# Computational design of serine hydrolases

# 丙氨酸水解酶的计算设计

Anna Lauko , Samuel J. Pellock , Ivan Anischanka , Kiera H. Sumida , David Juergens , Woody Ahern , Alex Shida , Andrew Hunt , Indrek Kalvet , Christoffer Norn , lan R. Humphreys , Cooper Jamieson , Alex Kang , Evans Brackenbrough , Asim K. Bera , Banumathi Sankaran , K. N. Houk , David Baker

安娜·劳科，塞缪尔·J·佩洛克，伊万·阿尼尚卡，基拉·H·苏米达，大卫·尤根斯，伍迪·阿赫恩，亚历克斯·希达，安德鲁·亨特，印德雷克·卡尔维特，克里斯托弗·诺恩，兰·R·汉弗莱斯，库珀·贾米森，亚历克斯·康，埃文·布拉肯布劳，阿西姆·K·贝拉，巴努马提·桑卡兰，K·N·霍克，大卫·贝克

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

1. 华盛顿大学生物化学系，西雅图，华盛顿州，美国

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

2. 华盛顿大学蛋白质设计研究所，西雅图，华盛顿州，美国

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

3. 华盛顿大学生物物理、结构与设计研究生项目，西雅图，华盛顿州，美国

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

4. 华盛顿大学分子工程研究生项目，西雅图，华盛顿州，美国

5. Department of Chemistry, University of Washington, Seattle, WA, USA

5. 华盛顿大学化学系，西雅图，华盛顿州，美国

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

6. 霍华德·休斯医学研究所，华盛顿大学，西雅图，华盛顿州，美国

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

7. 华盛顿大学计算机科学与工程保罗·G·艾伦学院，美国华盛顿州西雅图。

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

8. 加利福尼亚大学洛杉矶分校化学与生物化学系，美国加利福尼亚州洛杉矶。

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

9. 这些作者贡献相同:安娜·劳科，塞缪尔·J·佩洛克。

# Abstract

# 摘要

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 that catalyze ester hydrolysis with catalytic efficiencies up to , closely match the design models (Ca RMSDs ), and have folds distinct from natural serine hydrolases. 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.

通过多步反应机制进行反应的酶通常利用复杂的极性活性位点，以亚埃级的精度定位，以介导不同的化学步骤，这使得它们的全新构建极具挑战性。我们试图通过经典的催化三联体和丝氨酸水解酶的氧阴离子孔作为模型系统来克服这一挑战。我们使用RFdiffusion 生成具有日益复杂和不同几何形状的催化位点的蛋白质，并使用一种新开发的集合生成方法ChemNet来评估反应每一步的活性位点几何形状和预组织。实验表征揭示了新型丝氨酸水解酶，这些酶催化酯水解的催化效率 高达 ，与设计模型(Ca RMSDs )非常接近，并且其折叠结构与天然丝氨酸水解酶不同。基于反应坐标上活性位点预组织的设计的计算选择显著提高了成功率，使得在仅有20个设计的筛选中识别出新的催化剂成为可能。我们的全新构建方法提供了催化几何决定因素的见解，补充了从天然酶的结构和突变研究中获得的结果(在这些研究中，催化基团几何形状和活性位点组成无法如此系统地变化)，并为工业相关的丝氨酸水解酶的设计提供了路线图，更广泛地说，为设计催化多步转化的复杂酶提供了指导。

# Main Text

# 主要文本

Enzymes are exquisite 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 . 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 . 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. A further challenge of enzyme design is the preorganization of the active site 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 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.

