Computational Design of Metallohydrolases

Donghyo Kim[1,2‡], Seth M. Woodbury[1,2,3‡], Woody Ahern[1,2,4‡], Indrek Kalvet[1,2,7*], Nikita Hanikel[1,2], Saman Salike[1,2,5], Samuel J. Pellock[1,2], Anna Lauko[1,2,6], Donald Hilvert[8*], David Baker[1,2,7*]

1. Department of Biochemistry, University of Washington, Seattle, WA, USA.
2. Institute for Protein Design, University of Washington, Seattle, WA, USA.
3. Graduate Program in Bioengineering, University of Washington, Seattle, WA, USA.
4. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.
5. Graduate Program in Molecular Engineering, University of Washington, Seattle, WA, USA.
6. Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA, USA.
7. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.
8. Laboratory of Organic Chemistry, ETH Zurich, Zurich, Switzerland.

[‡]Contributed Equally
[*]To whom correspondence should be addressed

Abstract

New enzymes can be designed by starting from a description of an ideal active site composed of catalytic residues surrounding the reaction transition state(s) and identifying or generating a protein scaffold that supports the site[1–7], but there are a few current limitations. First, the catalytic efficiencies achieved by such efforts have generally been quite low, and considerable optimization by directed evolution has been required to reach activities typical of native enzymes.[8–10] Second, generative AI methods such as RFdiffusion[11,12] now enable the direct generation of proteins around active sites, but to date, such scaffolding has required specification of both the position in the sequence and the backbone coordinates of the catalytic residue, which complicates sampling. Here we introduce a generative AI method called RFam that overcomes these limitations and use it to design zinc metallohydrolases starting from a density functional theory description of active site geometry. Of 96 designs tested experimentally, the most active has a k_{cat}/K_M of 23,000 M^{-1} s^{-1}, orders of magnitude higher than previously designed metallohydrolases.[6,7,13,14] This 148 amino acid protein has a novel fold with an enclosed chamber which positions the reaction substrate nearly perfectly for attack by a catalytic water molecule activated by the bound metal and is predicted by ChemNet[15] to have a highly preorganized active site. The ability to generate high activity catalysts starting from quantum chemistry calculated active site geometries without experimental optimization should open the door to a new generation of potent designer enzymes.[16,17]

Main

Metallohydrolases catalyze some of the most difficult hydrolysis reactions in biology by using their bound metal ion(s) to activate a water molecule positioned adjacent to the substrate bond to be cleaved.[18–20] Engineering new metallohydrolases is currently of considerable interest for degrading human-generated environmental pollutants for which there has not been sufficient time for enzymes to evolve that efficiently hydrolyze them.[21–23] Protein engineering has expanded the scope of substrates that metallohydrolases can hydrolyze, but this often requires initial promiscuous activity.[24,25] De novo enzyme design has been used to generate new metallohydrolases[6,8,13], but these have had quite low activity and efficiency requiring extensive directed evolution to match those of native enzymes.[8] Given a description of an ideal metallohydrolase active site, de novo metalloenzyme design seeks to find or generate a protein scaffold that positions the catalytic residues, metal(s), and substrate/transition state with high accuracy.[26,27] RFdiffusion has been used successfully to scaffold sites, but the search has been complicated by the requirement for the specification of the sequence positions and conformations of the catalytic residues.[11,12,28]

We reasoned that a generative AI design method that required only specification of the positions of sidechain functional groups around the reaction transition state, and that could sample over all possible sequence positions and conformations of these residues, could more readily satisfy complex catalytic constraints.[16,17,29,30] We set out to develop such an approach and to use it to design new metallohydrolases starting from a quantum chemistry-generated active site description with a bound metal cofactor.

To enable sequence position and sidechain rotamer-agnostic enzyme design, we developed a generative AI flow-matching model called Rosetta Flow Atomic Motif (RFam). RFam extends the RFdiffusion capabilities of generating scaffolds that position a set of functional residues (a “motif”) in two key ways. First, it enables atomic substructure scaffolding: RFdiffusion can only scaffold backbone-level motifs (with the sidechain backbone atom N-C_α-C=O positions specified), while RFam can scaffold arbitrary atom-level motifs (any subset of amino acid heavy atoms). This is important for enzyme design as it enables specification of the positions of only the key functional groups interacting with the reaction transition transition state, rather than the full sidechain conformation. Second, RFam enables sequence-position-agnostic scaffolding: RFdiffusion requires specification of the primary sequence positions of the motif residues, but RFam can scaffold motifs whose primary sequence positions are unknown. RFam replaces diffusion with flow matching[31,32] and achieves sequence position-agnostic atomic substructure scaffolding by providing randomly selected native atomic coordinates (but not their sequence positions) during training in addition to the partially noised, sequence-labeled atomic coordinates. With these improvements, RFam generates diverse proteins starting directly from catalytic configurations consisting of input functional group positions and substrate coordinates. Allowing the model to

resolve the *a priori* unknown degrees of freedom (i.e., the primary sequence positions and sidechain rotamer conformations of the catalytic sidechains) is considerably more effective at generating self-consistent design solutions than randomly sampling (the space is far too large to exhaustively sample) those degrees of freedom before inference, as was necessitated with RFdiffusion. A detailed description of model training and benchmarking results for a wide range of active site scaffolding problems is described in Ahern et al.[33]

