Rational Design of a Peptide Ligase from HRV-3C Protease

Xuchao Zhou[a], Xiaoken Lin[a], Yuhang Sheng[a], Lishu Chen[a], Heming Li[a], Yijie Wang[a], Xuzhi Wang[a], Liya Ma[a], Keyi Yang[a], Jiwei Ye[a], Kedu Jin[a], Le Chang[a], Xinya Hemu[a*]

[a]State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing, 211198, China

Abstract

The development of efficient peptide ligases is a cornerstone of protein engineering and biopharmaceutical synthesis. However, existing ligases often suffer from limitations such as low catalytic efficiency, reaction reversibility, or stringent substrate pre-activation requirements. This study aimed to convert the HRV-3C protease, a highly specific cysteine hydrolase, into a novel peptide ligase through rational design. Our primary strategy involved computationally engineering a hydrophobic active-site pocket to suppress the enzyme's intrinsic hydrolytic activity by sterically hindering the access of nucleophilic water molecules. This work underscores a critical challenge in enzyme engineering and suggests that future design strategies must pivot from passively excluding water to actively enhancing the binding affinity and nucleophilicity of the desired acceptor substrate, enabling it to kinetically outcompete the hydrolysis reaction.

Key words: Peptide Ligation; Transamidation; Peptides And Proteins.

INTRODUCTION

In the vast landscape of biochemistry, transferases (EC 2) represent a major class of enzymes that catalyze the transfer of a functional group from one molecule (the donor) to another (the acceptor). Their catalytic activities are fundamental to metabolism and macromolecular synthesis in living organisms. For instance, glycosyltransferases are responsible for transferring activated sugar moieties to protein or lipid molecules, constructing intricate glycan networks that are crucial for cellular recognition and signaling. Similarly, kinases, which transfer phosphate groups, regulate the vast majority of cellular processes. Among all these transfer reactions, one of the most central and fundamental is the formation of the peptide bond. Within the magnificent molecular machine of the ribosome, the peptidyl transferase center catalyzes the sequential linkage of amino acids with remarkable precision and efficiency, synthesizing the proteins that form the basis of life's functions.

The ability to precisely form peptide bonds is not only fundamental to life's synthesis of proteins but also provides a powerful tool for biotechnology and chemical biology. Through controllable peptide transfer reactions (i.e., peptide ligation), researchers can achieve site-specific modification, labeling, cyclization, and fragment assembly of proteins, significantly expanding the scope of protein engineering and pharmaceutical development[1]. The fundamental prerequisite enabling these applications is specificity. An ideal peptide ligation tool must accurately recognize specific amino acid sequences to ensure the ligation reaction occurs exclusively at the predetermined site, avoiding non-specific modifications to other parts of the protein. This stringent requirement for specificity is an intrinsic property of many proteases. In the laboratory, tool enzymes such as Tobacco Etch Virus protease (TEV protease[2]), Human Rhinovirus 3C protease (HRV-3C protease[3]), and Thrombin are widely used for cleaving purification tags from fusion proteins. Their widespread adoption is precisely due to their high fidelity for specific recognition sequences
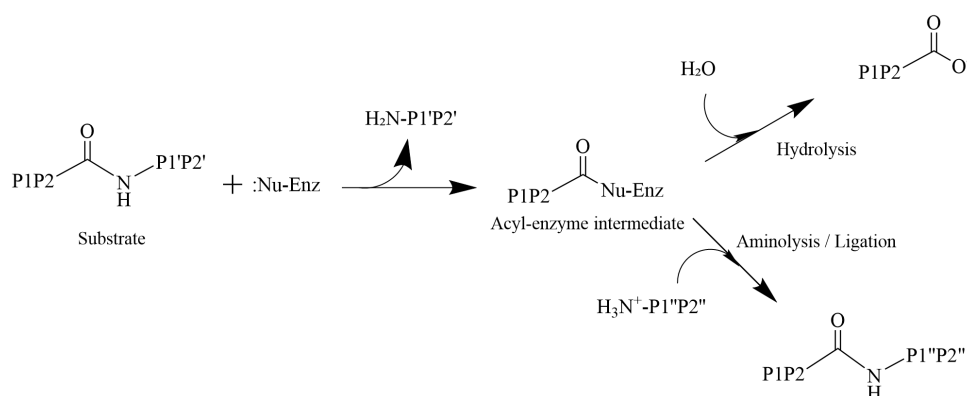

Figure 1. The two competing reaction pathways during the proteolytic stage.

(e.g., ENLYFQ/G for TEV, LEVLFQ/(G/P) for HRV-3C). This inherent high specificity provides a solid foundation for engineering them into ligases.

The catalytic mechanisms of cysteine and serine proteases typically follow a two-step process. First, the nucleophilic residue of the enzyme (the thiol group of cysteine or the hydroxyl group of serine) attacks the carbonyl carbon of the substrate's peptide bond. This results in the formation of a covalent acyl-enzyme intermediate and the release of the C-terminal peptide fragment. Subsequently, this intermediate is attacked by a second nucleophile, completing the catalytic cycle. This second step represents a critical "branching point." In the conventional hydrolytic reaction, this nucleophile is a ubiquitous water molecule; its attack leads to the hydrolysis of the acyl-enzyme intermediate, releasing the N-terminal peptide fragment and regenerating the free enzyme. However, if an external nucleophile with a free N-terminal α-amino group (such as another peptide chain) is present, it can compete with the water molecule to attack the acyl-enzyme intermediate. Successful competition by this nucleophile results in the formation of a new peptide bond, accomplishing a peptide transfer (or ligation) reaction instead of hydrolysis.

