De novo design of metalloproteases for targeted amyloid-β cleavage

Yannan Qu [1, 2, †], Chentong Wang [1, 2, †], Hongli Zhu[1,2], Yanjun Wang[1,2], Longxing Cao[1,2*]

1. State Key Laboratory of Gene Expression, School of Life Sciences, Westlake University, Hangzhou 310024, China.

2. Artificial Intelligence Drug Design Core Laboratory, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, 310024, China.

*Corresponding author. E-mail: caolongxing@westlake.edu.cn

† These authors are co-first authors.

Abstract

De novo protein design has not yet achieved the creation of proteases capable of selectively cleaving any desired peptide bond within a native protein with high precision. Here, we report the use of the flow-based generative model Proteus2 to design metalloproteases by generating enzyme–substrate complexes conditioned on a target sequence and a predefined catalytic motif. Our approach employs a two-step encapsulation strategy to create clamp-like metalloproteases that bind the target peptide in a manner that maximizes substrate sequence specificity. This generative process simultaneously optimizes the precise positioning of the target peptide bond in a catalytically competent configuration and accurately scaffolds the transition state catalytic residues—both essential for specific and efficient catalysis. Using this strategy, we designed zinc metalloproteases targeting three distinct cleavage sites within the aggregation-prone regions of amyloid-β (Aβ), a key pathogenic factor in Alzheimer's disease. Experimental characterization validated five enzymes, each capable of precise cleavage at the intended sites with high specificity and minimal or undetectable activity on non-cognate substrates. On average, these enzymes accelerated peptide bond hydrolysis by more than 10^7-fold relative to the uncatalyzed reaction. Together, these results demonstrate the potential of sequence-guided, generative approaches for developing programmable, sequence-specific proteolysis and lay the groundwork for future applications in basic research and therapeutic development.

Introduction

Proteases are among biology's most powerful regulators[1,2]. By cutting specific peptide bonds they activate hormones[3], terminate signalling cascades[4], remodel extracellular matrices[5], and control protein homostasis[1]. Their catalytic activity also underpins a wide spectrum of industrial and medical applications—from laundry detergents and food processing to wound debridement and thrombolytic therapy[6,7]. Yet the utility of natural proteases is circumscribed by their hard-wired substrate preferences[8,9]. An enduring objective is therefore to create "molecular scalpels" whose cleavage site can be programmed at will, enabling precise dissection of cellular pathways and highly selective therapeutic intervention[10,11]. Despite decades of effort, very few platform delivers custom proteases that combine robust catalytic efficiency with single-bond specificity[12].

Metalloproteases are prime candidates for such re-engineering, and their catalytic chemistry has been well studied[13-15]. In a typical zinc metalloprotease, the zinc ion is coordinated by three amino-acid ligands and positioned adjacent to a general base that activates a water molecule for nucleophilic attack on the peptide bond, while a neighbouring residue stabilises the tetrahedral oxyanion intermediate[15] (**Fig. 1a**). This minimalist engine is embedded in diverse scaffolds—including matrix metalloproteinases[16], ADAMs[17], astacins[18], thermolysins[19], and bacterial virulence factors[20]—that collectively drive processes from extracellular-matrix turnover to cytokine maturation. Attempts to retarget protease through directed evolution[12,21,22] or rational mutagenesis have met with limited success: substrate specificity is typically distributed across flexible loops that are conformationally coupled to the active site, so mutations that alter substrate recognition often compromise catalysis. Exhaustive screening can explore only a minute fraction of sequence space[23] and rarely achieves single-bond precision, leaving off-target

cleavage a persistent concern[24,25]. Consequently, repurposed natural metalloproteases have so far provided only incremental gains in selectivity and limited therapeutic potential[26].

De novo protein design holds the promise of sidestepping the intrinsic limitations of natural enzymes by constructing entirely new scaffolds around any chosen chemical task[27-31]. Recent breakthroughs in deep-learning based structure generation have indeed delivered artificial metallohydrolases whose catalytic efficiencies approach those of native enzymes[28,30,32]; however, these successes have been confined almost exclusively to model reactions such as the hydrolysis of activated ester bonds in small, hydrophobic substrates—reactions that present relatively low activation barriers and low specificity demands. Extending this progress to true proteolysis is vastly more challenging. Amide bonds are orders of magnitude more inert than esters[33,34], and a protease must not only accelerate their cleavage but also capture a highly flexible, polar peptide and orient one particular bond in a near-attack conformation. Moreover, practical applications require a high level of sequence selectivity, such that the engineered substrate binding groove preferentially accommodates the target sequence while minimizing activity toward closely related peptides. Natural proteases achieve this balance with folds that have been evolutionarily honed for cognate substrates; even slight structural perturbations can disrupt the sub-angstrom precision of the catalytic residues[9,35]. Consequently, conventional sequence redesign seldom retargets these enzymes while maintaining native-like catalytic efficiency[36]. The central challenge, therefore, is to design a new scaffold that simultaneously preserves the strict metal-dependent catalytic geometry, locks the scissile bond into position, and rigorously discriminates against off-target sequences—an objective that has so far remained out of reach.

Computational strategy for designing sequence specific metalloprotease

We set out to develop a general strategy for the de novo design of sequence-specific metalloproteases. To establish catalytic motifs for design, we focused on two representative zinc-dependent endopeptidases: PPEP-1 (PDB: 6r4w) [37] and Fungalysin (PDB: 4k90) [38]. Both enzymes are secreted by microorganisms and play critical roles in survival or pathogenicity by hydrolyzing extracellular host or cell-surface proteins. Despite sharing a common catalytic mechanism, these two enzymes exhibit distinct active site architectures. Notably, their oxyanion hole compositions differ: PPEP-1 employs a tyrosine, whereas Fungalysin utilizes a histidine residue. Structurally, PPEP-1 features a broad, clamp-like groove formed by two flexible loops that close over the substrate, conferring high sequence specificity for cleavage at Pro-Pro amide bonds. To reprogram this specificity, we developed a strategy that removes the peptide-binding loops from PPEP-1, while retaining the core protein structure (**Fig. 1b**). The excised regions were then rebuilt, enabling the design of new peptide-binding interfaces. Proteases designed through this approach are denoted “PP” (peptide-binding pocket designed protease) followed by a numerical identifier. In contrast, Fungalysin lacks an extended, pre-formed substrate-binding pocket and instead presents a relatively open substrate binding groove, which imparts a broader specificity for peptide bonds flanked by hydrophobic side chains. For this scaffold, we extracted only the minimalist catalytic motif—comprising the zinc-coordinating residues, the general base glutamate, and the oxyanion hole histidine (**Fig. 1b**). This minimalist approach necessitates de novo generation of the entire surrounding protein scaffold to accurately position the catalytic motif and design a novel peptide-binding interface for sequence-specific recognition. Proteases generated using this minimalist scaffold-based motif are designated “DP” (de novo protease).

