Computational design of serine hydrolases

Anna Lauko

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:

David Baker, Chair

Frank DiMaio

Barry L Stoddard

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Computational design of serine hydrolases

Anna Lauko

Chair of the Supervisory Committee:

David Baker

Biochemistry

Nature's enzymes are exceptionally powerful catalysts, exerting dramatic rate accelerations and exquisite control over a remarkable variety of chemical transformations. Since their initial discovery and characterization, the ability to generate artificial enzymes for chemical reactions involved in industrial processes, chemical synthesis, and therapeutic applications has been of considerable interest. Despite decades of effort, artificial enzymes continue to display lower catalytic activities than their native counterparts, even for well-understood model reactions. Here, we present a novel and general approach to computational enzyme design utilizing recent advances in tailored protein scaffold generation and active site conformational ensemble prediction. As a proof of concept, we have applied this method to the design of esterases that utilize the serine hydrolase enzymatic mechanism. Despite a deep understanding of the mechanism amassed through decades of study, previous attempts to design esterases acting through this mechanism have failed. To our knowledge, the designs made using our approach represent the first examples of accurately designed, de novo serine hydrolases spanning folds not found in natural hydrolases and exhibiting catalytic efficiencies on par with hydrolases in nature that act on

similar substrates. We believe our approach will not only enable the design of industrially relevant serine hydrolases but also be broadly applicable to accelerating a wider array of chemical reactions, including ones that do not occur in nature.

Acknowledgements

I would like to thank my family – Gabriella and István, who taught me to delight in the hidden logic of the everyday, Domi and Zoli, and my wife Corrah, for showing me that spirit is all one needs to create dazzling new things.

Thank you to my undergraduate professors, especially Dani Kohen and Chris Calderone, who ignited my fascination with computation and enzymes. In the Baker lab, my sincere gratitude goes to Kiera Sumida, Ian Humphreys, Phil Leung, Matthias Gloegl, Florence Hardy, Florian Praetorius, Jeremy Sims, Naveen Jasti, Aditya Krishnakumar, and all the others who made each day spent pursuing our personal crazy ideas worthwhile, and to lab members and collaborators Ivan Anishchenko, David Juergens, Christoffer Norn, Indrek Kalvet, Yakov Kipnis for support and direction. Thanks to my advisor, David Baker, and my committee members Frank DiMaio, Mike Gelb, Barry Stoddard, and Forrest Michael.

To my friend, mentor, and greatest collaborator, Dr. Samuel J. Pellock: thank you for making the last five years spent chipping away at an incredibly challenging problem so extraordinarily interesting and meaningful.

Table of Contents

Chapter 1: Introduction

Enzymes are the molecular catalysts that make life possible. In their absence, many of the chemical reactions that are critical to living systems are incomprehensibly slow at room temperature, with half times as high as approximately one billion years (*1*). Enzymatic rate enhancements (k_{cat}/k_{non}, where k_{cat} and k_{non} are the rate constants for the catalyzed and uncatalyzed reactions, respectively) range from 10^7 and 10^{19}-fold (*1*), illustrating the ability of enzymes to dramatically accelerate biological processes. Further, in the context of a crowded, heterogeneous cellular environment, enzymes are capable of exquisite selectivity, acting almost exclusively on their intended substrates and performing chemistry with precise regioselectivity and enantioselectivity (*2*). The remarkable power of these molecular machines has motivated considerable efforts to understand the structural principles that underlie enzyme function, and ultimately use them to understand how to design high-activity enzymes given an arbitrary chemical reaction of interest.

1.1 How do enzymes work?

The prevailing theory of enzyme catalysis, that enzymes lower the activation energy of the chemical reactions they catalyze by stabilizing the reaction transition state, was proposed by Linus Pauling in 1946 almost 20 years before the first atomic structure of an enzyme was solved by David Philips and co-workers (*3*). Exactly how this stabilization is achieved is the subject of continued research, but several factors are widely appreciated and well-supported by experimental evidence.

Prior to the visualization of enzyme structures using x-ray crystallography, enzymologists had already suggested several structural hypotheses about how enzymes catalyze chemical reactions. The observation of the extreme specificity of many enzymes prompted Emil Fischer to suggest the “lock and key” model of enzyme

activity, that enzymes must possess strong shape complementarity to their substrates to allow chemistry to occur only upon the substrate(s) of interest (*4*, *5*). Building on this theory, Daniel Koshland later suggested that binding of the substrate may induce conformational changes in the active site that prime the enzyme-substrate complex for catalysis (*6*, *7*). Linus Pauling hypothesized that enzymes bind not only their substrates, but also the transition state of the reaction being accelerated, leading to its stabilization (*3*, *8*). Importantly, in order to lower the activation energy barrier the stabilization of the transition state must be preferential; it must be stabilized to a higher degree than both the ground state and product states. Theoretically, this should be achievable if the enzyme adopts a conformation that is more complementary to the transition state than either the substrate or the product. An obvious alternative strategy therefore is to preferentially *destabilize* the ground state, which would also lead to lowering of the activation energy. This idea is often attributed to Jack Haldane, who described ground state destabilization as the exercise of "strain" on the substrate by binding it in a distorted, higher-energy conformation that is closer to the transition state structure (*9*).

The first atomic-resolution structure of an enzyme was that of lysozyme, determined by David Philips and co-workers in 1965 (*10*, *11*). Analysis of several structures, including unbound and inhibitor-bound structures of the enzyme, affirmed many of the proposed catalytic strategies. Assuming that the inhibitor bound similarly to the true enzyme substrate, Philips noted strong complementarity between the enzyme active site and the substrate, mediated by van der Waals interactions and a number of hydrogen bonds. The binding of the inhibitor to the active site caused a distortion of its sugar ring to an unfavorable conformation which mimicked that of a high energy intermediate expected to form during the reaction, suggesting that lysozyme indeed used strain to accelerate hydrolysis. Two catalytic residues, Asp52 and Glu35, were observed in proximity to the substrate, each occupying distinct chemical environments, leading Philips to suggest a mechanism in which one carboxylate acts as an acid, and the other a base (*10*, *11*). Later, site-directed mutagenesis was employed to confirm the essential role of these two catalytic residues in the reaction; single-point mutations removing either

of these residues led to inactive enzymes (*12*). Finally, the apo and inhibitor-bound structures displayed small differences, most notably in the position of one of the catalytic carboxylates, which shifts closer to the inhibitor in the bound conformation, supporting Koshland's "induced fit" hypothesis (*10*). The determination of the atomic structure of lysozyme was a landmark moment in the study of enzymes, confirming many of the leading theories about how enzymes accelerate chemical reactions and demonstrating that catalytic mechanisms could be inferred through a careful examination of the structure of the active site.

The success of lysozyme inspired structure determination efforts for a variety of enzymes, leading to additional insights regarding the origin of enzymatic rate enhancements. Subsequent structural studies of other enzymes revealed similar trends; enzymes were highly complementary to their substrates and the presumed transition state, and contained pre-positioned catalytic groups that directly participate in the chemical reaction, performing proton transfers or nucleophilic attack (*13*, *14*). Like the catalytic glutamates in lysozyme, these groups appeared to possess unusual chemical properties (pK_a, nucleophilicity) compared to the free functional group in solution, lending support to the idea that the enzyme microenvironment tunes these properties to achieve optimal chemical reactivity, particularly through the exclusion of water and/or the use of complex hydrogen bond networks (*15*). Removal of these groups by site-directed mutagenesis led to dramatic losses in activity, and these effects were not additive, showing that the catalytic groups in the active site act synergistically to catalyze the reaction (*16*, *17*). It was observed that the catalytic groups tend to be positioned with near-ideal conformations for catalysis often even in unbound enzymes, a feature referred to as "preorganization" (*18*, *19*).

The enzyme structures were the "smoking gun," clearly implicating the concepts of transition state stabilization and preorganization in enzymatic activity, but subsequent studies devised by chemists to test these proposals provided equally strong evidence for their involvement. Elegant chemical investigations approximated the effect of holding enzyme catalytic groups near the substrate in reactive conformations by measuring the rates of several intramolecular reactions of carefully designed molecules. These experiments showed dramatic rate

enhancements as high as 10^8-10^{12} resulting from tethering reacting groups together, suggesting that preorganization contributed significantly to enzymatic rate enhancement (*20*). Around this time, inspired by the evidence that enzymes are highly complementary to the transition state of the chemical reaction they catalyze, Wolfenden and co-workers devised a brilliant strategy to obtain inhibitors of enzymatic reactions by synthesizing transition state analogues, small molecules that mimic the structure and charge distribution of the transition state but could not be processed by the enzyme of interest. For successful transition state analogue inhibitors, extremely tight binding to the active site was observed, establishing a reliable method for discovering inhibitors against any enzyme of interest and lending further evidence in support of Pauling's original hypothesis that enzymes act by stabilizing the transition state of the reaction through strong binding interactions (*21–23*).

In the decades following, significant efforts were undertaken to use the above-summarized concepts to make artificial enzymes, giving rise to artificial protein catalysts for a variety of simple, well-understood model reactions.

1.2 Approaches to the generation of artificial enzymes

Given a chemical reaction of interest, undeniably the most reliable method for generating a catalyst to accelerate it is directed evolution. Much like the process used by Nature to evolve enzymes over billions of years, directed evolution is capable of amplifying vanishingly low levels of activity to highly effective catalysis through iterative mutation and selection (*24*). Other properties, such as thermostability, selectivity, solvent tolerance, etc., can also be optimized (*25*). The main limitations of directed evolution arise when initial activity on the reaction of interest is hard to access, or when larger structural changes are required to achieve desired functional properties. Despite the success of directed evolution, there remains an interest in combining the principles of enzyme function with molecular design to rationally generate novel enzymes. This section will review approaches to generating novel enzymes using rational design and their limitations.

Catalytic antibodies

Inspired by the concept that enzymes tightly bind the transition state of the reaction of interest, two research groups working independently showed that antibody catalysts could be obtained by raising antibodies against transition state analogous (*26*, *27*). Subsequently, catalytic antibodies appeared acting on a range of different model reactions with high specificity (*28*). Structure determination of antibody-transition state analogue complexes demonstrated not only precise complementarity to the transition state of the reaction but also the presence of polar residues likely involved in catalysis (*29*).

While there are some notable exceptions, catalytic efficiencies (k_{cat}/K_m) of catalytic antibodies fall in the range of 10^2-10^4 M^{-1} s^{-1}, a few orders of magnitude lower than the typical natural enzyme (*30*). The remaining gap has been attributed to a variety of limitations. Transition state analogues, while similar, necessarily differ from the true transition state in their steric and electronic structure, and thus catalytic antibodies raised against them are unlikely to be perfectly complementary to the transition state. Further, the process uses a single transition state analogue, while complex reactions can involve multiple transition states that may resemble each other but are ultimately different chemical species. Since product states often mimic the transition state geometry, catalytic antibodies tend to tightly bind them, leading to inhibition (*29*). Finally, optimizing only on binding affinity to a single transition state mimic may not necessarily lead to the inclusion of catalytic residues to mediate acid/base or covalent catalysis, and there is little that can be done to control the active site makeup of the resulting catalysts, making it difficult to target a specific catalytic mechanism. In contrast, protein engineering and design techniques offer the flexibility to rationally select and position these essential components according to the intended chemistry.

Protein engineering and computational design

The development of recombinant protein production strategies opened up infinite possibilities in the world of protein engineering. It was now possible to easily edit existing proteins and dream up new ones. Exploration began of a novel approach to enzyme design involving the re-engineering of existing proteins to generate catalytic activity. For many of the model reactions for which antibodies had been developed, it was shown that careful mutagenesis of an otherwise inert protein scaffold could generate catalysts, albeit with exceedingly low activities (*30*).

Design approaches generally involve mutation of existing protein scaffolds to impart new functions, which can be successfully done rationally through human intuition (*31–34*) or using computational active site and sequence design approaches (*35–38*). Rational approaches generally involve mutating helical bundles, which can be easily edited due to their stability and predictable structure, and have been successfully applied to generate hydrolytic enzymes, metal-dependent decarboxylases, acyltransferases, and other catalysts. A more involved approach uses computational design to install catalytic activity into a scaffold. Given a "theozyme," the conformation of the transition state and surrounding catalytic residues, scaffolds capable of accommodating the desired active site can be identified using conformational search algorithms such as RosettaMatch (*39*). The region of the scaffold near the new active site is then redesigned using sequence design algorithms such as Rosetta FastDesign (*40*) or more recently LigandMPNN (*41*), to simultaneously bind the transition state, stabilize the desired orientations of the catalytic residues, and maintain the integrity of the protein scaffold. This approach has been used to design a number of catalytic activities, including reactions not performed by natural enzymes.

1.3 Limitations of designed enzymes

Designed enzymes universally exhibit far lower initial catalytic activity than natural enzymes, although they can reach comparable activities after laboratory evolution (*30*). Significant efforts have been made to discover the origin of this activity gap, leading to a number of illuminating observations about differences in preorganization, active site complexity, and dynamics. MD simulations comparing designed and natural enzymes have demonstrated that catalytic residues in natural enzyme active sites are more preorganized than even the most active computational designs (*42*). In computational designs, key hydrogen bonds in the active site form and break constantly over the course of the simulation, while natural enzymes maintain these networks almost all of the time. A similar trend emerged from retrospective studies using MD of active and inactive designs, which suggested that the reason many of the inactive designs failed was because they rarely form the required hydrogen bonding interactions in solution (*36*, *42*). Experimental analysis of designed retroaldolases also revealed the absence of the hallmarks of natural enzymes, such as preorganized catalytic residues and highly specific substrate binding and positioning. For the most active designed retroaldolase RA61, it was shown that transient, non-specific enzyme-substrate interactions were almost exclusively responsible for the rate enhancement, while a designed catalytic motif did not contribute to activity (*43*). Taken together, both computational and experimental studies suggested that the initial successes of computational enzyme design differed significantly from Nature's catalysts, lacking precise substrate binding interactions and positioned catalytic groups.

The structural features that contribute to specific binding and preorganization are extremely challenging to design using current state of the art computational methods. In Nature, preorganization is achieved through a combination of hydrogen bonding interactions and precise steric packing interactions

surrounding and positioning the catalytic residues and substrate. These features, which are mediated predominantly by sidechains, are difficult to emulate in designed enzymes, which tend to contain fewer extended hydrogen bond networks in their active sites (*35*–*38*). Directed evolution efforts on designed enzymes often lead to higher activity enzymes that display increased active site complexity, as evidenced by the emergence of a complex hydrogen bonding arrangement over the evolution of a designed retroaldolase (*44*). Dynamic effects like backbone movements can also affect preorganization, as even small conformational shifts can break hydrogen bonds or disrupt transition state shape complementarity. A recent study of a designed Kemp eliminase catalogues these catalytically unproductive motions, identifying two conformational states of a loop near the active site, one inactive and one active, and showed that directed evolution of the enzyme led to mutations that increased activity by stabilizing the active loop conformation (*45*). These results illustrate that achieving preorganization is a formidable challenge in enzyme design: designed active sites must be supported by precise hydrogen bonds and hydrophobic packing, and to achieve this reliably requires extremely accurate *in silico* prediction of both backbone and sidechain conformations.

Since in most cases only a limited number of designs can be synthesized, purified, and screened for activity, it is necessary to develop methods that can accurately evaluate preorganization in potential designs. Protein structure prediction methods only attempt to capture the single most likely conformation of a protein and its sidechains; in contrast, ensemble methods are significantly more useful for evaluating preorganization. Given a reasonably accurate starting structure, which can be generated by structure prediction tools such as AlphaFold2 (*46*), the information captured in MD trajectories appears to be strongly predictive of design success (*30*, *42*). However, it is intractable to run MD simulations on a large set of designed enzymes for *in silico* screening, a limitation which is exacerbated as enzyme design efforts begin to target more complex enzymatic mechanisms that require preorganization across multiple chemical steps.

A second shortcoming of the current enzyme design pipeline is the use of existing protein scaffolds, which was previously necessary because of severe limitations in the ability to design new protein backbones from first principles. Most enzyme design efforts to date have started from a curated set of native scaffolds (*35*–*38*, *47*), simple helical bundle scaffolds as mentioned previously (*32*–*34*, *48*, *49*), or more recently a large set of deep-learning generated NTF2-fold scaffolds (*50*, *51*). While these have been sufficient for a number of reactions that require relatively simple active sites involving sidechain functional groups, as active site complexity increases, the likelihood that these finite sets of scaffolds can properly accommodate the desired active site positioning drops dramatically, and the computational cost of searching for compatible scaffolds becomes an additional bottleneck. For example, in an attempt to design cysteine hydrolases, an active site that contains five catalytic elements in Nature was simplified by the removal of two of the catalytic residues in order to successfully identify placements in a set of native scaffolds, likely at a detriment to overall enzymatic activity (*38*). When the desired catalytic site involves backbone functional groups or larger backbone-level motifs for binding cofactors or large charged substrates such as ATP, it is even more unlikely that a compatible scaffold can be identified. In response, researchers either a) reduce the complexity of the active site by removing the elements deemed to be the least important or most problematic in terms of scaffold compatibility, a decision almost guaranteed to ultimately reduce the level of activity that can be achieved, or b) choose an entirely different reaction that uses simpler active site machinery. This challenge has therefore severely limited both the activities achieved with designed enzymes and the scope of reactions that have been attempted using computational enzyme design.

1.4 New opportunities with recent advances in protein design

Previous enzyme design efforts have been seriously hampered by an inability to design scaffolds directly for the reaction of interest, and a lack of fast and accurate methods to evaluate the preorganization of designs *in*

silico. Recent advances in protein structure prediction and design have the potential to revitalize enzyme design. The advent of the highly accurate protein structure prediction tools such as AlphaFold2 (*46*) and RosettaFold (*52*, *53*) have made it possible to reliably evaluate the likelihood that a designed enzyme will fold to the desired structure, and they can be repurposed to generate novel protein backbones (*53*, *54*). Concurrently, the recently-developed sequence design tools ProteinMPNN (*55*) and LigandMPNN (*41*) can select high-quality sequences for these backbones, and are better at producing soluble, highly expressed proteins than previous sequence design tools based on physics-based energy functions. These advancements open up a new paradigm in enzyme design, in which diverse scaffolds are generated around a three-dimensional constellation of the substrate and catalytic residues in a transition state orientation. This guarantees that the resulting designs contain all of the required catalytic features in the desired orientation and a binding pocket for the substrate(s), thus allowing for the generation of designed enzymes harboring significantly more complex active sites for chemical reactions that were previously deemed too challenging.

ChemNet (*51*) is a recently-developed method that we have utilized to rapidly evaluate the degree of preorganization of a given design *in silico*. As input, ChemNet receives the backbone coordinates of a given protein in a 3D region of interest, along with the atom identities and connectivity for all sidechains and small molecules (if applicable) found in the region. It then stochastically predicts the coordinates of the sidechains and small molecules. Repeated runs produce conformational ensembles of active site sidechains and the substrate, which can be used in place of MD trajectories that are far too slow to generate predictions for large numbers of designs. ChemNet can be reasonably run on tens of thousands of designs, and can also predict chemical modifications of the active site, making it especially powerful for selecting enzymes capable of catalyzing multistep reactions, since the conformation of the enzyme at each mechanistic step of the reaction can be separately predicted and evaluated.

1.5 Dissertation overview

In this work, we combine recent advances in protein design methodologies to invert the traditional enzyme design pipeline, designing protein scaffolds around an active site of interest with RFdiffusion, and performing stringent *in silico* preorganization filtering with ChemNet. As a proof of concept, we use the well-understood but historically challenging ester hydrolysis mechanism of natural serine hydrolases as a model system (*56*). We show that the combination of RFdiffusion backbone generation, sequence design with LigandMPNN, structure prediction with AlphaFold2, and ChemNet filtering enables the design of enzymes for this previously unsolved challenge. The resulting enzymes exhibit higher initial catalytic activities than all previous designed enzymes, contain accurately designed yet complex active sites, and span a range of novel folds with no relation to natural ester hydrolases.