As an initial test of RFam, we chose to design a zinc metallohydrolase for a fluorogenic ester, 4-methylumbelliferyl phenylacetate (4MU-PA), as a model reaction (Fig. 1A). We began by using density functional theory to identify the geometry of the rate-determining Zn(II)-OH nucleophilic attack on the substrate ester. Four distinct catalytic arrangements based on the stereochemistry of the tetrahedral intermediate and the nature of the oxyanion hole were considered (Fig. 1B and Fig. S3; for further details see SI, section 2.1). These calculations provide the coordinates of the three Zn(II)-binding imidazole rings, the metal, and the transition state. Our previous RFdiffusion approach required the backbone coordinates and residue positions as inputs; this required upfront sampling of rotameric states and sequence position for each histidine. This cannot be done exhaustively: even with relatively coarse sampling around the sidechain chi angles χ_1, χ_2, and the backbone torsion angle ψ there are on the order 10^{18} possible choices for the sidechain conformations and sequence placements of our full catalytic site (Fig. 1C and Fig. S4). Whereas each RFdiffusion run has to be initialized with a specific (and generally randomly selected) choice from this enormous set of combinations, RFam as described above searches the entire space in each trajectory.

RFam inference trajectories were used to build protein scaffolds housing the DFT-generated minimal active site configurations, referred to as theozymes.[2,34] Several snapshots from a representative trajectory are shown in Fig. 1D, transforming random noise on the left into the final backbone on the right (Movie S1). The C_α atoms of each residue (shown in colored spheres representing final sequence position) are initially sampled from a Gaussian distribution, while the target functional atom positions (shown in sticks) stay fixed. As the trajectory proceeds from left to right, the global structure takes shape around the motifs, with the fixed histidine sidechains eventually connecting to C_α atoms of the protein backbone at sequence positions of the network's choosing. 5120 RFam inference trajectories were carried out starting from different random seeds and for each of the resulting protein scaffolds, sequences were generated using ProteinMPNN.[35] The catalytic geometry and interactions with the transition state of those designs for which the AlphaFold2[36] (AF2) predicted structure was close to the design model were further optimized using iterative LigandMPNN[37] and constrained Rosetta repacking and minimization[38] (Fig. S1; see SI, section 2.1). Designs containing a proposed general base positioned to activate the water molecule (i.e., Glu/Asp/His

within H-bonding distance of the Zn(II)-bound water) and sidechain hydrogen bonds stabilizing the transition state oxyanion (if applicable), and that AF2 predicted to adopt the target structure, were characterized with ChemNet to assess active site preorganization.[15] 96 designs were selected for experimental characterization based on predicted active site geometry and preorganization (Fig. S5; see SI, section 2.1).

Linear DNA fragments encoding the 96 designs were cloned into a plasmid encoding a C-terminal Strep-tag and used to transform *E. coli*, and the resulting proteins were purified using Streptactin affinity chromatography. 86/96 designs were expressed and soluble as judged by SDS-PAGE analysis of the eluants (Fig. S6). Purified designs were supplemented with zinc sulfate, and hydrolysis of 4MU-PA was monitored by fluorescence. Five designs (A1, A8, B9, C4, and F7) had activity well above background (Fig. 2B and Fig. S7). Sequence-verified single clones for each of these design hits were expressed and purified by affinity chromatography followed by size-exclusion chromatography to obtain pure, monomeric protein fractions (Fig. S8 and Fig. S9). Michaelis-Menten kinetic characterization of the purified variants revealed a k_{cat}/K_M of 23000 ± 3000 M^{-1} s^{-1} for the most active A1 design, and k_{cat}/K_M values in the range of 51-210 M^{-1} s^{-1} for the other four designs (Fig. 2, C,D; Fig. S10; Table S1). For comparison, the k_{cat}/K_M of previously designed metallohydrolases ranged from 3 to 60 M^{-1} s^{-1} (Table S3).[8] A1 is also quite a robust enzyme, retaining activity for at least 1000 turnovers (Fig. 3E). Additionally, A1 differs considerably from previously described structures: the most similar structures identified through TM alignment with the PDB and AFDB have template modeling (TM) scores[39] of 0.41 and 0.49, respectively, and do not have analogous arrangements of catalytic residues (Fig. S16).

