

Computational design of metalloproteases

Anqi Chen^{1,2,13}, Kejia Wu^{1,2,13*}, Hojae Choi^{1,2,3}, Preetham Venkatesh^{1,2,3}, Samuel J. Pellock^{1,2}, Nikita Hanikel^{1,2}, **Hojeong shin**^{1,2}, Declan Evans^{1,2}, Kieran Didi^{2,4,5}, Lars L. Schaaf^{6,7}, **Cristina Díaz-Perlas**⁸, Brian Coventry^{1,2,9}, Donghyo Kim^{1,2}, Seth M. Woodbury^{1,2,3}, Pengfei Ji^{1,2,10}, Shingo Honda^{1,2}, Asim K. Bera^{1,2}, Hannah Nguyen^{1,2}, Alex Kang^{1,2}, Yanqing Wang^{1,2}, Xinting Li^{1,2}, Stacey Gerben^{1,2}, Lemuel Chang^{1,2}, Xiao Yan¹¹, **Benjamí Oller-Salvia**⁸, Anthony A. Hyman¹, Donald Hilvert^{12*}, David Baker^{1,2,9*}

¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

²Institute for Protein Design, University of Washington, Seattle, WA 98195, USA

³Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA, USA

⁴Department of Computer Science, University of Oxford, Parks Rd, Oxford OX1 3QD, UK

⁵NVIDIA Corp., Santa Clara, USA

⁶Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK

⁷Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, U.K.

⁸**Institut Químic de Sarrià (IQS), Universitat Ramon Llull, Via Augusta 390, Barcelona 08017, Spain**

⁹Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

¹⁰Department of Chemistry, Zhejiang University, Hangzhou, China

¹¹Max Planck Institute of Molecular Cell Biology and Genetics; Dresden, Saxony, 01307, Germany

¹²Laboratory of Organic Chemistry, ETH Zurich, Zurich, Switzerland

¹³These authors contributed equally

*Corresponding authors: kejiawu@uw.edu, hilvert@ethz.ch, dabaker@uw.edu

Abstract

Although significant progress has been made in creating *de novo* metalloenzymes that hydrolyze activated esters^{1,2}, the energetically demanding cleavage of amide bonds has remained a major challenge for enzyme design: amide bonds are significantly more stable than ester bonds, the amine leaving groups in proteins are not activated, and peptide substrates are flexible making them difficult to bind precisely. Here, we report the *de novo* design of zinc proteases from minimal catalytic motifs using RoseTTAFold Diffusion 2 for Molecular Interfaces³, optimized

for both enzyme and protein-protein interaction design. Of 135 computational designs experimentally tested, 36% had activity and cleaved precisely at the intended site. The most active design accelerates peptide bond hydrolysis by more than 10^8 -fold relative to the uncatalyzed reaction, and by over 10^{10} -fold following the introduction of four point mutations that enhanced active-site preorganization⁴. **Building on these capabilities, we designed metalloproteases that specifically cleave human TDP-43, the amyloid- β peptide, and serum amyloid A with rate accelerations up to 9.2×10^8 -fold over background. We illustrate the potential of our approach for bio-orthogonal control over cell state by generating caged cytokines and caged receptor antagonists that are selectively unmasked by our designed proteases.** These results demonstrate that *de novo* enzyme design has advanced well beyond model reactions with activated substrates and open the door to design of proficient metallohydrolases for a wide range of applications in medicine and bioremediation.

Introduction

Metalloproteases utilize a metal ion as a cofactor to activate a water molecule to hydrolyze a peptide bond in a protein^{5,6} (Fig. 1a). These enzymes play essential roles in biological processes such as tissue remodeling^{7,8}, cell migration^{9,10}, and signaling¹¹. Beyond their roles in biology, proteases have widespread uses in industry and medicine in application areas ranging from food¹² to therapeutics^{13–14} and cosmetics¹⁵. Because of these widespread applications, there has been considerable interest in engineering variants of native enzymes with customized activities. *De novo* enzyme design is a potentially powerful approach to design new proteases. However, thus far computational design of *de novo* metallohydrolases has focused on small molecule ester substrates with activated leaving groups^{1,2}. Achieving efficient amide bond hydrolysis poses a far greater challenge. Amides are much less reactive than esters due to stronger resonance stabilization, resulting in higher activation barriers¹⁶. Furthermore, the amine leaving group in peptides is intrinsically poor, with a pK_a of >35 for the deprotonation of amines to amine anions¹⁷ compared to <8 for typical activated esters^{18,19}. Consequently, spontaneous peptide hydrolysis under physiological conditions is exceedingly slow, with half-lives on the order of hundreds of years⁴. Overcoming this barrier requires active sites that integrate precise metal coordination with strategically positioned general acid–base residues to facilitate transition state stabilization and proton transfer²⁰. Moreover, the protease-substrate interface must be carefully designed to ensure both binding and accurate positioning of the target amide bond at the active site.

We reasoned that recent developments in deep learning-based methods could address these challenges and enable the *de novo* design of metalloproteases. Diffusive denoising methods such as RoseTTAFold Diffusion (RFDiffusion) have proven powerful in generating a wide range of new functional proteins starting from random residue distributions and sets of conditions

specifying the functional constraints^{21–24}. The flow-matching-based RFDiffusion 2 (RFD2) method improved on RFDiffusion for enzyme design by eliminating the need for specifying the sequence positions of the catalytic residues and enabling conditioning directly on the positions of the catalytic functional groups in an ideal active site. As protease design requires not only positioning of catalytic groups, but also binding a peptide substrate, we chose to use a version of RFD2, RFD2 for Molecular Interfaces³ (RFD2-MI), optimized for design of both enzyme active sites and protein-protein interactions. We set out to explore the use of RFD2-MI for building up protein scaffolds around a minimalist metallohydrolase active site and an adjacent peptide binding site, positioning the substrate for cleavage of a defined amide bond.

Computational design of metalloproteases

We began by constructing an ideal active site for carrying out the metalloprotease reaction. In native metalloproteases such as aminopeptidase N²⁵, three residues—two histidines and a glutamate—constitute a Zn(II) binding center. The fourth ligand of Zn(II), a water molecule, is activated by another glutamate to attack the target carbonyl carbon, forming an unstable tetrahedral intermediate. This high energy oxyanionic species and its flanking transition states are stabilized by a hydrogen bond donor tyrosine. Breakdown of the intermediate and product release are promoted by proton transfer from the protonated glutamic acid to the amine leaving group (Fig. 1b). We chose to focus on active sites containing all five functional groups, and chose for their relative positions and orientations the placements of these groups in the crystal structure of aminopeptidase N in complex with a transition state analog (TSA) (PDB 4QHP)²¹. For substrate placement relative to these catalytic groups, we chose to use β -strand hydrogen bonding. As a guide, we used AlphaFold3²⁶ (AF3) to predict the structure of the enzyme-substrate (ES) complex for native astacin²⁷, where the substrate is bound to the protease cleft as an antiparallel β -strand. We superimposed the substrate P1-P2' coordinates (P1 is the residue before and P1' is the residue after the cleavage site) and the backbone atoms of three adjacent astacin residues on the catalytic residues to generate the full target active site (Fig. 1c; Supplementary Methods and Supplementary Discussion).

We used RFD2-MI to generate protein scaffolds around the constructed active site. We chose to allow the structure of both the protease and the peptide substrate to be widely sampled during the design calculations. Starting from a choice of the length of the designed enzyme (160-190AA) and the substrate (8-13AA), RFD2-MI generated random residue distributions for the two chains, and progressively denoised them, attempting to satisfy the conditioning information (the target catalytic site and substrate binding site geometry). Many of the RFD2-MI trajectories produced scaffolds with all the desired structural features (Supplementary Methods 1.2). We used a fine-tuned version of LigandMPNN, EnhancedMPNN²⁸, to generate amino acid sequences that encode the designed fold, preorganize the catalytic residues, and position the substrate properly in the binding cleft. To ensure precise positioning of the target amide bond with respect to the

active site, we carried out a two-sided design²⁹ of both the protease sequence and the substrate sequence; hence the substrate sequence is unique for each design. Designs that AF3 predicted with high confidence and high similarity to the design model (i.e., monomer pLDDT, monomer/complex/catalytic site RMSD) were fed back to RFD2-MI for two more rounds of finer backbone sampling and to MPNN for one round of finer sequence sampling, to achieve higher precision and interface quality (Supplementary Methods).

We modeled each resulting design in three states: 1) the apo protease, to confirm correct folding; 2) the ES complex, to assess substrate binding and positioning of the target amide bond; and 3) the TSA complex, to evaluate transition state stabilization. For the modeling of the ES complex, we first used Zn(II) ion alone (SMILES string: [Zn+2]) as the input for modeling the metal center, common for the prediction of Zn(II) binding complexes²⁶. However, this strategy omitted the essential nucleophilic water molecule, which is critical for catalysis in zinc proteases^{25,27,30,31}. In a second approach, we included both the Zn(II) ion and the coordinating water oxygen (SMILES string: [Zn+2].O), enabling more accurate assessment of the nucleophilic attack step in the ES complex. For modeling the transition state, we used a phosphoramidate TSA³²; many crystal structures have been solved for similar TSA complexes^{25,27,30,31}, so they are better modeled by machine learning software trained on the PDB than the actual transition state. Designs were assessed in all three states based on the confidence of prediction, interface quality, and active site geometry. Synthetic genes were obtained encoding 95 designs from the Zn(II)-only modeling strategy, and 40 designs from the Zn(II)-water strategy (See Extended Data Fig.1 and Supplementary Methods for the full details on the computational pipeline).