To enable precise co-design of enzyme functionality and substrate specificity, we employed Proteus2, a generative flow-based model significantly advanced over its predecessor[39]. To explicitly endow the structural generative process with a deep understanding of intrinsic sequence constraints and specificity, we integrated a customized protein language model pre-trained on UniRef50[40] with an auxiliary structural decoding objective (see Methods). Proteus2 was trained on experimental structures from the Protein Data Bank[41] and high-confidence AlphaFold2[42] predictions using a dual-modality motif scaffolding task. In this framework, sequence and structure conditions are treated as independent masking tracks, leveraging geometry-aware sequence semantics to control structure generation. This decoupled conditioning strategy empowers the model to process hybrid inputs comprising structurally defined catalytic motifs and coordinate-free substrate sequences, thereby generating metalloproteases with stabilized catalytic sites and precise substrate specificity.

We leveraged Proteus2 to co-generate metalloprotease–substrate complex structures using the extracted active motif configurations as templates. To determine the catalytically competent conformation of the substrate relative to the catalytic residues, we used AlphaFold3[43] to model the substrate peptide configuration within the template protease (**Extended Data Fig. 1a**). The predictions revealed that the substrate adopted a β-strand conformation in both enzymes. From these predicted structures, we extracted the P2–P1′ backbone atoms of the Fungalysin substrate peptide (where P1 is the residue immediately preceding, and P1′ immediately following, the cleavage site), and the P3–P1 backbone atoms of the PPEP-1 substrate. In both cases, these segments form strand pairing interactions with the enzyme's strand adjacent to the active site,

thereby enabling the construction of a complete target active site. To maximize structural complementarity and specificity for the substrate peptide, we developed a 'two-step encapsulation' generative diffusion strategy (**Fig. 1b**). Unlike conventional one-step diffusion approaches—which frequently yield suboptimal packing and limited control over overall enzyme topology—our method proceeds sequentially. The first step generates the protein backbone surrounding the catalytic glutamate, stabilizing the catalytic base and achieving partial pre-wrapping of the substrate. In the second step, the scaffold is constructed around the oxyanion hole to complete substrate encapsulation. This stepwise process minimizes steric clashes, substantially enhances the packing density at the enzyme–substrate interface, and consistently generates clamp-like enzymes that precisely grasp the substrate peptide—thereby improving sequence-specific recognition (**Fig. 1c**). Following backbone generation, protein sequences were optimized using LigandMPNN[44], and the resulting protease–substrate–zinc complex structures were validated with AlphaFold3 to ensure convergence to the intended catalytic geometry.

De novo proteases facilitate the direct hydrolysis of Aβ peptide

We selected the amyloid-β (Aβ) peptide as the substrate for design, given its central role in Alzheimer's disease (AD) pathology, where the aggregation of neurotoxic Aβ plaques is a hallmark event[45]. The pathological self-assembly of Aβ is primarily driven by intermolecular interactions within two hydrophobic regions: the central hydrophobic core (residues 17–21) and the C-terminal segment (residues 30–42) (**Fig. 2a**). These domains act as critical nucleation sites for β-sheet stacking, facilitating the formation of stable, neurotoxic fibrils[46,47]. Consequently, these regions represent key therapeutic targets for catalytic interventions aimed at destabilizing amyloid architecture and mitigating AD progression[48]. To demonstrate the generalizability of our

computational methodology, we selected three distinct cleavage sites within Aβ: one within the central hydrophobic core and two within the C-terminal hydrophogic region. We employed both the peptide-binding pocket (PP) and de novo protease (DP) strategies to design metalloproteases targeting each site. On average, 26 designs were selected from each strategy–substrate pair for experimental characterization (**Extended Data Table 2**). These designed enzymes were expressed in *Escherichia coli*, purified using Ni-NTA magnetic beads, with the His-tag subsequently removed via HRV3C protease. Final purification was performed by size exclusion chromatography. To assess catalytic activity, we engineered fusion substrates in which the targeted Aβ fragment was inserted between the mannose-binding protein (MBP) and a small ubiquitin-like modifier (SUMO) tag. Successful proteolytic cleavage results in two lower-molecular-weight fragments, readily detectable by SDS-PAGE (**Fig. 2b**). Upon overnight incubation of the designed proteases with their respective substrates in the presence of zinc, we observed that the DP strategy yielded active enzymes for all three targeted sites: DP720-S1, DP622-S2, and DP221-S3. In contrast, the PP strategy generated active designs only for the first site, resulting in PP507-S1 and PP532-S1 (**Fig. 2c**). We attribute the limited success of the PP designs at the remaining sites to steric constraints imposed by excessive preservation of the template backbone, which may be incompatible with the sequence and geometry of these cleavage sites. In contrast, the DP strategy allows broader sampling of the conformational space, facilitating the discovery of active variants. Proteolytic specificity was confirmed by mass spectrometry analysis of the cleavage products, with observed molecular weights matching the theoretical predictions and validating site-specific cleavage (**Extended Data Fig. 2a-e**). Notably, for DP221-S3, in addition to the major mass spectrometric peak corresponding to the intended cleavage products, a very minor peak—representing approximately 10% of the main product's

abundance—was also observed, indicating an alternative cleavage site shifted by one residue toward the C-terminus (**Extended Data Fig. 2e**). We next characterized their catalytic efficiency by monitoring the time course of substrate cleavage (**Extended Data Fig. 3**). At a 1:2 enzyme-to-substrate molar ratio, DP622-S2 exhibited the highest activity, achieving complete substrate degradation within 4 hours (**Extended Data Fig. 3b**). DP720-S1 and PP507-S1 achieved full digestion in 8 hours (**Extended Data Fig. 3a,c**), while PP532-S1 required nearly 12 hours for complete cleavage (**Extended Data Fig. 3e**). In contrast, DP221-S3 displayed slower kinetics, cleaving more than half of the substrate after 12 hours (**Extended Data Fig. 3d**). Collectively, these results indicate our design strategies enables the de novo generation of sequence-specific metalloproteases capable of targeting disease-relevant sites within Aβ.

Validation of catalytic mechanism

To elucidate the catalytic mechanism and confirm the functional roles of key active site residues, we conducted a series of site-directed mutagenesis experiments (**Fig. 3a**). For designs generated by both the DP and PP strategies, the catalytic glutamate and oxyanion hole residues (histidine in the DP designs and tyrosine in the PP designs) were mutated to alanine (**Fig. 3b**). Additionally, in the DP designs, we mutated the newly generated second-sphere residues that form hydrogen bonds with the zinc-coordinating residues, to evaluate their contribution to catalytic activity. Functional assays revealed that substitution of the catalytic glutamate with alanine universally abolished proteolytic activity across both DP and PP designs, highlighting its indispensable role as the general base in catalysis (**Fig. 3b**). Interestingly, in the DP designs, mutation of the oxyanion hole histidine to alanine had minimal or even positive effects on protease function; notably, the H191A variant of DP221-S3 exhibited enhanced substrate degradation. By contrast,

in the PP designs, alanine substitution of the oxyanion hole tyrosine resulted in partial reduction of activity. These findings suggest that the oxyanion stabilization function may be partially compensated by solvent molecules in the absence of the native residue. Furthermore, mutagenesis of the second coordination sphere residues in the DP designs underscored their importance in tuning catalytic efficiency. The D128A mutation in DP720-S1 and the N111A mutation in DP221-S3 each resulted in a marked reduction in activity. In DP622-S2, single mutations (D126A or Y91F) had minimal impact on catalytic activity, whereas the double mutant (D126A/Y91F) led to a dramatic decrease in activity (**Fig. 3b**). These results suggest that second-shell interactions play context-dependent roles in stabilizing the primary coordination environment and modulating the electrostatic landscape of the active site. Collectively, these findings support the conclusion that the observed catalytic activity arises from the designed active sites and highlight the importance of precise active site architecture for efficient catalysis.