酶是精妙的催化剂，能够在温和的水相条件下显著加速反应速率。构建能够催化任意化学反应的酶的能力在广泛的应用中具有巨大的实用价值，因此，酶设计一直是计算蛋白质设计的长期目标。自下而上的酶设计通常从催化残基围绕反应过渡态的排列规范(理论酶)开始，旨在识别该活性位点在现有结构中的位置。固定的骨架限制了催化几何形状的准确实现，这可能限制了许多设计酶在实验室进化优化之前的活性。酶设计的另一个挑战是以原子级精度预先组织活性位点。对于多步反应机制，实现预组织尤其困难，因为酶必须优先稳定多个过渡态和中间体。现有的计算方法评估设计预组织的准确性受到低精度或计算成本的限制，通常仅应用于一个反应状态。为了实现多步酶的准确设计，需要新的方法来生成专门针对特定活性位点的蛋白质骨架，并评估其在催化循环中的结构兼容性。

We reasoned that advances in deep learning for protein design and structure prediction could be used to design proteins from scratch to scaffold a given active site and assess compatibility across a proposed reaction coordinate. Recent advances in scaffolding functional sites with RFdiffusion have yielded improved in silico and experimental success rates across a range of design tasks ; we aimed to use the same approach to generate enzymes starting from geometric descriptions of an active site (Fig. 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. 1B).

我们推测，深度学习在蛋白质设计和结构预测方面的进展可以用于从头设计蛋白质，以支撑给定的活性位点，并评估在提议的反应坐标上的兼容性。最近，使用RFdiffusion对功能位点进行支撑的进展在一系列设计任务中提高了计算和实验的成功率；我们旨在采用相同的方法，从活性位点的几何描述出发生成酶(图1A)。为了评估催化循环每一步的预组织和功能相互作用，我们希望利用基于深度学习的蛋白质-小分子复合物预测的进展，通过建模催化中间体的结构集合(图1B)。

Ester hydrolysis has served as a model reaction for computational enzyme design for decades , and the catalytic triad and oxyanion hole of natural serine hydrolases utilize one of the most extensively studied enzymatic mechanisms to catalyze this reaction . The catalytic cycle can be divided into four steps (Fig. 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 , which tunes its acid/base function, is mediated by interaction with aspartate or glutamate, the final residue in the triad .

酯水解几十年来一直作为计算酶设计的模型反应， 自然丝氨酸水解酶的催化三联体和负离子孔利用了最广泛研究的酶促机制之一来催化这一反应。 催化循环可以分为四个步骤(图1C)。首先，底物与无酶(apo)结合，催化丝氨酸在催化组氨酸的去质子化作用下攻击酯的羰基碳，形成第一个四面体中间体(TI1)。其次，催化组氨酸使离去基团的氧质子化，促进其离去，留下活性位点的丝氨酸与底物的酰基共价连接(酰基-酶中间体，AEI)。第三，组氨酸去质子化一个水分子，该水分子攻击AEI生成第二个四面体中间体(TI2)。最后，通过组氨酸介导的丝氨酸质子化和酰基的释放来解决这个中间体，重新构成自由酶并完成催化循环。在整个过程中，带负电的过渡态和中间体通过一对氢键供体得到稳定，这对供体构成了负离子孔。组氨酸的扰动 ，调节其酸/碱功能，是通过与天冬氨酸或谷氨酸的相互作用介导的，后者是三联体中的最后一个残基 。

Despite extensive structural, mutational, and computational characterization of native serine hydrolases , de novo design efforts that have attempted to employ this mechanism have been largely unsuccessful, yielding proteins that harbor activated serines and cysteines but fail to catalyze turnover . We initially speculated that increasing scaffold diversity would help identify backbones that more accurately reconstruct the desired active site; and we carried out a preliminary design campaign searching for placements of a serine hydrolase active site in a library of deep-learning generated hallucinated NTF2 scaffolds that previously yielded catalysts for a luciferase reaction . 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), suggesting that key features important for catalysis were missing.