Experimental validation of two-sided designed proteases

As each design has a unique protease-substrate pair, we first screened for activity using a fusion construct in which the protease was covalently linked to the substrate so both could be encoded in the same synthetic gene. We also included an mScarlet tag to facilitate the readout of molecular weight change upon cleavage and a Strep-tag for purification (Fig. 2a; Extended Data Fig.2a); we refer to this as a “cis” screen as the cleavage in this case is intramolecular. The 135 designed enzyme and substrate constructs were overexpressed in *E. coli* and purified by Strep-tag affinity chromatography followed by size exclusion chromatography (SEC). Of the 135, 131 were soluble and 127 appeared at approximately the expected size for the full-length construct on SDS-PAGE gels (Fig. S1). To assess metalloprotease activity, we incubated the designs at 37°C in the presence and absence of Zn(II) and assayed catalytic cleavage by SDS-PAGE: active designs cleaved the constructs into the lower MW N-terminal protease and the C-terminal mScarlet domains, whereas inactive designs remained at the original MW of the fusion construct (Fig. 2b). We measured cleavage after 1 hour (h) incubation and again after 12 h incubation, and quantified the activity of each design by densitometry measurements of the product and substrate

bands (Fig. S2). For the designs selected using the Zn(II)-only strategy, 3/89 showed >10% cleavage after 1 h, and 14/89 showed >10% cleavage after 12 h. For the designs assessed by the Zn(II)-water strategy, 31/38 showed >10% cleavage after 1 h, and 35/38 showed >10% cleavage after 12 h (Fig. 2c). For the remainder of the paper, we refer to designs in the first class as “Zn1”, “Zn2”, etc. and designs in the second class as “ZnO1”, “ZnO2”, etc. The overall success rate for all 135 designs, as defined by >10% cleavage after 12 h incubation at 37°C, was 36% (15% for the Zn designs and 88% for the ZnO designs). In every case, mass spectrometry confirmed that the cleavage sites matched the design specifications (Table S1, Fig. S3).

To understand the difference between the Zn and ZnO series of designs, we compared the ES structure models of the inactive Zn56 design and the active ZnO7 design. In the Zn(II)-only models, the substrates for both Zn56 and ZnO7 were positioned correctly in their respective binding clefts, showing no distinction between active and inactive designs (Fig. 2d). In the Zn(II)-water models, however, the catalytic water in Zn56 sterically prevented productive substrate binding, displacing the target amide bond by several ångströms (Å). In contrast, the substrate in ZnO7 was predicted to bind correctly at the active site of ZnO7, with the scissile amide properly positioned for productive nucleophilic attack by the zinc-bound water (Fig. 2e). Thus, the Zn(II)-water models successfully distinguished the active design ZnO7 and would have eliminated the inactive design Zn56, suggesting that explicit inclusion of key molecules in the catalysis mechanism—such as the catalytic water—may improve design success.

Although the cis screen conveniently identified designs with functional active sites, it did not assess proper independent folding of the apo enzyme or intrinsic substrate binding, as covalent linkage greatly increases local substrate concentration. For assays under true catalytic conditions, we obtained synthetic genes for 22 hits for screening in trans, with protease and substrate expressed separately (Fig. 2f, Extended Data Fig. 2b-c). We appended a (GS)₁₅ linker to the N-terminus and an mScarlet tag at the C-terminus of the substrate peptide. The protease and substrate constructs were separately expressed in *E. coli* and purified using Strep-tag affinity chromatography followed by SEC. For each design, we co-incubated the protease and the substrate at 37°C for 1 h in the presence or absence of supplemented Zn(II). Successful cleavage resulted in the removal of an ~3 kDa N-terminal fragment from the substrate, detectable by SDS-PAGE (Fig. 2g; Fig. S4-5). Fourteen designs across five unique scaffolds exhibited >10% trans cleavage after 1 h at 37°C (Table S2). Activities in the cis and trans screens were strongly correlated with a Pearson’s *r* value of 0.91 (Fig. 2h), validating the screening workflow (See Table S3 for the sequences of all designs).

Validation of designed mechanism

For each scaffold, we chose one design for further characterization (Fig. 3a). To assess the contributions of individual catalytic residues to activity and zinc binding, site-directed mutagenesis was used to knock out each zinc-binding residue, the catalytic base, and the oxyanion-stabilizing tyrosine. For all five designs, knocking out one of the three zinc-binding residues, or the catalytic base, reduced the activity of the design by at least sevenfold as measured by densitometry quantification of SDS-PAGE images (Fig. 3b; Fig. S6-9). In four of the five designs, substituting the oxyanion-stabilizing tyrosine with phenylalanine also reduced activity by at least eightfold. Thus, the catalytic activity of the designs arises from the designed active sites.

Competition experiments using a ratiometric dye revealed dissociation constants (K_d) for the designed proteases with respect to the Zn(II) ion in the range between 10^{-10} and 10^{-8} M (Fig. 3c; Fig.S10). ZnO39 bound Zn(II) with a subnanomolar K_d —tighter than previously designed zinc enzymes^{2,33,34}, and comparable to some native zinc enzymes³⁵. The tight Zn(II) binding of these *de novo* zinc proteases could be useful for applications where supplementing additional Zn(II) is not practical, for example in therapeutic uses.

Design selectivity and activity

As we performed two-sided sequence design of both enzyme and substrate, each designed protease had a uniquely designed substrate sequence. We assessed the cross-reactivity of proteases from these five scaffolds. Four of the five designs exhibited preferential cleavage of their cognate substrate (Fig. 3d; Fig. S11-12). This is notable because substrate selectivity was not explicitly enforced, and designs predicted to cleave alternative substrates were not excluded. Instead, specificity appears to have emerged spontaneously from the two-sided design process, which produced highly complementary protease–substrate interfaces.

The fifth design, Zn45, exhibited high activity across all the substrates (Fig. 3d). This experimentally observed cross-reactivity is reflected in the AF3 metrics (Extended Data Fig.3) used in our design strategy. To quantify the cross-reactivity, we chose the protease-substrate pair of Zn45 (protease) and S_ZnO36 (cognate substrate for ZnO36). Zn45 digested the ZnO36 substrate almost completely in 12 h at concentrations ranging from 1.6 μ M to 100 μ M at 37°C (Fig. 3e). At the highest substrate concentration, each Zn45 molecule effectively turned over ~200 ZnO36 substrate molecules overnight. In the AF3 model of the Zn45 Michaelis complex with the ZnO36 substrate and an explicitly modeled zinc-bound water, the main-chain carbonyl of the seventh substrate residue is positioned precisely at the Zn45 active site, matching the experimentally validated cleavage site by MS (Fig.S13).

We characterized the kinetics for two protease-substrate pairs: ZnO7 on substrate ZnO7, the design that exhibited the highest activity for its designed substrate under the conditions of SDS-PAGE cleavage assays, as well as Zn45 on substrate ZnO36, the protease-substrate pair with the overall highest activity. To acquire continuous reaction measurements over time, we measured the Fluorescence Resonance Energy Transfer (FRET) signal from a substrate fused to an N-terminal GFP and a C-terminal mScarlet (Extended Data Fig.2d) as reaction progress curves (Fig. S14-16). Kinetic measurements revealed the k_{cat} of ZnO7 to be $0.0010 \pm 0.0001 \text{ s}^{-1}$, K_M to be $4.9 \pm 1.2 \text{ }\mu\text{M}$, and k_{cat}/K_M to be $210 \pm 60 \text{ M}^{-1}\text{s}^{-1}$ (Fig. 3f). For Zn45 cleaving substrate ZnO36, we measured a k_{cat} of $0.025 \pm 0.002 \text{ s}^{-1}$, K_M of $26 \pm 5 \text{ }\mu\text{M}$, and k_{cat}/K_M of $900 \pm 200 \text{ M}^{-1}\text{s}^{-1}$ (Fig. 3g). This k_{cat} represents a $>10^8$ -fold rate acceleration compared to uncatalyzed hydrolysis of peptide bonds in proteins⁴.

Structural validation

We determined X-ray crystal structures for designs in three distinct scaffolds (the scaffolds of Zn5, ZnO7 and ZnO36, Fig. 4). The structure of Zn5 with bound Zn(II) at 3.41 Å resolution (PDB 11CE) exhibited a small Cα RMSD of 0.82 Å, and the structure of ZnO7 with bound Zn(II) and substrate at 2.37 Å resolution (PDB 11CM) showed a Cα RMSD of 1.35 Å, both closely matching the design models (Fig.4a,b).

In the Zn5–Zn(II) complex, the five catalytic residues adopt conformations closely matching the design model (RMSD = 0.9 Å). A Zn(II) ion is clearly resolved at the designed coordination site with full occupancy (RMSD = 0.3 Å; Fig. 4e). The apo structure of Zn5 (resolution = 3.57 Å) likewise revealed a preorganized catalytic configuration in the absence of the Zn(II) cofactor (Extended Data Fig. 6a), confirming the accurate design of the catalytic residues and the metal coordination geometry.

To capture the Michaelis complex of ZnO7, we substituted the E32 general base with a glutamine to abolish activity and stabilize the substrate bound in the protease cleft. In the resulting crystal structure, the substrate peptide binds the protease in the designed β strand-pairing conformation (Fig.4f). A Zn(II) ion is resolved at the designed coordination site with full occupancy (RMSD = 0.3 Å), coordinated by H31, H35, and E111. A bound water molecule, corresponding to the nucleophilic water in the Michaelis complex that initiates catalysis, completes the tetrahedral coordination shell as the fourth ligand (Fig.4g). The substituted Q32 residue adopts a side-chain rotamer distinct from the catalytically competent rotamer predicted in the design model, suggesting stabilizing the correct general base rotamer as a direction for improvement for this design. Although this deviation may stem from the E32Q substitution (Supplementary Discussion 3.1 and Fig.S20), the same mutation did not induce rotamer flipping

in three other crystal structures we obtained (Extended Data Fig.6b-d), which all display a catalytically competent glutamate geometry and thus better preorganization.