The intrinsic specificity of proteases forms a foundation for engineering them into peptide ligases. Therefore, the "hydrolytic" and "ligation" activities of proteases are essentially two different outcomes resulting from the competition between different nucleophiles within the same catalytic pathway like Figure 1. However, under most physiological conditions, this reverse reaction involving nucleophilic attack by an amine on the ester intermediate is thermodynamically unfavorable. This process is governed by kinetic control. Strategies to shift the equilibrium of the protease-catalyzed reaction from hydrolysis towards ligation have been explored from multiple perspectives. These include altering the pH to modify the ionization states of reactive functional groups, employing organic co-solvents to reduce water activity, utilizing activated esters as substrates to accelerate nucleophilic attack, engineering proteases through site-specific modification or directed evolution, and combinations of the aforementioned approaches.

The idea of engineering proteases into peptide ligases has achieved remarkable success in practice, demonstrating the broad prospects of this approach. From serine proteases to ligases: the most classic example is the engineering of Subtilisin[4]. By mutating its catalytic serine to cysteine (S221C), supplemented by other key site mutations (such as P225A), scientists successfully developed the first-generation peptide ligase, Subtiligase[4]. This engineering strategy significantly reduced the hydrolysis rate while retaining the catalytic capability for aminonucleophiles, thereby enabling efficient peptide fragment ligation in aqueous solution. If the binding and reaction efficiency of the external nucleophilic peptide are simultaneously enhanced, it becomes possible to shift the reaction equilibrium toward the ligation product, thus transforming a natural "scissor" into a "glue". A similar strategy was applied to the engineering of Trypsin, resulting in Trypsiligase[5], which exhibits

stronger peptide aminolysis activity than hydrolysis and enables site-specific peptide linkage. In the design of enzymes such as Aqualigase[6], scientists promoted peptide condensation by weakening the efficiency of water molecules as nucleophiles, forming a hydrophobic reaction cavity that favors the release of water.

Nature provides perfect blueprints for this endeavor. For instance, Sortase A[7] from Staphylococcus aureus is a natural transpeptidase that specifically recognizes a C-terminal LPXTG motif and catalyzes its linkage to an N-terminal glycine residue. Furthermore, Butelase-1[8], a recently discovered peptide asparaginyl ligase from Clitoria ternatea, represents the most efficient peptide ligase known to date, whose remarkably high catalytic rate constant sets a new benchmark for enzymatic ligation. Asparaginyl Endopeptidases (AEPs), which also belong to the cysteine protease family, have been successfully engineered into powerful peptide ligation tools; these ligases are particularly useful for peptide head-to-tail cyclization and site-specific C-terminal protein labeling. Through targeted mutagenesis, their stringent substrate specificity has been relaxed while their ligation efficiency was optimized, leading to the development of a series of highly efficient enzymes known as Peptide Asparaginyl Ligases[9]. The successful engineering of AEPs[10] strongly validates the feasibility of developing ligase activity from cysteine proteases.

Based on the aforementioned background, this study aims to investigate strategies for converting a hydrolase into a ligase. We selected the HRV-3C as the engineering scaffold. As a cysteine protease, HRV-3C possesses several unique advantages: 1) Exceptionally high specificity: Its stringent sequence recognition provides a reliable foundation for precise ligation; 2) Well-established application protocols: It is already one of the most commonly used tool enzymes in biotechnology, with mature protocols for its expression, purification, and reaction conditions; 3) A unique recognition sequence: Its octapeptide recognition sequence is longer than those of many existing ligases, like the pentapeptide for Sortase A. Potentially offering higher addressing accuracy within complex protein substrates. Therefore, we hypothesize that via a rational design strategy, if the hydrolytic activity of HRV-3C can be suppressed while its peptide ligation activity is enhanced, a highly efficient peptide ligase with novel specificity could be created. The development of such a novel tool enzyme would provide a unique and powerful new option for site-specific protein conjugation, macromolecular drug synthesis, and chemical biology research.

METHOD

Construction of Mutant Plasmids

The pET-47b(+)-NT-HRV3CP plasmid was used as the template for site-directed mutagenesis. Primers containing the desired mutations were synthesized by Tsingke Biotechnology Co., Ltd. (Beijing, China). Point mutations were introduced using the Mut Express II Fast Mutagenesis Kit V2 (Vazyme, Cat: C214-02) according to the manufacturer's protocol. The mutagenesis reaction products were then transformed into E. coli DH5α chemically competent cells (Sangon Biotech, Shanghai, China). Transformants were selected on Luria-Bertani (LB) agar plates supplemented with 50 μg/mL kanamycin. Positive colonies were cultured, and the mutant plasmids were extracted using the FastPure Plasmid Mini Kit (Vazyme, Cat: DC201-01). The primers used to construct Mutant 1 and Mutant 2 are listed in Supplementary Table S1. The successful introduction of all mutations was verified by Sanger sequencing, performed by Tsingke Biotechnology Co., Ltd. (Beijing, China).

Recombinant Expression of HRV-3C Mutants

The NT*-HRV3CP[11] protein was expressed in E. coli BL21(DE3) cells (Sangon Biotech, Shanghai, China). To prepare the protein, the pET-47b(+) plasmid was transformed into E. coli BL21(DE3) cells, which were then cultured at 37 °C in LB medium containing 50 μg/mL kanamycin. An overnight culture of 2.5 mL was used to inoculate 250 mL of TB medium supplemented with 50 μg/mL

kanamycin. The culture was grown at 37 °C until the optical density at 600 nm (OD600) reached 0.6–1. Subsequently, the temperature was lowered to 18 °C, and protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 hours.