1.6 References

1. R. Wolfenden, M. J. Snider, The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **34**, 938–945 (2001).

2. L. Hedstrom, Enzyme Specificity and Selectivity, *Encyclopedia of Life Sciences* (2010). https://doi.org/10.1002/9780470015902.a0000716.pub2.

3. L. Pauling, Nature of forces between large molecules of biological interest. *Nature* **161**, 707–709 (1948).

4. E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985–2993 (1894).

5. F. W. Lichtenthaler, 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? *Angew. Chem., Int. Ed. Engl.* **33**, 2364–2374 (1995).

6. D. E. Koshland, Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **44**, 98–104 (1958).

7. D. E. Koshland Jr, The key–lock theory and the induced fit theory. *Angew. Chem., Int. Ed. Engl.* **33**, 2375–2378 (1995).

8. L. Pauling, Molecular Architecture and Biological Reactions: Answers to many basic problems of biology-nature of growth, mechanism of duplication of viruses and genes, action of enzymes, mechanism, of physiological activity of drugs, hormones, and vitamins, structure and action of nerve and brain tissue-may lie in knowledge of molecular structure and intermolecular reactions. *Chem. Eng. News Archive* **24**, 1375–1377 (1946).

9. J. B. S. Haldane, *Enzymes* (Longmans, Green and Co., London, 1930).

10. C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, V. R. Sarma, Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **206**, 757–761 (1965).

11. D. C. Phillips, The three-dimensional structure of an enzyme molecule. *Sci. Am.* **215**, 78–90 (1966).

12. B. A. Malcolm, S. Rosenberg, M. J. Corey, J. S. Allen, A. de Baetselier, J. F. Kirsch, Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 133–137 (1989).

13. G. N. Reeke, J. A. Hartsuck, M. L. Ludwig, F. A. Quiocho, T. A. Steitz, W. N. Lipscomb, The structure of carboxypeptidase a, vi. Some results at 2.0-å resolution, and the complex with glycyl-tyrosine at 2.8-å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 2220–2226 (1967).

14. B. W. Matthews, P. B. Sigler, R. Henderson, D. M. Blow, Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature* **214**, 652–656 (1967).

15. T. K. Harris, G. J. Turner, Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002).

16. P. Carter, J. A. Wells, Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564–568 (1988).

17. D. R. Corey, C. S. Craik, An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.* **114**, 1784–1790 (1992).

18. F. M. Menger, On the source of intramolecular and enzymatic reactivity. *Acc. Chem. Res.* **18**, 128–134 (1985).

19. J. Villà, A. Warshel, Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **105**, 7887–7907 (2001).

20. F. M. Menger, F. Nome, Interaction vs preorganization in enzyme catalysis. A dispute that calls for resolution. *ACS Chem. Biol.* **14**, 1386–1392 (2019).

21. R. Wolfenden, Transition state analogues for enzyme catalysis. *Nature* **223**, 704–705 (1969).

22. R. Wolfenden, Analog approaches to the structure of the transition state in enzyme reactions. *Acc. Chem. Res.* **5**, 10–18 (1972).

23. G. E. Lienhard, Enzymatic catalysis and transition-state theory: Transition-state analogs show that catalysis is due to tighter binding of transition states than of substrates. *Science* **180**, 149–154 (1973).

24. F. H. Arnold, Directed evolution: Bringing new chemistry to life. *Angew. Chem. Int. Ed Engl.* **57**, 4143–4148 (2018).

25. J. Kaur, R. Sharma, Directed evolution: an approach to engineer enzymes. *Crit. Rev. Biotechnol.* **26**, 165–199 (2006).

26. A. Tramontano, K. D. Janda, R. A. Lerner, Catalytic antibodies. *Science* **234**, 1566–1570 (1986).

27. S. J. Pollack, J. W. Jacobs, P. G. Schultz, Selective chemical catalysis by an antibody. *Science* **234**, 1570–1573 (1986).

28. P. G. Schultz, Catalytic antibodies. *Angew. Chem., Int. Ed. Engl.* **28**, 1283–1295 (1989).

29. D. Hilvert, Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751–793 (2000).

30. G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K. N. Houk, Computational enzyme design. *Angew. Chem. Int. Ed Engl.* **52**, 5700–5725 (2013).

31. K. S. Broo, L. Brive, P. Ahlberg, L. Baltzer, Catalysis of hydrolysis and transesterification reactions of p-nitrophenyl esters by a designed Helix−Loop−Helix dimer. *J. Am. Chem. Soc.* **119**, 11362–11372 (1997).

32. J. Kaplan, W. F. DeGrado, De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11566–11570 (2004).

33. S. Chakraborty, J. Yudenfreund Kravitz, P. W. Thulstrup, L. Hemmingsen, W. F. DeGrado, V. L. Pecoraro, Design of a three-helix bundle capable of binding heavy metals in a triscysteine environment. *Angew. Chem. Weinheim Bergstr. Ger.* **123**, 2097–2101 (2011).

34. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).

35. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas 3rd, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).

36. J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309–313 (2010).

37. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).

38. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).

39. A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).

40. J. B. Maguire, H. K. Haddox, D. Strickland, S. F. Halabiya, B. Coventry, J. R. Griffin, S. V. S. R. K. Pulavarti, M. Cummins, D. F. Thieker, E. Klavins, T. Szyperski, F. DiMaio, D. Baker, B. Kuhlman, Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).

41. J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock, D. Baker, Atomic context-conditioned protein sequence design using LigandMPNN, *bioRxiv* (2023)p. 2023.12.22.573103.

42. G. Kiss, D. Röthlisberger, D. Baker, K. N. Houk, Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).

43. J. K. Lassila, D. Baker, D. Herschlag, Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4937–4942 (2010).

44. R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).

45. R. Otten, R. A. P. Pádua, H. A. Bunzel, V. Nguyen, W. Pitsawong, M. Patterson, S. Sui, S. L. Perry, A. E. Cohen, D. Hilvert, D. Kern, How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).

46. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

47. S. Rajagopalan, C. Wang, K. Yu, A. P. Kuzin, F. Richter, S. Lew, A. E. Miklos, M. L. Matthews, J. Seetharaman, M. Su, J. F. Hunt, B. F. Cravatt, D. Baker, Design of activated serine-containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10**, 386–391 (2014).

48. C. M. Summa, M. M. Rosenblatt, J.-K. Hong, J. D. Lear, W. F. DeGrado, Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J. Mol. Biol.* **321**, 923–938 (2002).

49. E. A. Naudin, A. G. McEwen, S. K. Tan, P. Poussin-Courmontagne, J.-L. Schmitt, C. Birck, W. F. DeGrado, V. Torbeev, Acyl transfer catalytic activity in DE Novo designed protein with N-terminus of α-helix as oxyanion-binding site. *J. Am. Chem. Soc.* **143**, 3330–3339 (2021).

50. A. H.-W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk, D. Baker, De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).

51. I. Anishchenko, Y. Kipnis, I. Kalvet, G. Zhou, R. Krishna, S. J. Pellock, A. Lauko, G. R. Lee, L. An, J. Dauparas, F. DiMaio, D. Baker, Modeling protein-small molecule conformational ensembles with ChemNet. *bioRxivorg*, doi: 10.1101/2024.09.25.614868 (2024).

52. M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

53. R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. DiMaio, D. Baker, Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).

54. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, De novo design of protein structure and function with RFdiffusion. *Nature*, doi: 10.1038/s41586-023-06415-8 (2023).

55. J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, Robust deep learning based protein sequence design using ProteinMPNN, *bioRxiv* (2022)p. 2022.06.03.494563.

56. L. Hedstrom, Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4524 (2002).

Chapter 2: Computational design of serine hydrolases

At the time of writing, a version of this chapter is under review and currently available as a pre-print:
A. Lauko, S. J. Pellock, I. Anischanka, K. H. Sumida, D. Juergens, W. Ahern, A. Shida, A. Hunt, I. Kalvet, C. Norn, I. R. Humphreys, C. Jamieson, A. Kang, E. Brackenbrough, A. K. Bera, B. Sankaran, K. N. Houk, D. Baker, Computational design of serine hydrolases. *bioRxiv* (2024).

Authors: Anna Lauko[1,2,3,†], Samuel J. Pellock[1,2,†], Ivan Anischenko[1,2], Kiera H. Sumida[1,2,5], David Juergens[1,2,4], Woody Ahern[1,2,7], Jihun Jeung[1,2], Alex Shida[1,2], Andrew Hunt[1,2], Indrek Kalvet[1,2,6], Christoffer Norn[1,2], Ian R. Humphreys[1,2], Cooper Jamieson[8], Rohith Krishna[1,2], Yakov Kipnis[1,2], Alex Kang[1,2], Evans Brackenbrough[1,2], Asim K. Bera[1,2], Banumathi Sankaran[1,2], K. N. Houk[8], David Baker[1,2,6*]

Affiliations:

[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 Biological Physics, Structure and Design, University of Washington, Seattle, WA, USA

[4]Graduate Program in Molecular Engineering, University of Washington, Seattle, WA, USA

[5]Department of Chemistry, University of Washington, Seattle, WA, USA

[6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

[7]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.

[8]Department of Chemistry and Biochemistry, University of California, Los Angeles, California, USA

†These authors contributed equally: Anna Lauko, Samuel J. Pellock

*Corresponding author. Email: dabaker@uw.edu

2.1 Abstract

Enzymes that proceed through multistep reaction mechanisms often utilize complex, polar active sites positioned with sub-angstrom precision to mediate distinct chemical steps, which makes their de novo construction extremely challenging. We sought to overcome this challenge using the classic catalytic triad and oxyanion hole of serine hydrolases as a model system. We used RFdiffusion to generate proteins housing catalytic

sites of increasing complexity and varying geometry, and a newly developed ensemble generation method called ChemNet to assess active site geometry and preorganization at each step of the reaction. Experimental characterization revealed novel serine hydrolases with folds distinct from natural serine hydrolases that catalyze ester hydrolysis with catalytic efficiencies (k_{cat}/K_m) up to $3.9 \times 10^3\ M^{-1}\ s^{-1}$ and structures that closely match the design models (crystal structure Cα RMSDs < 1 Å). In silico selection of designs based on active site preorganization across the reaction coordinate considerably increased success rates, enabling identification of new catalysts in screens of as few as 20 designs. Our de novo buildup approach provides insight into the geometric determinants of catalysis that complements what can be obtained from structural and mutational studies of native enzymes (in which catalytic group geometry and active site makeup cannot be so systematically varied), and provides a roadmap for the design of industrially relevant serine hydrolases and, more generally, for designing complex enzymes that catalyze multi-step transformations.

2.2 Introduction

Enzymes are powerful catalysts that dramatically accelerate reaction rates in mild aqueous conditions. The ability to construct enzymes catalyzing arbitrary chemical reactions would have enormous utility across a wide range of applications, and hence, enzyme design has been a long-standing goal of computational protein design (*1*). De novo enzyme design has generally started from a specification of arrangements of catalytic residues around the reaction transition state (a theozyme), and sought to identify placements of this active site in pre-existing scaffolds (*2–7*). Fixed backbone scaffolds restrict how accurately the catalytic geometry can be realized, and this has likely limited the activities of many designed enzymes to date prior to optimization by laboratory evolution, as recent studies of designed kemp eliminases demonstrate (*8–10*). A further challenge of enzyme design is the preorganization of the active site such that the catalytic protein functional groups are

positioned relative to the transition state with atomic accuracy. Achieving preorganization is especially difficult for multistep reaction mechanisms, because the enzyme must preferentially stabilize multiple transition states and intermediates. Existing methods to evaluate design preorganization in silico (*7*, *11–14*) are limited by low accuracy or computational cost and are typically only applied to one reaction state. To enable the accurate design of multistep enzymes, new methods are needed for both the generation of protein backbones tailored specifically to a given active site and assessment of their structural compatibility throughout the catalytic cycle.

Ester hydrolysis has served as a model reaction for computational enzyme design for decades (*15–20*), and justifiably so: numerous mechanisms catalyze ester hydrolysis, enabling a range of distinct design approaches to target this reaction, activity is easily monitored by absorbance and fluorescence with reporter substrates, and esterases are highly valuable in industrial processes, most recently for their application in plastic recycling (*21–23*). The textbook example of enzymatic ester hydrolysis is the double-displacement reaction mechanism employed by serine hydrolases, in which a serine nucleophile undergoes acylation to form a stable acyl-enzyme intermediate (AEI) that is subsequently hydrolyzed by an activated water. Despite extensive structural, mutational, and computational characterization of the mechanism of serine hydrolases found in Nature (*24–35*), de novo design efforts attempting to employ this machinery have been unsuccessful, and to our knowledge, no previous efforts have successfully constructed a serine hydrolase that extends beyond the fold space found in nature. A major challenge in designing serine hydrolases is overcoming the stability of the AEI, the resolution of which is typically rate-limiting when activated esters are employed. Numerous previously designed enzymes and peptide-based systems inactivate or dramatically slow down after acylation (*6*, *15–17*). In addition to this chemical challenge, constructing the serine hydrolase active site combines some of the most difficult current challenges in protein design: 1) the catalytic site is very complex, requiring the scaffolding of at least four individual residues with atomic precision, a task that state-of-the-art design tools struggle to achieve

(*36*), 2) the serine nucleophile is otherwise inert, requiring activation by construction of intricate hydrogen bond networks, and 3) the design must accommodate subtle conformational changes throughout the multistep catalytic cycle, and while there is recent progress in multistate design (*37*), it remains challenging, particularly when the energetic differences between desired states are small.

Previous efforts to design esterases have circumvented the challenges presented by serine hydrolases by employing simpler, more easily designable active sites, leveraging nucleophiles more activated than serine, and by targeting reaction mechanisms that do not require the formation of stable covalent intermediates. For example, previously designed metallohydrolases skip the AEI by activating water to cleave esters in a single step (*18*, *38*), the non-canonical amino acid N_δ-methylhistidine has been elegantly employed to make the AEI less stable (*17*), and cysteine has been used in place of serine due to its greater nucleophilicity (*6*, *15*). Structural analysis of the resulting cysteine esterases indicated key interactions between the cysteine nucleophile and histidine base of the desired dyad or triad were not formed (*6*, *15*), suggesting that the inherent chemical reactivity of the residues employed, not their coordinated effort, may have been responsible for the observed steady-state rate enhancements. Even with these chemical interventions, the efficiency of the initial computational designs from these prior efforts remain low and far below the range observed for natural enzymes.

A potential explanation for the lack of designed serine hydrolases to date stems from incompatibility of the complex active site with the protein scaffolds previously used for hydrolase design (*6*). We first investigated whether increasing scaffold diversity could help identify backbones that more accurately reconstruct the desired active site, and carried out a preliminary design campaign searching for placements of a serine hydrolase active site in a library of NTF2 scaffolds that previously yielded catalysts for a luciferase reaction (*39*). As in previous studies, experimental characterization of the resulting designs revealed activated serines but no catalytic turnover

on activated ester substrates, despite a close match between the experimental and designed structures (Fig. S1). We suspect that an inability to install common catalytic features found in natural hydrolases into NTF2s, including at least one backbone oxyanion hole contact common to all serine hydrolases, limited the function of these designs.

We reasoned that advances in deep learning for protein design could enable the design of proteins from scratch to directly scaffold the serine hydrolase active site and assess design compatibility for the entire multistep catalytic cycle. Recent advances in scaffolding functional sites with RFdiffusion have yielded improved *in silico* and experimental success rates across a range of design tasks (*36*, *40*, *41*); we aimed to use the same approach to generate serine hydrolases starting from geometric descriptions of an active site (Fig. 2.1A). To assess preorganization and functional interactions in each step of the catalytic cycle, we sought to leverage advances in deep learning-based prediction of protein-small molecule complexes by modeling structural ensembles of catalytic intermediates (Fig. 2.1B).

2.3 Assessing reaction path compatibility with ChemNet

We set out to understand why previously designed serine hydrolases failed to appreciably catalyze ester hydrolysis and hypothesized that modeling states across the complete reaction coordinate could be critical for assessing the ability of a design to achieve catalytic turnover. To model the extent to which a designed enzyme can form each of the key states along the reaction cycle and to assess the preorganization of the active site residues in the desired catalytic geometries, we developed a deep neural network that, given 1) the backbone coordinates of a small molecule binding pocket or active site, 2) the identities of the amino acid residues at each position, and 3) the chemical structures of bound small molecules (but not their positions), generates the full

atomic coordinates of the binding site, comprising both protein sidechains and small molecules. We trained this network, called ChemNet (*42*), on protein-small molecule complexes in the PDB by randomizing the atomic coordinates of sidechains and small molecules within spherical regions with up to 600 heavy atoms, and seeking to minimize a loss function assessing the recapitulation of the atomic coordinates within the region. ChemNet rebuilds regions within native structures with an average RMSD of 1.1 Å. ChemNet is stochastic, and repeated runs from different random seeds yield an ensemble of models for the rebuilt region.

We used ChemNet to generate structural ensembles for each step of the catalytic cycle for a set of native and previously designed serine hydrolases. The catalytic cycle of serine hydrolases can be divided into four steps (Fig. 2.1C). First, the substrate binds to the apoenzyme (apo) and the catalytic serine, deprotonated by the catalytic histidine, attacks the carbonyl carbon of the ester to form the first tetrahedral intermediate (TI1). Second, the catalytic histidine protonates the leaving group oxygen promoting its departure, leaving the active site serine covalently linked to the acyl group of the substrate (acyl-enzyme intermediate, AEI). Third, the histidine deprotonates a water molecule, which attacks the AEI to generate a second tetrahedral intermediate (TI2). Finally, this intermediate is resolved by histidine-mediated protonation of serine and release of the acyl group, reconstituting the free enzyme and completing the catalytic cycle. Throughout, negatively charged transition states and intermediates are stabilized by a pair of hydrogen bond donors that constitute the oxyanion hole. Perturbation of the histidine pK_a, which tunes its acid/base function, is mediated by interaction with aspartate or glutamate, the final residue in the triad (*43–45*).

Modeling this catalytic cycle with ChemNet showed that native hydrolases are considerably more preorganized than previous designed systems (Fig. 2.1D, Fig. S2). In native systems, the catalytic residues at each step sample a very limited number of conformations in which all key hydrogen bonding interactions are

maintained, but in designed systems there were often wide variations in the ensembles at multiple steps. Since the reaction rate should be proportional to the fraction of the enzyme in the active state, the lack of preorganization of the designed active sites is expected to compromise catalysis. To quantify the extent of active site formation in the ChemNet ensembles, we compute the frequency of formation of key interactions between the catalytic functional groups and reaction intermediates over each step of the reaction (see Computational methods, filtering section), and use this metric, which clearly distinguishes between the native hydrolases and previous designed hydrolases (Fig. S2), to assess the new designs in the following sections.

2.4 Design and characterization of serine hydrolases

We next set out to design proteins with active sites of increasing complexity, using RFdiffusion to scaffold serine hydrolase active site motifs and ChemNet to assess their preorganization in each step of the reaction (Fig. 2.2A,B). We designed catalysts for the hydrolysis of 4-methylumbelliferone (4MU) esters (Fig. 2.2C) that fluoresce upon hydrolysis. To generate backbones to scaffold the catalytic machinery, we placed the catalytic sidechains around the substrate and starting from the backbone N, Cα, and C atoms of these key residues and their immediate neighbors (i.e. a contiguous three-residue segment), used RFdiffusion to build up backbones, starting from random noise, which have coordinates that exactly match the input motif and also form a binding pocket for the substrate (see Computational Methods, motif generation and backbone generation). To drive folding to the designed state, and to make favorable interactions with the substrate and active site residues, LigandMPNN (*46*) was used to design the sequence. Rosetta FastRelax (*47*) was used to refine the protein backbone and ligand pose, and the sequence was again designed with LigandMPNN with the new backbone as input (*48*). Following several iterations between LigandMPNN and FastRelax, the structures of the designs were predicted with AlphaFold2 (AF2) (*49*), and designs for which all catalytic residue Cα's were

positioned within 1.0 Å of the design models were selected for experimental characterization (*49*) (see Computational methods, sequence design for details).