Design A1 not only has remarkably high activity but was also the top-ranked design in our *in silico* ranking, hence the name. The structure in the absence of substrate was predicted to be very close to the design model by AlphaFold2 (Fig. S12 and Fig. S15A). Most strikingly, the designed active site of A1 was predicted to be highly preorganized by ChemNet, with the catalytic sidechains fixed in place and the substrate held closely in its designed position, adjacent to the proposed zinc ion position. ChemNet is a deep neural network that, given a protein backbone containing a substrate, fully randomizes the positions of the substrate and all side chains within a 600-atom sphere, and then generates new coordinates for these groups; repeated ChemNet trajectories generate an ensemble of possible sidechain conformations and small molecule docks. Design A1 stood out from the other designs in both the extent of catalytic site pre-organization (the catalytic sidechains were largely fixed in space in catalytically competent conformations) and the positioning of the substrate/transition state in the active site (in the A1 ensemble the substrate remained largely fixed in space in the active site, whereas in the inactive designs H7 and H8 it fluctuated severely) (Fig. 2, E-H; Movies S2-S5). Seven designs based on the same A1 backbone family were initially filtered out during the design selection phase as they had suboptimal ChemNet

metrics; we retrospectively expressed and purified these designs and found that they had very low or no activity, further highlighting the utility of ChemNet ensemble calculations for identifying active designs (Fig. S14). These findings suggest that combining global structure prediction with detailed ChemNet modeling of the active site provides a powerful approach to assessing the catalytic machinery and substrate binding geometry for design selection (Fig. S13).

The A1 active site consists of a primarily hydrophobic pocket with three histidines binding Zn(II) with their N_ε atoms, an aspartate as a potential general base, and an asparagine that forms a hydrogen bond to the coumarin ring (Fig. 3, A,B). As in the original theozyme model used to generate A1, the Zn(II) ion also acts as an oxyanion hole, stabilizing the forming negative charge at the transition state; there are no nearby sidechain H-bond donors (Fig. S15). The Zn(II) is absolutely critical for A1 activity: extraction of bound Zn(II) by dialysis in the presence of the Zn(II) chelator 1,10-phenanthroline completely eliminated activity, and activity was subsequently restored by addition of zinc to the solution (Fig. 3D). Zn(II) titration experiments measure the A1 binding affinity (K_D) for Zn(II) to be 41±5 nM which is comparable to previous designed Zn(II) enzymes[27,40], but higher than native zinc hydrolases which typically have K_D <10 nM.[20,41–43]

We carried out mutagenesis experiments on the proposed catalytic residues to evaluate their contributions to zinc binding and catalysis (Fig. 3, G-I). One of our hypotheses was that N17 may help position the substrate by hydrogen bonding with the ester carbonyl of the coumarin moiety and could stabilize the developing negative charge on the leaving group; the N17A mutation led to a 7.9-fold decrease in k_{cat}/K_M (Fig. S18). We further hypothesized that the catalytic histidines and aspartate were critical for activity. Mutation of all three metal coordinating histidine residues to alanine simultaneously (H118A;H130A;H134A), as well as two of the three single histidine-to-alanine substitutions (H118A and H134A), completely inactivated the enzyme, as expected. Interestingly, mutating the third Zn(II)-coordinating residue to alanine (H130A) only resulted in a decrease of 12-fold in k_{cat}/K_M, while mutation of the proposed general base D67 to alanine had little effect on k_{cat}/K_M and increased zinc binding affinity. These results suggest that H134/H118/H130 and H134/H118/D67 may be competing Zn(II)-binding sites due to the close proximity of the coordinating side chains of H130 and D67, corroborated by Chai-1[44] predictions of the protein-zinc-substrate complex (Fig. S15 B,C); the D67A mutation may confine the zinc to the originally designed coordination sphere with the three histidines, which is more catalytically competent. In the H130A mutant, coordination with D67 instead maintains Zn(II) binding, albeit in a less optimal binding geometry, lowering zinc affinity and activity.

These results suggest clear avenues to explore for further increasing the already high activity of A1. First, the general base residue could be relocated to another site

further from the Zn(II) ion and better positioned by additional hydrogen bond interactions to avoid interference with Zn(II) binding and enable more effective water activation. Second, additional hydrogen bonding interactions with the oxyanion could confer further transition state stabilization. These features could potentially be incorporated into A1 using the RFam analog of the partial diffusion approach used with RFdiffusion to increase target binding affinity[45]: the positions of the general base and/or newly added oxyanion-stabilizing functional groups could be specified, and the nearby backbone noised and then denoised with full exploration of the possible sequence position and rotameric conformations of the residues providing these functionalities. For a broader design search, the original RFam generation procedure could be repeated starting from the beginning, this time incorporating the general base and oxyanion stabilizing functional groups in the set of guiding catalytic functional groups.

Discussion:

We demonstrate that RFam can generate highly active metallohydrolases directly from active site configurations obtained by quantum chemistry calculations. The zero-shot design of an enzyme with a k_{cat}/K_M >10^4 M^{-1} s^{-1} – the top-ranked *in silico* design out of a total of 96 tested straight from the computer (with no experimental optimization) – is a considerable advance for de novo enzyme design which in the past has required extensive design screening and directed evolution to achieve activities at this level.[8,16] The catalytic efficiency of A1 is within the range observed for native metallohydrolases (k_{cat}/K_M = 10^4~10^8 M^{-1} s^{-1}) and greater than all previously designed hydrolases prior to evolutionary optimization (Table S3).[46–50] Experimental characterization of A1 provides strong evidence to suggest that it functions as designed, utilizing a bound zinc ion to activate a water molecule for nucleophilic attack and to stabilize the resulting oxyanion intermediate and flanking transition states. The ChemNet ensemble results suggest the key to obtaining k_{cat}/K_M >10^4 M^{-1} s^{-1} is precise substrate placement relative to the activated water and the zinc ion; going beyond this through positioning of the general base such that it cannot reconfigure to interact with the bound ion, and incorporating further sidechain oxyanion stabilization could increase activity into the range of the most active native metallohydrolases.

