme’s superior kinetic proficiency. Critically, this molecular innovation translated into process-level efficiency. The engineered biocatalyst enabled high-yield SFN recovery from heterogeneous solid biomass under mild conditions, exhibiting broad operational robustness against pH and temperature fluctuations essential for industrial scaling. Furthermore, comparative LCA quantified the environmental dividends of this approach, confirming a >70% reduction in carbon footprint and significant energy savings compared to conventional chemical and extractive routes. Overall, this work offers a scalable, low-carbon paradigm for the circular biomanufacturing of bioactive compounds. By seamlessly coupling AI-guided molecular design with rigorous process and sustainability evaluations, a transferable blueprint is provided for upgrading waste-to-value systems, advancing the frontier of green chemical engineering.

CRedit authorship contribution statement

Hao Luo: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Siyuan Guo:** Methodology, Investigation, Data curation, Conceptualization. **Yihong Bao:** Investigation, Formal analysis, Conceptualization. **Shan Lu:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Fengxia Lv:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Ping Zhu:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32200262).

Appendix A. Supplementary data

Supplementary data associated with this article include additional

kinetic datasets, analytical validation results, and life-cycle inventory details. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biortech.2026.134988>.

Data availability

Data will be made available on request.

References

- Aldaghi, S.A., Ubais, R., Schmitt, I., Wendisch, V.F., Costamagna, M., Perucca, M., 2023. Life cycle assessment of bacterial, algal, and synthetic approaches for astaxanthin production at a laboratory scale: comparative environmental analysis and sensitivity of energy sources. *Processes* 11 (10).
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44 (W1), W344–W350.
- Asif Ali, M., Khan, N., Kaleem, N., Ahmad, W., Alharethi, S.H., Alharbi, B., Alhassan, H. H., Al-Enazi, M.M., Razis, A.F.A., Modu, B., Calina, D., Sharifi-Rad, J., 2023. Anticancer properties of sulforaphane: current insights at the molecular level. *Front. Oncol.* 13, 1168321.
- Atwell, L.L., Hsu, A., Wong, C.P., Stevens, J.F., Bella, D., Yu, T.W., Pereira, C.B., Lohr, C. V., Christensen, J.M., Dashwood, R.H., Williams, D.E., Shannon, J., Ho, E., 2015. Absorption and chemopreventive targets of sulforaphane in humans following consumption of broccoli sprouts or a myrosinase-treated broccoli sprout extract. *Mol. Nutr. Food Res.* 59 (3), 424–433.
- Becke, A.D., 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98 (7), 5648–5652.
- Becker, M., Zieminska-Stolarska, A., Markowska, D., Lutz, S., Rosenthal, K., 2023. Comparative Life Cycle Assessment of Chemical and Biocatalytic 2’3’-Cyclic GMP-AMP Synthesis. *ChemSusChem* 16 (5), e202201629.
- Botti, M.G., Taylor, M.G., Botting, N.P., 1995. Studies on the mechanism of myrosinase. Investigation of the effect of glycosyl acceptors on enzyme activity. *J. Biol. Chem.* 270 (35), 20530–20535.