尽管对天然丝氨酸水解酶进行了广泛的结构、突变和计算特征分析，但尝试利用这一机制进行的从头设计工作大多未能成功，产生的蛋白质虽然含有活化的丝氨酸和半胱氨酸，但未能催化反应。我们最初推测，增加支架的多样性将有助于识别更准确重建所需活性位点的骨架；因此，我们进行了初步设计活动，寻找在之前产生荧光素酶反应催化剂的深度学习生成的幻觉NTF2支架库中放置丝氨酸水解酶活性位点的可能性。与之前的研究一样，结果设计的实验表征显示出活化的丝氨酸，但在活化酯底物上没有催化转化，尽管实验结构与设计结构之间非常接近，这表明催化所需的关键特征缺失。

# Assessing reaction path compatibility with ChemNet

# 评估与ChemNet的反应路径兼容性

We set out to understand why these and earlier computational designs failed to 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, 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 . ChemNet is stochastic, and repeated runs from different random seeds yield an ensemble of models for the rebuilt region.

我们着手理解为什么这些以及早期的计算设计未能催化酯水解，并假设在完整反应坐标上建模状态可能对评估设计实现催化转化的能力至关重要。为了建模设计的酶在反应周期中形成每个关键状态的程度，并评估活性位点残基在所需催化几何中的预组织，我们开发了一个深度神经网络，该网络在给定(1)小分子结合口袋或活性位点的主链坐标，(2)每个位置的氨基酸残基的身份，以及(3)结合的小分子的化学结构(但不包括它们的位置)时，生成结合位点的完整原子坐标，包括蛋白质侧链和小分子。我们在PDB中的蛋白质-小分子复合物上训练了这个名为ChemNet的网络，通过在最多600个重原子组成的球形区域内随机化侧链和小分子的原子坐标，并试图最小化评估该区域内原子坐标重现的损失函数。ChemNet以平均RMSD为1.1重建了原生结构中的区域。ChemNet是随机的，来自不同随机种子的重复运行产生了重建区域的模型集。

We used ChemNet to generate structural ensembles for each of the four reaction steps for a set of native and previously designed serine hydrolases. These calculations showed that native hydrolases are considerably more preorganized than previous designed systems (Fig. 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 can often be 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 Methods).

我们使用ChemNet为一组天然和先前设计的丝氨酸水解酶生成了四个反应步骤的结构集合。这些计算表明，天然水解酶的预组织程度明显高于先前设计的系统(图1D，图S2)。在天然系统中，每个步骤的催化残基采样的构象数量非常有限，所有关键的氢键相互作用都得以维持，但在设计的系统中，多个步骤的集合往往存在较大的变化。由于反应速率应与处于活性状态的酶的比例成正比，因此设计的活性位点缺乏预组织预计会影响催化效率。为了量化ChemNet集合中活性位点形成的程度，我们计算了催化功能团与反应中间体之间关键相互作用的形成频率，涵盖了反应的每个步骤(见方法)。

# Design and characterization of serine hydrolases

# 丝氨酸水解酶的设计与表征

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. 2A, B). We designed catalysts for the hydrolysis of 4- methylumbelliferone (4MU) esters (Fig. 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 , and atoms of these key residues and their adjacent 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 Methods). To drive folding to the designed state, and to make favorable interactions with the substrate and active site residues, LigandMPNN was used to design the sequence. Rosetta FastRelax was used to refine the protein backbone and ligand pose, and the sequence was again designed with LigandMPNN with the new backbone as input . Following several iterations between LigandMPNN and FastRelax, the structures of the designs were predicted with AlphaFold2 (AF2) , and designs for which all catalytic residue Ca’s were positioned within of the design models were selected for experimental characterization (see Methods for additional details of computational design).