To assess structural similarity with known proteins, we used Foldseek³⁶ to compare the crystal structures of the designs to those in the Protein Data Bank (PDB). There were no fold level matches, as the best TM scores were less than 0.42 (above 0.5 indicates fold similarity; Extended Data Fig.7). Sequences of active designs were searched against the NCBI non-redundant database using BLASTP. The best hits had E-values ranging from 0.34 to 33, indicating no sequence similarity with known proteins (below 0.05 indicates sequence similarity; Extended Data Table 2).

Origin of activity discrepancies between the designs and native metalloproteases

Despite the overall structural accuracy of our de novo metalloproteases, their activities still lag behind the native metalloproteases from which our theozyme was derived^{37,38}. Therefore, sub-ångström structural accuracy does not guarantee optimal catalysis. Subtle differences in the geometry of the catalytic residues, as well as the surrounding chemical environment may contribute significantly as the system progresses along the reaction coordinate. **To investigate the contribution of these effects in the activity difference between the designs and native metalloproteases**, we used machine learning force fields³⁹ (MLFF) to compute the enthalpic activation barriers for the likely rate-limiting proton abstraction and nucleophilic attack steps⁴⁰ (Fig. 1b) for Zn45, ZnO7, and native astacin. All catalytic residues, the substrate and all residues with at least one atom within 5 Å of the substrate were included in the simulation (Supplementary Discussion 4).

The computed activation energies for Zn45 and ZnO7 are 3.6 and 15.1 kcal/mol higher than that of native astacin (Fig. 5a,b), indicating less favorable reaction energetics in the designed active sites. Although the active site RMSD of ZnO7 is lower than that of Zn45, Zn45 exhibits a higher k_{cat} than ZnO7 (Fig. 5b). Structural precision, therefore, does not reliably predict catalytic efficiency. In contrast, the activation energy barriers obtained from simulations of the reaction pathway align with the experimental activity trends, even if they do not quantitatively reproduce the rate differences, suggesting additional contributing factors (Supplementary Discussion 4). Structural accuracy alone may thus serve only as a coarse indicator of activity. A productive active site must support smooth progression along the entire reaction coordinate; assessing the full energy landscape may enhance activity prediction and guide optimization of one-shot designs.

To identify sources of inefficiency beyond the enthalpic effects captured in the MLFF calculations, we systematically evaluated the sequence dependence of catalysis. Substitutions were introduced at 21 positions located within 4 Å of either the five catalytic residues or the

substrate P2, P1, and P1' residues in the Zn45 - S_ZnO36 model, replacing each with all other amino acids except cysteine (Extended Data Fig. 6). At eleven positions, the designed amino acid was already optimal, while for the remaining ten positions, we found single mutations that afforded at least 30% higher activity than the wild-type (Extended Data Fig. 6c), providing clues to shortcomings in the original design.

Substitution of F48, which is close to the general base E77 (Fig. 5c), with tryptophan increased k_{cat} tenfold (Extended Data Fig. 8d), likely by positioning the base more effectively for catalysis. Y152, which was introduced by MPNN during sequence design, forms additional hydrogen bonds with the backbone nitrogen and oxygen atoms of the substrate P1 residue, competing with those made by the designed oxyanion hole. Substituting Y152 with hydrophobic or bulky residues (I, L, M, V, W, R) improved activity (Fig. 5d), suggesting that the hydrogen bond introduced by MPNN overly stabilized the ground state rather than selectively stabilized the transition state. A similar effect may underlie the increased activity seen with the R79A mutation in ZnO36 (Supplementary Discussion 5).

A second-generation variant, Zn45_v2, which combines F48W and Y152I with two additional beneficial mutations (A79Q and Y83Q; Extended Data Fig. 6e, Extended Data Fig. 7), exhibits a k_{cat} of $0.75 \pm 0.04 \text{ s}^{-1}$, a K_{M} of $24 \pm 3 \text{ }\mu\text{M}$, and a $k_{\text{cat}}/K_{\text{M}}$ of $31000 \pm 5000 \text{ M}^{-1}\text{s}^{-1}$, corresponding to a 30-fold improvement in both k_{cat} and $k_{\text{cat}}/K_{\text{M}}$ over the wild-type enzyme (Fig. 5g-i). Zn45_v2 enhances the rate of peptide bond hydrolysis by 1.4×10^{10} fold relative to the spontaneous hydrolysis reaction, achieves over 3500 turnovers per active site (Extended Data Fig. 7d), and displays kinetic parameters comparable to some native metalloproteases (Extended Data Table 3), as well as proteases of other mechanisms⁴¹⁻⁴³ with well-characterized biological functions and broad applications. Nevertheless, neither AF3 nor MPNN captured the improvements in the single or recombined variants, highlighting an additional direction for computational method refinement (Extended Data Fig. 6f-g, Extended Data Fig. 7h, Supplementary Discussion 6).

In an attempt to elucidate the effects of the mutations, we performed molecular dynamics (MD) simulations on the Michaelis complex of Zn45 with the S_ZnO36 substrate, and compared the results with those of Zn45_v2 and the native protease astacin. In Zn45, the catalytic glutamate (E77) sampled a broad range of rotamer states, causing large fluctuations in the distance between its closest oxygen atom and the water hydrogen, hindering efficient proton abstraction. In contrast, both Zn45_v2 and astacin showed a preorganized glutamate rotamer at consistent oxygen-hydrogen distances of 1.75 Å and 1.88 Å, respectively, favoring catalysis (Fig. 5e). The substrate geometry also differed markedly. In Zn45, the scissile amide bond was poorly preorganized, forming an acute angle of 37.4° between the water oxygen, the target carbonyl carbon, and the carbonyl oxygen (Fig. 5f). In contrast, this angle averaged 99.4° in Zn45_v2, and 87.1° in astacin, both much closer to the ideal geometry for nucleophilic attack⁴⁴. Collectively, the mutagenesis and MD results suggest that suboptimal preorganization of the catalytic residues

and the substrate **may partially account for** the activity discrepancy between the designed zinc proteases and their natural counterparts, emphasizing preorganization as a strategy for further improvement.

Despite the improvements in Zn45_v2, its k_{cat} remains lower than the best native proteases^{37,38}. To assess the contribution of individual catalytic residues, we introduced targeted mutations and measured the activity of the resulting variants. Mutations of the Zn(II)-chelating residues and the general base led to substantial losses in activity. In contrast, mutation of the oxyanion-stabilizing tyrosine, Y122, resulted in only a ~10-fold reduction. Given that oxyanion-stabilizing residues typically contribute >50-fold to the activity of native proteases⁴⁵⁻⁴⁷, this modest effect suggests suboptimal positioning in the design; improving the placement of this residue may enhance activity, consistent with the increased activity observed for the V121I variant located in close proximity to the tyrosine (Extended Data Fig. 6, Fig. S21). To further investigate the origin of the activity discrepancy, we characterized the pH dependence of Zn45_v2 activity and compared it with that of native metalloproteases. The resulting pH profile displays a similar bell-shaped trend to those reported for thermolysin⁴⁸ and stromelysin-1⁴⁹; however, the transition point of the acidic limb is shifted higher by approximately 1–2 pH units. Because this transition is often associated with activation of the general base–Zn(II)–water system, the observed shift suggests differences in the local chemical environment compared to the native enzymes, potentially leading to less effective nucleophilic attack in the design.

Design of proteases cleaving **peptides in disease-relevant targets**

With the capability of designing proficient proteases in hand, we set out to explore possible applications. We investigated the possibility of designing proteases against potential therapeutic targets, and for synthetic protease-activated control systems. **We chose three human proteins as protease targets: TAR DNA-binding protein 43 (TDP-43), Serum Amyloid A (SAA), and the amyloid- β (A β) peptide, as their abnormal aggregation can cause diseases⁵⁰⁻⁵².** For TDP43, we pursued two complementary strategies: redesigning an existing scaffold and generating a new scaffold from scratch. To preserve essential functions while disrupting pathological assemblies, we focused on the C-terminal region of TDP-43⁵³ (residues 320–350; Extended Data Fig.8).

For the redesign approach, we applied the logos side-chain pocket scoring metric⁵⁴ to assess compatibility between substrate windows and designed active sites based on side-chain centered interaction potentials, identifying the most targetable patch and selecting ZnO36 as a suitable starting scaffold. The catalytic residues and two to four key polar contacts were retained, while the remaining backbone and side-chain regions were partially diffused and redesigned using EnhancedMPNN (Fig. 6a and Supplementary Methods). For the de novo approach, we used RFDiffusion 3⁵⁵ (RFD 3), which carries out diffusion at the atom level and hence could provide

more specific interactions with the target side chains and used the theozyme with the Zn(II)-bound water. Designs passing all in silico filters were screened in the trans format. The top designs from both strategies, TDP_r3 from the redesign approach and TDP_n3 from de novo design, were selected for further characterization (Supplementary Methods, Fig.S17-19).

We first characterized substrate preferences by monitoring FRET signals from sfGFP–mScarlet reporters containing the TDP-43 target sequence and the reference substrates shown in Fig.3d. Although TDP_r3 cleaved the TDP-43 target sequence at detectable levels, its highest activity was observed on the ZnO36 substrate from which it was derived (Fig. 6b). The substantially higher activity on the parent substrate highlights a key limitation of the redesign strategy, namely persistence of the intrinsic substrate bias of the original scaffold. We solved the crystal structure of the substrate complex of TDP_r3, which exhibited low RMSD in the protease chain with the design model (Fig. 4c,d), yet a shifted substrate registration, consistent with suboptimal interface design (Fig. 4h). In contrast, the best de novo design, TDP_n3 (Fig. 6c), exhibited much higher activity on the TDP-43 target sequence than on other substrates (Fig. 6d).