In the first two rounds of design, we built relatively simple active sites consisting of Ser-His dyads with a single oxyanion hole contact from the backbone amide of the serine (Fig. 2.2A,B), and explicitly evaluated the utility of ChemNet to select designs for experimental characterization; round 1 designs were filtered with AF2 alone, while round 2 designs that passed the AF2 filter were selected for experimental screening if ChemNet ensembles of the apo state indicated the key Ser-His hydrogen bond was formed (see Computational Methods, filtering). Only 1.6% of round 2 designs that passed AF2 filtering were predicted to be preorganized by ChemNet. For experimental testing, we obtained synthetic genes encoding 129 and 192 designs for rounds 1 and 2, respectively, for *E. coli* overexpression and screening.

We used a fluorophosphonate (FP) activity-based probe and fluorescent 4MU-acetate (4MU-Ac) and 4MU-butyrate (4MU-Bu) ester substrates to identify designs with activated serines and esterase activity, respectively (Fig. 2.2C). The fraction of designs labeling with the FP probe in *E. coli* lysate increased nearly 5-fold from 3% to 17% from round 1 to round 2 (Fig. 2.2B, Fig. S3). Designs that reacted with the FP probe were purified and incubated with 4MU esters, and two round 1 designs (1.6%) and 10 round 2 designs (5.2%) showed catalytic activity. Retrospective ChemNet analysis of the round 1 designs revealed that the Ser-His H-bonds in the two catalytically active designs were predicted to be among the most preorganized (Fig. S4). ChemNet filtering of round 2 designs on the extent of formation of the key Ser-His H-bond not only increased the fraction of designs exhibiting FP probe labeling and enzymatic activity, but also resulted in higher activities (Fig. 2.2E,F). The progress curves for these round 1 and 2 designs plateau after approximately one enzyme equivalent of fluorescent product is formed (Fig. 2.2E), suggesting they catalyze initial nucleophilic attack but

fail to hydrolyze the AEI, the rate-limiting step in the cleavage of activated esters (*32*). When incubated with substrate, mass spectra of these designs revealed a mass shift corresponding to acylation, further supporting protein inactivation following formation of the acylated intermediate (Fig. S5).

We hypothesized that incorporating a histidine-stabilizing catalytic acid and a second oxyanion hole H-bond donor in a third round of designs (round 3) and filtering for ChemNet preorganization in both the apo and AEI states could generate designs capable of catalytic turnover via hydrolysis of the AEI. For round 3 designs, we required all catalytic triad and oxyanion hole H-bonds to be highly preorganized in ChemNet ensembles of both the apo and AEI states. Of 132 round 3 designs, 111 (84%) displayed FP probe labeling, 20 hydrolyzed 4MU substrates (18%), and two designs (1.5%) displayed multiple turnover activity (Fig. 2.2B,E). Active designs from all three rounds showed significantly reduced activity upon mutation of any one of the catalytic residues (Ser, His, Asp/Glu, and oxyanion sidechain contact) (Fig. 2.2E), suggesting that the observed activities are dependent on the designed active site. To determine the kinetic parameters of the active designs, initial or steady-state rates were measured to determine k_2/K_m or k_{cat}/K_m for single-turnover and multiple-turnover designs, respectively (Fig. 2.2E, Fig. S6). For the two designs that displayed catalytic turnover, called **'super'** and **'win**,' k_{cat}/K_m values were 22 M^{-1} s^{-1} (k_{cat} = 0.00137 ± 0.00005 s^{-1}, K_m = 64 ± 6 μM) and 410 M^{-1} s^{-1} (k_{cat} = 0.00117 ± 0.00003 s^{-1}, K_m = 2.8 ± 0.3 μM), respectively for the more preferred of the two 4MU substrates (**win** and **super** preferentially hydrolyzed 4MU-Ac and 4MU-Bu, respectively (Fig. S7)). Despite the low K_m observed for **win**, we were unable to reach saturation of the initial burst phase of the reaction by increasing substrate concentration up to 100 μM (Fig. S8), suggesting that $K_s >> K_m$ and that the low apparent K_m observed for **win** is a result of rapid acylation and not tight substrate binding.

2.5 Structural characterization of designed serine hydrolases

We pursued x-ray crystallography to determine the accuracy with which **super** and **win** were designed. We were able to solve crystal structures of both **super** and **win**, and found that they had very low Cα RMSDs of 0.8 Å over 165 residues and 0.83 Å over 160 residues (Fig. 2.3A,D), respectively, to the design models. The very close agreement between experimental and designed structures extends to the geometry of the active site: the sidechain conformations of the catalytic residues are in atomic agreement for **super** (all-atom RMSD = 0.38 Å over 22 atoms) and for **win** (all-atom RMSD = 0.86 Å over 20 atoms) except for a rotamer shift in the sidechain oxyanion contact, T99 (Fig. 2.3B,E). In the active site of **super**, a water molecule sits above the nucleophilic serine and forms hydrogen bonds with the oxyanion hole contacts, which likely mimics the positioning of the carbonyl oxygen of its ester substrate (Fig. 2.3B). Similarly, in **win**, an acetate molecule is positioned at the catalytic center and hydrogen bonds to the Oγ of the catalytic serine S142, the oxygen of T99, and the histidine acid/base residue H17 (Fig. 2.3E).

While the structures were solved in the absence of bound small molecule substrate or transition state analogue, overlay of the design model and crystal structure of **super** reveals high shape complementarity to the butyrate acyl group of its preferred substrate (Fig. 2.3C and S7). At the same time, the 4MU moiety is largely exposed, corroborating the selectivity of **super** for 4MU-Bu over 4MU-Ac and suggesting that substrate binding, in this case, is largely driven by binding to the acyl group. For **win**, a rotamer shift in F98 in the crystal structure would clash with the butyrate moiety, and indeed, **win** is selective for the smaller substrate 4MU-Ac that avoids this clash (Fig. 2.3F and S7).

The structures of **super** and **win** are very different from known structures; the closest matches found

with TM-align to the PDB and larger AlphaFold database have TM-scores of 0.41/0.46 (PDB/AlphaFold database) and 0.46/0.51 (at or below the 0.5 cutoff below which structures are considered to have different topological folds), are proteins of unknown function, and have no similarity to known hydrolases at the fold or active site level (Fig. S9), demonstrating that the design method employed here can find protein structural solutions that extend well beyond those found in nature, and to our knowledge demonstrates the first example of a designed serine hydrolase that extends beyond the structural space of that found in nature.

2.6 Filtering for preorganization across the reaction coordinate improves catalysis

We next sought to generate and compare designs filtered explicitly with ChemNet for preorganization over two states (apo and AEI) or over all four states of the reaction path by carrying out additional iterations of LigandMPNN and FastRelax starting from the active design **win** (fixing only the identities of the four catalytic residues) (Fig. 2.4A). We obtained genes encoding 45 two-state filtered designs for experimental characterization, all of which were diverse in sequence compared to the original designs (mean sequence identity to the parent design of 58% and 61% within the active site), and found 38 (84%) labeled with FP-probe. Three of these, **win1**, **win11**, and **win31**, displayed higher k_{cat} values compared to the starting design: **win** has a k_{cat} of 0.00117 s^{-1}, which increases 15-fold in **win1** (0.018 s^{-1}), 17-fold in **win11** (0.0197 s^{-1}), and 9-fold in **win31** (0.0105 s^{-1}) (Fig. 2.4B and Fig. S6). Of the 11 four-state filtered designs tested, 10 (91%) labeled with FP-probe (Fig. S10). Two of these, **dad_t1** and **win_t4**, displayed higher catalytic efficiencies than the starting design, with k_{cat}/K_m values of 3800 M^{-1} s^{-1} and 640 M^{-1} s^{-1}, driven by increases to k_{cat} and decreases in K_m relative to win (Fig. 2.4B and S6). Catalytic triad residue knockouts for all designs showed significant reductions in activity. In **win11** and **win31**, mutation of stabilizing residues in the second shell of the active site that H-bond to the catalytic aspartate also significantly reduced activity (Fig. S11). The two redesigns with the highest k_{cat} values (**win1** and **win11**) do not

display burst phase kinetics, suggesting that deacylation is no longer rate-limiting (Fig. S6).

We determined the crystal structures of **win1** and **win31** and comparison to the design models revealed Cα RMSDs of 1.42 Å and 0.7 Å, respectively (Fig. 2.4E,F). For **win1**, the active site closely matches the designed architecture (mean all-atom RMSD = 0.54 Å) (Fig. 2.4E), and T99, the oxyanion hole contact, occupies the designed rotamer, which may account for the 15-fold increase in k_{cat} compared to **win**, in which T99 is flipped. In one of the two chains in the **win1** structure, the catalytic serine partially occupies a flipped conformer with an occupancy of 0.33, suggesting there is still room for improved stabilization of the catalytically productive conformation. For **win31**, five chains are present in the asymmetric unit, all of which closely match the design model (average Cα RMSD = 0.7 Å) at the backbone level (Fig. 2.4F and S11). Analysis of the active site across all chains in the asymmetric unit revealed mobility in the catalytic serine, sidechain oxyanion threonine, and a second shell tyrosine (Fig. S12), but overall a very close match to the design model with a mean all-atom RMSD of 0.7 Å. Tartrate, derived from the crystallization solution, fit the electron density present in the active site of all five chains, and forms hydrogen bonds with the serine, histidine, and oxyanion hole contacts (Fig. 2.4F), likely mimicking key contacts employed throughout the catalytic cycle.

We next explored whether stringent ChemNet filtering for optimal catalytic geometry and preorganization across the reaction coordinate could generate active esterases with novel backbone topologies, active sites, and substrates. We carried out extensive sequence redesign and filtering based on catalytic geometry in all four states starting from round 3 backbones that had not previously displayed esterase activity, and of 20 designs tested, two (**charlie_t2** and **ken_t1**) displayed significant esterase activity, with catalytic efficiencies of 180 M^{-1} s^{-1} and 1400 M^{-1} s^{-1} (Fig. 2.4G,H,I,J).

To test the generality of ChemNet filtering, we applied it to a different substrate, 4MU-phenylacetate (4MU-PhAc), and a different active site configuration in which the oxyanion hole consists of two backbone amides, rather than a backbone amide and a sidechain, and the first backbone donor was the residue following the catalytic serine rather than the catalytic serine itself (Fig. 2.4K). We used the design pipeline described above to generate 66 designs for this new substrate and catalytic site. The most active of these, **momi**, displayed a k_{cat}/K_m of 1240 M^{-1} s^{-1} and k_{cat} of 0.1 s^{-1}, a 5-fold faster rate than **win11**, the previous best design in terms of turnover number. The distribution of folds generated by RFdiffusion for this active site geometry differed from that for the original geometry, with more α/β fold solutions (as in the case of **momi**), showing how the RFdiffusion buildup approach crafts overall protein structure topology to the specific active site of interest. The high activity achieved without any prior experimental characterization for this new substrate and catalytic site shows that filtering for preorganization across the reaction cycle can yield novel catalysts in one shot.

Several experimental results identify areas to address for improved function. First, ken_t1 inactivates after roughly 10 turnovers, and mass spectra of the catalyst and the serine knockout incubated with substrate reveal stable acylated species (Fig. S13), indicating that designs that hydrolyze the AEI are still susceptible to inactivation, potentially from off-mechanism acylation events in the active site or acylation-induced conformational changes. Second, mutation of the sidechain oxyanion hole contact had variable effects on activity. In three designs (**dad_t1**, **charlie_t2**, **ken_t1**) from later design rounds made with stringent ChemNet filtering, mutation of the sidechain oxyanion hole residue has minimal effect on activity, suggesting limited contribution to catalysis (Fig. S11). Analysis of the oxyanion hole geometries in these designs reveals in-plane H-bonds to the sp2 ester carbonyl of the substrate (Fig. S14, Supplementary Text), in opposition to the ones found in nature, which are perpendicular to the carbonyl plane, a geometry that selectively stabilizes the transition state and likely destabilizes the ground state (*34*, *50*, *51*). Designing oxyanion hole geometries that

explicitly favor the transition state over the ground state could significantly improve k_{cat} in future design efforts.

2.7 Acyltransferase activity of designed hydrolases

Several native serine hydrolases exhibit promiscuous acyltransferase activity, reacting with small-molecule nucleophiles that compete with hydrolysis to break down the AEI (*52*). Due to the long-lived nature of the AEI in these designed hydrolases and the hydrophobicity of their substrate binding pockets, we hypothesized they may also catalyze acyl transfer to aromatic alcohols (Fig. S15). To assess acyl transfer, we incubated designs with their cognate 4MU-ester substrates in the presence of an acyl acceptor, 2-phenylethanol (PEA). For several designs (**win**, **win31**, **win_t4**, and **dad_t1**), the addition of PEA significantly increased the rate of ester hydrolysis, suggesting these designs catalyzed acyl transfer (Fig. S15). Incubation with PEA and substrate alone or with catalytic serine to alanine knockout mutants of **win_t4** and **dad_t1** did not exhibit rate increases, suggesting observed rate enhancements are enzyme dependent (Fig S14). Acyltransferase activity was higher for designs with lower K_m: for example, **win1** (4MU-Ac K_m = 110 μM) was inhibited by PEA, while **win** (4MU-Ac K_m = 2.8 μM) had a 3.6-fold maximal rate increase upon addition of PEA, suggesting that transesterification activity may be driven by tighter binding of the acyl acceptor displacing water.

2.8 Structural determinants of catalysis

The high structural conservation of catalytic geometry in native serine hydrolases suggests that it is close to optimal for catalysis (*33*, *53*), but it is difficult to assess how activity depends on the detailed geometry of the interactions of the transition states with the catalytic serine, histidine, and oxyanion hole functional groups since while the identities of the catalytic residues can be readily changed by mutation, it is not straightforward to

systematically vary backbone geometry. In contrast, our de novo buildup approach samples a wide range of catalytic geometries. To investigate how active site geometry and preorganization influence catalytic activity, we generated ChemNet ensembles of all 812 experimentally characterized designs, categorized as inactive, FP probe labeling, acylation, and catalytic turnover, for each reaction step in the hydrolysis of 4MU-acetate (including design rounds 1-3 and previous NTF2-based designs). The following features were associated with activity.

Increased preorganization and bending of the Ser-His H-bond were associated with higher rates of probe-labeling, acylation, and turnover. All designs capable of catalyzing turnover displayed highly preorganized Ser-His H-bonds across all four states, while inactive designs often displayed rotamer shifts causing loss of the interaction (Fig. 2.5A,B). Designs that catalyzed turnover had Ser(Oγ):His(Nε-Cε) bond angles that were more acute (median, all states = 94°) than inactive designs (median, all states = 108°), which were more similar to serine-histidine hydrogen bonds across the PDB (~125°) (*34*) (Fig. 2.5C). This acute H-bond is consistent with the reaction mechanism, as this geometry allows histidine to participate, without changing conformation, in all of the necessary proton transfers involving serine, the leaving group oxygen in TI1, and the hydrolytic water (*35*, *54*). This compromise in positioning is observed not only in our active designs but also in many of those found in nature (*34*, *54*, *55*).

The geometry of the serine rotamer throughout the catalytic cycle was also strongly correlated with experimental outcome. For designs that display acylation or turnover, we found that serine largely occupies the active *g-* rotamer (*53*) in the apo state. Designs that display turnover retain the *g-* serine conformer upon formation of the AEI, but designs that irreversibly acylate switch to the *g+* rotamer in the AEI (Fig. 2.5H,I,J). The *g+* serine rotamer is catalytically incompetent in these designs because it leads to an acyl group conformation that occludes interaction of the hydrolytic water with histidine (Fig. 2.5G), increases the median

Ser-His H-bond distance (Fig. 2.5G), and reduces the frequency that the Ser-His and oxyanion hole-acyl group H-bonds form (Fig. 2.5E). The same retention of the *g-* rotamer in the AEI is observed in native crystal structures (*35*). ChemNet analysis also revealed that the presence of a second oxyanion hole residue favors the active *g-* serine rotamer: those designs with only one oxyanion hole H-bond (from the backbone amide of the serine nucleophile) shift from *g-* to *g+* upon acylation, while designs with two oxyanion hole H-bonds predominantly occupy *g-* Ser rotamers (Fig. 2.5J, right). The second oxyanion hole contact in serine hydrolases thus not only stabilizes the transition state but likely helps orient intermediates in catalytically productive conformations.

Differential preorganization may also explain activity trends in the **win**, **win1**, and **win31** series. ChemNet analysis of the crystal structures of these designs revealed that in the AEI state, the more active **win1** and **win31** sample the designed T99 oxyanion hole rotamer in 56 and 60% of predictions, respectively, while **win** never adopts this rotamer (Fig. 2.5K). Although both observed rotamers place T99 Oγ within hydrogen bonding distance of the oxyanion, the designed rotamer-oxyanion dihedral angle (91°) much more closely matches the angles observed in native serine hydrolases, suggesting it is likely more optimal for selective transition state stabilization (*34*, *50*, *51*). We also observed differences in the serine rotameric state and the preorganization of the acyl group in the AEI state. Both **win** and **win31** occupy the catalytically unfavorable *g+* rotamer across the entire AEI ensemble, while **win1** displays a less pronounced rotameric shift, which leads to shorter Ser-His H-bond distances (2.8 Å in **win1** compared to 3.1 Å in **win** and **win31**). Overall, the acyl groups of **win1** and especially **win31** display significantly less conformational heterogeneity than that of **win**, which presumably increases the likelihood of histidine-mediated water attack (Fig. 2.5K).

2.9 Conclusions

The substantial catalytic efficiencies, the complexity of the active sites, and the atomic accuracy of the designs described here represent significant advances in computational enzyme design. The serine catalytic triad plus oxyanion hole mechanism involves complex machinery that is challenging to scaffold (compared to for example the Kemp eliminase which requires only a general base in a hydrophobic environment (*2*)), necessitates chemical activation of an otherwise inert residue (serine), and proceeds through a complex multistep mechanism that traverses a chemically stable AEI. To our knowledge, the designs herein represent the first examples of de novo designed serine hydrolases, and with k_{cat}/K_m ranging up to 3900 M^{-1} s^{-1}, are the most efficient designed esterases from direct computation. The previous most efficient designed esterase by direct computation is OE1 (k_{cat}/K_m = 210 M^{-1} s^{-1}), which employed the non-canonical amino acid N_δ-methylhistidine as the nucleophile and only reached a comparable efficiency (3190 M^{-1} s^{-1}) to the most active designs described here after four rounds of directed evolution and screening over 12,000 clones (*17*). The closest comparable de novo design in terms of mechanism, in which a cysteine-based catalytic triad was mutated into a peptide-based helical barrel that proceeds via a more activated thioester intermediate (*15*), has a k_{cat}/K_m of 3.7 M^{-1} s^{-1} and k_{cat} of 0.0005 s^{-1}, 1000x and 200x less than the most efficient (**win11**) and highest turnover (**momi**) designs characterized in this study, respectively. The ability to accelerate the hydrolysis of a chemically stable acyl-enzyme intermediate has been a decades-old challenge in enzyme design. To approximate the deacylation rate enhancement achieved by our fastest design, we compare the uncatalyzed rate of hydrolysis of the unactivated ester ethyl acetate (2.5–5.0)x10^{-10} s^{-1}, (*56*)) to the lower limit of the deacylation rate constant of **momi** (k_{cat}, 0.076 s^{-1}, pH 7.0, 25°C), yielding a rate enhancement of over 10^8. Taken together, the design of serine hydrolases spanning five novel folds, significant improvement over previous designed esterases, and acceleration of deacylation represent key advances in enzyme design.