Characterization of the catalytic kinetics

To quantitatively assess the kinetics of our designed proteases, we developed a specialized fluorescence-based assay. The intrinsic hydrophobicity of the Aβ target sequences posed significant challenges for synthesizing soluble peptides compatible with standard fluorophore-quencher systems, due to marked solubility and aggregation issues. As an alternative, we engineered a FRET-based biosensor by inserting the targeted Aβ sequences directly between the fluorescent protein pair mTurquoise2 (donor) and mVenus (acceptor) (**Extended Data Table 3**). In this system, enzymatic cleavage of the substrate separates the fluorescent proteins, resulting in a measurable decrease in FRET efficiency. Kinetic parameters were determined by fitting initial reaction velocities to the Michaelis-Menten model (**Fig. 3c**). Analysis revealed that the turnover

numbers (k_{cat}) for our designs were generally on the order of 10^{-4} s^{-1}, with Michaelis constants (K_m) in the micromolar range. These k_{cat} values represent an average > 10^7-fold rate enhancement relative to the uncatalyzed hydrolysis of peptide bonds[33]. Among the designed proteases, DP622-S2 demonstrated the highest catalytic efficiency, achieving a k_{cat}/K_m of 160.8 $M^{-1}s^{-1}$, driven by a favorable combination of a high k_{cat} (0.0009 s^{-1}) and a low K_m (5.36 μM). The PP507-S1 and PP532-S1 exhibited moderate efficiencies of 28.4 and 24.3 $M^{-1}s^{-1}$, respectively, while DP221-S3 and DP720-S1 displayed the lowest efficiencies at 4.2 and 3.8 $M^{-1}s^{-1}$. With the exception of DP720-S1, these kinetic findings were in close concordance with our time-course degradation assays. The discrepancy observed for DP720-S1—which showed effective substrate degradation comparable to PP507-S1 in previous assays but exhibited low catalytic efficiency in the FRET system—may be attributable to steric hindrance, as the bulky fluorescent proteins could restrict substrate access to the active site. Notably, the K_m value of our lead design, DP622-S2, closely parallels those of naturally occurring Aβ-degrading proteases such as Insulin-Degrading Enzyme, which also exhibit K_m values in the low micromolar range[49]. Although the turnover rates of our designs are lower than those of natural enzymes, achieving micromolar substrate affinity demonstrates that our computational strategy can successfully encode robust recognition for disease-relevant targets.

To further elucidate the functional roles of the oxyanion hole and second coordination sphere residues, we characterized the catalytic kinetics of the corresponding mutants, including single mutants of DP622-S2 (Y91F, D126A, or H172A), DP221-S3 (H191A), as well as the DP622-S2 double mutant (Y91F/D126A) (**Extended Data Fig. 4a-e**). Intriguingly, all tested single mutations in the DP622-S2 series (Y91F, D126A, and H172A) (**Extended Data Fig. 4a-c**)

showed enhanced catalytic efficiency (k_{cat}/K_m) compared to the parental design, whereas the double mutant Y91F/D126A decreased catalytic efficiency (**Extended Data Fig. 4d**), consistent with the SDS-PAGE results. The observed improvements were predominantly driven by increased turnover numbers (k_{cat}); notably, the DP622-S2 D126A variant exhibited a threefold increase in k_{cat} (**Extended Data Fig. 4b**). This suggests that optimizing the catalytic microenvironment can significantly modulate the activation energy barrier of the rate-limiting step. Most mutations led to a modest decrease in substrate affinity, as reflected by an increase in K_m; however, the DP622-S2 Y91F variant maintained a K_m nearly identical to the original, resulting in a twofold enhancement in overall catalytic efficiency **(Extended Data Fig. 4a)**. The DP221-S3 H191A mutation had a synergistic effect, increasing k_{cat} while simultaneously decreasing K_m, resulting in a 2.7-fold improvement in catalytic efficiency **(Extended Data Fig. 4e)**. Together, these findings indicate that second-shell coordination and oxyanion hole residues have a substantial impact on catalytic efficiency, largely by modulating k_{cat}—a parameter that remains challenging to accurately model in current design protocols.

Characterization the substrate specificity

We next assessed the substrate specificity of the designed metalloproteases by evaluating their cleavage activity against three distinct MBP–Aβ_segment–SUMO fusion constructs (Substrates 1, 2, and 3) (**Fig. 4a-e**). SDS-PAGE analysis of the cleavaged products revealed that PP507-S1 exhibited the highest degree of selectivity, catalyzing cleavage exclusively of Substrate 1, with no detectable activity against the other two substrates (**Fig. 4c**). In contrast, PP532-S1 displayed broader substrate tolerance, efficiently processing all three substrates (**Fig. 4e**), while PP720-S1 showed measurable, though reduced, activity against Substrate 2 (**Fig. 4a**). DP622-S2 and

DP221-S3 exhibited partial cross-reactivity with each other's substrates, albeit at considerably lower efficiencies compared to their cognate targets (**Fig. 4b,d**). We speculate that the observed cross-reactivity for DP622-S2 and DP221-S3 may be attributed to extensive sequence overlap between the target sequences and the use of glycine-serine (GS) linkers flanking the inserted Aβ segments, which may facilitate recognition of similar sequence motifs across different constructs (**Extended Data Table 1**). Notably, the flanking sequence of the S1 cleavage site contains a high proportion of bulky aromatic and large hydrophilic residues, whereas the regions flanking the S2 and S3 sites are primarily composed of small aliphatic amino acids and glycines. This composition may underlie the high specificity of PP507 and likely prevents the S2/S3 enzymes from cleaving the S1 substrate due to steric occlusion.

We next evaluated the substrate specificity of the active site mutants (**Extended Data Fig. 5a-c**). The second-shell and oxyanion hole mutants in DP622 and DP720 exhibited specificity profiles similar to those of their respective parent designs (**Extended Data Fig. 5a,b**). Notably, the DP221-S3 H191A variant showed a remarkable increase in selectivity, virtually abolishing cleavage activity toward non-target substrates (**Extended Data Fig. 5c**). However, mass spectrometry analysis of the cleavage products revealed that the scissile bond shifted almost entirely to the original minor cleavage site—one residue closer to the C-terminus—compared to the parent design (**Extended Data Fig. 6b**). In contrast, the DP622-S2 Y91F mutant retained the intended cleavage site, consistent with the parental enzyme (**Extended Data Fig. 6a**). Nonetheless, the stringent specificity demonstrated by PP507-S1 underscores the potential of our approach to generate custom, sequence-specific proteases. We anticipate that future iterations

incorporating explicit negative design strategies during the design and filtering stages will further enhance substrate specificity.