接下来，我们着手设计具有日益复杂的活性位点的蛋白质，使用RFdiffusion来支架丝氨酸水解酶活性位点基序，并使用ChemNet评估反应每一步的预组织(图2A, B)。我们设计了用于水解4-甲基伞形酮(4MU)酯的催化剂(图2C)，在水解时会发出荧光。为了生成支架催化机制的骨架，我们将催化侧链放置在底物周围，并从这些关键残基及其相邻邻居的骨架 和 原子开始，使用RFdiffusion从随机噪声构建骨架，这些骨架的坐标与输入基序完全匹配，并形成底物的结合口袋(详见方法)。为了驱动折叠到设计状态，并与底物和活性位点残基形成有利的相互作用，使用LigandMPNN 设计序列。使用Rosetta FastRelax 来优化蛋白质骨架和配体姿态，并再次使用LigandMPNN以新的骨架作为输入设计序列 。经过LigandMPNN和FastRelax之间的几次迭代，设计的结构通过AlphaFold2(AF2) 进行预测，并选择所有催化残基Ca的位置在设计模型的 范围内的设计进行实验表征 (有关计算设计的更多细节，请参见方法)。

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. 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 Methods). Only 1.6% of round 2 designs passing 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.

在前两轮设计中，我们构建了相对简单的活性位点，由具有单一氧阴离子孔接触的Ser-His二聚体组成(图2A, B)，并明确评估了ChemNet在选择实验表征设计中的实用性；第一轮设计仅通过AF2过滤，而第二轮设计在AF2过滤通过后，如果ChemNet对apo状态的集合表明关键的Ser-His氢键已形成，则被选为实验筛选(见方法)。只有1.6%的第二轮设计在AF2过滤通过后被ChemNet预测为预先组织。为了进行实验测试，我们获得了编码129和192个设计的合成基因，分别用于第一轮和第二轮的E. coli过表达和筛选。

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. 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. 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. 1E, F). The progress curves for these round 1 and 2 designs plateau after approximately one enzyme equivalent of fluorescent product is formed (Fig. 2E), suggesting they catalyze initial nucleophilic attack but fail to hydrolyze the AEI, the rate-limiting step in the cleavage of activated esters . 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).

我们使用了一种氟磷酸酯(FP)活性探针和荧光4MU-醋酸酯(4MU-Ac)及4MU-丁酸酯(4MU-Bu)底物，以识别具有活化丝氨酸和酯酶活性的设计(图2C)。在大肠杆菌裂解液中，标记FP探针的设计比例从第一轮的3%增加到第二轮的17%，几乎增加了5倍(图2B，图S3)。与FP探针反应的设计被纯化并与4MU酯孵育，结果显示两种第一轮设计(1.6%)和十种第二轮设计(5.2%)表现出催化活性。对第一轮设计的回顾性ChemNet分析显示，两个具有催化活性的设计中的丝氨酸-组氨酸氢键被预测为最具预组织性(图S4)。对第二轮设计在关键丝氨酸-组氨酸氢键形成程度的ChemNet筛选，不仅增加了表现出FP探针标记和酶活性的设计比例，还导致了更高的活性(图1E, F)。这些第一轮和第二轮设计的进展曲线在大约形成一个酶当量的荧光产物后趋于平稳(图2E)，这表明它们催化了初始的亲核攻击，但未能水解AEI，这是活化酯裂解中的速率限制步骤。当与底物孵育时，这些设计的质谱显示出与酰化相对应的质量变化，进一步支持在形成酰化中间体后蛋白质失活的观点(图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 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. 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. 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 or for single-turnover and multiple-turnover designs, respectively (Fig. 2E, Fig. S6). For the two designs that displayed catalytic turnover, called ’super’ and ’win,’ values were 22 and ), respectively for the more preferred of the two 4MU substrates (win and super preferentially hydrolyzed 4MU-Ac and 4MU-Bu, respectively (Fig. S7)).