After expression, the cells were harvested by centrifugation and resuspended in Buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole) supplemented with 0.1% Triton X-100, followed by lysis using a sonicator. The cell lysate was clarified by centrifugation at 12,000 rpm for 1 hour. The resulting supernatant was loaded onto a 1 mL HisTrap FF affinity chromatography column connected to an ÄKTA start chromatography system (Cytiva, USA). The column was washed with 20 column volumes of Buffer B (identical to Buffer A, but with 20 mM imidazole) and the protein was then eluted with 3 column volumes of Buffer C (identical to Buffer A, but with 500 mM imidazole). The eluted protein was buffer-exchanged into buffer D (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM DTT) and concentrated using an Amicon ultrafiltration centrifugal tube (Merck Millipore, USA) with a molecular weight cut-off of 10 kDa. Afterwards, the protein was stored in 10% glycerol at −80 °C after snap freezing in liquid nitrogen. Protein concentration was measured with a Unano-1000 Micro-Spectrophotometer (Hangzhou UMI Instrument Co., Ltd., Hangzhou, China) using the 280 nm extinction coefficient of 11,460 $M^{-1}cm^{-1}$ calculated for NT*-HRV3CP.

Functional Validation of Mutants for Peptide hydrolytic and ligation

The enzymatic activity of NT*-HRV3CP and its variants was evaluated in a 100 μL reaction system. For the hydrolysis reaction, the mixture contained 20 mM of the peptide substrate LEVLFQAPG and 0.2 mM of the enzyme, dissolved in Buffer D. For the ligation reaction, the mixture was additionally supplemented with 640 mM of the nucleophilic peptide GITRR. All reactions were incubated at 25 °C for 24 hours and subsequently stopped by the addition of 0.1% formic acid. The quenched reactions were then centrifuged at 12,000 rpm for 30 minutes, and 80 μL of the supernatant was transferred to an HPLC vial for analysis.

Reaction products were analyzed by high-performance liquid chromatography (HPLC) on a C8 column. The injection volume for each run was 20 μL. The mobile phase consisted of solvent A (water with 0.1% trifluoroacetic acid) and solvent B (acetonitrile with 0.1% trifluoroacetic acid). Separation was performed at a flow rate of 1.0 mL/min using the following gradient: a linear increase from 2% to 95% solvent B over 35 minutes, followed by a 4-minute wash with 95% solvent B, and finally, a 3-minute equilibration with 2% solvent B. The elution signal was monitored by absorbance at 220 nm.

Corresponding Mutation Changing Energies Calculation

The HRV-3C wild-type protein structure was obtained from the Protein Data Bank (PDB ID: 2B0F). We used all these structures to estimate the effects of amino acid substitutions on protein stability using FoldX. The RepairPDB function was run 5 times to repair incorrect torsion angles, VanderWaals clashes and total energy of the structure. We then used the MutateX[12] pipeline to perform an in silico deep mutational scanning of the whole protein, running the 3 rounds BuildModel function from FoldX and computing the average difference in Gibbs free energy between the mutant and the wild type.

Structure Preparation for Simulations

All protein monomers and protein-hydrolytic substrate complex structures were modeled using AlphaFold3[13]. The protonation states of amino acid residues were determined using the H++ webserver. Prior to the production simulations, a 20 ns stability validation was conducted. Molecular docking was performed with Autodock Vina[14] to obtain an ester intermediate-linked substrate complex. The topology and force field parameters for the enzyme-substrate ester intermediate

structures were described via additional parametrization, which modeled the deprotonated catalytic residue (SER or CYS) forming an ester intermediate with the hydrolytic substrate.[15], [16]

Conventional Molecular Dynamics

Conventional molecular dynamics (cMD) simulations were conducted using the GROMACS 2025.2[17] package. Unless otherwise specified, the Amber ff19SB[18] force field along with the OPC[19] water model were employed for all simulations. All system were neutralized with an appropriate number of ions. On top of neutralization, additional $K+$ and $Cl-$ ions were added to bring the final ion concentration to 150 mM. Periodic boundary conditions and dispersion corrections for energy and pressure were applied throughout all the minimization and cMD simulations.A nonbonded interaction cutoff of 10 Å was implemented, with long range electrostatic interactions treated via the Particle Mesh Ewald method.

First, the system was minimized using the steep method for 100000 steps,ensuring that the maximum force was reduced to 100 kJ·mol¹·nm¹. No restraints were applied in minimization. After the mininimization, 2ns NPT simulation was carried out at target temperature and 1 atm to further equilibrate the system and correct the density. The temperature coupling was controlled using the V-rescale method (τ_p = 0.5), and pressure coupling was initially managed using the C-rescale method (τ_p = 2.0). A 1-fs timestep was used. No restraints were applied. During the equilibration phase, subsequent NPT equilibration employed pressure coupling via the C-rescale barostat with isotropic scaling (τ_p = 2.0). Temperature coupling was maintained using the V-rescale algorithm (τ_t = 0.2).

Then each system underwent NPT equilibration for a 20 ns duration. All bonds involving hydrogen atoms were constrained with the LINCS algorithm, enabling a 2 fs integration time step. Finally, three independent replicates of 100 ns simluations were carried out for production and this trajectory was used for data analysis.

The force field parameters for the non-standard residue and additional bonded parameters derived from the covalently-linked ester intermediate were obtained from the GROMACS database. Detailed parameters for all atoms, bonds, angles, and dihedrals are provided in the Supporting Information.

Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD) Simulations

The AMBER24[20] package was used to preform LiGaMD3[21]. LiGaMD3 (igamd = 28) was performed to sample the interaction from ester intermediate-linked substrate complex. Simulations utilized the Amber ff19SB force field and OPC3 water model. Temperature was maintained at their respective physiological temperatures using the Langevin thermostat (ntt = 3, and gamma_ln = 5.0). Pressure was maintained isotropically (ntp = 1, and taup = 2.0). Nonbonded interactions were handled with a cutoff of 10.0 Å (cut = 10.0).