Our RFam-ChemNet design approach has many advantages over previous de novo enzyme design methods and should be broadly applicable for generating efficient catalysts for a wide diversity of chemical reactions. By enabling direct scaffolding of sidechain functional groups, rather than backbone N-C_α-C=O coordinates as required in enzyme design calculations with RFdiffusion[1,28], RFam bypasses the need for explicit enumeration of catalytic sidechain rotamer conformations and placement of the catalytic residues along the linear sequence–this enables each design trajectory to sample from the enormous space of possibilities rather than being confined to a small (and perhaps unrealizable) subspace. Assessment of active site preorganization and

substrate/transition state positioning with ChemNet proved remarkably effective at identifying the most active designs; this result, together with similar observations with de novo designed serine hydrolases[1] and retroaldolases[15], suggests that ChemNet and related deep learning approaches will be widely useful for design ranking. In our laboratory, we have found the RFam-ChemNet design approach described here to yield catalysts for several uni- and bimolecular reactions, and we look forward to seeing what the broader design community can generate with these tools which we are making freely available open source.

Acknowledgments

We thank Y. Wang, A. Hunt, J. Gershon, J. Min, C. Demakis, A. Shida, and Y. Kipnis for discussions about zinc hydrolases, A. Swartz, P. Lund-Anderson, N. Ennist, Y. Hsia, S. Sharma, and S. Honda for valuable discussions about metal binding, small molecule binding, and general discussions about protein design, D. Evans for guidance with performing quantum chemistry, M. Bauer for software assistance, D. Tischer for the provision of a customized cloning vector, K. VanWormer, R. Ticzon, H. Nunez-Ortega for managing the wet-lab resources, and L. Goldschmidt for maintaining the computational resources at the Institute for Protein Design. This work was funded by the National Science Foundation Graduate Research Fellowship Program, Grant No. DGE-2140004 (S.M.W), The Howard Hughes Medical Institute (D.B., I.K.), The Audacious Project at the Institute for Protein Design (A.L., N.H.), The Open Philanthropy Project Improving Protein Design Fund (I.K., S.J.P.), the ETH Zurich (D.H.), The Bill and Melinda Gates Foundation INV-010680 (D.K., W.A.), and the Schmidt Futures Foundation (A.L., S.J.P., S.S.).

Author Contributions

S.M.W., I.K., and D.B. conceptualized the metallohydrolase design project. W.A. and D.B. conceptualized RFam. W.A. developed, trained, and optimized RFam. S.S. added inference capabilities to RFam. S.M.W. and I.K. performed DFT calculations. S.M.W., I.K., and D.K. developed the computational design pipeline. S.J.P. and A.L. contributed ideas, code, and functions. S.M.W. and D.K. produced and characterized the designs and analyzed the results. N.H. performed the Zn(II)-binding affinity experiments. I.K., S.J.P., A.L., and D.H. provided guidance for optimizing experimental protocols and helped interpret key results. D.B. and D.H. supervised the research. The manuscript and supplemental information were drafted by S.M.W., D.K., I.K., and D.B. All authors reviewed and commented on the manuscript.

Competing Interests: The authors declare no competing interests.

Data and Materials Availability: All data are available in the main text or as supplementary materials. Design scripts and RFam will be made available on GitHub.