我们假设在第三轮设计(第3轮)中引入一种组氨酸稳定的催化酸和第二个氧阴离子孔氢键供体，并在apo和 状态下过滤ChemNet预组织，能够生成通过水解AEI实现催化转化的设计。在第3轮设计中，我们要求所有催化三联体和氧阴离子孔氢键在apo和AEI状态的ChemNet集合中高度预组织。在132个第3轮设计中，111个(84%)显示了FP探针标记，20个水解了4MU底物(18%)，两个设计(1.5%)显示了多次转化活性(图2B, E)。来自三轮的活跃设计在任何一个催化残基(Ser、His、Asp/Glu和氧阴离子侧链接触)突变后显示出显著降低的活性(图2E)，这表明观察到的活性依赖于设计的活性位点。为了确定活跃设计的动力学参数，测量了初始或稳态速率，以确定单次转化和多次转化设计的 或 (图2E，图S6)。对于显示催化转化的两个设计，称为“super”和“win”， 值分别为22 和 ，对于两个4MU底物中更优选的一个(win和super分别优先水解4MU-Ac和4MU-Bu)(图S7)。

# Structural characterization of designed serine hydrolases

# 设计的丝氨酸水解酶的结构表征

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 Ca RMSDs of 0.8 Å over 165 residues and 0.83 Å over 160 residues (Fig. 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. 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. 3B). Similarly, in win, an acetate molecule is positioned at the catalytic center and hydrogen bonds to the Oy and backbone amide nitrogen of the catalytic serine S142, the oxygen of T99, and the histidine acid/base residue H17, key hydrogen bonds in the catalytic cycle (Fig. 3E).

我们采用X射线晶体学来确定super和win的设计准确性。我们成功解析了super和win的晶体结构，发现它们与设计模型的钙根均方根偏差(Ca RMSD)分别为0.8 Å(165个残基)和0.83 Å(160个残基)(图3A, D)。实验结构与设计结构之间的高度一致性延伸到活性位点的几何形状:催化残基的侧链构象在super(全原子RMSD = 0.38 Å，22个原子)和win(全原子RMSD = 0.86 Å，20个原子)中是原子级一致的，除了侧链氧阴离子接触的旋转异构体变化T99(图3B, E)。在super的活性位点上，一个水分子位于亲核性丝氨酸上方，并与氧阴离子孔接触形成氢键，这可能模拟了其酯底物的羰基氧的位置(图3B)。类似地，在win中，一个乙酸分子位于催化中心，并与催化丝氨酸S142的氧和主链酰胺氮、T99的氧以及组氨酸酸/碱残基H17形成氢键，这些是催化循环中的关键氢键(图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. 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. 3F and S7).

虽然在没有结合的小分子底物或过渡态类似物的情况下解决了这些结构，但设计模型与超分子的晶体结构叠加显示出与其优选底物的丁酸酰基(butyrate acyl group)之间高度的形状互补性(图3C和S7)。与此同时，4MU部分大部分暴露，证实了超分子对4MU-Bu的选择性高于4MU-Ac，并且表明在这种情况下，底物结合主要是由与酰基的结合驱动的。对于win，晶体结构中F98的旋转异构体位移会与丁酸部分发生冲突，实际上，win对较小的底物4MU-Ac具有选择性，避免了这种冲突(图3F和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. S8), demonstrating that the design method employed here can find protein structural solutions that extend well beyond those found in nature.

超蛋白和胜利蛋白的结构与已知结构非常不同；与PDB和更大的AlphaFold数据库通过TM-align找到的最接近匹配的TM分数为0.41/0.46(PDB/AlphaFold数据库)和0.46/0.51(在0.5的截止值或以下，低于该值的结构被认为具有不同的拓扑折叠)，是功能未知的蛋白质，并且在折叠或活性位点水平上与已知水解酶没有相似性(图S8)，这表明这里采用的设计方法可以找到远远超出自然界中发现的蛋白质结构解决方案。

Filtering for preorganization across the reaction coordinate improves catalysis

在反应坐标上进行预组织过滤可以改善催化反应

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 of the active design win (fixing only the identities of the four catalytic residues) (Fig. 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 values compared to the starting design: win has a of , which increases 15-fold in win -fold in win , and 9-fold in win31 (0.0105 s ) (Fig. 4B and Fig. S6). Of the 11 four-state filtered designs tested,10 (91%) labeled with FP-probe (Fig. S9). Two of these, dad\_t1 and win\_t4, displayed higher catalytic efficiencies than the starting design, with values of and , largely driven by improvements to (Fig. 4B and S6). Catalytic triad residue knockouts for all designs showed significant reductions in activity. In win11 and win31, mutation of preorganizing residues in the second shell of the active site that H-bond to the catalytic aspartate also significantly reduced activity (Fig. S10).