Finally, we evaluated the cleavage activity of our most effective designs and select variants—including PP507, DP622 Y91F, and DP221 H191A—against a full-length synthetic Aβ42 peptide. Using a cocktail approach, all three enzymes were incubated together with the substrate. Mass spectrometry analysis revealed the presence of the expected cleavage fragments—DAEFRHDSGYEVHHQKL, VFFAEDVGSNKGAIIGL, and VVIA—demonstrating that our designed proteases successfully mediated site-specific proteolysis of Aβ42 as designed (**Extended Data Fig. 7**). These results confirm that our de novo metalloproteases are capable of catalyzing the hydrolysis of the full-length Aβ42 peptide.

Discussion

By explicitly integrating target sequence information into the flow-based structure generation process, Proteus2 enables the direct design of sequence-specific metalloproteases. Our two-step encapsulation strategy facilitates the creation of clamp-like proteases that envelop the substrate peptide, thereby maximizing both specificity and discrimination. The successful design of metalloproteases capable of cleaving three distinct native Aβ peptide sequences extends de novo protein design beyond stable binding, enabling functional proteolysis of physiological targets. Achieving strict substrate exclusivity in de novo scaffolds remains a formidable challenge, far surpassing the establishment of catalytic activity alone. The development of PP507-S1—a design that displays high specificity toward its intended substrate—demonstrates that our computational approach can indeed achieve precise discrimination. Our specificity characterization also

indicates that, for future enzyme design, careful selection of the target site is critical; regions enriched in distinctive side chains, such as bulky aromatic or highly charged residues, may facilitate the development of enzymes with higher specificity. While a certain degree of nonspecific cleavage was seen in other variants, the clamp-like architecture provides a robust starting point for further sequence redesign or structure-guided refinement to enhance specificity.

Our successful de novo design of metalloproteases, starting from minimalist catalytic motifs, represents a major advance in the de novo enzyme design field by bridging the gap between model reactions and catalysis relevant for therapeutic intervention. On average, our designs accelerate peptide bond cleavage by more than 10^7-fold compared to the uncatalyzed reaction in aqueous solution[33]. Importantly, these metalloproteases exhibit substrate affinities (K_m) comparable to those of natural enzymes, indicating that our design strategies can accommodate the structural and chemical complexity of physiological polypeptides[49]. However, our mutational analysis also indicates that key microenvironmental factors affecting turnover rates (k_{cat}) remain challenging to accurately model with current approach, and suggest that further improvements may require consideration of solvent effects, second-shell coordination, and the dynamic catalytic process to enhance catalytic proficiency.

The aggregation of Aβ into neurotoxic plaques is a defining event in Alzheimer's disease (AD) pathology, making the development of molecular tools to counteract Aβ accumulation a high priority. Current strategies for Aβ clearance, predominantly based on monoclonal antibodies[50], act by sequestering Aβ for immune-mediated removal rather than direct peptide degradation[51]. In contrast, catalytic degradation offers a compelling alternative, as a single enzyme can process

multiple Aβ molecules through direct hydrolysis. Although endogenous proteases like neprilysin and insulin-degrading enzyme can target Aβ[52,53], their broad specificity presents a risk for undesirable off-target effects when therapeutically enhanced[54,55]. Our work provides a fundamentally distinct strategy—fully de novo design of programmable, sequence-specific proteases for Aβ peptide degradation—enabling the selective hydrolysis of pathogenic sequences and facilitating efficient monomer clearance. By targeting cleavage within the hydrophobic, aggregation-prone regions of Aβ, these proteases have the potential to inhibit the formation of toxic fibrils. Notably, the cleavage site for our most specific design, PP507-S1, is only one residue away from the α-secretase site responsible for the non-pathogenic processing of APP, showcasing the potential for precision engineering[56]. As de novo designed enzymes can target different sites, our multi-enzyme cocktail approach may reveal new possibilities for developing therapeutic interventions for AD.

Overall, our work establishes a robust and generalizable framework for the de novo design of programmable, sequence-specific metalloproteases with wide-ranging applications in research and medicine. Our design approach can be readily adapted to other protease classes, such as serine or cysteine proteases, opening new possibilities for selective protein manipulation and targeted protein degradation that are unattainable with naturally occurring enzymes. Looking ahead, the integration of explicit negative design strategies, machine learning-guided optimization, and advanced transition-state modeling will likely yield next-generation proteases with even higher specificity and catalytic efficiency. Ultimately, this technology sets the stage for the rational development of “designer” proteases for a broad spectrum of diseases and empowers

both basic research and therapeutic innovation by providing unprecedented control over protein function in living systems.

Code availability

The Rosetta macromolecular modelling suite (https://www.rosettacommons.org) is freely available to academic and commercial users. Design protocols, analysis scripts, and design models used in this paper will be made available upon final publication.

Acknowledgements

We extend our sincere thanks to the Dang Lab at Westlake University for their valuable guidance and technical expertise in peptide synthesis. We would like to thank the Mass Spectrometry Facility of Westlake University for help in sample analysis; and the Westlake University HPC Center for computation assistance.

Author contributions

Y.Q., C.W. and L.C. designed the research. Y.Q., C.W. and L.C. developed the design pipeline and made the designs. Y.Q., H.Z. and Y.W. pefromed the experiments. All authors analysed data. L.C. supervised research. Y.Q., C.W. and L.C. wrote the manuscript with the input from the other authors. All authors revised the manuscript.

Competiting interests

A provisional patent application will be filed prior to publication.

Funding sources

This work was supported by grants from the National Key R&D Program of China (2022YFA1303700, 2021YFC2301401), the Ministry of Science and Technology

(2020YFA0909200), the National Natural Science Foundation of China (32370989, 324B2052), and Westlake University Center of Synthetic Biology and Integrated Bioengineering and Westlake Education Foundation to L.C.