Figures

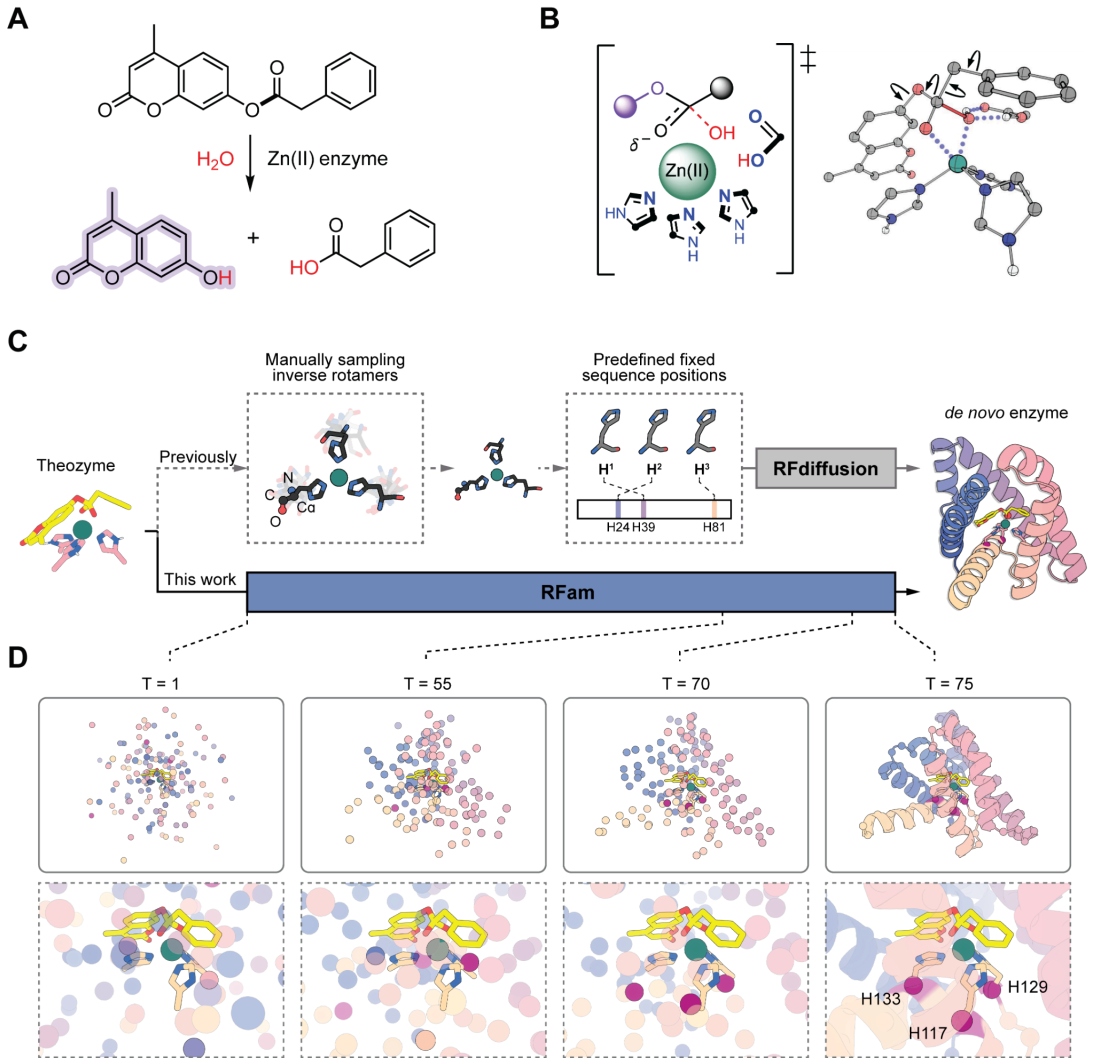

Fig. 1. RFam design method (**A**) Hydrolysis of 4MU-PA yields phenylacetic acid and a fluorescent coumarin product. (**B**) Example theozyme for zinc-hydroxide nucleophilic attack of the 4MU-PA ester. 2D representation (left) and 3D DFT model (right). Arrows on the 3D model represent sampled conformational flexibility. (**C**) Comparison of scaffold generation around an input theozyme using previous backbone centric RFdiffusion (top row) versus interaction functional group centric RFam. RFdiffusion requires explicit upfront sampling of side chain conformations and residue sequence positions, whereas RFam only requires the transition state complex and the catalytic side chain functional groups, implicitly sampling sequence space and rotameric space during inference. (**D**) Snapshots of the global structure and active site from model X_T during an RFam inference trajectory. The coordinates of the input transition state complex and catalytic functional groups stay fixed during inference while the overfall backbone structure and sequence positions and unspecified atoms of the catalytic side chains are sampled by RFam. The C_α atoms that host the catalytic histidines at the end of the trajectory are retrospectively highlighted as red spheres; these C_α atoms are not predetermined but rather move into the frame to host the fixed sidechains as the global structure forms around the motifs.