我们接下来寻求生成并比较通过ChemNet明确过滤的设计，以实现两个状态(apo和AEI)或反应路径的所有四个状态的预组装，方法是对活跃设计的LigandMPNN和FastRelax进行额外迭代(仅固定四个催化残基的身份)(图4A)。我们获得了编码45个两状态过滤设计的基因用于实验表征，所有这些设计在序列上与原始设计相比都具有多样性(与母体设计的平均序列同一性为58%，在活性位点内为61%)，并发现38个(84%)标记为FP探针。这三种设计，win1、win11和win31，与起始设计相比显示出更高的 值:win的 为 ，在win 中增加了15倍，在win 中增加了 倍，在win31中增加了9倍(0.0105 s )(图4B和图S6)。在测试的11个四状态过滤设计中，10个(91%)标记为FP探针(图S9)。其中两个，dad\_t1和win\_t4，显示出比起始设计更高的催化效率，具有 值为 和 ，主要是由于对 的改善(图4B和S6)。所有设计的催化三联体残基缺失显示出活性显著降低。在win11和win31中，活性位点第二壳层中与催化天冬氨酸形成氢键的预组装残基的突变也显著降低了活性(图S10)。

We determined the crystal structures of win1 and win31 which revealed very close matches to the design models, with RMSDs of 1.42 Å and 0.7 Å, respectively (Fig. 4E, F). For win 1, the active site, including the oxyanion hole sidechain contact, more closely matches the designed conformation (mean all-atom RMSD = 0.54 Å) than the parent design win (Fig. 4E), which may be partly responsible for the 15 -fold increase in . For win31, five chains are present in the asymmetric unit, all of which closely match the design model (average RMSD ) at the backbone level (Fig. 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 preorganizing tyrosine (Fig. S10), but still a very close match to the design model with a mean all-atom RMSD of . Tartrate, derived from the crystallization solution, satisfied 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. 4F), likely mimicking key contacts employed throughout the catalytic cycle.

我们确定了win1和win31的晶体结构，显示出与设计模型非常接近的匹配，RMSD分别为1.42 Å和0.7 Å(图4E，F)。对于win1，活性位点，包括氧阴离子孔侧链接触，更接近设计构象(平均全原子RMSD = 0.54 Å)而不是母体设计win(图4E)，这可能部分解释了 的15倍增加。对于win31，非对称单元中存在五条链，所有链在主链水平上与设计模型(平均 RMSD )非常接近(图4F和S11)。对非对称单元中所有链的活性位点分析显示，催化丝氨酸、侧链氧阴离子苏氨酸和一个预组织的酪氨酸具有一定的活动性(图S10)，但仍与设计模型非常接近，平均全原子RMSD为 。来自结晶溶液的酒石酸满足了所有五条链活性位点中的电子密度，并与丝氨酸、组氨酸和氧阴离子孔接触形成氢键(图4F)，可能模拟了在催化循环中使用的关键接触。

We then 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 and 1400 (Fig. 4G, H, I, J).