TDP_n3 cleaved the TDP-43 target peptide with a k_{cat} of $0.033 \pm 0.003 \text{ s}^{-1}$, a K_M of $52 \pm 9 \mu\text{M}$, and a k_{cat}/K_M of $640 \pm 130 \text{ M}^{-1}\text{s}^{-1}$ (Fig. 6e), comparable to the best two-sided design, Zn45, which exhibited broad substrate specificity (Fig. 3d). These results confirm that appreciable activity and substrate specificity can be achieved simultaneously through one-shot de novo design. When incubated with full length TDP-43 (monomer and oligomeric forms), TDP_n3 produced cleavage products of the expected size confirmed by mass spec (Supplementary Table S1, Supplementary Table S4) at the 6 h and 12 h time points (Fig. 6f), whereas buffer or base-knockout controls showed no cleavage (Extended Data Fig.8d). However, cleavage was slower on full-length TDP-43 compared to the synthetic sfGFP-mScarlet substrate, potentially due to complications of liquid-liquid phase separation or precipitation⁵⁶ (Extended Data Fig.8c).

To demonstrate the potential generalizability of our method, we further designed metalloproteases targeting the peptide sequences SSRSFFSFLG from SAA and KGAIIGLMVG from A β peptide. Both proteases exhibited highest activity on the on-target substrate as compared to other substrates discussed in this work, as shown in Extended Data Fig. 9.

Conditional activation of a cytokine mimic and a receptor antagonist by a designed metalloprotease

The potential use of de novo proteases extends beyond targeting human proteins to synthetic protease-conditioned circuits for precision medicine, as in Antibody-Directed Enzyme Prodrug

Therapy (ADEPT)⁵⁷. To investigate the applicability of *de novo* metalloprotease for the conditional activation of therapeutics, we designed a masked construct of Neo2, a previously reported *de novo* IL2 mimic⁵⁸, and explored if Neo2 can be conditionally activated by the designed proteases. A Neo2 minibinder generated by RFD2 was used to mask the IL2R-binding interface and fused to Neo2 through a linker containing the S_ZnO36 cleavage sequence. The resulting masked construct was termed Neo2-M-Z (Fig. 6g). To assess Neo2 activation and signaling through IL2R, we performed a HEK-Blue IL-2/IL-15 reporter assay at a range of Neo2-M-Z concentrations, with or without protease activation. In the absence of protease treatment, Neo2-M-Z exhibited an EC50 of ~0.5 nM. In contrast, treatment with Zn45 or Zn45_v2 shifted the EC50 to ~0.01 nM, comparable to unmasked Neo2 and corresponding to a ~50-fold activation window. In contrast, the Neo2-M-Z construct could not be activated by matrix metalloprotease 2 (Fig 6h), a tumor-enriched human protease commonly used as an activation trigger in prodrug therapies^{59,60}. Further, we validated this circuit using another masked *de novo* EGFR antagonist system⁶¹. When the masked EGFR antagonist (EGFRmb-M-Z) was treated with *de novo* proteases Zn45 and Zn45_v2, the antagonist was activated and prevented EGFR signaling and the downstream phosphorylation of ERK. The unmasking effect was abolished with the catalytic residue knockout of Zn45_v2, or with MMP2 (Fig. 6i,j, Fig. S22). Adoptive cell therapies presenting *de novo* designed bioorthogonal proteases on their surfaces could enable more precise activation of caged biologics at the desired disease sites.

Discussion

Our results show that *de novo* enzyme design has advanced far beyond model reactions on activated substrates. Using a Zn(II) ion to activate a water in an active site comprising three Zn(II) coordinating residues and additional water-activating and oxyanion-stabilizing residues, we achieved >10⁸-fold rate accelerations for peptide bond cleavage relative to uncatalyzed hydrolysis in aqueous solution. Codesigning the substrate and protease, or designing a protease for a prescribed substrate sequence, both yielded considerable substrate selectivity without explicit negative design against alternative substrates.

Despite these advances, there remains considerable scope to improve our design methodology. Although the turnover rates of enzymes straight out of the computer are still 10³-10⁴-fold lower than those of the most efficient natural proteases, characterization of our metalloprotease designs highlights strategies for further enhancement: increasing the precision of active site placement, integrating energetic assessments alongside geometric criteria, improving preorganization of the catalytic residues and substrate, **and refining the local electrostatic environment of the active site**. We show that increasing preorganization with a few mutations can narrow the activity gap with top-performing natural metalloproteases to two orders of magnitude.

With the capability of targeting prescribed epitopes, it is exciting to now consider the many applications where the ability to design selective proteases could be transformative. Unlike many native proteases that recognize short 2-5 residue motifs^{62,63}, our designs bind 8–13-residue substrates, creating an extended recognition window capable of encoding substantial sequence and conformational selectivity. Such tunable specificity could be used to target viral polyprotein junctions, selectively dismantle disease-associated assemblies while sparing essential protein pools, rewire signaling pathways or engineer cassettes in synthetic biology applications—greatly reducing off-target risks. Beyond therapeutics, programmable proteases could enable safe and sustainable solutions across agriculture, food, and cosmetics, including the control of pests and pathogens and improvements to product texture and stability. Together, these advances point toward a future in which proteolysis becomes a programmable modality for biotechnology, medicine, and engineered living systems.

Figures

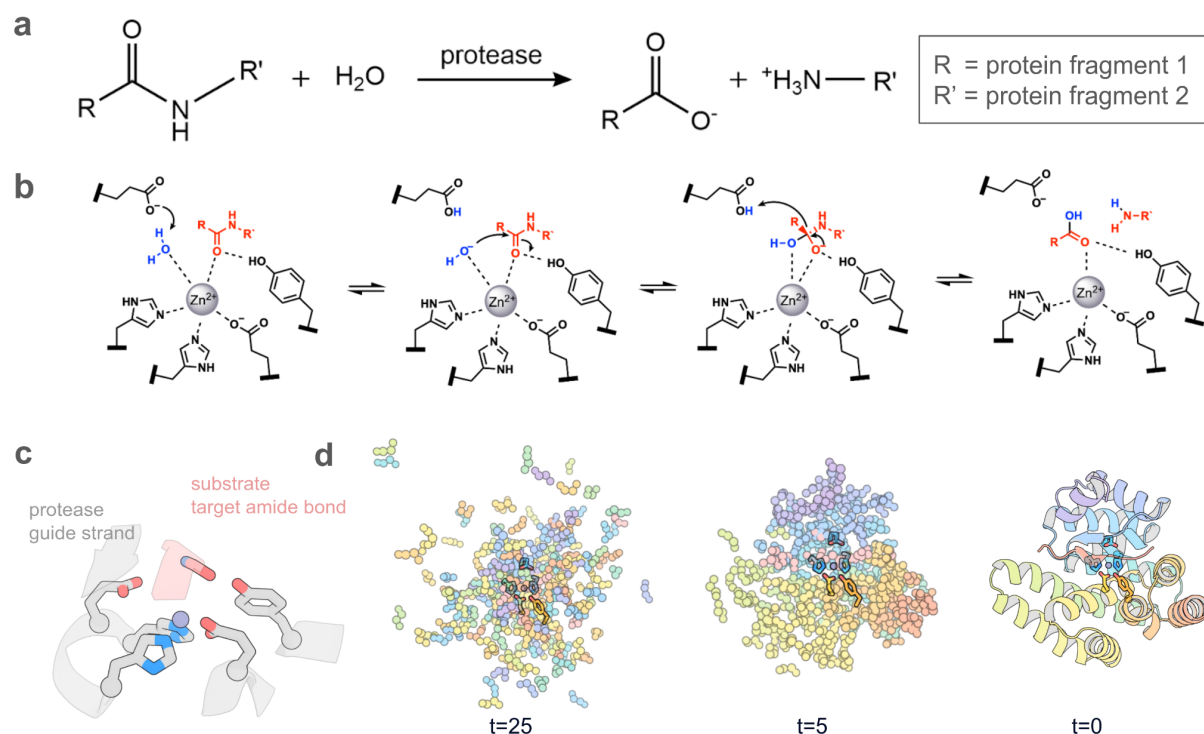


Fig. 1: Computational design of metalloproteases. **a**, Chemical equation for the proteolytic hydrolysis of a peptide bond in a protein substrate. R and R' represent polypeptide fragments. **b**, Reaction mechanism for the hydrolysis of peptide bonds by zinc proteases. **c**, Input motif for RFD2-MI. Catalytic residues were extracted from the crystal structure of aminopeptidase N in complex with a transition state analog inhibitor (pdb: 4qhp). The substrate conformation and strand pairing with the guide strand in the enzyme were extracted from an AF3 prediction of native astacin in complex with a substrate peptide. **d**, Representative RFD2-MI trajectory for backbone generation.