In LiGaMD3, three distinct boosts are applied: one on the non-bonded interaction energy of the substrate, the second one on the remaining non-bonded potential energy of the system, and a third one on the system bonded potential energy. For upper limit of the standard deviation of the potential boost, σP were set to 3.0 kcal/mol (The Bu2g(V/GA) system is 4.5 kcal/mol). Default values were used for other input parameters. GaMD consisted of the pre-equilibration stage, equilibration stage and production stage. The previously equilibrated snapshot prepared from the unbiased conventional MD was used as the initial configuration of GaMD. The pre-equilibrated system was subjected to additional unbiased cMD of 6 ns as pre-equilibration before collection of potential data for 20 ns for the determination of boost potential. Three independent replicas of 100 ns GaMD simulations were done for the production stage using fixed boost statistics and coordinates being saved every 2 ps.
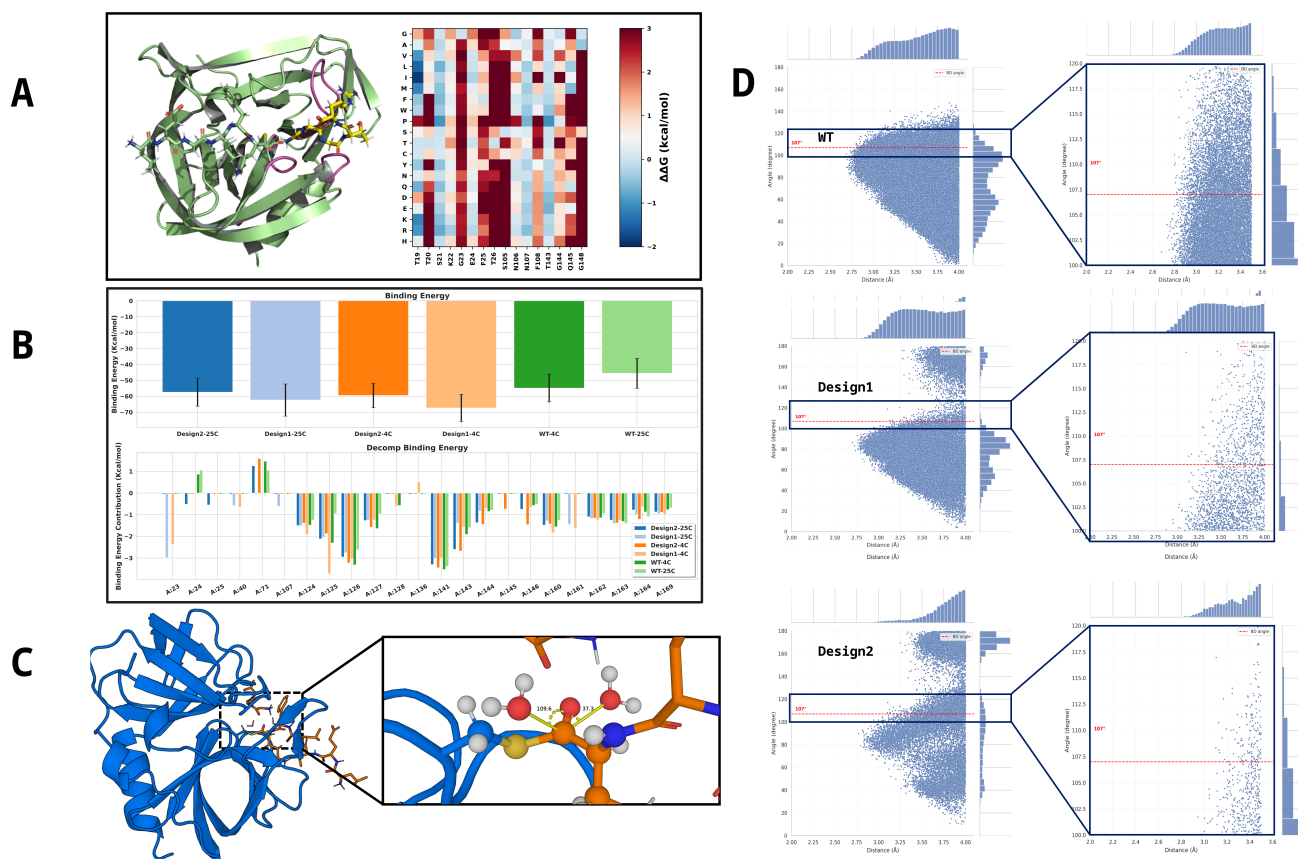
Figure 2. Utilization of PocketGen for pocket design and evaluation of binding and hydrophobic properties. (A) Results of the ΔΔG Scan for the Target Design Region. (B) Comparison of substrate binding energies for two high-order mutants and the wild-type(WT) enzyme at different temperatures. (C) Schematic diagram illustrating the attack angle formed between the thioester intermediate and surrounding water molecules. (D) Statistical distribution of the attack distance and attack angle for water molecules within 4 Å of the thioester intermediate. The red line indicates the canonical Bürgi-Dunitz (BD) angle. The statistical profiles of the two mutants exhibit significant differences compared to the WT.

Design targeting the hydrophobic pocket

To computationally identify mutations capable of converting HRV-3C protease into a peptide ligase, a three-tiered hierarchical screening protocol was implemented. The workflow was designed to first generate a targeted virtual library, then filter for variants that retain substrate binding capacity, and finally select candidates predicted to exhibit impaired hydrolytic activity. This strategy was based on the hypothesis that increased hydrophobicity of the active site would sterically disfavor the precise positioning of the nucleophilic water molecule required for hydrolysis of the acyl-enzyme intermediate.