我们接着探讨了严格的ChemNet过滤是否能够针对反应坐标上的最佳催化几何形状和预组织生成具有新颖骨架拓扑、活性位点和底物的活性酯酶。我们在所有四种状态下进行了广泛的序列重新设计和基于催化几何形状的过滤，从第3轮骨架开始，这些骨架之前未显示酯酶活性，在测试的20个设计中，有两个(charlie\_t2和ken\_t1)显示出显著的酯酶活性，催化效率分别为 和1400 (图4G, H, I, J)。

To test the generality of this ChemNet filtering approach, 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 amide hydrogen bond donors, rather than a backbone donor and a sidechain donor, and the first backbone donor was the residue following the catalytic serine rather than the catalytic serine itself (Fig. 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 of and of , a 5 -fold faster rate than win 11, 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 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 combination shows that filtering for preorganization throughout the reaction cycle can yield novel catalysts in one shot.

为了测试这种ChemNet过滤方法的普遍性，我们将其应用于不同的底物4MU-苯乙酸酯(4MU-PhAc)以及不同的活性位点构型，其中的氧阴离子孔由两个主链氨基的氢键供体组成，而不是一个主链供体和一个侧链供体，并且第一个主链供体是催化丝氨酸后面的残基，而不是催化丝氨酸本身(图4K)。我们使用上述设计流程为这个新底物和催化位点生成了66个设计。其中最活跃的设计momi显示出 为 和 为 ，其速率比之前的最佳设计win 11快5倍，后者在转化数方面表现最佳。RFdiffusion为该活性位点几何形状生成的折叠分布与原始几何形状不同，产生了更多的 折叠解决方案(如momi的情况)，显示出RFdiffusion构建方法如何将整体蛋白质结构拓扑塑造为特定的活性位点。对于这个新底物和催化位点组合，在没有任何先前实验表征的情况下所达到的高活性表明，在反应周期中对预组织的过滤可以一次性产生新型催化剂。

While the catalytic efficiencies of our designed serine hydrolases are far higher than previously reported catalytic triad-based designs, they are still orders of magnitude slower than native hydrolases. Several experimental results identify clear areas for improvement. 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. S12), indicating that designs that hydrolyze the are still susceptible to inactivation, potentially from off-mechanism acylation events in the active site, which will be important to avoid in future design efforts. Second, in three designs (dad\_t1, charlie\_t2, ken\_t1) from later design rounds made with stringent Chemnet filtering, mutation of the second sidechain oxyanion hole residue has a smaller effect on activity than in the earlier design rounds and compared to analogous mutations in native enzymes (Fig. S10). To investigate the structural effect of the oxyanion hole, we made ChemNet predictions of wild-type and oxyanion hole alanine knockout mutants for all active designs. In the case of super, predictions of Q71A exhibit a clear conformational change of the acylated serine in the AEI which lengthens its distance from the histidine, providing a structural explanation for the loss in activity (Fig. S13). In contrast, wild-type and oxyanion hole knockout predictions were indistinguishable for other designs, including win and high-activity redesigns of win (Fig. S13). Our analysis suggests that the improvements in catalysis achieved throughout our design rounds may derive primarily from improvements in catalytic triad organization and intermediate positioning; future work will focus on optimally placing the oxyanion hole residues to more preferentially stabilize the transition state over the ground state.

尽管我们设计的丝氨酸水解酶的催化效率远高于之前报道的基于催化三联体的设计，但它们的速度仍然比天然水解酶慢几个数量级。一些实验结果明确指出了改进的方向。首先，ken\_t1在大约10个反应循环后失活，催化剂和与底物孵育的丝氨酸敲除体的质谱显示出稳定的酰化物种(图S12)，这表明水解 的设计仍然容易受到失活的影响，可能是由于活性位点中的非机制酰化事件，这在未来的设计工作中需要避免。其次，在后续设计轮次中使用严格的Chemnet过滤的三个设计(dad\_t1、charlie\_t2、ken\_t1)中，第二个侧链负离子孔残基的突变对活性的影响小于早期设计轮次以及与天然酶中类似突变的比较(图S10)。为了研究负离子孔的结构效应，我们对所有活跃设计进行了野生型和负离子孔丙氨酸敲除突变体的ChemNet预测。在super的情况下，Q71A的预测显示出酰化丝氨酸在AEI中的明显构象变化，使其与组氨酸的距离变长，为活性损失提供了结构解释(图S13)。相比之下，野生型和负离子孔敲除的预测在其他设计中是无法区分的，包括win和win的高活性重设计(图S13)。我们的分析表明，在整个设计轮次中实现的催化改进可能主要源于催化三联体的组织和中间体定位的改善；未来的工作将集中在最佳放置负离子孔残基，以更优先地稳定过渡态而非 基态。