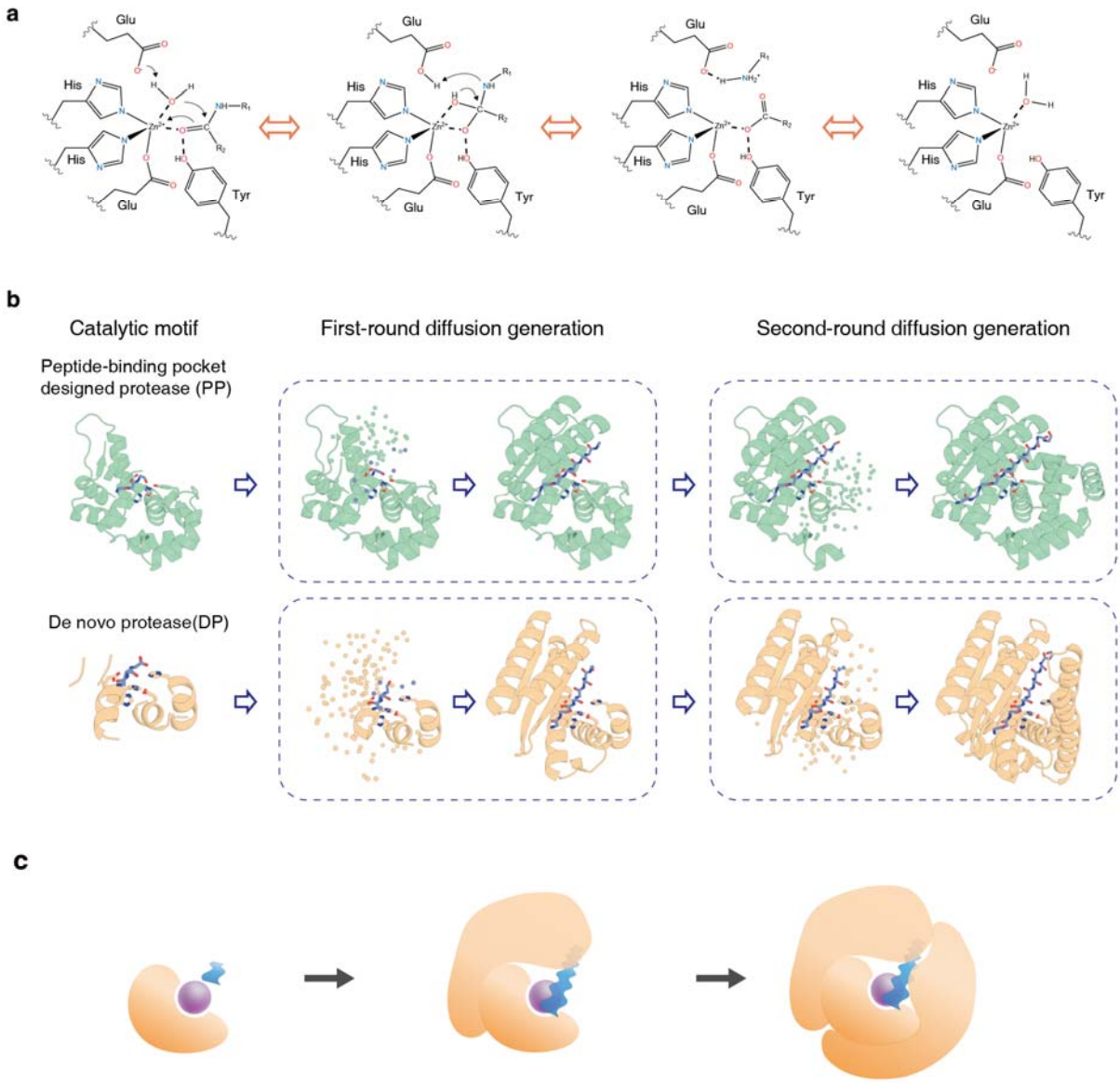

Fig. 1: Catalytic mechanism of zinc-dependent metalloproteases and two-step de novo design of clamp-like protease scaffolds. a, The catalytic cycle of zinc-dependent metalloproteases involves a precisely orchestrated sequence of molecular events at the active site. The reaction initiates as the scissile carbonyl oxygen of the substrate coordinates with the catalytic zinc ion, polarizing the carbonyl and increasing its electrophilicity. Concurrently, a conserved glutamate residue acts as a general base, deprotonating a zinc-bound water molecule to trigger nucleophilic attack on the carbonyl carbon. This generates a transient tetrahedral

intermediate, with its negatively charged oxyanion stabilized via hydrogen bonds within the oxyanion hole—typically formed by a tyrosine or histidine. The catalytic glutamate then functions as a general acid, transferring a proton to the amide nitrogen of the leaving group, thereby facilitating intermediate collapse and peptide bond cleavage. The hydrolyzed products subsequently dissociate, and a new water molecule occupies the zinc coordination sphere, restoring the enzyme for further catalytic cycles. **b,** De novo enzyme design utilizes a two-step encapsulation pipeline based on generative diffusion. In the first stage, diffusion modeling constructs the protein backbone immediately around the catalytic glutamate and the target substrate, stabilizing the general base and initiating "pre-wrapping" of the peptide. A second diffusion stage builds the scaffold around the oxyanion hole, completing molecular encapsulation and locking the substrate into a catalytically competent geometry for proteolysis. **c**, Stepwise generation of a clamp-like architecture. The scaffold is assembled in two sequential stages to achieve complete substrate encapsulation, minimizing steric clashes and ensuring the clamp precisely grasps the target peptide for sequence-specific recognition.