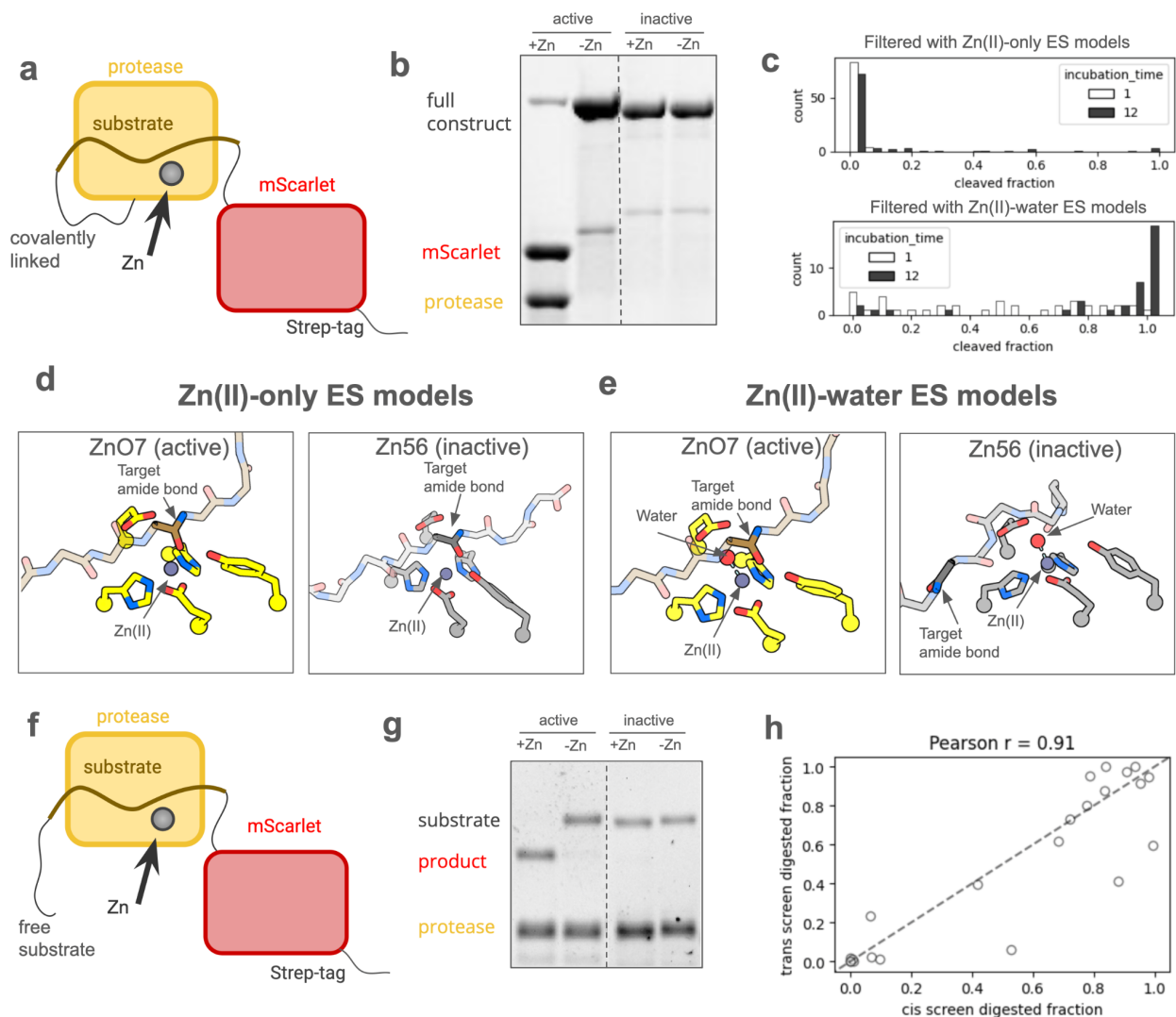


Fig. 2: Experimental screening. **a**, Schematic illustration of the fusion construct used in the cis screen. The proteases were expressed as a genetic fusion with their designed substrate peptide sequence, followed by a C-terminal mScarlet. **b**, Representative screening outcomes for an active and an inactive design after 1 h at 37 °C in the cis assay measured by protein SDS-PAGE. An active design exhibits Zn-dependent cleavage at the correct site as assessed by SDS-PAGE gel and mass spectrometry (Table S1). In contrast, an inactive design remained uncleaved. **c**, Digested fraction of the substrate measured by SDS-PAGE gel densitometry. The hit rate significantly improved when the nucleophilic water was explicitly modeled in the ES complex. **d**, Zn(II)-only models of the ES complexes did not differentiate between active (ZnO7) and inactive (Zn56) designs. **e**, Zn(II)-water models of the ES complexes clearly differentiated active (ZnO7) from inactive (Zn56) designs, based on the displaced substrate in the latter. **f**, Schematic illustration of the protein constructs used in the trans screen. The proteases and their designed substrates were expressed as separated chains. The substrate sequence was expressed as a genetic fusion with an N-terminal (GS)₁₅ linker and a C-terminal mScarlet. **g**, Representative screening

outcomes for an active and an inactive design in the trans assay after 1 h at 37 °C measured by protein SDS-PAGE. An active design exhibited Zn-dependent cleavage at the correct site as assessed by SDS-PAGE gel. In contrast, an inactive design exhibited no cleavage. **h,** Relationship between the digested fractions of the cis and trans assays, both performed at 37 °C for 1 h.

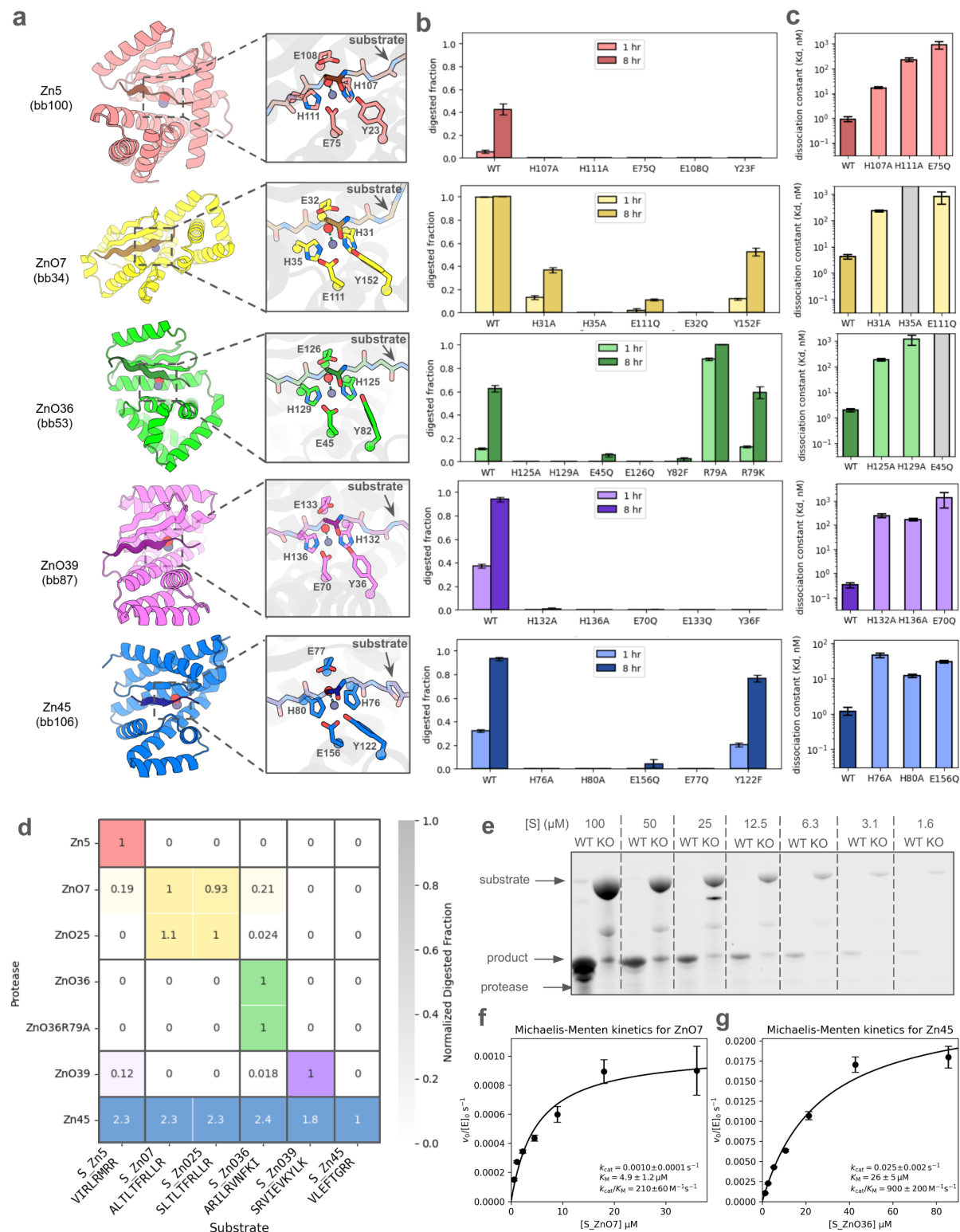


Fig. 3: Functional validation of two-sided metalloprotease designs. **a**, AF3 models of active protease designs with close-up views of the active sites from five unique scaffolds (TM-score < 0.23 between any pair). **b**, Impact of mutating individual catalytic residues on design activity at

short (1 h) and long (8 h) time points, measured by SDS-PAGE gel densitometry. In each bar plot, the catalytic residues are in the following order for their functional roles from left to right: first zinc-chelating residue, second zinc-chelating residue, third zinc-chelating residue, the general base, and the oxyanion stabilizer. For ZnO36, two additional substitutions, R79A and R79K, were investigated due to the proximity of R79 to the scissile carbonyl oxygen and tetrahedral intermediate/transition state oxyanion (Fig. S9). Measurements were made in triplicate. Bars and error bars represent mean $\pm$ std. **c**, Assessment of Zn(II) dissociation constant for the designed proteases and the respective knockout variants. Gray bars with heights above the upper boundary of the plots indicate weak Zn(II) binders that are beyond the dynamic range of the assay, suggesting K_d values at least above 2 μ M. Bars and error bars represent the fitted dissociation constant $K_d \pm$ std. **d**, Cross-reactivity heatmap of 7 proteases with 6 designed substrates. In the heatmap, each row represents a protease and each column represents a substrate. The substrate sequences were named by prepending “S_” to the name of the proteases designed with this substrate. Designs from the same backbone were labeled with the same color. ZnO25 is a one-shot design with the same backbone as ZnO7 but differing in sequence at 39 positions. Cleavage was quantified as the digested fraction of the substrate by SDS-PAGE densitometry normalized to the cognate substrate in triplicate. The mean of three measurements is shown in the heatmap. **e**, SDS-PAGE gel analysis of 500 nM Zn45-catalyzed cleavage of the substrate designed for ZnO36 at different substrate concentrations, [S], for 12 hrs at 37 °C. For the highest substrate concentration (100 μ M), each Zn45 molecule effectively turns over ~200 substrate molecules. By contrast, the general base knockout of the Zn45 (KO) showed only minor cleavage, likely due to residual native protease from the expression and purification process. **f**, Kinetics of ZnO7-catalyzed cleavage of its cognate substrate, and **g**, of Zn45-catalyzed cleavage of the substrate designed for ZnO36. v_0 : reaction rate in the initial linear region (nM/s). $[E_0]$: enzyme concentration, 500 nM. Measurements were made in triplicate. Black dots and error bars represent mean $\pm$ std.