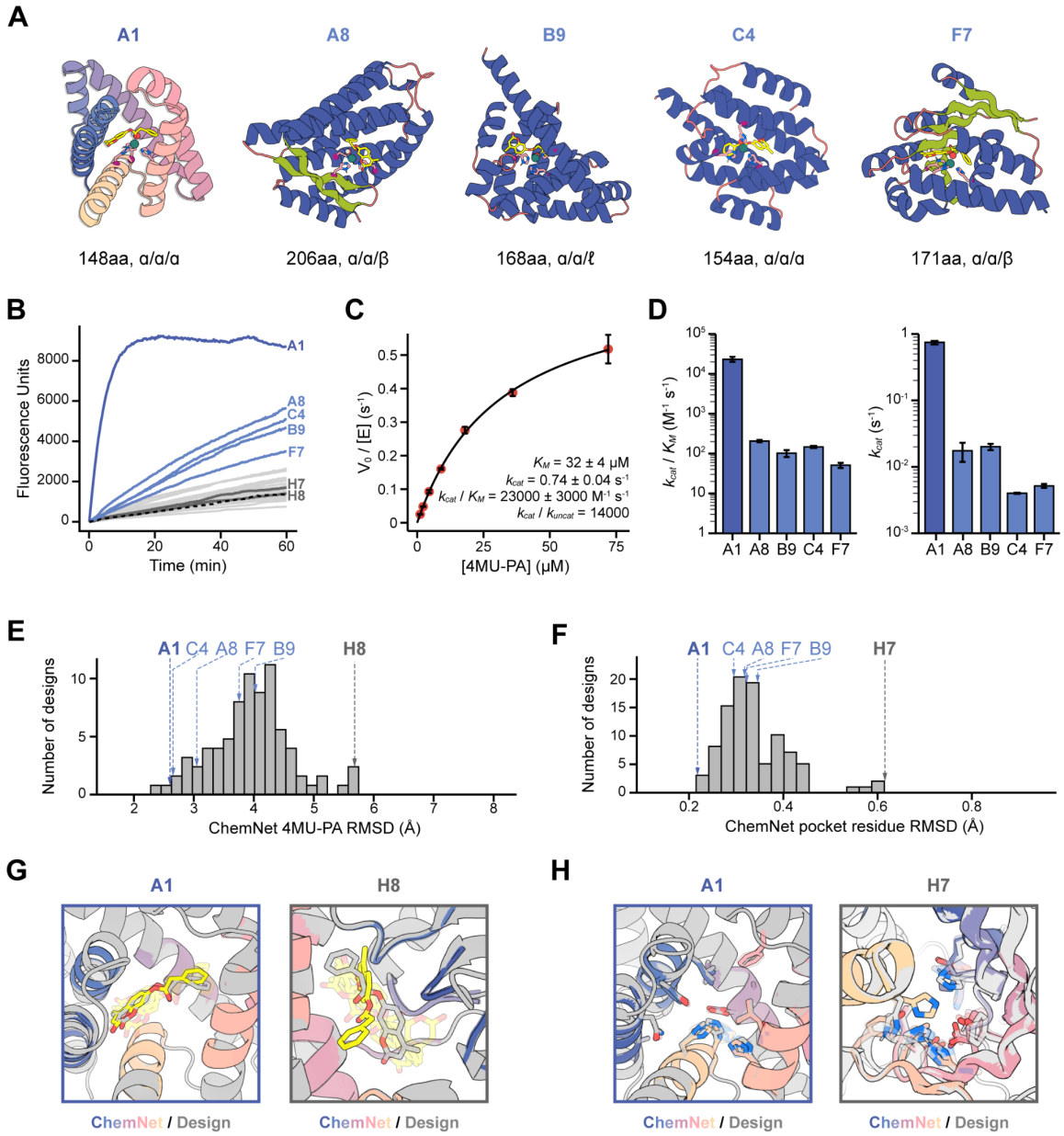

Fig. 2. Activity characterization and ChemNet preorganization assessment.
(**A**) Design models of most active designs. Sequence length and the secondary structure of each catalytic histidine are indicated below. (**B**) Reaction progress curves. Dotted black line is the buffer background. (**C**) Michaelis-Menten characterization of A1. The y-axis is the initial rate divided by the total enzyme concentration. (**D**) Michaelis-Menten parameters of most active designs. (**E,F**) Distribution of ChemNet active site pre-organization ensemble metrics for the ordered designs. Average design-prediction substrate RMSD (E), and the average catalytic and binding residue design-prediction RMSD (F) across all predicted ensembles generated by ChemNet for each design. (**G**) For A1, the substrate position in ChemNet ensembles is close to the design model, while in the inactive design H7 the substrate position fluctuates widely. (**H**) For A1, the sidechains surrounding the active site are largely fixed in positions close to the original design model, while in the inactive H8 design, the sidechain positions vary considerably. Note that only the first 5 randomly generated, unranked ensemble predictions are shown for ChemNet in (G,H).