The designs described here are not as efficient as native serine hydrolases with their cognate substrates, but they have efficiencies comparable to natural proteases for activated esters (**win11**: 3900 M^{-1} s^{-1}; α-chymotrypsin with *p*-nitrophenyl acetate: 3530 M^{-1} s^{-1}) (*57*), are within the distribution of efficiencies observed in nature (*58*), and to our knowledge, represent the most efficient de novo designed enzymes by direct computation that do not utilize a chemically reactive cofactor or metal. Higher designed k_{cat} could likely be achieved through optimization of the catalytic geometry, further preorganization of the active site (*8*, *9*), and increasing active site complexity. Acetylcholinesterase, which operates near the diffusion limit, employs three backbone amide hydrogen bonds to the oxyanion and an additional network of hydrogen bonds to stabilize the catalytic aspartate (*59*, *60*). The current designs do not employ such sophisticated machinery, and comparison of catalytic triad and oxyanion hole geometries to those found in highly efficient native serine hydrolase active sites highlights differences that could be responsible for the remaining activity gap (see Supplementary Text) (*34*, *50*, *51*). Our de novo buildup approach using RFdiffusion coupled with ChemNet ensemble analysis to ensure design accuracy and preorganization should allow us to test these hypotheses by direct construction, which should further complement more traditional approaches based on structural examination and mutation of highly evolved native enzymes.

Previous efforts to design catalytic triad-based designs have failed to achieve multiple turnover; in some cases, such as our preliminary NTF2-based designs, a backbone amide oxyanion hole was impossible to achieve due to scaffold limitations, while in others based on native scaffolds, the histidine geometry was difficult to control which likely limited activation of the leaving groups and water (Fig. S16) (*7*). De novo backbone generation building outward from a specified active site with RFdiffusion overcomes these limitations by enabling generation of almost any desired catalytic geometry. We further show that the deep neural network

ChemNet can rapidly generate ensembles for a series of reaction intermediates to predict preorganization, and provide insights that would otherwise require labor-intensive structural studies. For example, ChemNet revealed pervasive off-target conformational changes in the acyl-enzyme intermediate, providing feedback on design flaws that would go unnoticed when considering only a single state in the catalytic cycle. The value of this approach is evident in the dramatic improvement in experimental success rate upon filtering with ChemNet, suggesting that such ensemble generation will be broadly useful for enzyme design moving forward. While the designs described here do utilize a known mechanism, the geometries sampled and the folds that scaffold them are distinct from those found in native proteins, and the insights provided by chemnet for these unique geometries suggests that the approach should prove valuable for assessing catalytic geometries for which no native precedent exists. We anticipate that the ability to precisely position multiple catalytic groups with sub-angstrom precision using RFdiffusion, and to assess active site organization throughout a complex reaction cycle using ChemNet should enable the design of a wide variety of new catalysts, such as PETases, amidases, and ligases, in the near future.

2.10 Figures

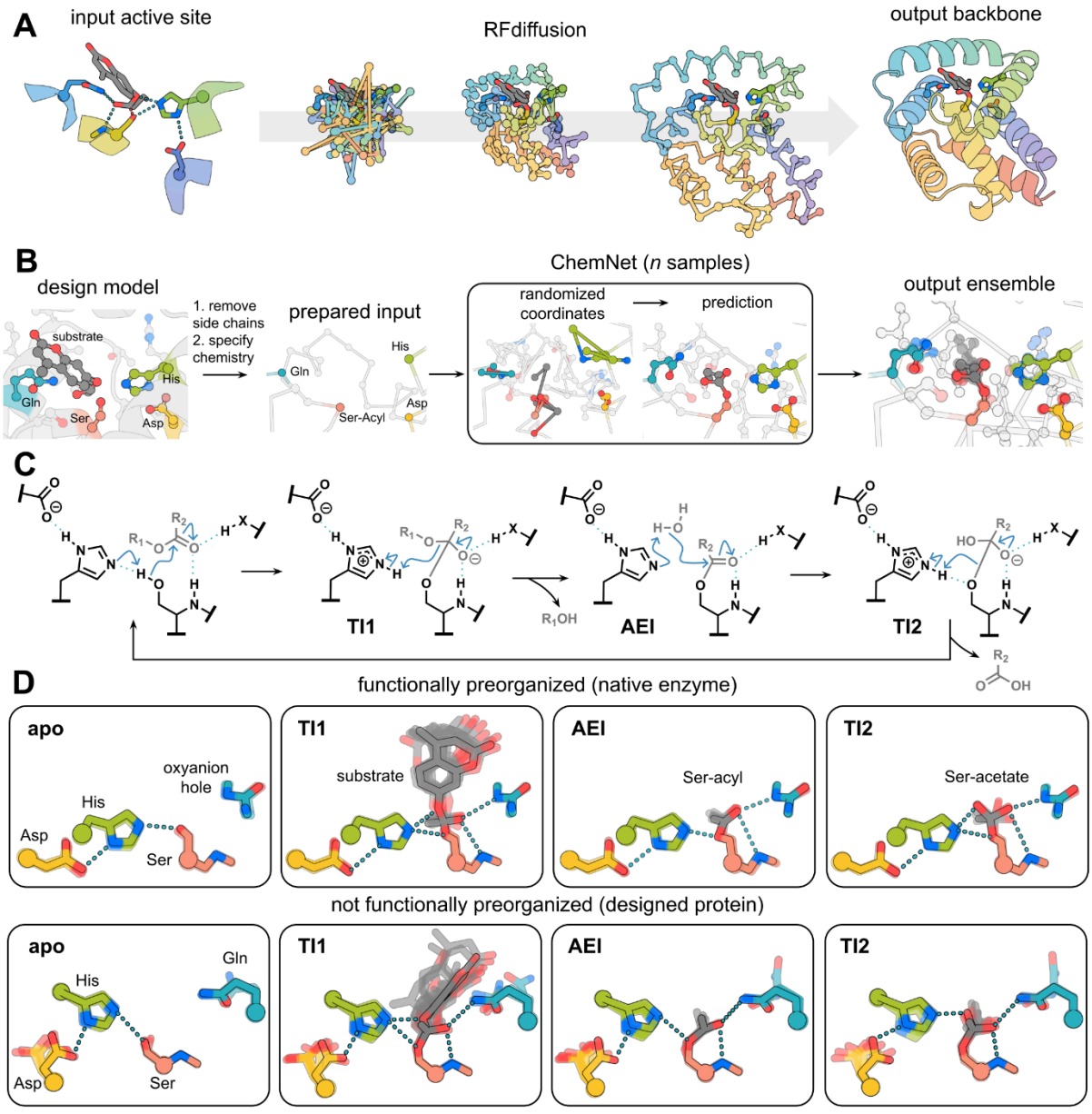

Figure 2.1. Design methods. (**A**) Active site specific backbone generation with RFdiffusion. Given the geometry of a possible active site configuration, RFdiffusion denoising trajectories generate backbone coordinates which scaffold the site. (**B**) Generation of active site ensembles with ChemNet. The coordinates of the sidechains around the active site and any bound small molecule for the step in the reaction being considered are randomized, and n samples are carried out to generate an ensemble of predictions. (**C**) Mechanism of ester hydrolysis by serine hydrolases. (**D**) Chemnet ensembles for distinct states along the reaction coordinate for hydrolysis of 4MU-Ac for a native serine hydrolase (top, PDB: 1IVY) and a designed serine hydrolase (bottom, josie).

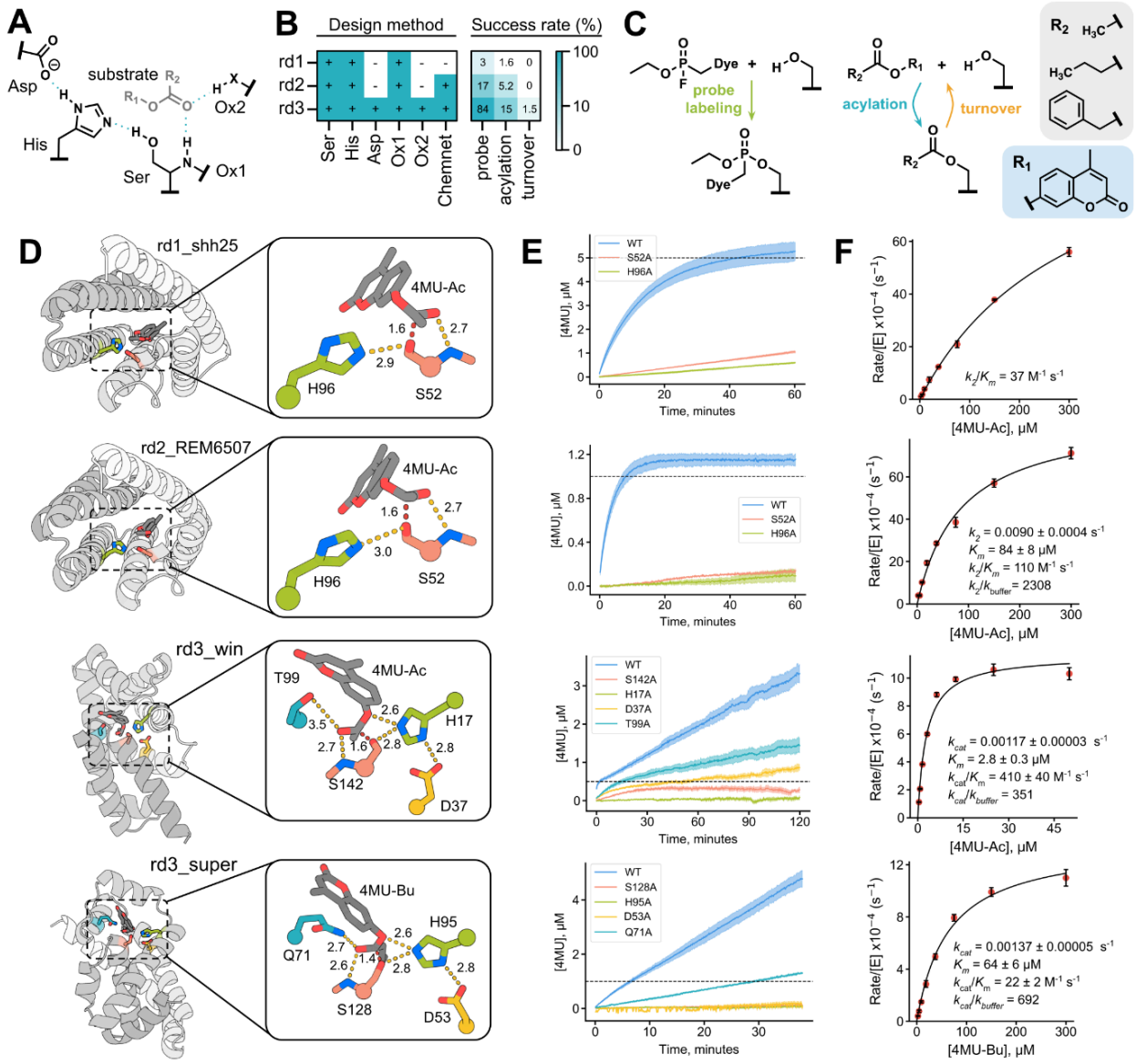

Figure 2.2. Functional characterization of designed serine hydrolases. (**A**) Chemical schematic of a serine hydrolase active site. (**B**) Summary of design method and experimental success rate for probe labeling, single turnover acylation, and catalytic turnover for each design round. (**C**) Chemical schematic depicting probe labeling, acylation, and catalytic turnover. (**D**) Fold (left) and active site (right) of serine hydrolase design models. (**E**) Reaction progress curves for the parent design and catalytic residue knockouts. Dashed line represents the enzyme concentration and shaded areas represent standard deviation of three technical replicates. (**F**) Michaelis-Menten plots derived from initial (rd1, rd2) or steady state velocities (rd3). Error bars represent standard deviation of three technical replicates.

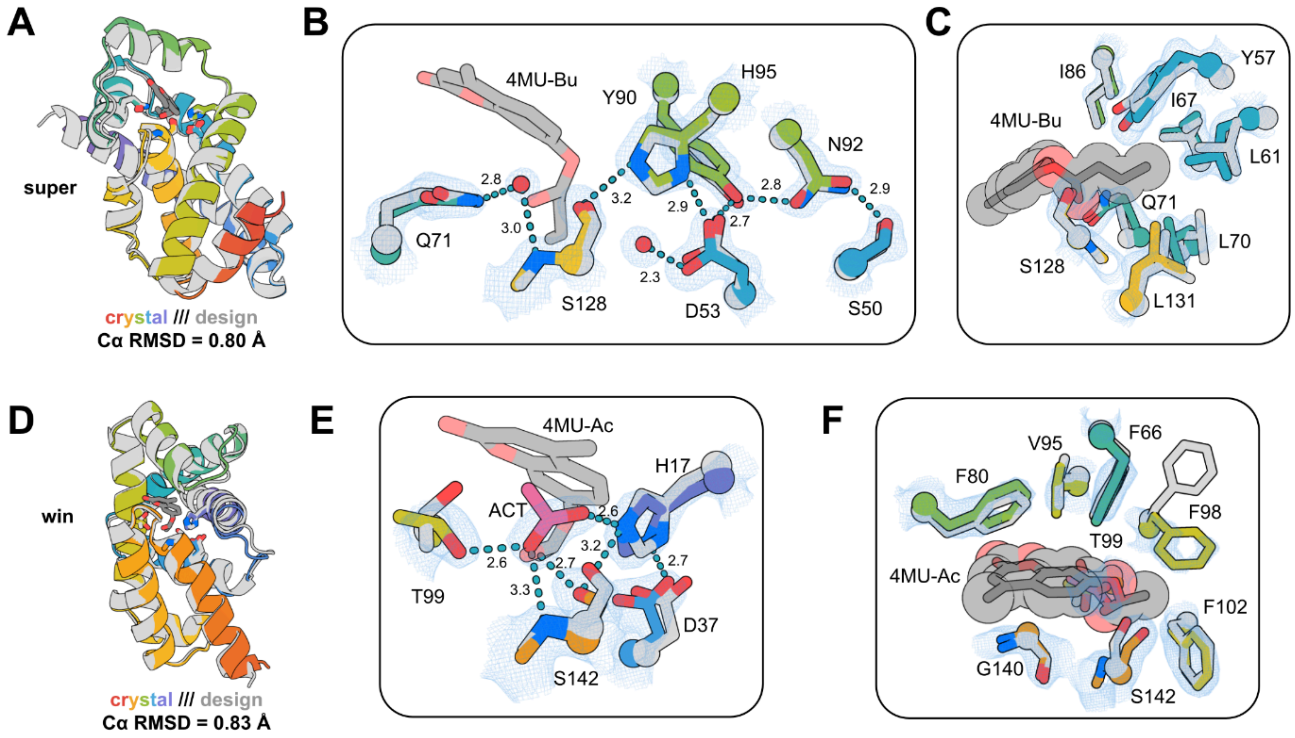

Figure 2.3. Structural characterization of designed serine hydrolases. (**A,D**) Structural superposition of design models (gray) and crystal structures (rainbow) for super (A) and win (D). (**B**, **E**) Active site overlays of design models (gray) and crystal structures (rainbow) of super (B) and win (E) with 2Fo-Fc map shown at 1σ (blue mesh). (**C**, **F**) Superposition of substrate binding sites of the design models (gray) and crystal structures (rainbow) with 2Fo-Fc map shown at 1 σ (blue mesh). Distances shown are in Å.

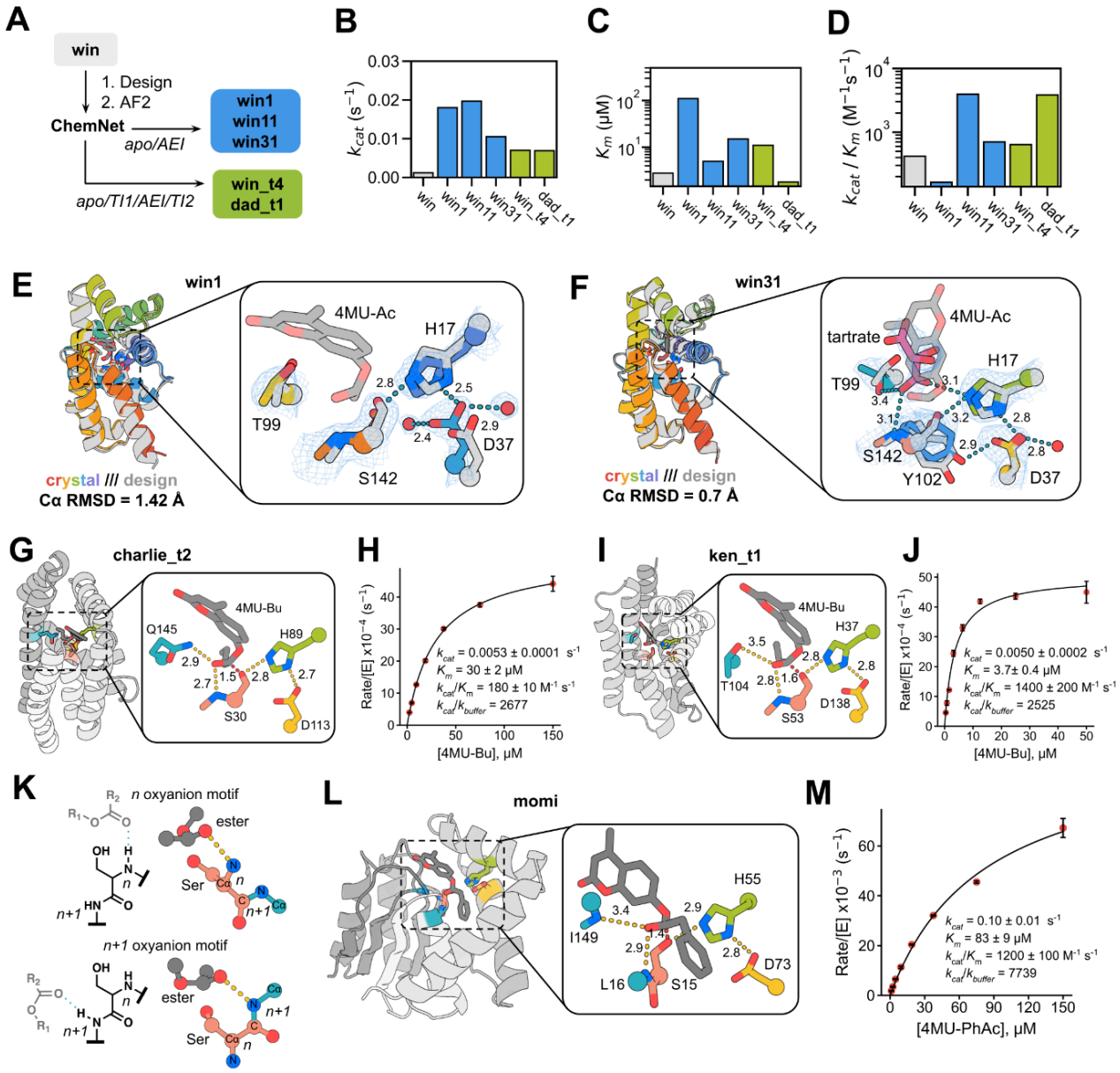

Figure 2.4. Computational redesign and more complex folds improve catalysis. (**A**) Computational pipeline for redesign of rd3_win. (**B,C,D**) k_{cat} (**B**), K_m (**C**), and k_{cat}/K_m (**D**) of parent rd3_win compared to computational redesigns. (**E,F**) Structural superposition of win1 (**E**) and win31 (**F**) design and crystal structure. (**G,H,I,J**) Design models (**G,I**) and Michaelis-Menten plots (**H,J**) for active designs with distinct folds and active site structures. (**K**) Chemical and structural comparison of *n* and *n+1* oxyanion hole motifs. (**L**) Design model of active design that utilizes two backbone amide oxyanion hole contacts, one from an n+1 backbone amide. (**M**) Michaelis-Menten plot of active design momi. Error bars represent standard deviation of three technical replicates.