- Chen, A., Wang, D., Ji, R., Li, J., Gu, S., Tang, R., Ji, C., 2021. Structural and catalytic characterization of TSBGL, a beta-glucosidase from *Thermophilum* sp. ex4484.79. *Front. Microbiol.* 12, 723678.
- Chong, W., Zhang, Z., Li, Z., Meng, S., Nian, B., Hu, Y., 2024. Hook loop dynamics engineering transcended the barrier of activity-stability trade-off and boosted the thermostability of enzymes. *Int. J. Biol. Macromol.* 278 (Pt 4), 134953.
- Christopher Selvam, D., Raja, T., Vaghela, D.R., Meena, M., Tripathy, S., Kumar, B., Manjunath, H.R., Mehar, K., Devarajan, Y., 2026. Artificial intelligence in process safety: a review of opportunities, challenges, and future directions for the chemical process industries. *J. Loss Prev. Process Ind.* 100, 105917.
- Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I. M., Courbet, A., de Haas, R.J., Bethel, N., Leung, P.J.Y., Huddy, T.F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A.K., King, N.P., Baker, D., 2022. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378 (6615), 49–56.
- Dmytriv, T.R., Lushchak, O., Lushchak, V.I., 2025. Glucoraphanin conversion into sulforaphane and related compounds by gut microbiota. *Front. Physiol.* 16.
- Fahey, J.W., Holtzclaw, W.D., Wehage, S.L., Wade, K.L., Stephenson, K.K., Talalay, P., 2015. Sulforaphane Bioavailability from Glucoraphanin-Rich Broccoli: Control by active Endogenous Myrosinase. *PLoS One* 10 (11), e0140963.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., et al., 2016. Gaussian 16, Revision C.01. Gaussian, Inc., Wallingford, CT.
- Galadova, H., Polozsanyi, Z., Breier, A., Simkovic, M., 2022. Sulphoraphane affinity-based chromatography for the purification of myrosinase from *lepidium sativum* seeds. *Biomolecules* 12 (3).
- Goldaraz-Salamero, N., Baima, L., Moles, S., Sierra-Perez, J., Martinez, I., Murillo, R., Brun, F., Blanc, S., 2025. Opportunities for biomass valorisation through sorption-enhanced gasification: assessing environmental and economic aspects on biochar production. *Bioresour. Technol.* 434, 132772.
- Hao, X., Wang, S., Zhang, X., Ma, Z., Zhang, M., Shi, H., Yang, H., 2024. Engineering enzyme conformation within liquid-solid hybrid microreactors for enhanced continuous-flow biocatalysis. *Nat. Commun.* 15 (1), 10440.
- Jafari, S., Ryde, U., Irani, M., 2024. QM/MM study of the catalytic reaction of aphid myrosinase. *Int. J. Biol. Macromol.* 262 (Pt 2), 130089.
- Jafari, S., Ryde, U., Irani, M., 2021. QM/MM study of the catalytic reaction of myrosinase; importance of assigning proper protonation states of active-site residues. *J. Chem. Theory Comput.* 17 (3), 1822–1841.
- James, D., Devaraj, S., Bellur, P., Lakkanna, S., Vicini, J., Boddupalli, S., 2012. Novel concepts of broccoli sulforaphanes and disease: induction of phase II antioxidant and detoxification enzymes by enhanced-glucoraphanin broccoli. *Nutr. Rev.* 70 (11), 654–665.
- Johnson, T.A., Holyoak, T., 2010. Increasing the conformational entropy of the Omega-loop lid domain in phosphoenolpyruvate carboxykinase impairs catalysis and decreases catalytic fidelity. *Biochemistry* 49 (25), 5176–5187.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M.,

- Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589.
- Lai, R.H., Miller, M.J., Jeffery, E.H., 2009. Glucoraphanin is hydrolyzed by lactobacilli in vitro and rat cecal microbiota in vitro and in situ. *FASEB J.* 23 (765), 7.
- Lee, C.T., Yang, W.T., Parr, R.G., 1988. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* 37 (2), 785–789.
- Li, Q., Pan, Y., Li, H., Alhalhooly, L., Li, Y., Chen, B., Choi, Y., Yang, Z., 2020. Size-Tunable metal-organic framework-coated magnetic nanoparticles for enzyme encapsulation and large-substrate biocatalysis. *ACS Appl. Mater. Interfaces* 12 (37), 41794–41801.
- Li, X., Wang, Y., Zhao, G., Liu, G., Wang, P., Li, J., 2022. Microorganisms-an effective tool to intensify the utilization of sulforaphane. *Foods* 11 (23).
- Liang, X., Niu, C., Jiang, C., Li, D., Zou, L., Zhang, Y., Benito, M.J., Cheng, Y., Wu, Y., 2025. Characterization of myrosinase-mediated glucosinolate degradation pathways in *Lactiplantibacillus plantarum* ZUST49. *J. Agric. Food Chem.* 73 (12), 7256–7269.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D. E., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Struct. Funct. Bioinform.* 78 (8), 1950–1958.
- Liu, B.B., Qu, G., Li, J.K., Fan, W.C., Ma, J.A., Xu, Y., Nie, Y., Sun, Z.T., 2019. Conformational dynamics-guided loop engineering of an alcohol dehydrogenase: capture, turnover and enantioselective transformation of difficult-to-reduce ketones. *Adv. Synth. Catal.* 361 (13), 3182–3190.
- Lozupone, C.A., Hamady, M., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Gordon, J.I., Knight, R., 2008. The convergence of carbohydrate active gene repertoires in human gut microbes. *PNAS* 105 (39), 15076–15081.
- Mohsin, I., Poudel, N., Li, D.C., Papageorgiou, A.C., 2019. Crystal Structure of a GH3 beta-Glucosidase from the Thermophilic Fungus *Chaetomium thermophilum*. *Int. J. Mol. Sci.* 20 (23).
- Mokhtari, R.B., Baluch, N., Homayouni, T.S., Morgatskaya, E., Kumar, S., Kazemi, P., Yeger, H., 2018. The role of Sulforaphane in cancer chemoprevention and health benefits: a mini-review. *J. Cell Commun. Signaling* 12 (1), 91–101.
- Munoz-Munoz, J., Cartmell, A., Terrapon, N., Henrissat, B., Gilbert, H.J., 2017. Unusual active site location and catalytic apparatus in a glycoside hydrolase family. *PNAS* 114 (19), 4936–4941.
- Namchuk, M.N., Withers, S.G., 1995. Mechanism of *Agrobacterium* beta-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions. *Biochemistry* 34 (49), 16194–16202.
- Okunade, O., Niranjana, K., Ghawi, S.K., Kuhnle, G., Methven, L., 2018. Supplementation of the diet by exogenous myrosinase via mustard seeds to increase the bioavailability of sulforaphane in healthy human subjects after the consumption of cooked broccoli. *Mol. Nutr. Food Res.* 62 (18), e1700980.
- Pegan, S.D., Rukser, K., Capodagli, G.C., Baker, E.A., Krasnykh, O., Franzblau, S.G., Mesecar, A.D., 2013. Active site loop dynamics of a class IIa fructose 1,6-bisphosphate aldolase from *Mycobacterium tuberculosis*. *Biochemistry* 52 (5), 912–925.
- Pereira, S.G., Martins, A.A., Mata, T.M., Pereira, R.N., Teixeira, J.A., Rocha, C.M.R., 2024. Life cycle assessment and cost analysis of innovative agar extraction technologies from red seaweeds. *Bioresour. Technol.* 414, 131649.
- Piao, X., Li, Y., Liu, L., Li, B., Ding, J., Su, G., 2026. Low-carbon sustainable bio-electrochemical system for upgrading oilfield wastewater treatment: Comparative life cycle assessment. *Bioresour. Technol.* 439, 133338.
- Picart, P., de Maria, P.D., Schallmey, A., 2015. From gene to biorefinery: microbial beta-etherases as promising biocatalysts for lignin valorization. *Front. Microbiol.* 6, 916.
- Plaza-Vinuesa, L., Hernandez-Hernandez, O., Sanchez-Arroyo, A., Cumella, J.M., Corzo, N., Munoz-Labrador, A.M., Moreno, F.J., Rivas, B.L., Munoz, R., 2022. Deciphering the myrosinase-like activity of *Lactiplantibacillus plantarum* WCFS1 among GH1 family glycoside hydrolases. *J. Agric. Food Chem.* 70 (49), 15531–15538.
- Quirk, J.C.K., Crich, D., 2021. GH47 and other glycoside hydrolases catalyze glycosidic bond cleavage with the assistance of substrate super-arming at the transition state. *ACS Catal.* 11, 10308–10315.