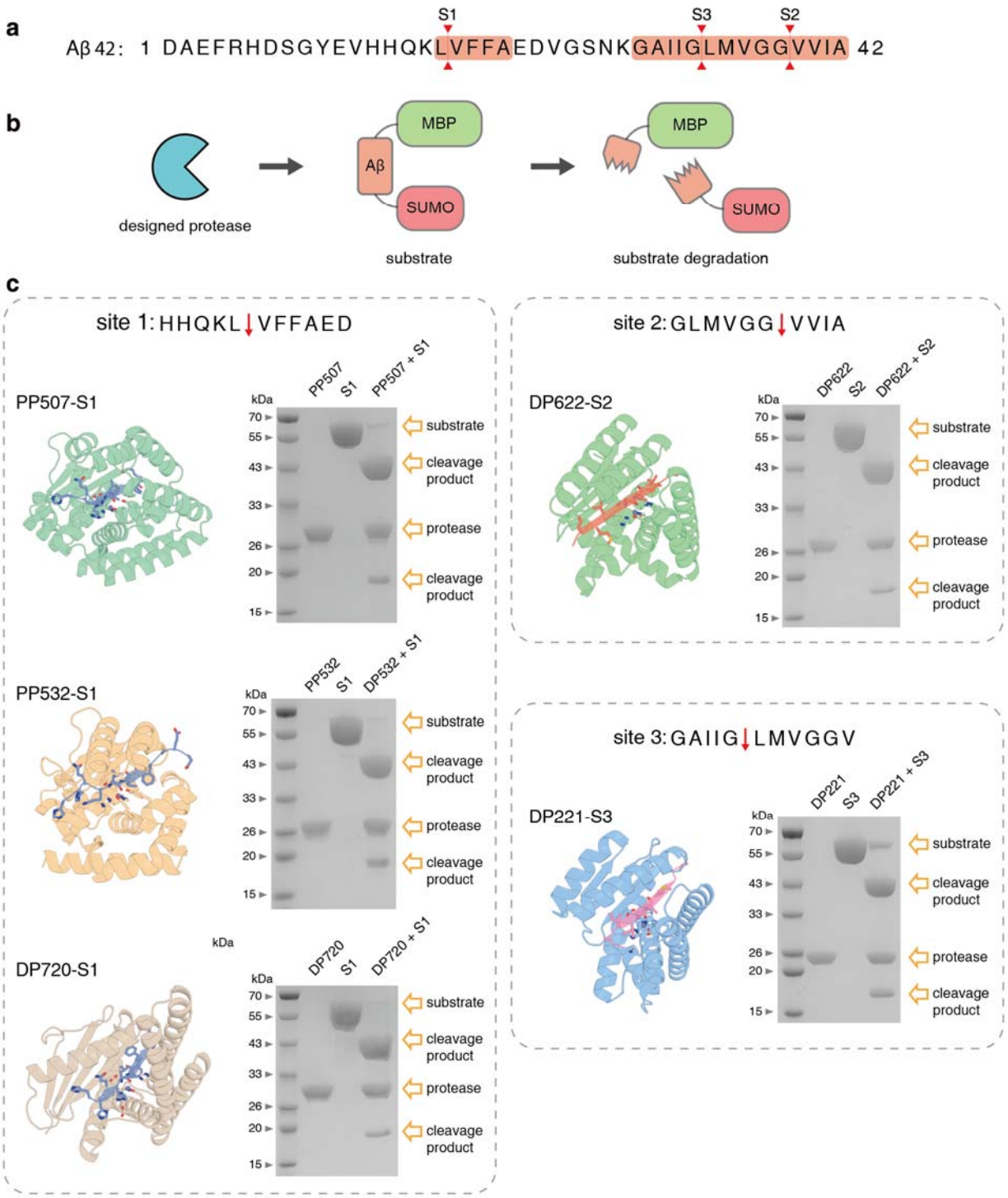

Fig. 2: Experimental validation of designed metalloproteases targeting Aβ. a, Primary sequence of Aβ42. The complete amino acid sequence of Aβ42 is shown, with two main hydrophobic domains highlighted in orange and three selected cleavage sites marked in red. **b**,

Schematic of the proteolysis assay. To facilitate expression and protelytic analysis, Aβ segment was engineered as a fusion protein (MBP-Aβ_segment-SUMO). The diagram illustrates that specific cleavage within the Aβ sequence by the designed proteases results in separation of the MBP and SUMO fragments, which are readily detected via molecular weight shifts in SDS-PAGE experiments. **c**, Design models and proteolytic screening of designed proteases. Design models of five designed proteases are displayed alongside SDS-PAGE analysis, which confirms the successful cleavage of the MBP-Aβ-SUMO substrate.

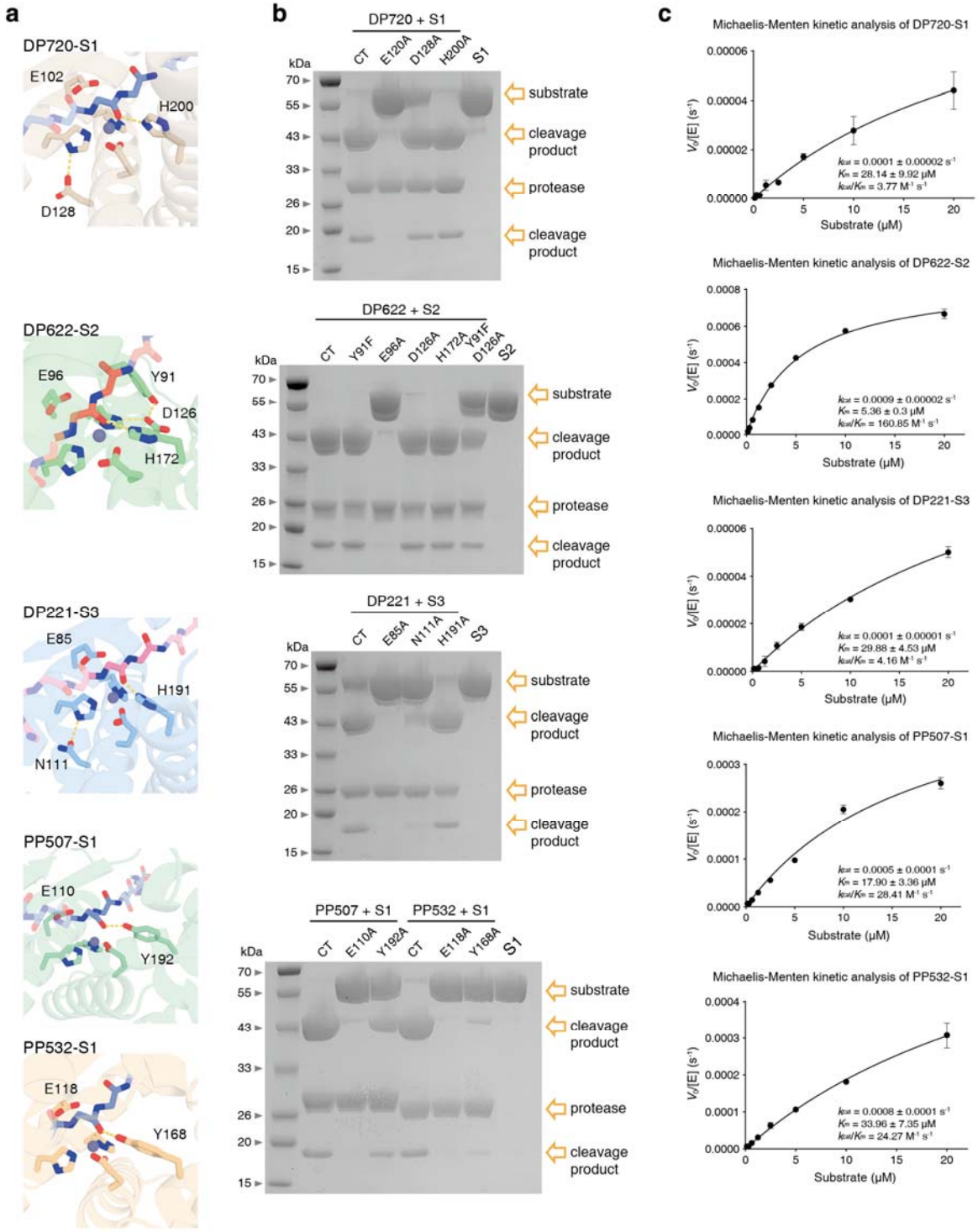

Fig. 3: Structural and kinetic characterization of the designed metalloproteases. a, Atomic details of active site architectures. High-resolution structural models of the designed proteases are shown, highlighting the precise spatial arrangement of key catalytic residues. The

coordination of the zinc ion by histidines and glutamates, the positioning of the general base, and the configuration of the oxyanion hole are illustrated. **b,** Mutational analysis of catalytic residues. SDS-PAGE results of various catalytic residue mutants are presented to verify the functional roles of the designed active site. **c,** Kinetic characterization of the designed proteases. Proteolytic kinetics were assessed by measuring initial reaction velocities at varying substrate concentrations. Data were fitted to the Michaelis-Menten equation to determine key kinetic parameters, including the Michaelis constant (K_m) and turnover number (k_{cat}). The calculated k_{cat}/K_m values quantify the catalytic efficiency of the de novo designed enzymes, demonstrating their ability to recognize and cleave substrates. Data represent mean ± standard deviation (black dots and error bars).