A virtual library of HRV-3C variants was generated using the deep learning-based protein design tool PocketGen[22]. The AlphaFold3-predicted structure of the wild-type HRV-3C in complex with its hydrolytic substrate LEVLFQAPG was used as the design template. The design process was restricted to residues within 4-5 Å of the substrate's C-terminal APG-binding groove, particularly those constituting the S1′ and S2′ pockets. A key constraint was introduced into the generative model to prioritize substitutions with hydrophobic amino acids (Val, Leu, Ile, Phe, Trp, Ala, Met), enforcing a statistical bias. In addition to using ESM[23] (evolutionary scale modeling)-2 15B

throughout our experiments. Folding stability was ensured by screening mutation schemes based on a full-sequence saturation mutagenesis ΔΔG library, applying a threshold (Σ(ΔΔG) ≤ 2.0).

To ensure that the designed mutations did not disrupt essential substrate recognition, all variants from the generated library were subjected to a binding affinity screen. For each mutant, a non-covalent Michaelis complex was constructed in silico with its cognate peptide substrate LEVLFQGGG. Each complex was solvated in an OPC3 water box, neutralized with counterions, and subjected to three independent replicates of 100 ns cMD simulations using the AMBER ff19SB force field for binding conformation sampling. Subsequently, the binding free energy was calculated for each variant using the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA)[24] method, analyzing snapshots from the final 80 ns of each MD trajectory. Only mutants exhibiting binding free energies comparable to or more favorable than that of the wild-type enzyme were advanced to the next stage.

The final screening stage aimed to identify candidates in which the geometry of the deacylation step is unfavorable for hydrolysis. For each variant that passed the binding affinity filter, a covalent thioester acyl-enzyme intermediate was modeled by forming a bond between the catalytic Cys146 and the carbonyl carbon of the P1 Gln residue of the substrate. The water attack properties are described by two parameters: The attack distance (Distance), defined by the distance between the nucleophilic water oxygen and the thioester carbonyl carbon; The water attack angle (Angle) ,defined by the angle formed by the nucleophilic water oxygen, the thioester carbonyl carbon, and the carbonyl oxygen. Those were sampled during three independent replicates of 100 ns classical MD simulations. Variants were prioritized if the statistical distribution of their BD angle showed a significant deviation from the canonical range considered optimal for nucleophilic attack (107° ± 7°). Candidates exhibiting a persistently non-optimal BD angle were selected as promising ligase designs for subsequent experimental validation.

Design focusing on binding affinity optimization

The ligation reaction catalyzed by Trypsiligase were usually complete within minutes and requires 0.1 molar equivalents of enzyme with an excess of the corresponding acyl acceptor substrate (ligation substrate) because it needs to compete with the RH leaving group. Inspired by this, we propose to expand the pocket around the C-terminal peptide fragment released in the first process to facilitate its retention. The AlphaFold3-predicted structure of the wild-type HRV-3C in complex with its hydrolytic substrate LEVLFQAPG served as the design template. Around the C-terminus of the hydrolytic substrate, the original amino acid residues at positions 105-108 (gray region of the protein) underwent Motif Scaffolding via RFdiffusion[25], with the substrate C-terminal motif (APG) designated as Hotspots. This motif was extended to lengths of 8, 9, and 10 amino acids, generating 30 backbone structures for each design. Subsequently, using the hydrolytic substrate C-terminus (APG) as all-atom ligand information, a modified version of LigandMPNN[26] was employed to assign 200 sequences to each backbone structure, with a statistical bias applied towards hydrophobic amino acids. The design from each set with the highest overall_confidence was selected. The enzyme-substrate intermediate for this selected design was then subjected to 50 ns of cMD simulation at 4 ℃ and 25 ℃ using the AMBER ff19SB force field and the OPC3 water model for interaction sampling.

RESULT

Expression and Purification of Enzyme Variants

To investigate the effects of site-directed mutagenesis, the wild-type (WT) and three mutant enzymes (Design1, Design2_IM(The intermediate construct from Design 2) and Design2) were expressed and purified. Quantification of the purified proteins revealed that the mutations
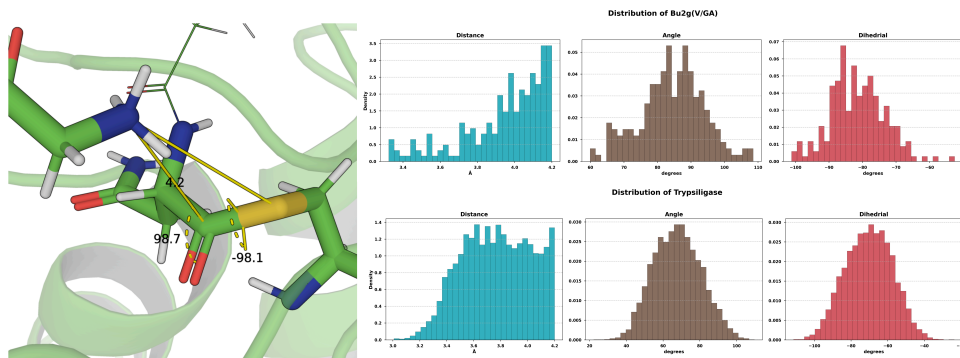
Figure 3. Characterization of the two ligation systems within a 4 Å proximity.

significantly impacted expression levels. The WT enzyme exhibited a robust expression yield of 3.27 mg/L. In stark contrast, all three mutants showed a substantial reduction in yield. The final yields for Design2_IM, Design1, and Design2 were 0.788mg/L, 1.32 mg/L, and 1.08 mg/L, respectively. These levels correspond to approximately 24%, 40%, and 33% of the WT expression level.