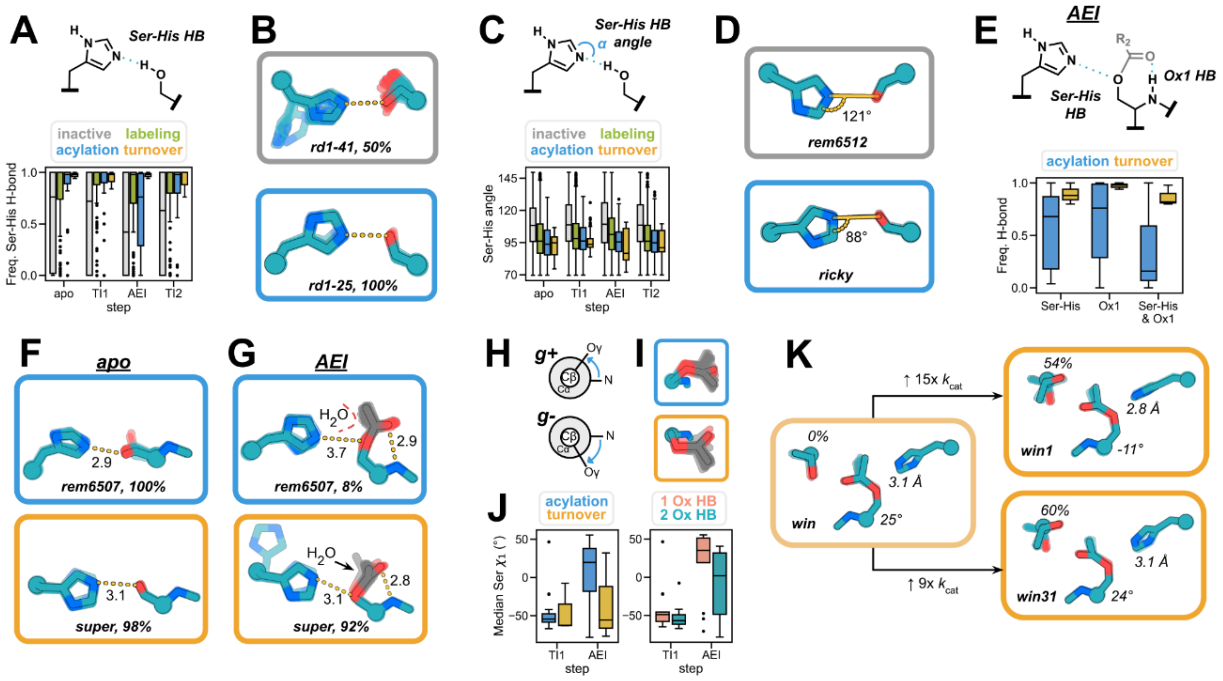

Figure 2.5. ChemNet ensembles reveal geometric determinants of catalysis. (**A**) Frequencies of catalytic Ser-His H-bond formation in ChemNet ensembles of each reaction intermediate, grouped by experimental outcome. (**B**) Apo ChemNet ensembles of representative inactive (top) and acylating (bottom) designs. (**C**) Median angle (α) between serine Oγ, histidine Nϵ and Cϵ across Chemnet ensembles of inactive and acylating designs. (**D**) Apo ChemNet ensembles of representative inactive (top) and acylating (bottom) designs, angle indicates median α. (**E**) AEI ChemNet ensemble H-bond frequencies for designs that undergo acylation or full turnover. (**F**) ChemNet ensembles of the apo state for an acylating (top) and multiple turnover design (bottom). (**G**) ChemNet ensembles of the AEI state for a representative design that undergoes acylation (top) and a design that catalyzes turnover (bottom). Measurements shown represent median distances (Å) of key H-bonds indicated for each ensemble and percentages represent frequency of H-bond formation across all ChemNet trajectories. (**H**) Newman projections of serine *g*+ and *g*- rotameric states (left). (**I**) Chemnet ensembles of an acylating design (top) and a design that catalyzes turnover (bottom). (**J**) Median serine χ_1 angle across TI1 and AEI state Chemnet ensembles for designs that catalyze acylation or turnover (left). Median serine χ_1 angle across TI1 and AEI state Chemnet ensembles for the same designs grouped by number of oxyanion hole hydrogen bonds. (**K**) AEI state Chemnet ensembles for win, win1, and win31, with percent of frames with correct oxyanion hole rotamer, Ser χ_1 angle, and catalytic Ser-His H-hbond distance shown. Boxplots represent median, upper and lower quartiles; whiskers extend 1.5×IQR above and below the upper and lower quartiles (respectively). Observations falling outside these ranges plotted as outliers.

2.11 Methods

Motif generation

Motifs were built in an iterative process. First, a substrate rotamer in a transition state geometry (either 4MU-Bu or 4MU-Ac) was placed in accordance with geometries in ref 7 in relation to a 3-residue stub of the serine and local oxyanion hole from one of two natural serine hydrolase crystal structures (1scn, residues 220-222, and 1lns, residues 347-349, in which all residues other than the serine were mutated to alanine). The transition state geometry of the substrate ester group was determined by DFT geometry optimization (B3LYP-D3(BJ)/6-31G(d)). Next, positions and rotamers of histidine on 3-residue helical or strand stubs flanked by alanine were sampled around the catalytic serine and filtered for those structures in which the histidine simultaneously formed hydrogen bonds with the catalytic serine and the substrate leaving group oxygen. This process resulted in 108 unique round 1 motifs. For the round 3 motifs, initially the aspartate or glutamate residue and second oxyanion hole hydrogen bond were added in a similar manner using geometric sampling of hydrogen-bonding conformations and rotamers. However, backbones produced from these motifs had exceedingly low AF2 success rates, presumably due to the generation of incompatible combinations of backbone conformations. To ensure that the remaining catalytic residue stubs were placed in realizable geometries, we generated 10,000 backbones with RFdiffusion using the simple substrate-Ser-His motifs as input, and then searched these backbones using Rosetta for positions on secondary structure that could accommodate the aspartate or glutamate triad residue to hydrogen bond to histidine. These stubs were then extracted, and in a final step, the same process was repeated to generate stubs for the second oxyanion hole, considering all hydrogen bond donating sidechains, ultimately producing 2238 unique round 3 motifs with Ser-His-Asp/Glu catalytic triads, and Ser/Thr/Tyr/His/Trp oxyanion holes.

Backbone generation

For a detailed description of CA diffusion which was employed to generate backbones to scaffold motifs, see the preprint cited at the beginning of this chapter.

Sequence design

We performed three cycles of LigandMPNN (*46*) and Rosetta FastRelax (*61*) to design sequences for backbones generated from RFdiffusion. To encourage formation of hydrogen bond contacts to the catalytic histidine (for round 1 motifs) and to the catalytic aspartate/glutamate (round 3 motifs), the log probabilities used by LigandMPNN to select residues were biased toward polar amino acids for all residues with Cα within 8 Å of the active site. Catalytic residues were kept fixed and Rosetta enzyme constraints (*62, 63*) were applied during the relax steps to maintain the catalytic geometry during cycles of design. Constraints were defined for each hydrogen bonding interaction using the starting motif geometry with tolerances of 0.1 Å for distances and 5° for angles and dihedrals.

Filtering

After sequence design, designs were filtered on the recapitulation of the motif catalytic geometry after FastRelax and the shape complementarity of the binding site to the substrate. Sequences of 3 passing designs were used as input to AF2 (*49*) for single sequence structure prediction. AF2 was run using model 4 with three recycles. Designs were filtered for a global Cα RMSD < 1.5 Å, pLDDT > 75, and catalytic residue Cα RMSD < 1.0 Å. In the case of final round N+1 oxyanion hole designs, a modified version of Initial Guess AF2 was used to predict designs with sparse template information provided. Designs that passed AF2 filters were subsequently analyzed using ChemNet. ChemNet is a denoising neural network which was trained on high- and medium-resolution X-ray and EM structures from the PDB to recapitulate the correct atom positions from partially

corrupted input structures provided that all the chemical information about the system being modeled is known from the start. ChemNet predictions were done for a spatial crop of 600 atoms closest to the active site. The inputs to the network included the protein backbone coordinates within the crop and the amino acid sequence with side chain coordinates randomly initialized around the respective C-alpha atoms. For proteins without a crystal structure, the AF2 model was used. For every designed protein, we modeled 5 reaction states differing in the chemical modifications the catalytic serine undergoes in the course of the reaction: 1) apo, 2) substrate bound, 3) tetrahedral intermediate 1, 4) acylenzyme intermediate, and 5) tetrahedral intermediate 2. We used 50 different seeds to generate an ensemble of 50 ChemNet models for each reaction state (apo, substrate bound, TI1, AEI, and TI2). These ensembles were then individually analyzed for the preservation of hydrogen bonding patterns in the active site. For each of the 50 predictions in each ensemble, geometries of each hydrogen bonding interaction in that step (see Methods) were measured. To analyze native hydrolases with Chemnet, a set of native crystal structures was collected (*34*) (PDB IDs: 1ACB_E, 1C5L_H, 1H2W_A, 1IC6_A, 1IVY_A, 1PFQ_A, 1QNJ_A, 1QTR_A, 1ST2_A, 2H5C_A, 2QAA_A, 3MI4_A, 5JXG_A), the active site locations identified, and the abovedescribed process was applied.

In-gel fluorescence screening with activity-based probes

DNA encoding the designed proteins was ordered from IDT as eblocks and cloned into vector LM627 (addgene), which contains a C-terminal SNAC and hexahistidine tag. Resulting plasmid was transformed into BL21(DE3) cells and grown overnight in 1 mL of LB supplemented with 50 μg/ml kanamycin. For expression, 100 μL of overnight culture was used to inoculate 1 mL of LB media and grown for 1.5 hours at 37°C on a Heidolph shaker and then 10 μL of 100 mM IPTG was added and cultures were shaken at 37°C for an additional 3 hours. Cultures were centrifuged at 4000g for 10 minutes and supernatant removed. Cell pellets were resuspended in 200 μL 20 mM HEPES (pH 7.4), containing 50 mM NaCl, 0.1 mg/mL lysozyme, and 0.01

mg/mL DNaseI. After 15 minutes, lysates were frozen in liquid nitrogen and subsequently thawed. 10 μL of lysate was incubated with 1 μM FP-TAMRA probe (10 μL of 2 μM stock in lysis buffer) for 1 hour at room temperature before quenching using 2x Laemmli sample buffer. Labeled samples were heated at 95°C for 5 minutes and 10 μL of each sample was separated on a BioRad AnykD Criterion precast gel and in-gel fluorescence was visualized using a LI-COR Odyssey M imager. Gels were subsequently stained with coomassie blue to visualize the molecular weights and levels of expression of each design.

Lysate screening

DNA encoding the designed proteins was ordered from IDT as eblocks and cloned into vector pCOOL1 which contains a C-terminal mScarlet-i3 fusion and His tag to enable normalization of 4 activity in lysate by mScarlet-i3 fluorescence. Cultures were grown overnight at 1 mL scale in 96-well plates on a Heidolph shaker at 1300 rpm and 37 °C. For expression, 50 μL of the overnight cultures were used to inoculate 1 mL of autoinduction media in 96-well round bottom plates and incubated at 1300 rpm and 37 °C for approximately 24 hours. Cultures were centrifuged at 4000g for 10 minutes and supernatant decanted, followed by a wash with buffer (20 mM HEPES, 50 mM NaCl, pH 7.4) and incubation on a Heidolph shaker at 1300 rpm at room temp for 5 minutes to resuspend. Plates were centrifuged again at 4000g for 10 minutes and supernatant decanted. For lysis, cell pellets were resuspended with 500 μL of lysis buffer (20 mM HEPES, 50 mM NaCl, 0.01 mg/mL DNAseI, 0.01 mg/mL lysozyme, 1 mM EDTA, 0.1% triton X-100) and incubated for 2 hours on a Heidolph shaker at 1300 rpm and 37 °C. Plates were centrifuged at 4300g for 30 minutes and supernatant collected for screening. For activity screening, 4 or 6 μL of lysate was aliquoted into microtiter plates and reactions initiated by addition of 36 or 54 μL of buffer containing 111.1 μM 4MU-Ac or 4MU-Bu, 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO.

Protein expression and purification

Genes encoding the designed proteins were ordered from IDT as eblocks and cloned via the Golden Gate method into vector LM627 as previously described (*64*). Resulting plasmid was transformed into BL21(DE3) cells and grown overnight in 1 mL of LB supplemented with 50 μg/ml kanamycin, after which 500 μL of overnight was used to inoculate 50 mL of autoinduction media (*65*), which was grown 4-6 hours at 37 °C and then overnight at 18 °C. Cultures were spun down at 4000g for 15 minutes, and supernatant decanted. Cell pellets were resuspended in 25 mL of cold wash buffer (40 mM imidazole, 500 mM NaCl, 50 mM sodium phosphate, pH 7.4) with 1 mg/mL lysozyme and 0.1 mg/mL DNAse I. Cell slurries were sonicated on ice for 2.5 minutes at 80% amplitude, 10s on 10s off. The resulting lysate was centrifuged at 14000g for 30 minutes and the supernatant was applied to 1 mL of Ni-NTA resin equilibrated with wash buffer. The resin was subsequently washed with 15 mL of wash buffer 3 times and once with 400 μL of elution buffer (400 mM imidazole, 500 mM NaCl, 50 mM sodium phosphate, pH 7.4) followed by elution with 1.3 mL elution buffer. The eluate was purified by size-exclusion chromatography on a Superdex 75 Increase 10/300 GL with running buffer of 20 mM HEPES, 50 mM NaCl, pH 7.4. Samples were either used immediately in downstream experiments or snap frozen in liquid nitrogen and stored at -80 C. Protein molecular weight was confirmed by LCMS.

Kinetic analysis

To characterize hits identified from in-gel fluorescence and lysate screens for catalytic turnover, we incubated purified protein samples with fluorogenic substrates 4MU-Ac, 4MU-Bu and 4MUPhAc. Kinetic screens were either performed in 40 μL reaction volumes in 96-well half area plates or 60 μL reaction volume in 96-well full-area plates. Protein and substrate were prepared in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO. Either 4 or 6 μL of enzyme was added to microtiter plates and the reactions were initiated by addition of substrate (36 or 54 μL). Generation of the fluorogenic product 4MU was monitored continuously (excitation

365 nm, emission 445 nm). Analysis of the resulting data were carried out using custom scripts (see computational methods). In cases where single-turnover activity was observed, initial velocities were used to determine k_1/K_m. For those designs that displayed a clear burst phase followed by a 5 slower steady-state rate, straight-line fits of the steady-state velocities were used to determine Michaelis-Menten catalytic parameters. To determine the uncatalyzed reaction rate in assay buffer (20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO), substrate was diluted in buffer alone and rates determined at multiple substrate concentrations, after which the rate was determined from fitting *[S]* versus rate with an equation of the form *rate = kbuffer[S]*.

Crystallography

Proteins for crystallography were prepared as described above, but SEC was done with SNAC tag cleavage buffer (*66*). After SEC, protein eluate was incubated with 500 mM guanidinium hydrochloride and 2 mM NiCl2 overnight at room temperature to remove the C-terminal His tag. The SNAC cleavage reaction was applied to a nickel column equilibrated with wash buffer to remove any uncleaved product and resulting eluate applied to a Superdex 75 Increase 10/300 GL column with 20 mM HEPES, 50 mM NaCl, pH 7.4 as the running buffer. Samples were concentrated and stored at -80° C or immediately used for crystallization. Crystallization screening was performed using a Mosquito LCP by STP Labtech and resulting crystals were harvested directly from the screening plate. Crystallization conditions for each design were as follows: slap215.8 (15 mg/mL) in 0.1 M Bis-Tris pH 5.5, 25% (w/v) PEG 3350, super (50 mg/mL) in 0.2 M Potassium fluoride, 20% (w/v) PEG 3350, win (42 mg/mL) in 0.1 M Sodium acetate pH 4.6, 8% (w/v) PEG 4000, win1 (54 mg/mL) in 60% v/v Tacsimate pH 7.0, and win31 (60 mg/mL) in 0.2 M di-Ammonium tartrate and 20% (w/v) PEG 3350. Data were processed with XDS(*67*), phased and refined with Phenix(*68*), and model building performed with COOT(69). Percent Ramachandran favored, allowed, and outliers for each structure are as follows: slap215.8 (98.21, 1.79, 0.00), super (99.37, 0.63, 0.00), win (97.99, 2.01, 0.00), win1 (99.68, 0.32, 0.00), and

win31 (99.36, 0.64, 0.00). Coordinates are deposited in the PDB with PDB IDs of 9DED (slap215.8), 9DEE (super), 9DEF (win), 9DEG (win1), and 9DEH (win31).

Mass spectrometry

Intact mass spectra of protein samples were obtained by reverse-phase LC/MS on an Agilent G6230B TOF after desalting using an AdvanceBio RP-Desalting column. Deconvolution using a total entropy algorithm was performed using Bioconfirm. In some cases, protein samples (1 mg/mL) were incubated overnight with substrate (300 μM) in SEC running buffer at room temperature prior to mass spectrometry analysis.

Acyltransferase activity screening

Enzymes (1 μM) were incubated with 100 μM cognate substrate in assay buffer in the presence of varying concentrations of acyl acceptor, PEA (50, 25, 12.5, 6.3, 3.1, 1.6, 0.8, 0 mM), and substrate hydrolysis were monitored for 1 hour as described above. Initial velocities were obtained by fitting the beginning of each progress curve and divided by the hydrolysis rate in the absence of PEA to obtain relative rates of hydrolysis.

Structural similarity search of the PDB and AFDB

To assess the structural novelty of our designed enzymes, we used TMalign (*70*) to compare our crystal structures against the Protein DataBank (PDB) and AlphaFold database (*71*). We downloaded all protein polymers from the PDB solved by X-ray crystallography or Cryo-EM on April 4, 2024 and extracted all protein chains from each entry. Models of AFDB50 (*72*) (version 4) proteins were fetched April, 2024. We report the average TM-score for the top hit.

2.12 Supplementary Text

Motif generation via conformational sampling

To generate constellations of histidine rotamers around the serine backbone motif, distances, angles, and torsions were first specified in enzyme constraint file format along with sampling ranges and frequencies for the serine-substrate and serine-histidine interactions. These were input to a script which 1) sampled probable histidine rotamers from the Dunbrack library at the specified geometries with respect to the serine nucleophile and substrate, 2) built out protein stubs in helical or strand conformation flanking the histidine using Rosetta, 3) and filtered for clashes between the substrate, serine stub, and histidine stub before outputting these conformations as PDB files. The code and a detailed description of this script can be found here: https://github.com/ikalvet/invrotzyme.

Evaluation of hydrogen bonds in ChemNet predictions

Hydrogen-bonding interactions in ChemNet predictions were evaluated by the measuring of distances, angles, and torsions between the acceptor and donor heavy atoms. Upper and lower bound cutoffs for each parameter (distance, angle, dihedral) were used to determine if the interaction constituted a hydrogen bond, depending on the hybridization of the participating heavy atoms.

Active site composition and geometric features

To identify potential strategies for improving k_{cat} in our designed hydrolases, we performed a comparative analysis of the active sites of our designs and natural hydrolases, identifying a number of deviating features in the catalytic triad, oxyanion hole, and surroundings.