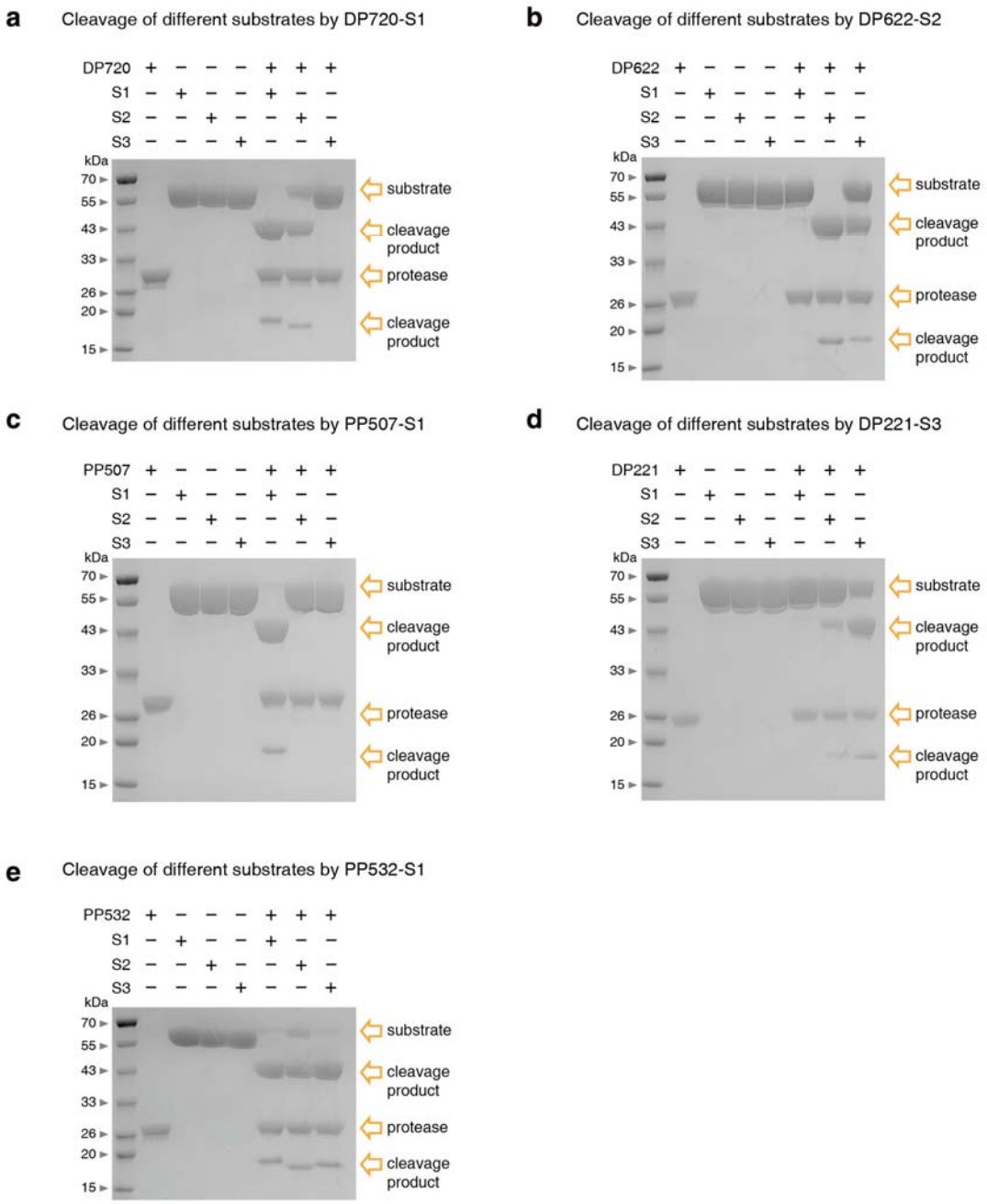

Fig. 4: Proteolytic specificity and substrate selectivity of designed proteases. The substrate recognition profiles of five designed proteases (a, DP720-S1; b, DP622-S2; c, PP507-S1; d, DP221-S3; e, PP532-S1) were systematically evaluated using a panel of three structurally distinct MBP-Aβ_segment-SUMO fusion constructs (Substrates 1, 2, and 3). Results

demonstrate the sequence and site selectivity of each protease toward the corresponding substrates.

Reference

1 Lopez-Otin, C. & Bond, J. S. Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **283**, 30433-30437 (2008). https://doi.org:10.1074/jbc.R800035200

2 Taylor, R. C., Cullen, S. P. & Martin, S. J. Apoptosis: controlled demolition at the cellular level. *Nat Rev Mol Cell Biol* **9**, 231-241 (2008). https://doi.org:10.1038/nrm2312

3 Seidah, N. G. & Prat, A. The biology and therapeutic targeting of the proprotein convertases. *Nat Rev Drug Discov* **11**, 367-383 (2012). https://doi.org:10.1038/nrd3699

4 Turk, B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5**, 785-799 (2006). https://doi.org:10.1038/nrd2092

5 Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**, 786-801 (2014). https://doi.org:10.1038/nrm3904

6 Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr Opin Biotechnol* **13**, 345-351 (2002). https://doi.org:10.1016/s0958-1669(02)00328-2

7 Li, Q., Yi, L., Marek, P. & Iverson, B. L. Commercial proteases: present and future. *FEBS Lett* **587**, 1155-1163 (2013). https://doi.org:10.1016/j.febslet.2012.12.019

8 Perona, J. J. & Craik, C. S. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem* **272**, 29987-29990 (1997). https://doi.org:10.1074/jbc.272.48.29987

9 Hedstrom, L. Serine protease mechanism and specificity. *Chem Rev* **102**, 4501-4524 (2002). https://doi.org:10.1021/cr000033x

10 Craik, C. S., Page, M. J. & Madison, E. L. Proteases as therapeutics. *Biochem J* **435**, 1-16 (2011). https://doi.org:10.1042/BJ20100965

11 Dyer, R. P. & Weiss, G. A. Making the cut with protease engineering. *Cell Chem Biol* **29**, 177-190 (2022). https://doi.org:10.1016/j.chembiol.2021.12.001

12 Blum, T. R. *et al.* Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* **371**, 803-810 (2021). https://doi.org:10.1126/science.abf5972

13 Bode, W., Gomis-Ruth, F. X. & Stockler, W. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett* **331**, 134-140 (1993). https://doi.org:10.1016/0014-5793(93)80312-i

14 Visse, R. & Nagase, H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ Res* **92**, 827-839 (2003). https://doi.org:10.1161/01.RES.0000070112.80711.3D

15 Spyroulias, G. A., Galanis, A. S., Pairas, G., Manessi-Zoupa, E. & Cordopatis, P. Structural features of angiotensin-I converting enzyme catalytic sites: conformational studies in solution, homology models and comparison with other zinc metallopeptidases. *Curr Top Med Chem* **4**, 403-429 (2004). https://doi.org:10.2174/1568026043451294

16 Nagase, H. & Woessner, J. F., Jr. Matrix metalloproteinases. *J Biol Chem* **274**, 21491-21494 (1999). https://doi.org:10.1074/jbc.274.31.21491

17 Blobel, C. P. ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol* **6**, 32-43 (2005). https://doi.org:10.1038/nrm1548

18 Bond, J. S. & Beynon, R. J. The astacin family of metalloendopeptidases. *Protein Sci* **4**, 1247-1261 (1995). https://doi.org:10.1002/pro.5560040701

19 Tronrud, D. E., Roderick, S. L. & Matthews, B. W. Structural basis for the action of thermolysin. *Matrix Suppl* **1**, 107-111 (1992).