Kinetic characteristics of hydrolase and ligase

To elucidate the key determinants of the covalent modifications on reactivity, the hydrolytic and ligation activities of several enzyme systems were evaluated (summarized in Table 2). Herein, the acyl donor refers to the portion of the hydrolysis substrate that remains covalently attached to the catalytic residue following the formation of the ester intermediate, whereas the acyl acceptor is the ligation substrate. Figure 4 shows the distribution of the attack angle between water molecules and the ester intermediate across the five simulated hydrolytic reaction systems. Perhaps due to the significant deviation of the initial structure from the native binding pose, the simulated ligation systems of Bu2g(V/GA) and Trypsiligase exhibited mutual approximation between the N-terminus of the acyl-acceptor and the electrophilic center of the acyl-donor. Furthermore, the resulting attack angles and attack dihedrals exhibited a defined distribution pattern.

Engineering a Ligase Requires Additional Information

To evaluate the catalytic activity of the wild-type (WT) and mutant enzymes (19, 105, 143), we sought to measure their performance in a transpeptidation (ligation) reaction. Reactions were conducted in the presence of potential nucleophiles: a peptide (32GITRR, at a 32:1 molar ratio to the substrate) and an amino acid (1000Gly, at a 1000:1 molar ratio). Notably, HPLC analysis revealed no detectable formation of the expected ligation product in any of the tested conditions. The only observed product corresponded to the hydrolytic cleavage of the substrate. Consequently, our analysis focused on the hydrolytic activity of the enzymes under these different conditions, quantified as the ratio of the hydrolytic product peak area to the remaining substrate peak area. The results after 12 and 24 hours of reaction are summarized in Table 1.

In our pursuit of engineering HRV-3C wild-type (WT) and its variants into efficient peptide ligases, we assayed their catalytic activities in the presence of potential external nucleophiles—the GITRR peptide and glycine. Contrary to our design objective, the desired ligation product was not detected under the tested aqueous conditions, with its concentration remaining consistently below the limit of detection of the HPLC method. Instead, the predominant catalytic activity observed was substrate hydrolysis. Consequently, our investigation pivoted to a systematic characterization of how these additives influence the intrinsic hydrolytic function of the enzymes. The activity, quantified by the peak area ratio of the hydrolysis product to the remaining substrate, is presented in Table 1. In the control (pure hydrolysis), mutant Design2 was identified as the most hydrolytically active variant, exhibiting a 24-hour activity ratio (12.13) markedly higher than that of the WT (5.87)

Enzyme	Condition	Hydrolysis Ratio (12h)	Hydrolysis Ratio (24h)
WT	Pure Hydrolysis	4.882871	5.870103
	32GITRR	6.287923	7.013444
	1000Gly	9.710145	8.505435
Design1	Pure Hydrolysis	4.831418	6.653915
	32GITRR	4.254176	6.278571
	1000Gly	7.637177	8.229491
Design2_IM	Pure Hydrolysis	6.467882	7.574692
	32GITRR	5.53022	7.127273
	1000Gly	5.215517	7.909449
Design2	Pure Hydrolysis	5.21548	12.1319
	32GITRR	4.037259	7.263775
	1000Gly	6.441385	8.73125

Table 1. Hydrolytic activity of WT and mutant enzymes in the presence of different nucleophiles. Values represent the ratio of the peak area of the hydrolysis product to the peak area of the remaining substrate, as determined by HPLC. "Pure Hydrolysis" serves as the control. "32GITRR" and "1000Gly" represent reactions supplemented with the GITRR peptide (peptide:substrate ratio of 32:1) and glycine (glycine:substrate ratio of 1000:1), respectively. All reactions were conducted under identical conditions.

and other mutants. Intriguingly, rather than facilitating ligation, these nucleophilic compounds served as modulators of hydrolytic activity in an enzyme-specific manner. For the highly active mutant Design2, both 32GITRR and 1000Gly exerted an inhibitory effect, reducing its activity ratio to 7.26 and 8.73, respectively. In sharp contrast, the WT enzyme was consistently activated by these additives, with its 24-hour activity ratio increasing from 5.87 to 7.01 and 8.50 in the presence of 32GITRR and 1000Gly, respectively. These findings suggest that despite their inefficacy as ligation substrates, these molecules can interact with the enzyme and allosterically modulate its canonical hydrolytic pathway.