Differences in active site makeup could contribute to decreased k_{cat} values in our designs. In nature, the

catalytic aspartate or glutamate is supported by a network of additional hydrogen bonds from surrounding sidechains and backbone amides (*60*), a feature rarely seen in our designs, which have at most one sidechain hydrogen bond to the catalytic aspartate. This highly polar environment helps aspartate remain in a charged state to stabilize the catalytic histidine, and clearly plays an important role in our designs as well, as evidenced by the significant effect of knockout mutants of these second shell contacts on activity (Fig. S11). Although some notable exceptions exist, such as the serine protease subtilisin, in most natural serine hydrolases, at least 7 two backbone amide groups comprise the oxyanion hole, while all but one of the active designs herein contain one backbone amide and one sidechain hydrogen bond. Sidechains are expected to be significantly more challenging to preorganize than backbone amides due to additional rotatable bonds, and crystal structures of our designs often show deviations from the designed oxyanion hole sidechain conformations. Indeed, the design with the highest k_{cat}, **momi**, is the only active design with two backbone amide hydrogen bonds.

Geometric deviations within the catalytic triad and oxyanion hole could also account for the differences in k_{cat} between our designs and native serine hydrolases. In nature, the catalytic aspartate and histidine form a particularly ideal hydrogen bond, characterized by short distances and near linear bond angles (*33*), while crystal structures of our designs display longer distances and deviations from ideal angles. Alignment of the active sites of natural enzymes and our active designs also reveals a difference in the angle of approach of the histidine to the serine in our designs (Fig. S14). This shift may affect the role of the histidine in shuffling protons between the serine, substrate and hydrolytic water. Finally, the oxyanion hole hydrogen bonds of natural hydrolases are positioned to make out-of-plane interactions with the ester carbonyl of the substrate, which has been hypothesized to lead to selective stabilization of the oxyanion-containing transition states and intermediates over the ground state (*34, 50, 51*). Our designs exhibit hydrogen bonds that are closer to in-plane hydrogen bonds that may over-stabilize the sp^2 ground state of the substrate carbonyl.

2.13 References

1. S. L. Lovelock, R. Crawshaw, S. Basler, C. Levy, D. Baker, D. Hilvert, A. P. Green, The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).

2. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).

3. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas 3rd, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).

4. J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309–313 (2010).

5. H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3790–3795 (2012).

6. F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).

7. S. Rajagopalan, C. Wang, K. Yu, A. Kuzin, F. Richter, S. Lew, A. E. Miklos, M. L. Matthews, J. Seetharaman, M. Su, J. F. Hunt, B. F. Cravatt, D. Baker, Design of activated serine-containing catalytic triads with atomic level accuracy. *Nat. Chem. Biol.* **10**, 386–391 (2014).

8. R. Otten, R. A. P. Pádua, H. A. Bunzel, V. Nguyen, W. Pitsawong, M. Patterson, S. Sui, S. L. Perry, A. E. Cohen, D. Hilvert, D. Kern, How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).

9. R. V. Rakotoharisoa, B. Seifinoferest, N. Zarifi, J. D. M. Miller, J. M. Rodriguez, M. C. Thompson, R. A. Chica, Design of efficient artificial enzymes using crystallographically enhanced conformational sampling. *J. Am. Chem. Soc.* **146**, 10001–10013 (2024).

10. A. Broom, R. V. Rakotoharisoa, M. C. Thompson, N. Zarifi, E. Nguyen, N. Mukhametzhanov, L. Liu, J. S. Fraser, R. A. Chica, Ensemble-based enzyme design can recapitulate the effects of laboratory

directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).

11. G. Kiss, D. Röthlisberger, D. Baker, K. N. Houk, Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).

12. S. J. Fleishman, S. D. Khare, N. Koga, D. Baker, Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20**, 753–757 (2011).

13. H. A. Bunzel, J. L. R. Anderson, D. Hilvert, V. L. Arcus, M. W. van der Kamp, A. J. Mulholland, Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat. Chem.* **13**, 1017–1022 (2021).

14. M. P. Frushicheva, J. Cao, Z. T. Chu, A. Warshel, Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16869–16874 (2010).

15. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).

16. D. N. Bolon, S. L. Mayo, Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14274–14279 (2001).

17. A. J. Burke, S. L. Lovelock, A. Frese, R. Crawshaw, M. Ortmayer, M. Dunstan, C. Levy, A. P. Green, Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219–223 (2019).

18. S. Studer, D. A. Hansen, Z. L. Pianowski, P. R. E. Mittl, A. Debon, S. L. Guffy, B. S. Der, B. Kuhlman, D. Hilvert, Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).

19. B. S. Der, D. R. Edwards, B. Kuhlman, Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940 (2012).

20. Y. S. Moroz, T. T. Dunston, O. V. Makhlynets, O. V. Moroz, Y. Wu, J. H. Yoon, A. B. Olsen, J. M. McLaughlin, K. L. Mack, P. M. Gosavi, N. A. J. van Nuland, I. V. Korendovych, New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **137**, 14905–14911 (2015).

21. V. Tournier, C. M. Topham, A. Gilles, B. David, C. Folgoas, E. Moya-Leclair, E. Kamionka, M.-L. Desrousseaux, H. Texier, S. Gavalda, M. Cot, E. Guémard, M. Dalibey, J. Nomme, G. Cioci, S. Barbe, M. Chateau, I. André, S. Duquesne, A. Marty, An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* **580**, 216–219 (2020).

22. S. Yoshida, K. Hiraga, T. Takehana, I. Taniguchi, H. Yamaji, Y. Maeda, K. Toyohara, K. Miyamoto, Y. Kimura, K. Oda, A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**, 1196–1199 (2016).

23. E. L. Bell, R. Smithson, S. Kilbride, J. Foster, F. J. Hardy, S. Ramachandran, A. A. Tedstone, S. J. Haigh, A. A. Garforth, P. J. R. Day, C. Levy, M. P. Shaver, A. P. Green, Directed evolution of an efficient and thermostable PET depolymerase. *Nature Catalysis* **5**, 673–681 (2022).

24. D. M. Blow, Structure and mechanism of chymotrypsin. *Acc. Chem. Res.* **9**, 145–152 (1976).

25. P. Carter, J. A. Wells, Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564–568 (1988).

26. P. Carter, J. A. Wells, Functional interaction among catalytic residues in subtilisin BPN'. *Proteins* **7**, 335–342 (1990).

27. L. Polgár, The catalytic triad of serine peptidases. *Cell. Mol. Life Sci.* **62**, 2161–2172 (2005).

28. P. Bryan, M. W. Pantoliano, S. G. Quill, H. Y. Hsiao, T. Poulos, Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3743–3745 (1986).

29. D. R. Corey, C. S. Craik, An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.* **114**, 1784–1790 (1992).

30. B. Zerner, R. P. M. Bond, M. L. Bender, 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).

31. J. Kraut, Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* **46**, 331–358 (1977).

32. L. Hedstrom, Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4524 (2002).

33. A. J. T. Smith, R. Müller, M. D. Toscano, P. Kast, H. W. Hellinga, D. Hilvert, K. N. Houk, Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J. Am. Chem. Soc.* **130**, 15361–15373 (2008).

34. S. Du, R. C. Kretsch, J. Parres-Gold, E. Pieri, V. W. D. Cruzeiro, M. Zhu, M. M. Pinney, F. Yabukarski, J. P. Schwans, T. J. Martínez, D. Herschlag, Conformational Ensembles Reveal the Origins of Serine Protease Catalysis, *bioRxiv* (2024)p. 2024.02.28.582624.

35. E. S. Radisky, J. M. Lee, C.-J. K. Lu, D. E. Koshland Jr, Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U. S. A.*

103, 6835–6840 (2006).

36. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).

37. F. Praetorius, P. J. Y. Leung, M. H. Tessmer, A. Broerman, C. Demakis, A. F. Dishman, A. Pillai, A. Idris, D. Juergens, J. Dauparas, X. Li, P. M. Levine, M. Lamb, R. K. Ballard, S. R. Gerben, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, B. F. Volkman, J. Nivala, S. Stoll, D. Baker, Design of stimulus-responsive two-state hinge proteins. *Science* **381**, 754–760 (2023).

38. M. L. Zastrow, A. F. A. Peacock, J. A. Stuckey, V. L. Pecoraro, Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118–123 (2011).

39. A. H.-W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk, D. Baker, De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).

40. R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. DiMaio, D. Baker, Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).

43. P. A. Frey, A. D. Hegeman, *Enzymatic Reaction Mechanisms* (Oxford University Press, 2007).

44. C. Walsh, *Enzymatic Reaction Mechanisms* (W. H. Freeman, 1979).

45. W. P. Jencks, *Catalysis in Chemistry and Enzymology* (Courier Corporation, 1987).

46. J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock, D. Baker, Atomic context-conditioned protein sequence design using LigandMPNN, *bioRxiv* (2023)p. 2023.12.22.573103.

47. R. Das, D. Baker, Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).

48. N. R. Bennett, B. Coventry, I. Goreshnik, B. Huang, A. Allen, D. Vafeados, Y. P. Peng, J. Dauparas, M. Baek, L. Stewart, F. DiMaio, S. De Munck, S. N. Savvides, D. Baker, Improving de novo protein binder design with deep learning. *Nat. Commun.* **14**, 2625 (2023).

49. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B.

Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

50. L. Simón, J. M. Goodman, Hydrogen-bond stabilization in oxyanion holes: grand jeté to three dimensions. *Org. Biomol. Chem.* **10**, 1905–1913 (2012).

51. L. Simón, J. M. Goodman, Enzyme catalysis by hydrogen bonds: the balance between transition state binding and substrate binding in oxyanion holes. *J. Org. Chem.* **75**, 1831–1840 (2010).

52. H. Müller, A.-K. Becker, G. J. Palm, L. Berndt, C. P. S. Badenhorst, S. P. Godehard, L. Reisky, M. Lammers, U. T. Bornscheuer, Sequence-based prediction of promiscuous acyltransferase activity in hydrolases. *Angew. Chem. Weinheim Bergstr. Ger.* **132**, 11704–11709 (2020).

53. A. R. Buller, C. A. Townsend, Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E653–61 (2013).

54. E. Zakharova, M. P. Horvath, D. P. Goldenberg, Structure of a serine protease poised to resynthesize a peptide bond. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11034–11039 (2009).

55. G. Dodson, A. Wlodawer, Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347–352 (1998).

56. R. Wolfenden, Y. Yuan, The "neutral" hydrolysis of simple carboxylic esters in water and the rate enhancements produced by acetylcholinesterase and other carboxylic acid esterases. *J. Am. Chem. Soc.* **133**, 13821–13823 (2011).

57. F. J. Kezdy, M. L. Bender, The kinetics of the α-chymotrypsin-catalyzed hydrolysis of p-nitrophenyl acetate*. *Biochemistry* **1**, 1097–1106 (1962).

58. A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, R. Milo, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).

59. J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, I. Silman, Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. *Science* **253**, 872–879 (1991).

60. A. Warshel, S. Russell, Theoretical correlation of structure and energetics in the catalytic reaction of trypsin. *J. Am. Chem. Soc.* **108**, 6569–6579 (1986).

61. M. D. Tyka, D. A. Keedy, I. André, F. Dimaio, Y. Song, D. C. Richardson, J. S. Richardson, D. Baker, Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**,

607–618 (2011).

62. F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, De novo enzyme design using Rosetta3. *PLoS One* **6**, e19230 (2011).

63. A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).

64. B. I. M. Wicky, L. F. Milles, A. Courbet, R. J. Ragotte, J. Dauparas, E. Kinfu, S. Tipps, R. D. Kibler, M. Baek, F. DiMaio, X. Li, L. Carter, A. Kang, H. Nguyen, A. K. Bera, D. Baker, Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).

65. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

66. B. Dang, M. Mravic, H. Hu, N. Schmidt, B. Mensa, W. F. DeGrado, SNAC-tag for sequence-specific chemical protein cleavage. *Nat. Methods* **16**, 319–322 (2019).

67. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

68. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).

69. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

70. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

71. M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Žídek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).

72. M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).

2.14 Supplementary figures

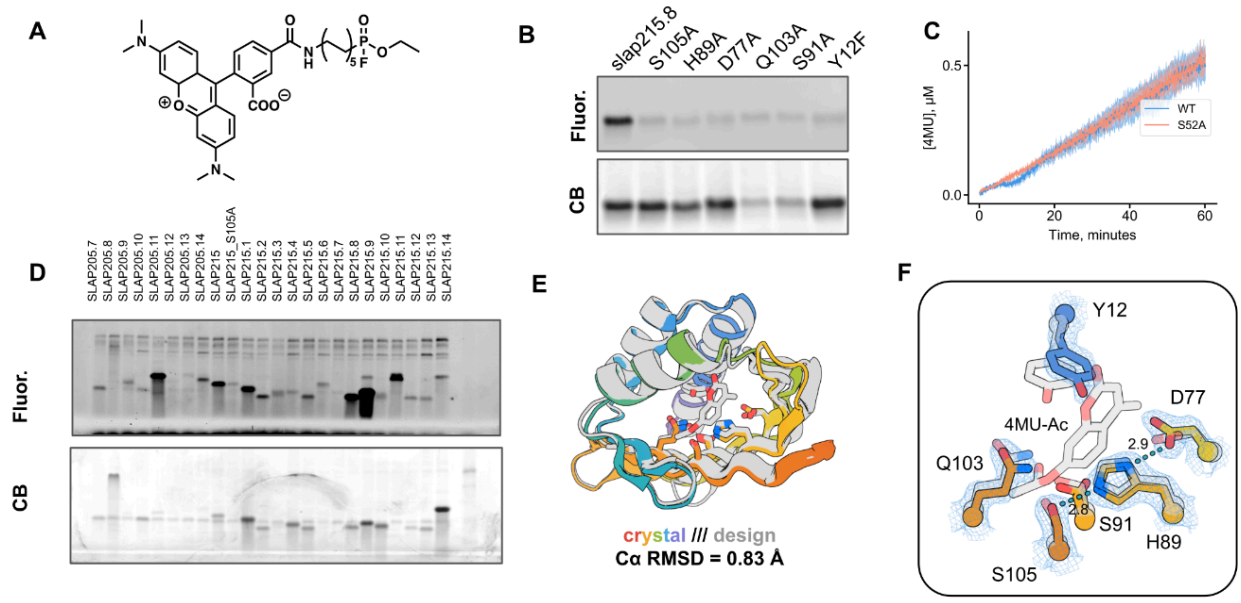

Figure S1. Serine hydrolase design and characterization with hallucinated NTF2s. (**A**) Chemical structure of TAMRA-conjugated fluorophosphonate probe (FP-probe). (**B**) In-gel fluorescence of catalytic residue alanine knockouts in probe-reactive design SLAP215.8. (**C**) Reaction progress curves of SLAP215.8 incubated with 4MU ester substrates. Shaded area represents standard deviation of three technical replicates. (**D**) In-gel fluorescence of NTF2-based serine hydrolase design cell lysates after 1 hour incubation with 1 µM FP-TAMRA. (**E,F**) Structural superposition of slap215.8 crystal structure and design model.

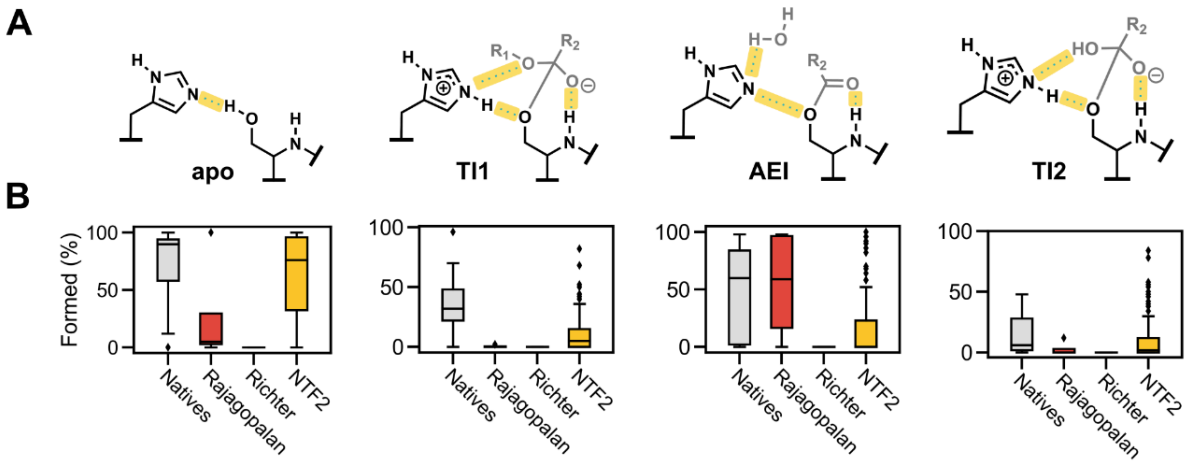

Figure S2. Comparing preorganization of native and designed serine hydrolases with ChemNet. (**A**) Reaction states modeled with ChemNet. Key hydrogen bonding interactions are highlighted in yellow. (**B**) Percent ChemNet ensemble frames in which all of the key hydrogen bonding interactions are simultaneously formed for each step (apo, TI1, AEI, TI2 from left to right). Designs labeled "Richter" and "Rajagopalan" were reported in references 7 and 8, respectively. Boxplots represent median, upper and lower quartiles; whiskers extend 1.5×IQR above and below the upper and lower quartiles (respectively). Observations falling outside these ranges plotted as outliers.

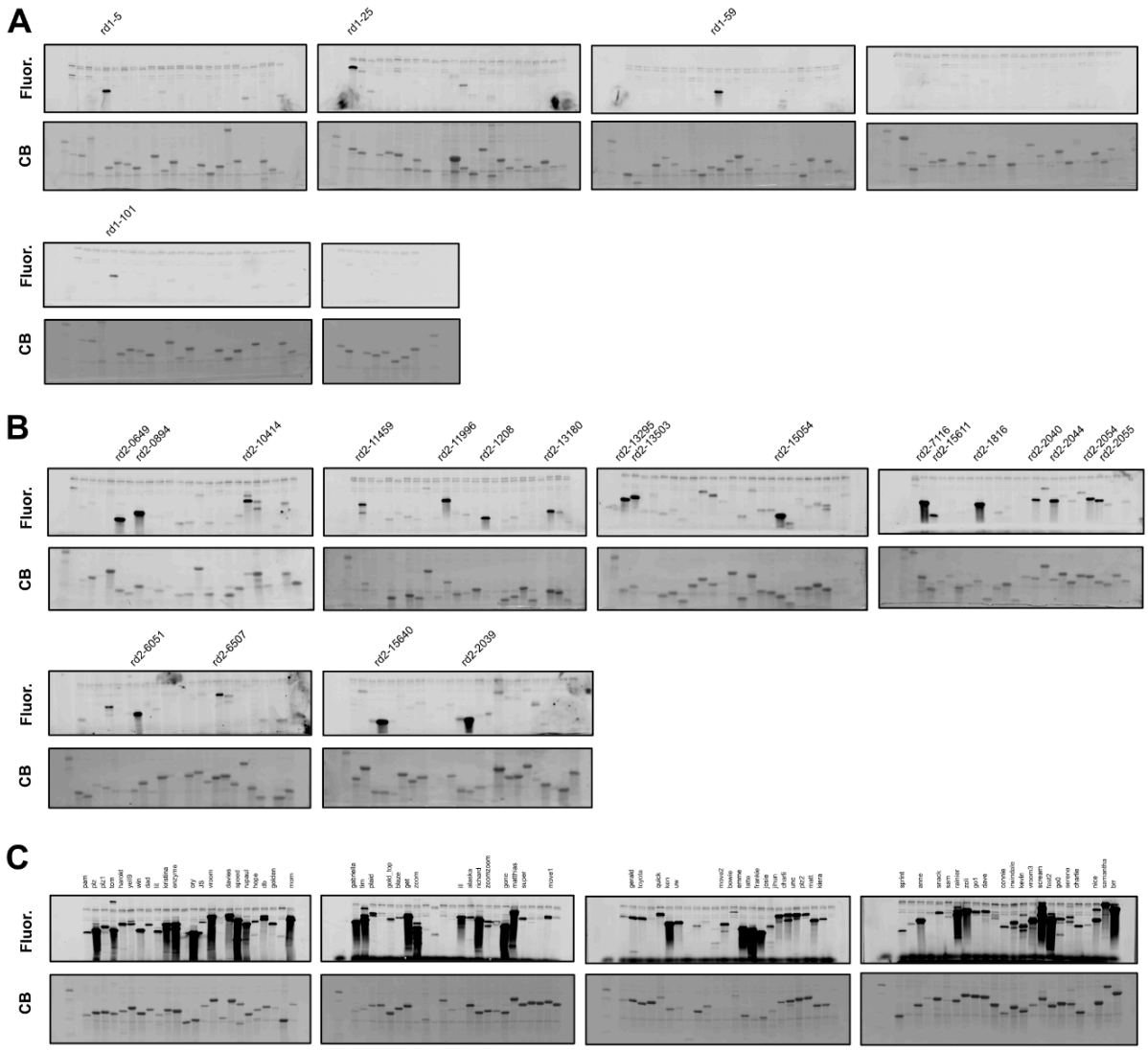

Figure S3. In-gel fluorescence imaging with fluorescently labeled fluorophosphonate activity-based probe. In-gel fluorescence of (**A**) round 1, (**B**) round 2, and (**C**) round 3 designs after 1 hour incubation of cell lysate with 1 μM FP-TAMRA.