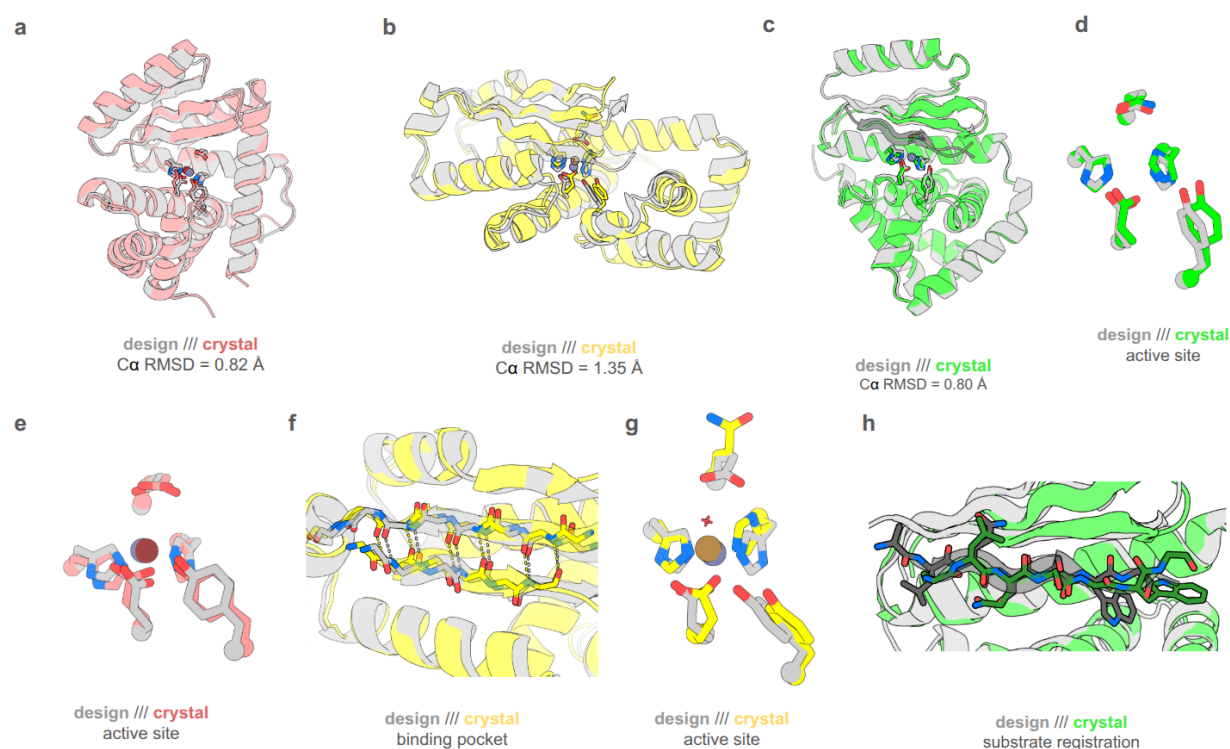


Fig. 4: Structural validation of metalloprotease designs. a-c, structural superposition of design models (gray) and crystal structures for Zn5 (a, crystal structure in red), ZnO7 (b, crystal structure in yellow) and TDPr3 (c, crystal structure in green). The active sites are shown as sticks. The Zn(II) ions are shown in dark gray in the design models, dark red in the crystal structure of Zn5 and dark yellow in the crystal structure of ZnO7. d,e,g Active site overlays of the crystal and the design model for TDPr3 (d), Zn5 (e), ZnO7 (g). The catalytic water in the active site of the ZnO7 crystal in f is shown as a red cross. f, Zoomed-in view of the β -strand pair between the substrate and the protease in the superposition of the ZnO7 crystal and design model. h, Zoomed-in view of substrate registration in TDPr3. The substrate solved in the crystal is shifted by one amino acid between the design and the crystal.

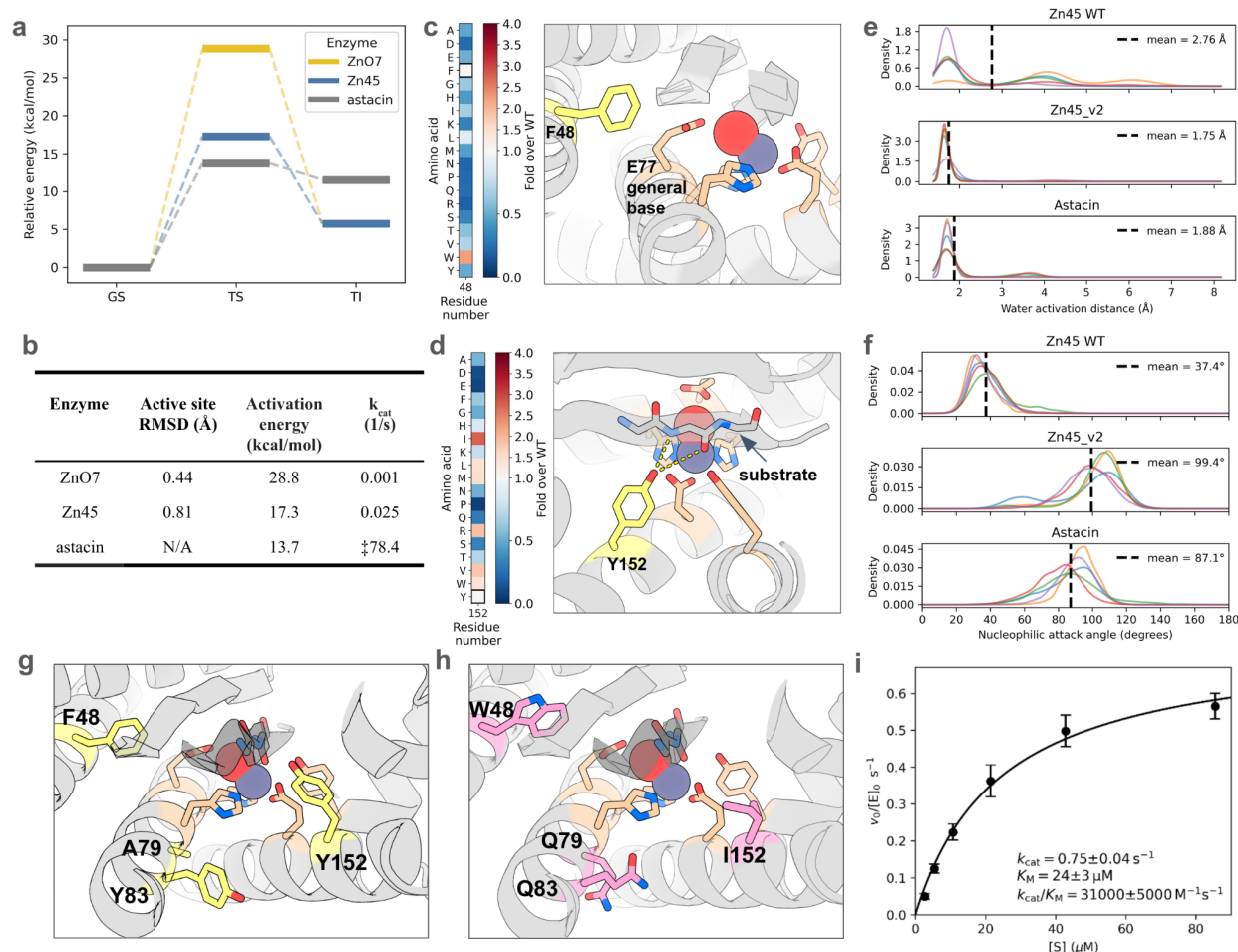


Fig. 5: Investigation of activity discrepancy between the designs and the native metalloproteases. **a**, relative energies calculated using MLFF with respect to the ground state (GS) of the proton abstraction and nucleophilic attack steps, for the transition state (TS) and the tetrahedral intermediate (TI). **b**, Active site RMSD between the design model and theozyme, activation energies and the k_{cat} values for the proteases involved in the MLFF calculations. ‡ Value reported in ref.40. **c-d**, activity heatmap (left panel) of residues in the active site mutagenesis screen for Zn45, and the positions of these residues (yellow) with respect to the catalytic residues (wheat) in the AF3 prediction. **e**, The kernel density distribution of the distance between the proton-abstracting oxygen in the E77 general base and the closest water hydrogen, in the MD simulation of wild-type Zn45 (top), astacin (middle) and Zn45_v2 (bottom). **f**, The kernel density distribution of the angle between the water oxygen, the substrate target carbonyl carbon, and the substrate target carbonyl oxygen, in the MD simulation of wild-type Zn45 (top), astacin (middle) and Zn45_v2 (bottom). For both e and f, five replicas are shown in solid curves of different colors. The mean of all five replicas for each group is shown as a black dashed line. **g,h**, The residues where beneficial mutations are combined to yield Zn45_v2 in the AF3 model of wild-type Zn45 (g, wild-type residues shown in yellow) and Zn45_v2 (h, beneficial mutations shown in pink). In all AF3 models, the catalytic residues are shown in wheat and the mutated

residues are shown in yellow. The Zn(II) ion is shown as a dark gray sphere and the Zn(II)-bound water is shown as a red sphere. **i**, Kinetics of Zn45_v2-catalyzed cleavage of the substrate of ZnO36. v_0 : reaction rate in the initial linear region (nM/s). $[E_0]$: enzyme concentration, 50 nM. Measurements were made in triplicate. Black dots and error bars represent mean $\pm$ std.