20 Miyoshi, S. Extracellular proteolytic enzymes produced by human pathogenic vibrio species. *Front Microbiol* **4**, 339 (2013). https://doi.org:10.3389/fmicb.2013.00339

21 Varadarajan, N., Rodriguez, S., Hwang, B. Y., Georgiou, G. & Iverson, B. L. Highly active and selective endopeptidases with programmed substrate specificities. *Nat Chem Biol* **4**, 290-294 (2008). https://doi.org:10.1038/nchembio.80

22 Gao, Z. *et al.* A gasdermin-based life-death evolution system for reprogramming protease specificity. *Nat Chem Biol* (2025). https://doi.org:10.1038/s41589-025-02063-3

23 Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379-394 (2015). https://doi.org:10.1038/nrg3927

24 Khersonsky, O. & Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* **79**, 471-505 (2010). https://doi.org:10.1146/annurev-biochem-030409-143718

25 Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9**, 690-701 (2010). https://doi.org:10.1038/nrd3053

26 Vandenbroucke, R. E. & Libert, C. Is there new hope for therapeutic matrix metalloproteinase inhibition? *Nat Rev Drug Discov* **13**, 904-927 (2014). https://doi.org:10.1038/nrd4390

27 Yeh, A. H. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774-780 (2023). https://doi.org:10.1038/s41586-023-05696-3

28 Ahern, W. *et al.* Atom-level enzyme active site scaffolding using RFdiffusion2. *Nat Methods* (2025). https://doi.org:10.1038/s41592-025-02975-x

29 Braun, M. *et al.* Computational enzyme design by catalytic motif scaffolding. *Nature* (2025). https://doi.org:10.1038/s41586-025-09747-9

30 Kim, D. *et al.* Computational design of metallohydrolases. *Nature* (2025). https://doi.org:10.1038/s41586-025-09746-w

31 Listov, D. *et al.* Complete computational design of high-efficiency Kemp elimination enzymes. *Nature* **643**, 1421-1427 (2025). https://doi.org:10.1038/s41586-025-09136-2

32 Studer, S. *et al.* Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285-1288 (2018). https://doi.org:10.1126/science.aau3744

33 Radzicka, A. & Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society* **118**, 6105-6109 (1996). https://doi.org:10.1021/ja954077c

34 Kiss, G., Celebi-Olcum, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew Chem Int Ed Engl* **52**, 5700-5725 (2013). https://doi.org:10.1002/anie.201204077

35 Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* **19**, 596-604 (2009). https://doi.org:10.1016/j.sbi.2009.08.003

36 Nestl, B. M. & Hauer, B. Engineering of Flexible Loops in Enzymes. *ACS Catalysis* **4**, 3201-3211 (2014). https://doi.org:10.1021/cs500325p

37 Pichlo, C. *et al.* Molecular determinants of the mechanism and substrate specificity of Clostridium difficile proline-proline endopeptidase-1. *J Biol Chem* **294**, 11525-11535 (2019). https://doi.org:10.1074/jbc.RA119.009029

38 Fernandez, D., Russi, S., Vendrell, J., Monod, M. & Pallares, I. A functional and structural study of the major metalloprotease secreted by the pathogenic fungus Aspergillus fumigatus. *Acta Crystallogr D Biol Crystallogr* **69**, 1946-1957 (2013). https://doi.org:10.1107/S0907444913017642

39 Wang, C. *et al.* in *Proceedings of the 41st International Conference on Machine Learning* Vol. 235 (eds Salakhutdinov Ruslan *et al.*) 51376--51395 (PMLR, Proceedings of Machine Learning Research, 2024).

40 Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926-932 (2014). https://doi.org:10.1093/bioinformatics/btu739

41 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000). https://doi.org:10.1093/nar/28.1.235

42 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

43 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024). https://doi.org:10.1038/s41586-024-07487-w

44 Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *Nat Methods* **22**, 717-723 (2025). https://doi.org:10.1038/s41592-025-02626-1

45 Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* **8**, 595-608 (2016). https://doi.org:10.15252/emmm.201606210

46 Tjernberg, L. O. *et al.* Arrest of beta-amyloid fibril formation by a pentapeptide ligand. *J Biol Chem* **271**, 8545-8548 (1996). https://doi.org:10.1074/jbc.271.15.8545

47 Jarrett, J. T., Berger, E. P. & Lansbury, P. T., Jr. The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32**, 4693-4697 (1993). https://doi.org:10.1021/bi00069a001

48 Saido, T. & Leissring, M. A. Proteolytic degradation of amyloid beta-protein. *Cold Spring Harb Perspect Med* **2**, a006379 (2012). https://doi.org:10.1101/cshperspect.a006379

49 Mukherjee, A. *et al.* Insulysin hydrolyzes amyloid beta peptides to products that are neither neurotoxic nor deposit on amyloid plaques. *J Neurosci* **20**, 8745-8749 (2000). https://doi.org:10.1523/JNEUROSCI.20-23-08745.2000

50 van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *N Engl J Med* **388**, 9-21 (2023). https://doi.org:10.1056/NEJMoa2212948

51 Sevigny, J. *et al.* The antibody aducanumab reduces Abeta plaques in Alzheimer's disease. *Nature* **537**, 50-56 (2016). https://doi.org:10.1038/nature19323

52 Qiu, W. Q. *et al.* Insulin-degrading enzyme regulates extracellular levels of amyloid beta-protein by degradation. *J Biol Chem* **273**, 32730-32738 (1998). https://doi.org:10.1074/jbc.273.49.32730

53 Iwata, N. *et al.* Metabolic regulation of brain Abeta by neprilysin. *Science* **292**, 1550-1552 (2001). https://doi.org:10.1126/science.1059946

54 Tang, W. J. Targeting Insulin-Degrading Enzyme to Treat Type 2 Diabetes Mellitus. *Trends Endocrinol Metab* **27**, 24-34 (2016). https://doi.org:10.1016/j.tem.2015.11.003

55 Turner, A. J., Isaac, R. E. & Coates, D. The neprilysin (NEP) family of zinc metalloendopeptidases: genomics and function. *Bioessays* **23**, 261-269 (2001). https://doi.org:10.1002/1521-1878(200103)23:3<261::AID-BIES1036>3.0.CO;2-K

56 Esch, F. S. *et al*. Cleavage of amyloid beta peptide during constitutive processing of its precursor. *Science* **248**, 1122-1124 (1990). https://doi.org:10.1126/science.2111583