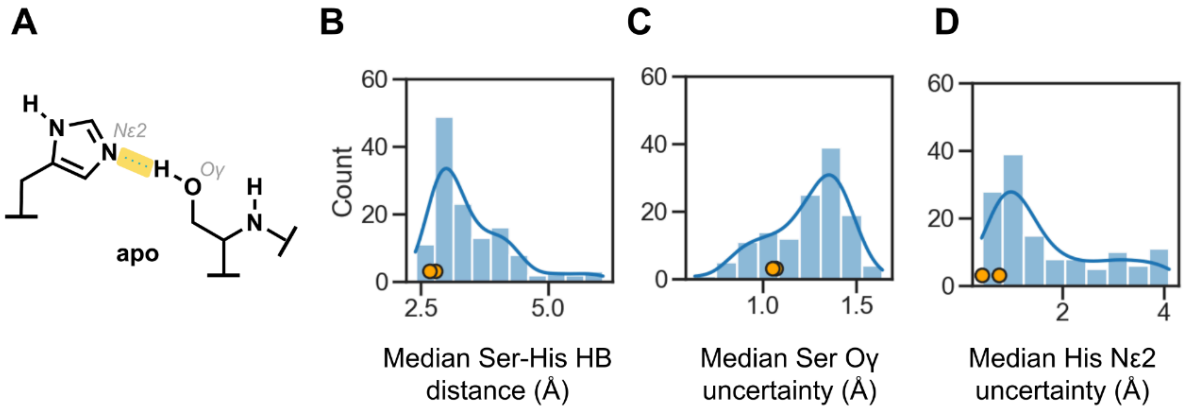

Figure S4. ChemNet analysis suggests single-turnover designs are more preorganized. (**A**) Schematic depicting catalytic serine-histidine hydrogen bond interaction. (**B**) Median Ser-His H-bond distance, (**C**) median serine Oγ uncertainty, and (**D**) median histidine Nε2 uncertainty calculated from ChemNet ensembles of the apo state for 130 round 1 designs. Yellow points represent the two single-turnover designs, shh25 and shh59.

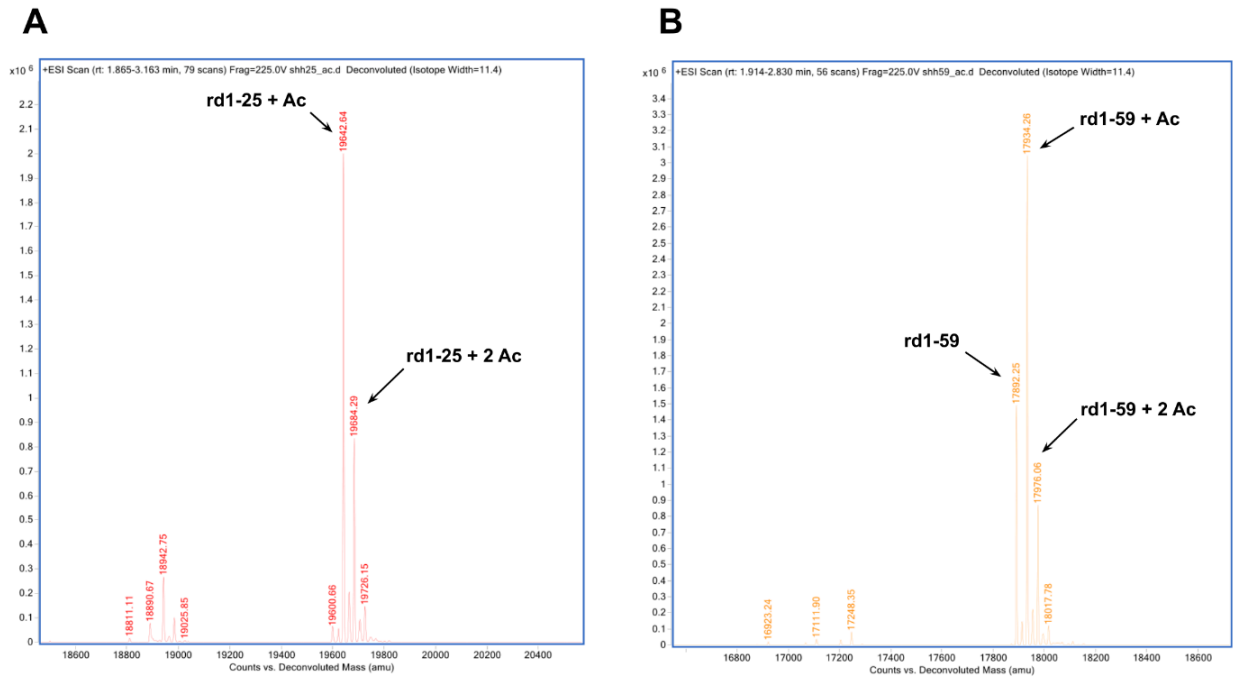

Figure S5. Mass spectra of single-turnover designs incubated with cognate ester substrates. Mass spectra of rd1-25 (**A**) and rd1-59 (**B**) after overnight incubation with 100 μM of 4MU-acetate.

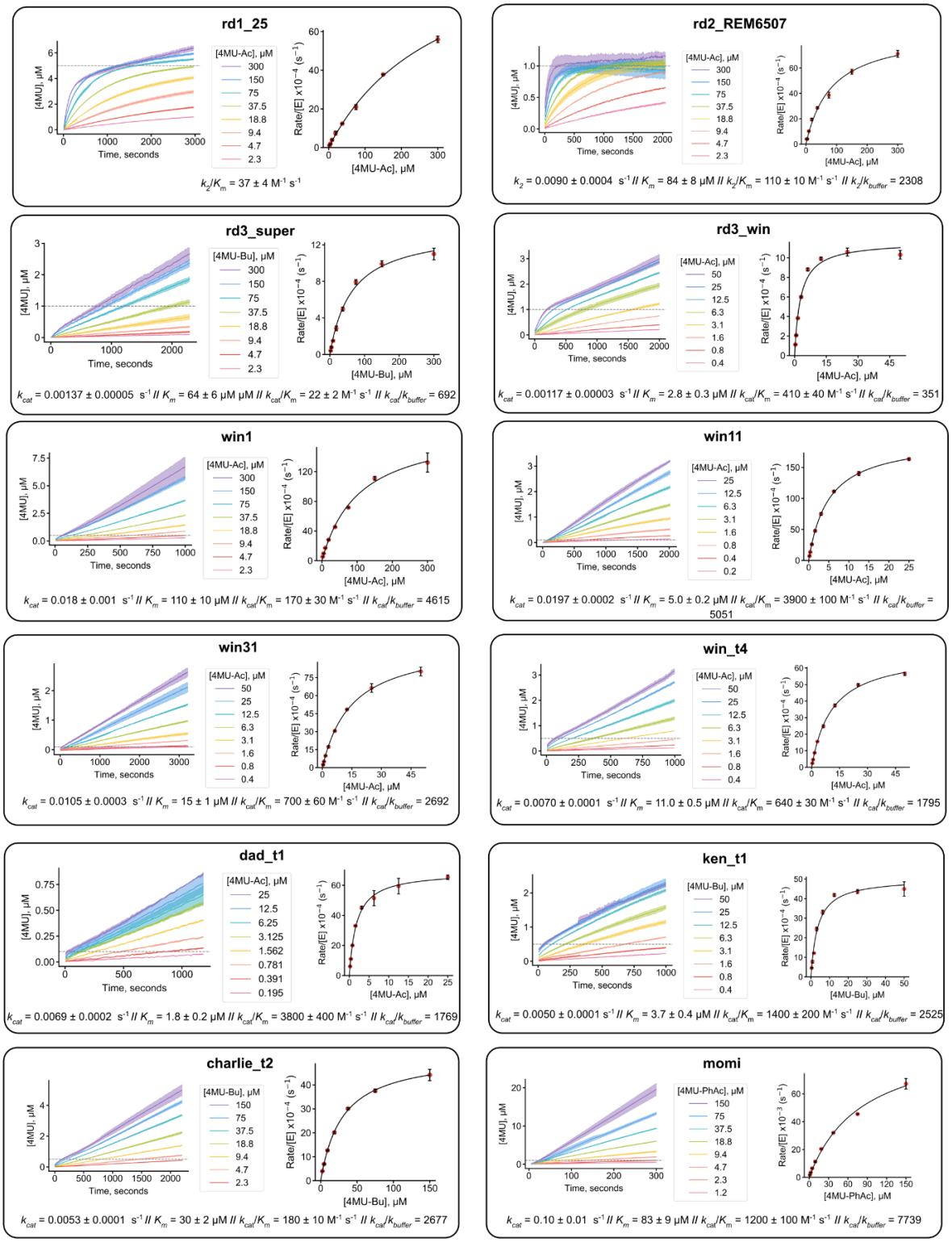

Figure S6. Michaelis-Menten kinetics of designed serine hydrolases. Reactions were performed at 30°C in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO assay buffer. Shaded area in progress curves and error bars in Michaelis-Menten plots represent standard deviation of three technical replicates.

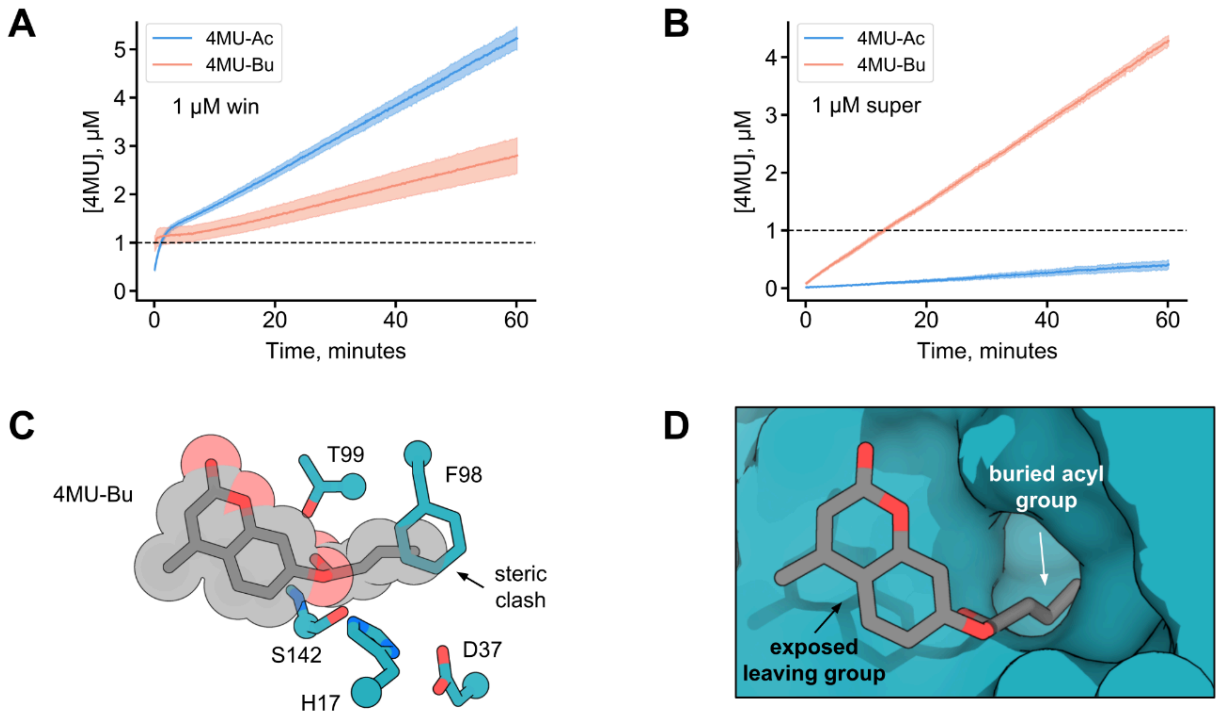

Figure S7. Substrate selectivity of win and super. Progress curves of (**A**) win and (**B**) super incubated with 50 μM of either 4MU-Ac or 4MU-Bu in 20 mM HEPES, 50 mM NaCl, pH 7.4 at 30°C. Dashed line indicates enzyme concentration and shaded area represents the standard deviation of three technical replicates. (**C**) win crystal structure with 4MU-Bu modeled in the active site. (**D**) Surface representation of super crystal structure with 4MU-Bu docked into the active site.

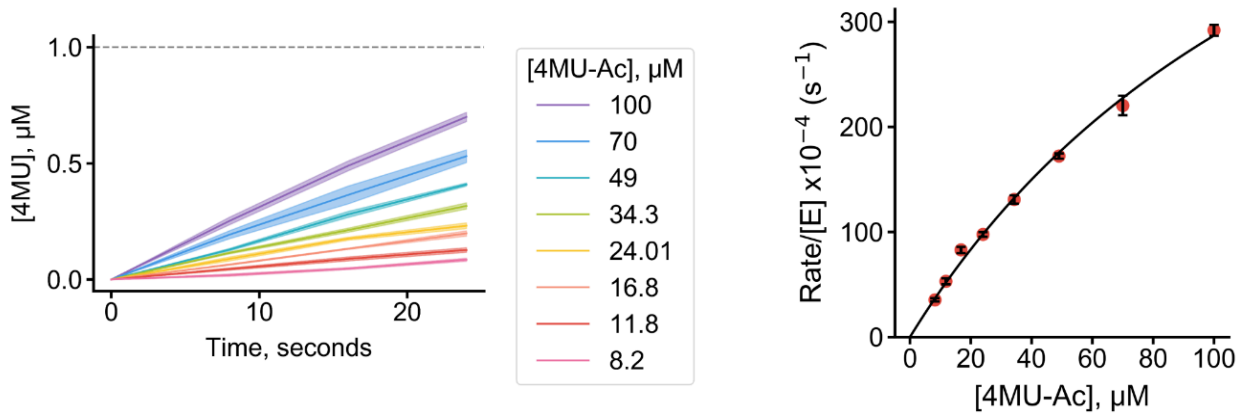

Figure S8. Burst phase portion of win reveals lack of saturation, suggesting K_s >> K_m. Dashed line represent enzyme concentration and shaded area represents standard deviation of three technical replicates.

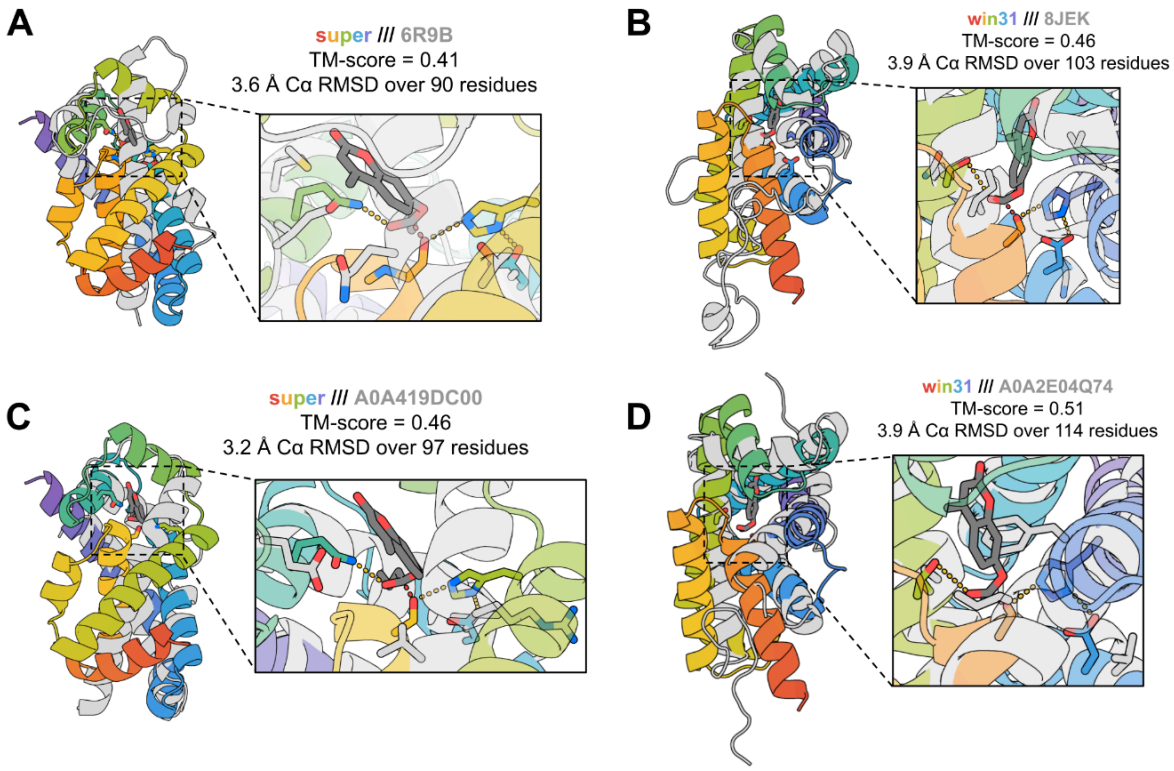

Figure S9. Structural novelty of designed esterases. (**A,B**) Overlay of super and win31 to their most similar structures in the PDB by TM-align**.** (**C,D**) Overlay of super and win31 to their most similar structure in the AlphaFold database by TM-align.