DISCUSSIONS

The strategy of rationally designing the HRV-3C protease to construct its hydrophobic pocket for developing novel ligases stems from a thorough evaluation of existing mainstream ligase technologies. First, PAL family ligases derived from asparaginyl endopeptidase (represented by Butelase 1) face bottlenecks in expression and purification. Despite their high catalytic efficiency and short recognition sequence (the tripeptide motif D/N-HV), these enzymes originate from the plant butterfly pea, making recombinant expression challenging[4]. The complex activation process required for their proenzyme form is difficult to precisely control, affecting experimental reproducibility. While the short recognition sequence facilitates handling, it increases the risk of non-specific reactions. Practical implementation also requires high substrate concentrations to drive the reaction, potentially triggering substrate inhibition and escalating costs. Additionally, Subtiligase derived from Bacillus subtilis protease faces constraint: substrate pretreatment requirements.[27] While this enzyme demonstrated the feasibility of serine protease-catalyzed ligation, it requires pre-activation of the substrate C-terminus into a high-energy ester. This additional step complicates operations, increases costs, and may induce side reactions like racemization, limiting its application in peptide modification. Finally, Staphylococcus aureus Sortase A suffers from core defects: thermodynamic reversibility and low kinetic efficiency. Its reversible catalytic mechanism limits
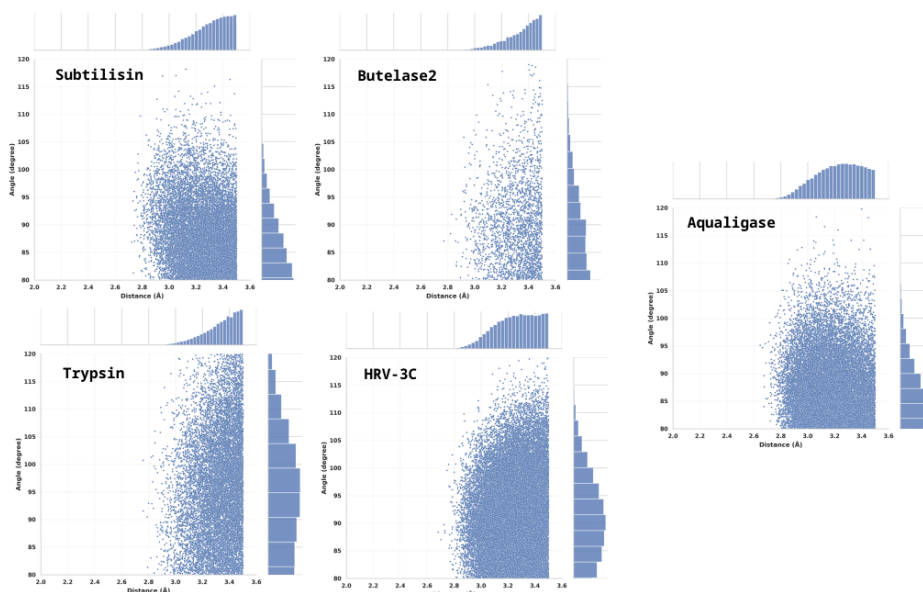
Figure 4. Statistical Distribution of the local Water-Oxygen Attack Angle within 3.5 Å of the Ester Intermediate for Four Wild-Type Hydrolases (Subtilisin,Butelase2,Trypsin,HRV-3C) and the Aqualigase Variant with a Water-Shielding Pocket Design.

Reaction	System	environment	acyl donor	acyl accpetor
Hydrolytic	Subtilisin	pH8, 25°C	Suc-AAPF	---
	Aqualigase	pH8, 40°C	Suc-AAPF	---
	HRV-3C	pH8, 4°C	LEVLFQ	---
	Butelse2	pH7, 37°C	ISYRN	---
	Trypsin	pH8, 37°C	GGGY	---
Ligation	Aqualigase	pH8, 40°C	Suc-AAPF	AF-NH2
	Subtiligase	pH8, 25°C	Suc-AAPF	AF-NH2
	HRV-3C	pH8, 4°C	LEVLFQ	GITRR
	Bu2g(V/GA)	pH7, 37°C	GISTKSIPPISYRN	GISTKSIPPISYRN
	Trypsiligase	pH8, 37°C	GGGY	RNGGG

Table 2. Systems and construction conditions used in this study

conversion rates, necessitating excessive use of donor substrates, which increases costs and purification complexity. The reaction exhibits sensitivity to conditions and poor reproducibility, making standardized workflow establishment challenging[28]. Slow catalytic rates often necessitate high substrate concentrations, compromising economic viability. Its application in cyclic peptide synthesis is particularly limited due to the generation of lengthy "scar" sequences and the requirement for substrates exceeding 19 amino acids to avoid oligomerization side reactions. In summary, existing ligases exhibit significant shortcomings in expression complexity, substrate pretreatment requirements, and reaction reversibility. Consequently, there is an urgent need to develop novel ligases that can be efficiently expressed in E. coli, require no preactivation, directly utilize native carboxylate substrates, and exhibit irreversible reactions. The HRV-3C protease offers advantages such as high prokaryotic expression yield, excellent stability, and strong specificity (recognizing the long sequence LEVLFQ-G), making it an ideal starting point for engineering. Through rational design to construct a hydrophobic pocket, its binding capacity with hydrophobic

substrates can be enhanced, optimizing the reaction pathway and thereby achieving a functional transformation from a hydrolase to an efficient ligase.

Predicated on the assumption that a portion of the mutant proteins retained correct folding, the lack of observed ligation activity indicates that the strategy of constructing a "water-proof pocket" solely through the introduction of hydrophobic residues is inadequate to surmount the potent hydrolytic background. Water acts as a ubiquitous and highly competitive nucleophile in aqueous environments, with an effective concentration reaching approximately 55 M. Furthermore, the active sites of natural hydrolases have undergone extensive evolutionary optimization to exploit this abundance efficiently. While our cMD simulations revealed that the distribution of water attack angles in the mutants deviated from that of the WT, this alteration was likely insufficient to fundamentally preclude the hydrolytic reaction. Consequently, dismantling this evolutionarily refined "hydrolysis machinery" may necessitate structural remodeling that is more drastic or intricate than previously envisioned.

This insight pivots our strategic focus for future designs. Our initial design of the hydrophobic pocket prioritized the direct exclusion of water; however, future work should specifically target the stage of competition with hydrolysis: the attack by the peptide nucleophile substrate. Characterizing the kinetics of this peptide's attack will provide a direct and far more meaningful metric for assessing our progress in creating a truly competitive ligation pathway.

DATA AND CODE AVAILABILITY

The modified versions of PocketGen and LigandMPNN used in this study will be made publicly available on the project Wiki. Software including AMBER and ORCA can be obtained free of charge through academic licensing. GROMACS is freely available at https://www.gromacs.org/.

ACKNOWLEDGEMENTS

Bibliography

[1] T. M. S. Tang and L. Y. P. Luk, "Asparaginyl endopeptidases: enzymology, applications and limitations," *Org. Biomol. Chem.*, vol. 19, no. 23, pp. 5048–5062, 2021, doi: 10.1039/D1OB00608H.