- Ribeiro, A.J.M., Tyszack, J.D., Borkakoti, N., Holliday, G.L., Thornton, J.M., 2020. A global analysis of function and conservation of catalytic residues in enzymes. *J. Biol. Chem.* 295 (2), 314–324.
- Rogowski, A., Basle, A., Farinas, C.S., Solovyova, A., Mortimer, J.C., Dupree, P., Gilbert, H.J., Bolam, D.N., 2014. Evidence that GH115 alpha-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. *J. Biol. Chem.* 289 (1), 53–64.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–W388.
- Sharma, P., Gaur, V.K., Sirohi, R., Varjani, S., Hyoun Kim, S., Wong, J.W.C., 2021. Sustainable processing of food waste for production of bio-based products for circular bioeconomy. *Bioresour. Technol.* 325, 124684.
- Shawky, E., Gibbons, S., Selim, D.A., 2025. Bio-sourcing from byproducts: a comprehensive review of bioactive molecules in Agri-Food Waste (AFW) streams for valorization and sustainable applications. *Bioresour. Technol.* 431, 132640.
- Sivapalan, T., Melchini, A., Saha, S., Needs, P.W., Traka, M.H., Tapp, H., Dainty, J.R., Mithen, R.F., 2018. Bioavailability of glucoraphanin and sulforaphane from high-glucoraphanin broccoli. *Mol. Nutr. Food Res.* 62 (18), e1700911.
- Sumida, K.H., Nunez-Franco, R., Kalvet, I., Pellock, S.J., Wicky, B.I.M., Milles, L.F., Dauparas, J., Wang, J., Kipnis, Y., Jameson, N., Kang, A., De La Cruz, J., Sankaran, B., Bera, A.K., Jimenez-Oses, G., Baker, D., 2024. Improving protein expression, stability, and function with ProteinMPNN. *J. Am. Chem. Soc.* 146 (3), 2054–2061.
- Vega, A., Planas, A., Biarnes, X., 2025. Electrostatic potential as a reactivity scoring function in computer-assisted enzyme engineering. *FEBS J.* 292 (16), 4211–4231.
- Walker, E.J., Hamill, C.J., Crean, R., Connolly, M.S., Warrender, A.K., Kraakman, K.L., Prentice, E.J., Steyn-Ross, A., Steyn-Ross, M., Pudney, C.R., van der Kamp, M.W., Schipper, L.A., Mulholland, A.J., Arcus, V.L., 2024. Cooperative conformational transitions underpin the activation heat capacity in the temperature dependence of enzyme catalysis. *ACS Catal.* 14 (7), 4379–4394.
- Wang, L., Althoff, E.A., Bolduc, J., Jiang, L., Moody, J., Lassila, J.K., Giger, L., Hilvert, D., Stoddard, B., Baker, D., 2012. Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of de novo enzyme design. *J. Mol. Biol.* 415 (3), 615–625.
- Wang, L., Jiang, H., Qiu, Y., Dong, Y., Hamouda, H.I., Balah, M.A., Mao, X., 2022. Biochemical Characterization of a novel Myrosinase Rmyr from *Rahnella inusitata* for high-level preparation of sulforaphane and sulforaphane. *J. Agric. Food Chem.* 70 (7), 2303–2311.
- Wang, Q., Zhao, Y., Zhang, W., Deng, J., Yang, H., 2025. Valorization of broccoli waste: Unlocking its potential as a functional food ingredient for sustainable nutrition. *S2090-1232 J. Adv. Res.* (25), 00580–00586.
- Wen, S.X., Zheng, W., Bornscheuer, U.T., Wu, S.K., 2025. Generative artificial intelligence for enzyme design: recent advances in models and applications. *Curr. Opin. Green Sustain. Chem.* 52, 101010.
- Xia, W., Xu, X., Qian, L., Shi, P., Bai, Y., Luo, H., Ma, R., Yao, B., 2016. Engineering a highly active thermophilic beta-glucosidase to enhance its pH stability and saccharification performance. *Biotechnol. Biofuels* 9, 147.
- Ye, Q., Fang, Y., Li, M., Mi, H., Liu, S., Yang, G., Lu, J., Zhao, Y., Liu, Q., Zhang, W., Hou, X., 2022. Characterization of a novel myrosinase with high activity from marine bacterium *Shewanella baltica* Myr-37. *Int. J. Mol. Sci.* 23 (19).
- Youseif, S.H., Abdel-Fatah, H.M.K., Khalil, M.S., 2022. A new source of bacterial myrosinase isolated from endophytic bacillus sp. NGB-B10, and its relevance in biological control activity. *World J. Microbiol. Biotechnol.* 38 (11), 215.
- Wu, Y.F., Lv, C.Z., Zou, L.G., Sun, J., Song, X.J., Zhang, Y., Mao, J.W., 2021. Approaches for enhancing the stability and formation of sulforaphane. *Food Chem.* 345, 128771.
- Zhang, X., Chen, F., He, C., Fang, W., Fang, Z., Zhang, X., Wang, X., Xiao, Y., 2020. Improving the thermostability of a GH97 dextran glucosidase by rational design. *Biotechnol. Lett.* 42 (11), 2211–2221.
- Zhang, Y., Makaza, N., Jiang, C., Wu, Y., Nishanbaev, S.Z., Zou, L., Sun, J., Song, X., Wu, Y., 2022. Supplementation of cooked broccoli with exogenous moringa myrosinase enhanced isothiocyanate formation. *Food Chem.* 395, 133651.