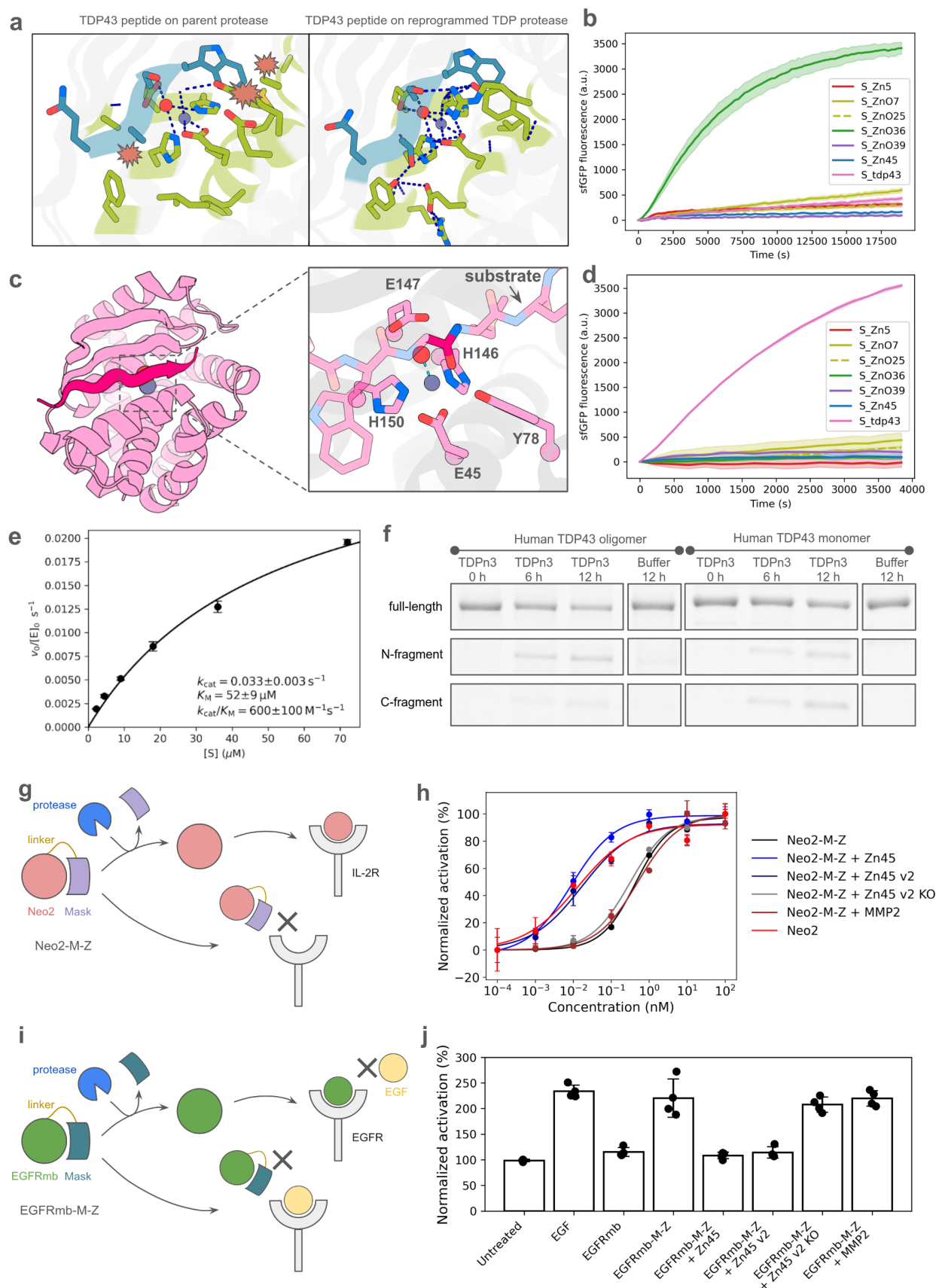


Fig. 6: De novo protease for cleavage of human protein and for conditional activation of masked cytokine mimic. **a**, Reprogramming ZnO36R79A for TDP-43 cleavage. (left) The redesigned protease has a well-packed interface and hydrogen bonding network with the new substrate, while (right) binding of the TDP substrate to the parent protease ZnO36R79A results in sidechain clashes and fewer hydrogen bonds with the polar residues; all structures shown are design models. **b**, Substrate specificity of TDPPr3 against the sfGFP-mScarlet substrates of all designs presented in this work. Measurements were made in triplicate. Solid lines and shaded regions represent the mean $\pm$ std. **c**, AF3 design model of TDPn3 with a zoomed-in view of the active site. **d**, Substrate specificity of TDPPr3 against the sfGFP-mScarlet substrates of all designs presented in this work. Measurements were made in triplicate. Solid lines and shaded regions represent the mean $\pm$ std. **e**, Kinetics of TDPn3-catalyzed cleavage of the target TDP-43 epitope. v_0 : reaction rate in the initial linear region (nM/s). $[E_0]$: enzyme concentration, 500 nM. Measurements were made in triplicate. Black dots and error bars represent mean $\pm$ std. **f**, Cleavage of full-length human TDP-43 in vitro. **g**, Schematic illustration of the conditional activation of a masked cytokine mimic, Neo2-M-Z, by de novo proteases. **h**, Normalized activation percentage of HEK-Blue IL-2 reporter cells by Neo2-M-Z treated with Zn45, Zn45_v2, Zn45_v2_KO, MMP2 and by Neo2. Measurements were made in triplicate. Data points and error bars represent mean $\pm$ std. Solid lines indicate sigmoidal fits to the raw data. **i**, Schematic illustration of the conditional activation of a masked EGFR antagonist, EGFRmb-M-Z, by de novo proteases. **j**, Normalized phospho-ERK percentage of cells treated with Zn45, Zn45_v2, Zn45_v2_KO, MMP2 and by EGFP-mb. Measurements were made in quadruplicate. Data points and error bars represent mean $\pm$ std.

Author Contributions

A.C., K.W., P.V., D.H., and D.B. conceptualized the project. A.C., K.W., H.C., P.V., S.J.P. developed the computational design pipeline. N.H. performed the Zn(II)-dissociation constant experiments. **H.S. designed and performed the conditional cytokine mimic activation experiments. B.O.-S. and C.D.-P. designed and performed the masked conditional antagonist experiments.** B.C. trained RFD2-MI. D.E. performed the MD simulations. K.D. and L.L.S performed the MLFF calculations. A.K.B, K.N. ,A.K., Q.W. performed protein crystallography. P.J.,D.K.,S.M.W.,and S.H. contributed ideas and code. A.C., K.W., and L.C. produced and characterized the designs. A.C. and K.W. analyzed the experimental results. X.L. performed intact protein mass spectrometry. S.G. provided training on equipment and advice on experimental protocols. X.Y and A.A.H produced and characterized full-length human TDP-43. K.W., D.H and D.B. supervised the project. A. C., K.W., D.H., D.B. wrote the manuscript. All authors reviewed, commented and edited the manuscript.

Competing Interests

A provisional patent application will be filed prior to publication, listing A.C., K.W., H.C., P.V., S.J.P., D.H., and D.B. as inventors or contributors.

Code and Data Availability

All code and data will be made available upon final publication.

Funding Sources

This work was funded by the following: Howard Hughes Medical Institute. Bill and Melinda Gates Foundation INV-043758. Defense Threat Reduction Agency Grant HDTRA1-19-1-0003. The Audacious Project at the Institute for Protein Design. The Open Philanthropy Project Improving Protein Design Fund. The National Institutes of Health's National Institute of Allergy and Infectious Disease, grant R0AI160052. The National Institutes of Health's National Institute on Aging, grant R01AG063845. Grantham Foundation for the Protection of the Environment. Schmidt Family Foundation. MICIU/AEI/10.13039/501100011033 and FEDER, EU (PID2023-151988OB-I00). "la Caixa" Foundation (LCF/BQ/PR25/12110008).

Acknowledgements

We thank L. Goldschmidt for maintaining the computational, and K. Van Wormer, R. Ticzon, H. Nunez-Ortegaand, E. Weston for maintaining the wet lab resources at the Institute for Protein Design. We thank I.Kalvet, H. Yu, B. Belmont, A. Broerman, B. Liu, L.T. Yu, C. Liu, W. Ahern, F. Xue, C. Wang for helpful discussions during the development of the project. We thank D. Whittington and R. Lundeen for helping with mass spectrometry. We thank I. Kalvet and Y. Liu for providing reagents. We thank M. A. Kennedy for reading and editing drafts of the manuscript. We thank L. Stewart for managing the Institute for Protein Design and securing funding resources.

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used beamtime awards (DOI: doi.org/10.46936/APS-190993/60014759) from the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This research used resources 17-ID-1 (AMX)/ 17-ID-2 (FMX) of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1605010).

References

1. Studer, S. *et al.* Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).
2. Kim, D. *et al.* Computational design of metallohydrolases. *Nature* **649**, 246–253 (2026).
3. Bauer, M. S. *et al.* *De novo* design of phosphotyrosine peptide binders. Preprint at <https://doi.org/10.1101/2025.09.29.678898> (2025).
4. Radzicka, A. & Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.* **118**, 6105–6109 (1996).
5. Lipscomb, W. N. & Sträter, N. Recent Advances in Zinc Enzymology. *Chem. Rev.* **96**, 2375–2434 (1996).
6. Hooper, N. M. Families of zinc metalloproteases. *FEBS Lett.* **354**, 1–6 (1994).
7. De Almeida, L. G. N. *et al.* Matrix Metalloproteinases: From Molecular Mechanisms to Physiology, Pathophysiology, and Pharmacology. *Pharmacol. Rev.* **74**, 714–770 (2022).
8. Page-McCaw, A., Ewald, A. J. & Werb, Z. Matrix metalloproteinases and the regulation of tissue remodelling. *Nat. Rev. Mol. Cell Biol.* **8**, 221–233 (2007).