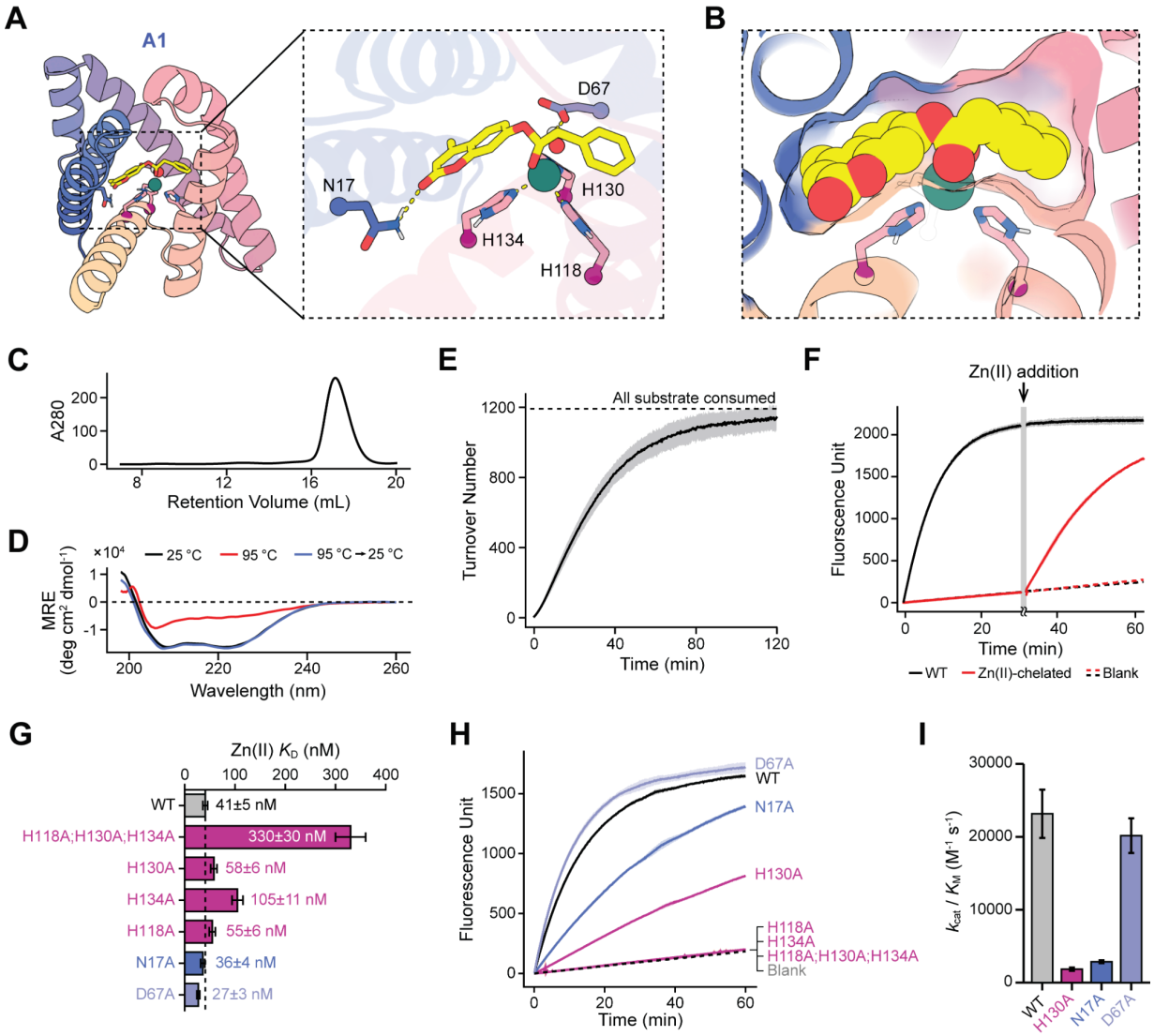

Fig. 3. Characterization of A1 activity. (**A,B**) A1 design model with magnification of the active site showing (**A**) the hypothesized catalytic residues and (**B**) a surface view of the pocket revealing high shape-complementarity to the substrate. (**C**) Size-exclusion chromatogram of A1 showing a single peak corresponding to monomeric protein. (**D**) Circular dichroism spectra for A1 at 25 °C (black), 95 °C (red), and a cycle from 25 °C to 95 °C to 25 °C (blue). The spectra suggest that A1 has an alpha-helical secondary structure and that it can refold after heating and partial unfolding. (**E**) [Product]:[enzyme] turnover curve measuring the lower bound of the total turnover of A1 demonstrating full hydrolysis of 4MU-PA and >1000 turnovers within 2 hours. Note that the background reaction has been subtracted from the spectra so that every turnover can be attributed to the enzyme (Fig. S11; more details in SI, section 2.2). (**F**) Reaction progress curves for the Zn(II) *holo*- and zinc-chelated *apo*-A1 proteins showing zinc-dependent activity. Adding excess Zn(II) to the *apo*-A1 sample after 30 minutes reestablishes the activity demonstrating that zinc is essential for the catalytic mechanism of A1. (**G**) Zinc affinity of the wild-type A1 and its knockout mutants, measured as the dissociation constant (K_D) where a lower value indicates tighter binding. (**H**) Fluorescence progress curves comparing the activity of the wild-type A1 to its different knockout mutants. (**I**) Michaelis-Menten k_{cat}/K_M for the active A1 mutants compared to the wild-type.

References

1. Lauko, A. *et al.* Computational design of serine hydrolases. *bioRxiv* 2024.08.29.610411 (2024) doi:10.1101/2024.08.29.610411.
2. Tantillo, D. J., Jiangang, C. & Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **2**, 743–750 (1998).
3. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309–313 (2010).
4. Yeh, A. H.-W. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
5. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
6. Zastrow, M. L., Peacock, A. F. A., Stuckey, J. A. & Pecoraro, V. L. Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118–123 (2011).
7. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940 (2012).
8. Studer, S. *et al.* Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).
9. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418–421 (2013).
10. Obexer, R. *et al.* Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).

11. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
12. Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).
13. Rufo, C. M. *et al.* Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **6**, 303–309 (2014).
14. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **346**, 1525–1528 (2014).
15. Anishchenko, I. *et al.* Modeling protein-small molecule conformational ensembles with ChemNet. *bioRxiv* 2024.09.25.614868 Preprint at https://doi.org/10.1101/2024.09.25.614868 (2024).
16. Lovelock, S. L. *et al.* The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).
17. Nam, K., Shao, Y., Major, D. T. & Wolf-Watz, M. Perspectives on computational enzyme modeling: From mechanisms to design and drug development. *ACS Omega* **9**, 7393–7412 (2024).
18. Basile, L. A., Raphael, A. L. & Barton, J. K. Metal-activated hydrolytic cleavage of DNA. *J. Am. Chem. Soc.* **109**, 7550–7551 (1987).
19. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **74**, 1–20 (1997).
20. Lipscomb, W. N. & Sträter, N. Recent advances in zinc enzymology. *Chem. Rev.* **96**, 2375–2434 (1996).
21. Chamas, A. *et al.* Degradation rates of plastics in the environment. *ACS Sustain.*

Chem. Eng. **8**, 3494–3511 (2020).

22. Marten, E., Müller, R.-J. & Deckwer, W.-D. Studies on the enzymatic hydrolysis of polyesters. II. Aliphatic–aromatic copolyesters. *Polym. Degrad. Stab.* **88**, 371–381 (2005).
23. Ellis, L. D. *et al.* Chemical and biological catalysis for plastics recycling and upcycling. *Nat. Catal.* **4**, 539–556 (2021).
24. Gould, S. M. & Tawfik, D. S. Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **44**, 5444–5452 (2005).
25. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).
26. Schwizer, F. *et al.* Artificial metalloenzymes: Reaction scope and optimization strategies. *Chem. Rev.* **118**, 142–231 (2018).
27. Hoffnagle, A. M. & Tezcan, F. A. Atomically accurate design of metalloproteins with predefined coordination geometries. *J. Am. Chem. Soc.* **145**, 14208–14214 (2023).
28. Braun, M. *et al.* Computational design of highly active de novo enzymes. *bioRxiv* 2024.08.02.606416 (2024) doi:10.1101/2024.08.02.606416.
29. Hossack, E. J., Hardy, F. J. & Green, A. P. Building enzymes through design and evolution. *ACS Catal.* 12436–12444 (2023).
30. Ming, Y. *et al.* A review of enzyme design in catalytic stability by artificial intelligence. *Brief. Bioinform.* **24**, bbad065 (2023).
31. Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M. & Le, M. Flow Matching for Generative Modeling. in *The Eleventh International Conference on Learning Representations* (2022).

32. Yim, J. *et al.* Improved motif-scaffolding with SE(3) flow matching. *arXiv [q-bio.QM]* (2024).
33. Ahern et al, preprint in preparation.
34. Tantillo, D. J. & Houk, K. N. Theozymes and Catalyst Design. in *Stimulating Concepts in Chemistry* 79–88 (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, 2005).
35. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
36. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
37. Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *bioRxiv* 2023.12.22.573103 (2023) doi:10.1101/2023.12.22.573103.
38. Tyka, M. D. *et al.* Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618 (2011).
39. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
40. 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).
41. Kiefer, L. L. & Fierke, C. A. Functional characterization of human carbonic anhydrase II variants with altered zinc binding sites. *Biochemistry* **33**, 15233–15240 (1994).
42. Coleman, J. E. & Vallee, B. L. Metallocarboxypeptidases. *J. Biol. Chem.* **235**,

390–395 (1960).

43. Feder, J., Garrett, L. R. & Kochavi, D. Studies on the inhibition of neutral proteases by 1,10-phenanthroline. *Biochim. Biophys. Acta* **235**, 370–377 (1971).
44. Discovery, C. *et al.* Chai-1: Decoding the molecular interactions of life. *bioRxiv* 2024.10.10.615955 (2024) doi:10.1101/2024.10.10.615955.
45. Gloegl, M. *et al.* Target-conditioned diffusion generates potent TNFR superfamily antagonists and agonists. *bioRxiv* 2024.09.13.612773 (2024) doi:10.1101/2024.09.13.612773.
46. Hingorani, D. V. *et al.* Impact of MMP-2 and MMP-9 enzyme activity on wound healing, tumor growth and RACPP cleavage. *PLoS One* **13**, e0198464 (2018).
47. Johansson, I. M. & Forsman, C. Kinetic studies of pea carbonic anhydrase. *Eur. J. Biochem.* **218**, 439–446 (1993).
48. Stöcker, W., Sauer, B. & Zwilling, R. Kinetics of nitroanilide cleavage by astacin. *Biol. Chem. Hoppe Seyler* **372**, 385–392 (1991).
49. Stoczko, M., Frère, J.-M., Rossolini, G. M. & Docquier, J.-D. Functional diversity among metallo-beta-lactamases: characterization of the CAR-1 enzyme of Erwinia carotovora. *Antimicrob. Agents Chemother.* **52**, 2473–2479 (2008).
50. Kamat, S. S. *et al.* The catalase activity of diiron adenine deaminase. *Protein Sci.* **20**, 2080–2094 (2011).