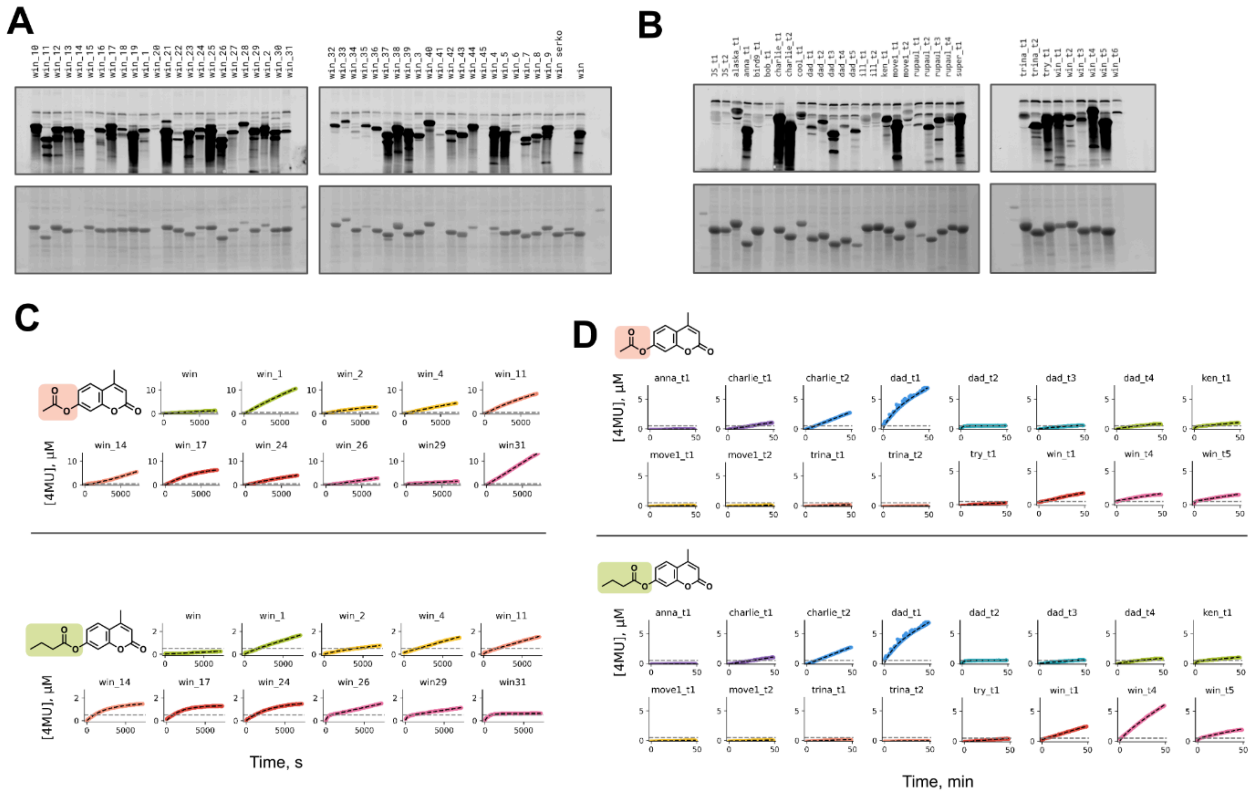

Figure S10. Screening and kinetic analysis of resampled designs. In-gel fluorescence imaging after incubation of cell lysates of (**A**) win redesigns and (**B**) round 3 redesigns with 1 µM FP-TAMRA. Progress curves of (**C**) win redesigns and (**D**) round 3 redesigns incubated at 30°C with 100 µM 4MU-Ac and 4MU-Bu in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO assay buffer. Dashed line represents enzyme concentration (0.5 µM).

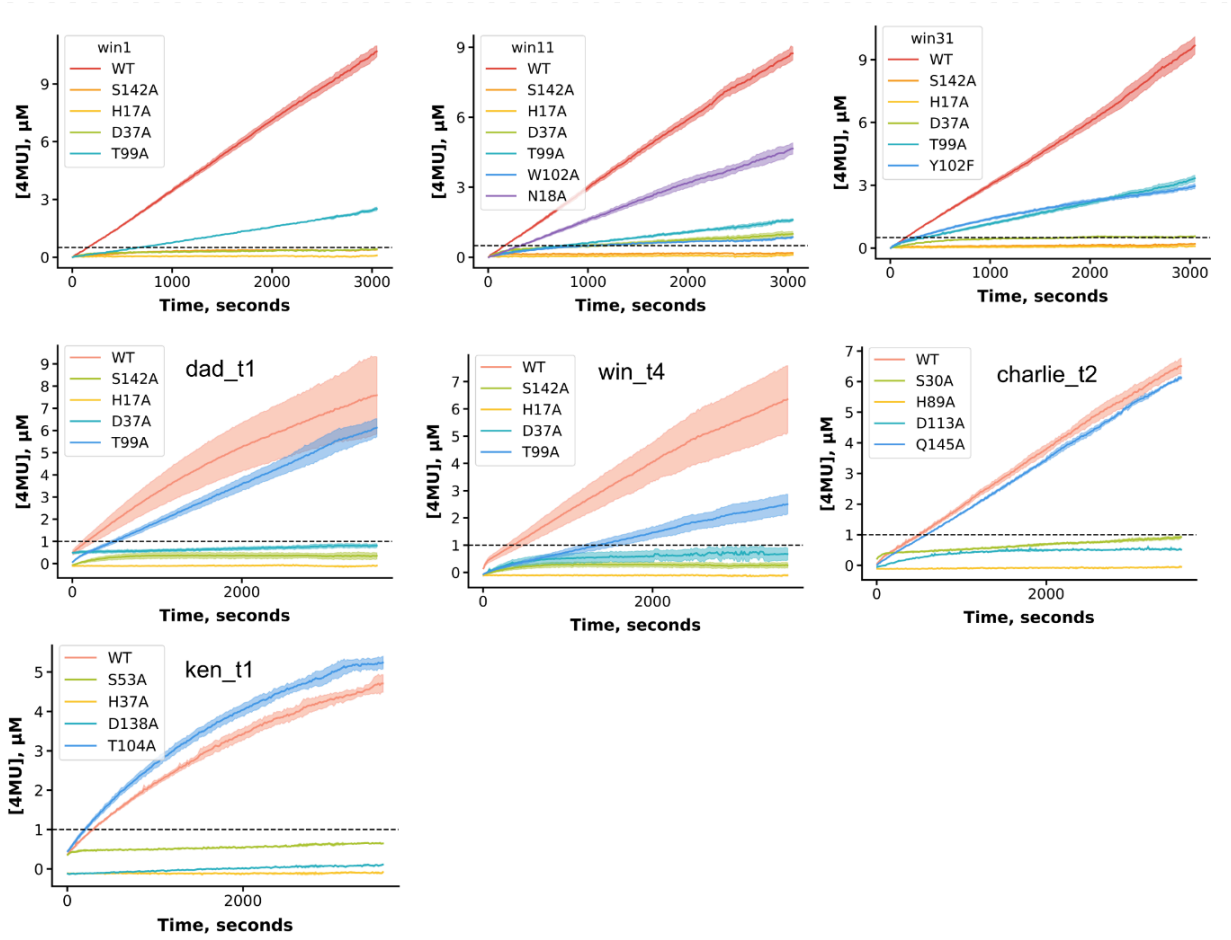

Figure S11. Progress curves for active site residue mutants of designed serine hydrolases. Reactions were performed at 30°C with 1 µM or 0.5 µM (win1, win11, win31) enzyme and 100 µM 4MU-acetate or 4MU-butyrate (charlie_t2, ken_t1) in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO assay buffer. Shaded areas represent standard deviation of three technical replicates.

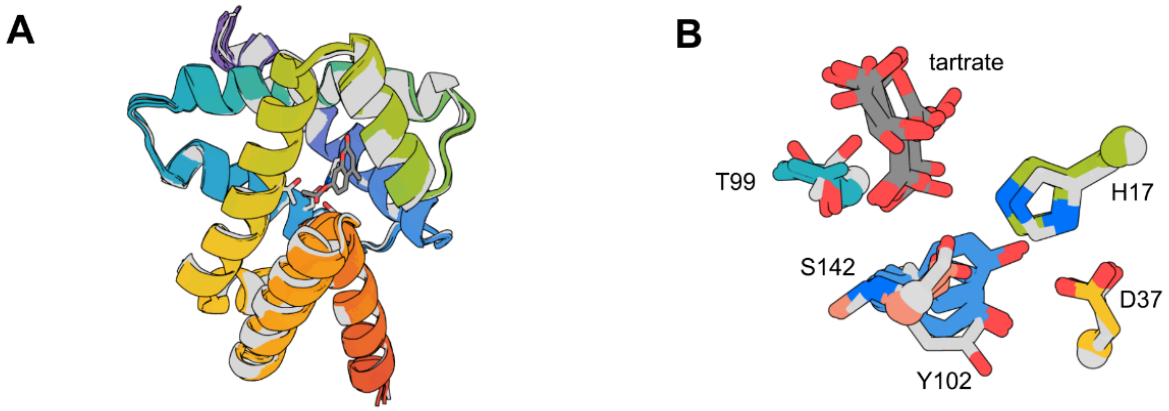

Figure S12. Analysis of win31 crystal structure. (**A**) Structural superposition of overall fold of all five chains in the asymmetric unit of win31 (color) and win31 design model (gray) (**B**) Active site zoom-in of structural superposition of the five chains in the win31 asymmetric unit (color) overlaid with design model (gray).

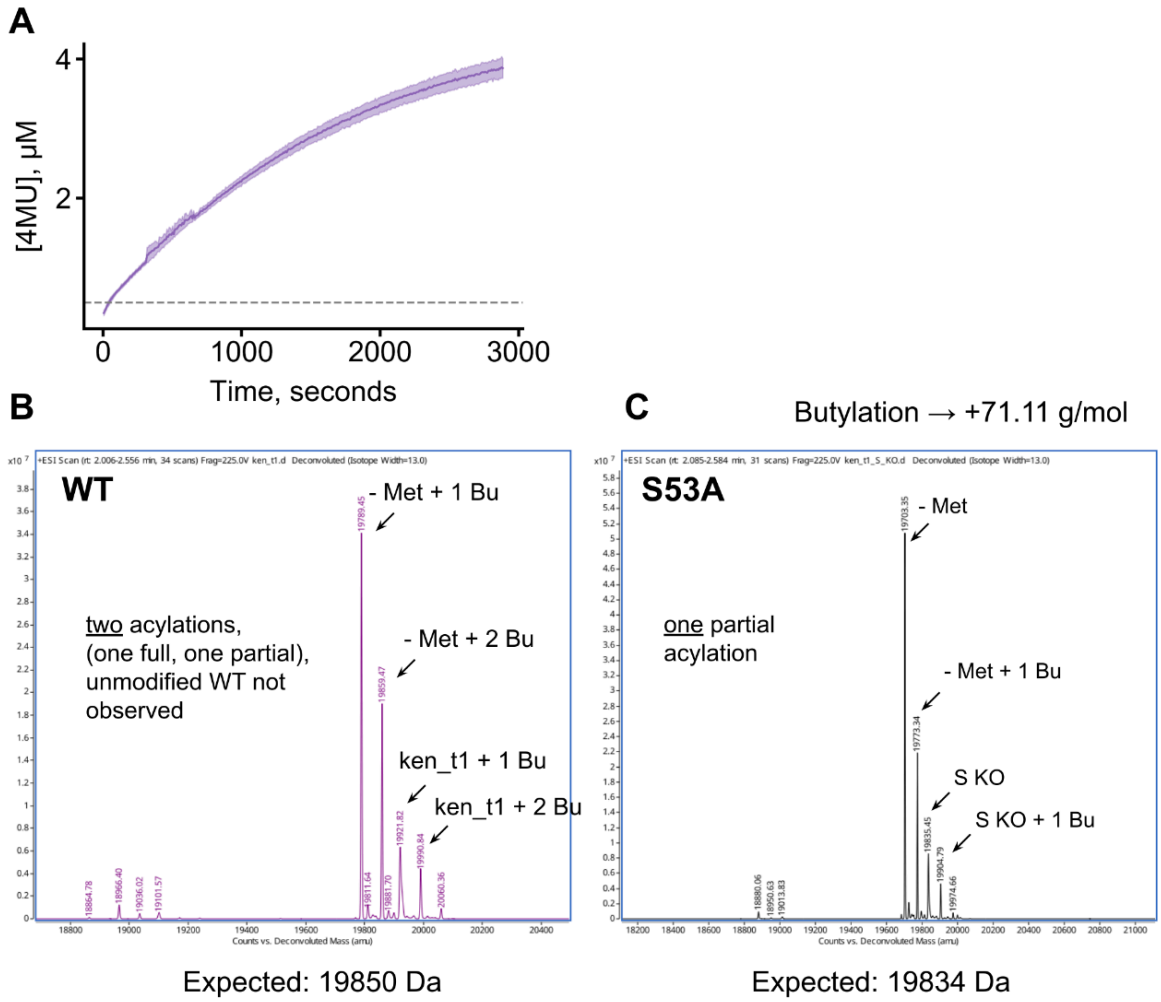

Figure S13. Kinetic and mass spectrometric analysis of ken_t1 reveal inactivation over time and stable acylated species. (**A**) Progress curve of 0.5 µM ken_t1 incubated with 50 µM 4MU-Bu in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO assay buffer at 30°C. Dashed line represents the enzyme concentration and shaded area represents standard deviation of three technical replicates. (**B**) Mass spectra of WT and catalytic serine knockout (S53A) of 50 µM ken_t1 after overnight incubation at room temperature with 100 µM 4MU-Bu.

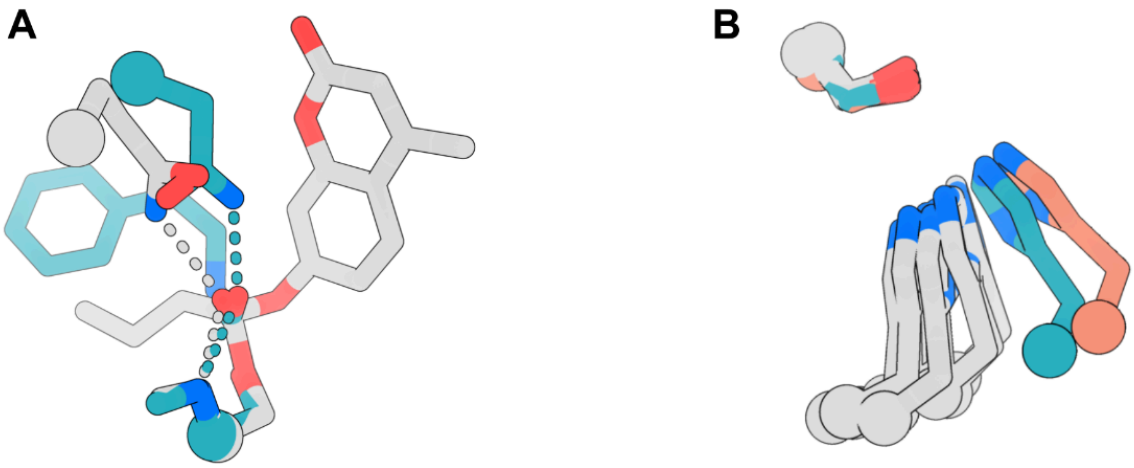

Figure S14. Potentially non-ideal catalytic geometries in designed serine hydrolases. (**A**) Structural alignment of super design model (gray) and crystal structure of subtilisin (teal) (PDB: 1scn) highlighting distinct oxyanion hole geometries of side chain aspargine (super) and glutamine (subtilisin) relative to the substrate carbonyl. Super's oxyanion may stabilize the ground state and slow catalysis. (**B**) Comparison of His-Ser interactions found among native hydrolases and those in **super** (teal) and **win** (peach).

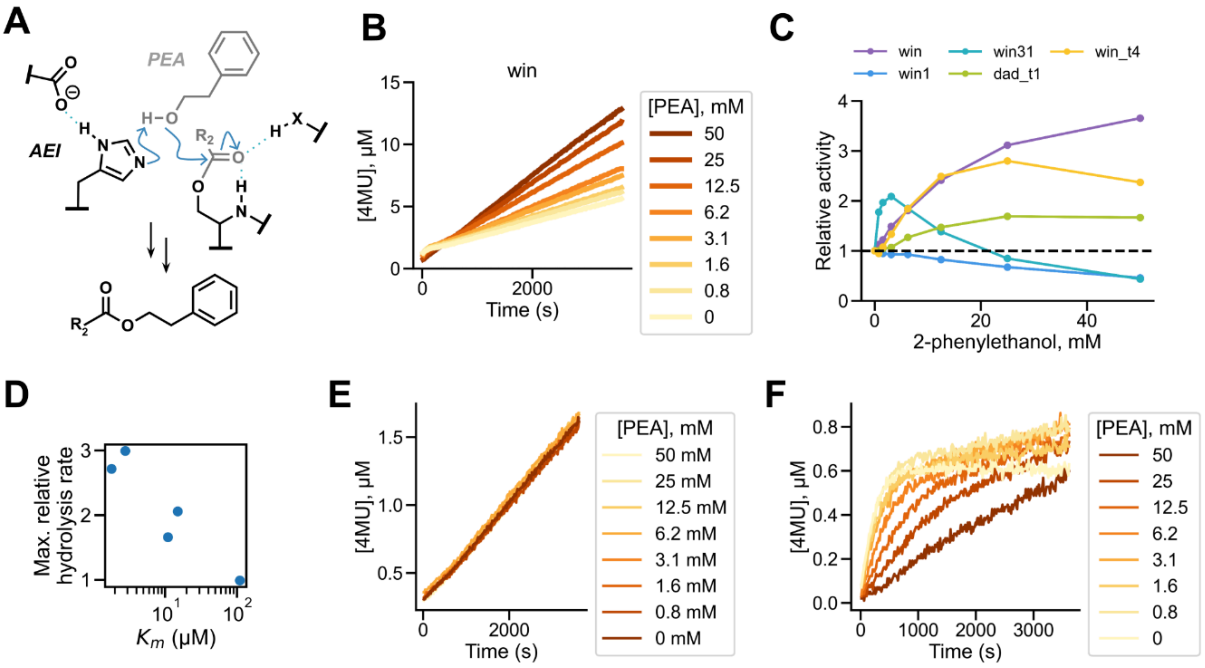

Figure S15. Acyltransferase activity. (**A**) Mechanism of transesterification via acyl acceptor attack on the acylenzyme intermediate. (**B**) Progress curves of 4MU-acetate (100 μM) hydrolysis in the presence of varying amounts of 2-phenylethanol (PEA) for a representative design (win, 1 μM). All reactions conducted in 20 mM HEPES, 50 mM NaCl, pH 7.4, 5% DMSO assay buffer at 30°C. (**C**) Relative acyltransfer to hydrolysis activity plotted against PEA concentration for five designs. (**D**) Maximal relative rate plotted against 4MU-acetate K_m for designs in (C). (**E**) Progress curves of background hydrolysis of 4MU-acetate in increasing concentrations of PEA. (**F**) Progress curves for 4MU-acetate (100 μM) hydrolysis in the presence of varying amounts of 2-phenylethanol (PEA) for the S142A variant of win_t4 (1 μM).

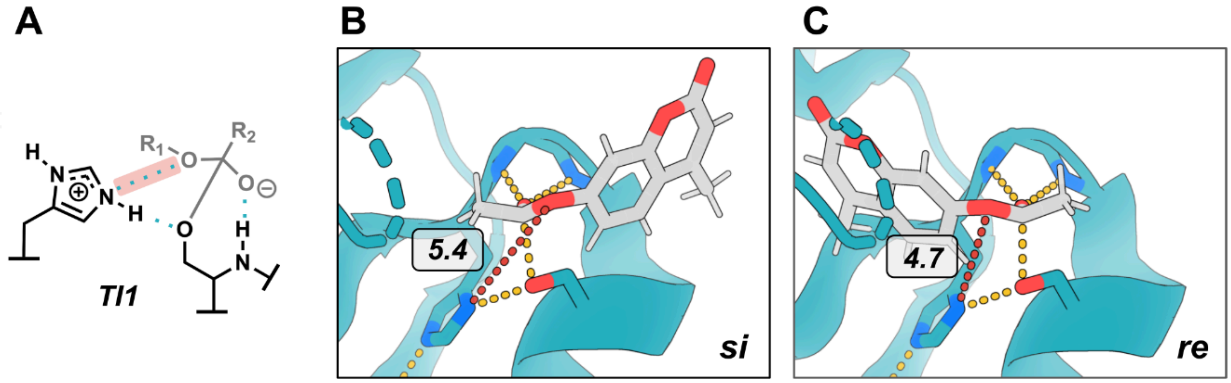

Figure S16. Lack of reaction compatibility in previous serine hydrolase design OSH55. (A) Key interaction in tetrahedral intermediate 1 between histidine Nε2 and substrate leaving group oxygen (red). (**B,C**) Crystal structure of OSH55 design (E6D/L155R variant, PDB ID 4ess) shown in complex with computationally docked 4MU-Ac esterase substrate for both re and si faces of attack. The key histidine-leaving group interaction is depicted with red dashes. Distances in Å.