9. Edwards, D., Handsley, M. & Pennington, C. The ADAM metalloproteinases. *Mol. Aspects Med.* **29**, 258–289 (2008).
10. Wang, Z., Li, W., Chen, S. & Tang, X. X. Role of ADAM and ADAMTS proteases in pathological tissue remodeling. *Cell Death Discov.* **9**, 447 (2023).
11. Iwai, M. & Horiuchi, M. Devil and angel in the renin–angiotensin system: ACE–angiotensin II–AT1 receptor axis vs. ACE2–angiotensin-(1–7)–Mas receptor axis. *Hypertens. Res.* **32**, 533–536 (2009).
12. Wang, J., Xu, A., Wan, Y. & Li, Q. Purification and Characterization of a New Metallo-Neutral Protease for Beer Brewing from *Bacillus amyloliquefaciens* SYB-001. *Appl. Biochem. Biotechnol.* **170**, 2021–2033 (2013).
13. Jabeen, A., Kandadai, R., Kannikannan, M. & Borgohain, R. Guidelines for the use of botulinum toxin in movement disorders and spasticity. *Ann. Indian Acad. Neurol.* **14**, 31 (2011).
14. Leippold, T., Reitz, A. & Schurch, B. Botulinum Toxin as a New Therapy Option for Voiding Disorders: Current State of the Art. *Eur. Urol.* **44**, 165–174 (2003).
15. Satriyasa, B. K. Botulinum toxin (Botox) A for reducing the appearance of facial wrinkles: a literature review of clinical use and pharmacological aspect. *Clin. Cosmet. Investig. Dermatol.* **Volume 12**, 223–228 (2019).
16. Zerner, Burt., Bond, R. P. M. & Bender, M. L. Kinetic Evidence for the Formation of Acyl-Enzyme Intermediates in the α -Chymotrypsin-Catalyzed Hydrolyses of Specific Substrates. *J. Am. Chem. Soc.* **86**, 3674–3679 (1964).
17. Bordwell, F. G. Equilibrium acidities in dimethyl sulfoxide solution. *Acc. Chem. Res.* **21**, 456–463 (1988).
18. Nagy, N. *et al.* 4-Methylumbelliferone Treatment and Hyaluronan Inhibition as a Therapeutic Strategy in Inflammation, Autoimmunity, and Cancer. *Front. Immunol.* **6**, (2015).
19. Bailey-Darland, S., Krueger, T. D. & Fang, C. Ultrafast Spectroscopies of Nitrophenols and Nitrophenolates in Solution: From Electronic Dynamics and Vibrational Structures to Photochemical and Environmental Implications. *Molecules* **28**, 601 (2023).
20. Mahesh, S., Tang, K.-C. & Raj, M. Amide Bond Activation of Biological Molecules. *Molecules* **23**, 2615 (2018).
21. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
22. Glögl, M. *et al.* Target-conditioned diffusion generates potent TNFR superfamily antagonists and agonists. *Science* **386**, 1154–1161 (2024).
23. Vázquez Torres, S. *et al.* De novo designed proteins neutralize lethal snake venom toxins. *Nature* **639**, 225–231 (2025).
24. Bennett, N. R. *et al.* Atomically accurate de novo design of antibodies with RFdiffusion. *Nature* **649**, 183–193 (2026).
25. Vassiliou, S. *et al.* Structure-Guided, Single-Point Modifications in the Phosphinic Dipeptide Structure Yield Highly Potent and Selective Inhibitors of Neutral Aminopeptidases. *J. Med. Chem.* **57**, 8140–8151 (2014).
26. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
27. Grams, F. *et al.* Structure of astacin with a transition-state analogue inhibitor. *Nat. Struct. Mol. Biol.* **3**, 671–675 (1996).

28. Xue, F. *et al.* Improving Protein Sequence Design through Designability Preference Optimization. Preprint at <https://doi.org/10.48550/ARXIV.2506.00297> (2025).
29. Sahtoe, D. D. *et al.* Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* **375**, eabj7662 (2022).
30. Tronrud, D. E., Monzingo, A. F. & Matthews, B. W. Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin. *Eur. J. Biochem.* **157**, 261–268 (1986).
31. Kim, H. & Lipscomb, W. N. Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by x-ray crystallography. *Biochemistry* **30**, 8171–8180 (1991).
32. Bartlett, P. A. & Marlowe, C. K. Phosphoramidates as transition-state analog inhibitors of thermolysin. *Biochemistry* **22**, 4618–4624 (1983).
33. Hoffnagle, A. M. & Tezcan, F. A. Atomically Accurate Design of Metalloproteins with Predefined Coordination Geometries. *J. Am. Chem. Soc.* **145**, 14208–14214 (2023).
34. Der, B. S. *et al.* Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. *J. Am. Chem. Soc.* **134**, 375–385 (2012).
35. Schullek, J. R. & Wilson, I. B. The binding of zinc to angiotensin-converting enzyme. *Arch. Biochem. Biophys.* **265**, 346–350 (1988).
36. Van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
37. Wolz, R. L. A Kinetic Comparison of the Homologous Proteases Astacin and Meprin A. *Arch. Biochem. Biophys.* **310**, 144–151 (1994).
38. Ryan, J. W. *et al.* A Radiochemical Assay for Aminopeptidase N. *Anal. Biochem.* **210**, 27–33 (1993).
39. Batatia, I. *et al.* MACE-POLAR-1: A Polarisable Electrostatic Foundation Model for Molecular Chemistry. Preprint at <https://doi.org/10.48550/ARXIV.2602.19411> (2026).
40. Chen, S.-L., Li, Z.-S. & Fang, W.-H. Theoretical investigation of astacin proteolysis. *J. Inorg. Biochem.* **111**, 70–79 (2012).
41. Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294**, 949–955 (2002).
41. Mei, M. *et al.* A combinatorial strategy for HRV 3C protease engineering to achieve the N-terminal free cleavage. *Int. J. Biol. Macromol.* **265**, 131066 (2024).
43. Fang, B., Boross, P. I., Tozser, J. & Weber, I. T. Structural and Kinetic Analysis of Caspase-3 Reveals Role for S5 Binding Site in Substrate Recognition. *J. Mol. Biol.* **360**, 654–666 (2006).
44. Radisky, E. S. & Koshland, D. E. A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci.* **99**, 10316–10321 (2002).
45. Yiallourous, I., Große Berkhoff, E. & Stöcker, W. The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. *FEBS Letters* **484**, 224–228 (2000).
46. Beaumont, A., O'Donohue, M. J., Paredes, N., Rousselet, N., Assicot, M., Bohuon, C., Fournié-Zaluski, M.-C. & Roques, B. P. The role of histidine 231 in thermolysin-like enzymes: a site-directed mutagenesis study. *Journal of Biological Chemistry* **270**, 16803–16808 (1995).
47. Phillips, M. A., Fletterick, R. & Rutter, W. J. Arginine 127 stabilizes the transition state in carboxypeptidase. *Journal of Biological Chemistry* **265**, 20692–20698 (1990).
48. Kunugi, S., Hirohara, H. & Ise, N. pH and temperature dependences of thermolysin catalysis. *European Journal of Biochemistry* **124**, 157–163 (1982).

49. Johnson, L. L., Pavlovsky, A. G., Johnson, A. R., Janowicz, J. A., Man, C.-F., Ortwine, D. F., Purchase, C. F., White, A. D. & Hupe, D. J. A rationalization of the acidic pH dependence for stromelysin-1 (matrix metalloproteinase-3) catalysis and inhibition. *Journal of Biological Chemistry* **275**, 11026–11033 (2000).
50. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
51. Westermarck GT, Fändrich M & Westermarck P. AA amyloidosis: pathogenesis and targeted therapy. *Annu. Rev. Pathol.* **10**, 321–344 (2015).
52. Selkoe DJ & Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608 (2016).
53. Guenther, E.L., Cao, Q., Trinh, H. *et al.* Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat Struct Mol Biol* **25**, 463–471 (2018).
54. Wu, K. *et al.* Design of intrinsically disordered region binding proteins. *Science* **389**, eadr8063 (2025).
55. Butcher, J. *et al.* De novo Design of All-atom Biomolecular Interactions with RFDiffusion3. Preprint at <https://doi.org/10.1101/2025.09.18.676967> (2025).
56. Babinchak, W. M. *et al.* The role of liquid–liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* **294**, 6306–6317 (2019).
57. Sharma SK, Bagshawe KD. Antibody directed enzyme prodrug therapy (ADEPT): trials and tribulations. *Adv Drug Deliv Rev.* 2017;118:2–7. doi:10.1016/j.addr.2017.09.009.
58. Silva, DA., Yu, S., Ulge, U.Y. *et al.* De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
59. Mansour AM, Dreves J, Esser N, Hamada FM, Badary OA, Unger C, Fichtner I & Kratz F. A new approach for the treatment of malignant melanoma: enhanced antitumor efficacy of an albumin-binding doxorubicin prodrug that is cleaved by matrix metalloproteinase 2. *Cancer Res.* **63**, 4062–4066 (2003).
60. Page-McCaw, A., Ewald, A. & Werb, Z. Matrix metalloproteinases and the regulation of tissue remodelling. *Nat Rev Mol Cell Biol* **8**, 221–233 (2007).
61. Escobar-Rosales, M., Montaner, C., Expòsit, M., Lucchi, R., Díaz-Perlas, C., Baker, D. & Oller-Salvia, B. De novo design of peptide masks enables rapid generation of conditionally active miniprotein binders. *J. Am. Chem. Soc.* **147**, 45495–45505 (2025)
62. Broomfield, G. R., Kleffmann, T., Saul, D. J. & Wilbanks, S. M. Thermostable neutral metalloprotease from *Geobacillus* sp. EA1 does not share thermolysin's preference for substrates with leucine at the P1 ' position. *FEBS Lett.* **599**, 3332–3343 (2025).
63. Ratnikov, B. I. *et al.* Basis for substrate recognition and distinction by matrix metalloproteinases. *Proc. Natl. Acad. Sci.* **111**, (2014).