[2] R. B. Kapust *et al.*, "Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency," *Protein Engineering*, vol. 14, no. 12, pp. 993–1000, 2001, doi: 10.1093/protein/14.12.993.

[3] X. Fan *et al.*, "Quantitative Analysis of the Substrate Specificity of Human Rhinovirus 3C Protease and Exploration of Its Substrate Recognition Mechanisms," *ACS Chemical Biology*, vol. 15, no. 1, pp. 63–73, 2020, doi: 10.1021/acschembio.9b00539.

[4] A. M. Weeks and J. A. Wells, "Subtiligase-Catalyzed Peptide Ligation," *Chemical Reviews*, vol. 120, no. 6, pp. 3127–3160, 2020, doi: 10.1021/acs.chemrev.9b00372.

[5] S. Liebscher *et al.*, "N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis," *Angewandte Chemie International Edition*, vol. 53, no. 11, pp. 3024–3028, 2014, doi: https://doi.org/10.1002/anie.201307736.

[6] Y. Feng *et al.*, "Aqualigase: A Star Enzyme for One-Step Peptide Bond Dehydration Condensation in a Nature Aqueous Phase," *ACS Catalysis*, vol. 15, no. 13, pp. 11594–11607, 2025, doi: 10.1021/acscatal.5c01532.

[7] N. Suree *et al.*, "The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized*," *Journal of Biological Chemistry*, vol. 284, no. 36, pp. 24465–24477, 2009, doi: https://doi.org/10.1074/jbc.M109.022624.

[8] G. K. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian, and J. P. Tam, "Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis," *Nature chemical biology*, vol. 10, no. 9, pp. 732–738, 2014.

[9] X. Hemu *et al.*, "Structural determinants for peptide-bond formation by asparaginyl ligases," *Proceedings of the National Academy of Sciences*, vol. 116, no. 24, pp. 11737–11746, 2019, doi: 10.1073/pnas.1818568116.

[10] X. Hemu *et al.*, "Turning an Asparaginyl Endopeptidase into a Peptide Ligase," *ACS Catalysis*, vol. 10, no. 15, pp. 8825–8834, 2020, doi: 10.1021/acscatal.0c02078.

[11] E. H. Abdelkader and G. Otting, "NT*-HRV3CP: An optimized construct of human rhinovirus 14 3C protease for high-yield expression and fast affinity-tag cleavage," *Journal of Biotechnology*, vol. 325, pp. 145–151, 2021, doi: https://doi.org/10.1016/j.jbiotec.2020.11.005.

[12] M. Tiberti *et al.*, "MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles," *Briefings in bioinformatics*, vol. 23, no. 3, p. bbac74, 2022.

[13] J. Abramson *et al.*, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, no. 8016, pp. 493–500, 2024.

[14] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2010, doi: https://doi.org/10.1002/jcc.21334.

[15] T. Lu, "A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn," *The Journal of Chemical Physics*, vol. 161, no. 8, p. 82503, 2024, doi: 10.1063/5.0216272.

[16] T. Lu, "Tian Lu, Sobtop, Version 1.0 (dev 5) http://sobereva.com/soft/Sobtop (accessed on 2 10,2025)," vol. 0, p. , 2025.

[17] M. J. Abraham *et al.*, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, pp. 19–25, 2015, doi: https://doi.org/10.1016/j.softx.2015.06.001.

[18] C. Tian *et al.*, "ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution," *Journal of Chemical Theory and Computation*, vol. 16, no. 1, pp. 528–552, 2020, doi: 10.1021/acs.jctc.9b00591.

[19] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building Water Models: A Different Approach," *The Journal of Physical Chemistry Letters*, vol. 5, no. 21, pp. 3863–3871, 2014, doi: 10.1021/jz501780a.

[20] D. A. Case *et al.*, "AmberTools," *Journal of Chemical Information and Modeling*, vol. 63, no. 20, pp. 6183–6191, 2023, doi: 10.1021/acs.jcim.3c01153.

[21] J. Wang and Y. Miao, "Ligand Gaussian Accelerated Molecular Dynamics 3 (LiGaMD3): Improved Calculations of Binding Thermodynamics and Kinetics of Both Small Molecules and Flexible Peptides," *Journal of Chemical Theory and Computation*, vol. 20, no. 14, pp. 5829–5841, 2024, doi: 10.1021/acs.jctc.4c00502.

[22] Z. Zhang, W. X. Shen, Q. Liu, and M. Zitnik, "Efficient generation of protein pockets with PocketGen," *Nature Machine Intelligence*, pp. 1–14, 2024.

[23] Z. Lin *et al.*, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, no. 6637, pp. 1123–1130, 2023, doi: 10.1126/science.ade2574.

[24] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno, "gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS," *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6281–6291, 2021, doi: 10.1021/acs.jctc.1c00645.

[25] J. L. Watson *et al.*, "De novo design of protein structure and function with RFdiffusion," *Nature*, vol. 620, no. 7976, pp. 1089–1100, 2023.

[26] J. Dauparas *et al.*, "Atomic context-conditioned protein sequence design using LigandMPNN," *Biorxiv*, pp. 2023–2012, 2023.

[27] T. Nuijens, A. Toplak, M. Schmidt, A. Ricci, and W. Cabri, "Natural occurring and engineered enzymes for peptide ligation and cyclization," *Frontiers in Chemistry*, vol. 7, p. 829, 2019.

[28] R. Warden-Rothman, I. Caturegli, V. Popik, and A. Tsourkas, "Sortase-Tag Expressed Protein Ligation: Combining Protein Purification and Site-Specific Bioconjugation into a Single Step," *Analytical Chemistry*, vol. 85, no. 22, pp. 11090–11097, 2013, doi: 10.1021/ac402871k.