# Acyltransferase activity of designed hydrolases

# 设计水解酶的酰基转移酶活性

Several native serine hydrolases exhibit promiscuous acyltransferase activity, reacting with small-molecule nucleophiles that compete with hydrolysis to break down the . Due to the long-lived nature of the in these designed hydrolases and the hydrophobicity of their substrate binding pockets, we hypothesized they may also catalyze acyl transfer to aromatic alcohols (Fig. S14). 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. S14). Incubation with PEA and substrate alone or with catalytic serine to alanine knockout mutants of win\_t4 and dad\_t1 did not exhibit increases in the rate, suggesting observed rate enhancements are enzyme dependent (Fig S14). Acyltransferase activity appears to be anti-correlated with : for example, win1 (4MU-Ac ) was inhibited by PEA, and win (4MU-Ac ) 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.

几种天然丝氨酸水解酶表现出多重酰基转移酶活性，与小分子亲核试剂反应，这些亲核试剂与水解反应竞争以分解 。由于这些设计的水解酶中 的持久性以及其底物结合口袋的疏水性，我们假设它们也可能催化酰基转移到芳香醇上(图 S14)。为了评估酰基转移，我们在酰基受体2-苯乙醇(PEA)的存在下，将设计与其对应的4MU-酯底物一起孵育。对于几个设计(win、win31、win\_t4和dad\_t1)，添加PEA显著提高了酯水解的速率，表明这些设计催化了酰基转移(图 S14)。与PEA和底物单独孵育或与win\_t4和dad\_t1的催化丝氨酸到丙氨酸敲除突变体一起孵育并未表现出速率的增加，表明观察到的速率增强是酶依赖性的(图 S14)。酰基转移酶活性似乎与 呈反相关:例如，win1(4MU-Ac )被PEA抑制，而win(4MU-Ac )在添加PEA后最大速率增加了3.6倍，这表明转酯化活性可能是由酰基受体的更紧密结合驱动的。

# Structural determinants of catalysis

# 催化的结构决定因素

The high structural conservation of catalytic geometry in native serine hydrolases suggests that it is close to optimal for catalysis , 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.

天然丝氨酸水解酶中催化几何结构的高度结构保守性表明，它接近于催化的最佳状态，但很难评估活性如何依赖于过渡态与催化丝氨酸、组氨酸和负氧离子孔功能团的相互作用的详细几何形状，因为催化残基的身份可以通过突变轻易改变，但系统地变化主链几何形状并不简单。相比之下，我们的从头构建方法采样了广泛的催化几何形状。为了研究活性位点几何形状和预组织如何影响催化活性，我们生成了所有812个实验表征设计的ChemNet集合，这些设计被分类为不活跃、FP探针标记、酰化和催化转化，涵盖了4MU-醋酸酯水解中每个反应步骤(包括设计轮次1-3和之前基于NTF2的设计)。以下特征与活性相关。

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. 5A, B). Designs that catalyzed turnover had Ser(Oy):His(Nε-Cε) bond angles that were more acute (median, all states = 94 ) than inactive designs (median, all states ), which were more similar to serine-histidine hydrogen bonds across the PDB ( ) (Fig. 5C). This acute H-bond is chemically intuitive given the reaction mechanism, in which 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 . This compromise in positioning is observed not only in our active designs but also in many of those found in nature .

Ser-His氢键的预先组织和弯曲增加与探针标记、酰化和周转的更高速率相关。所有能够催化周转的设计在所有四种状态下都显示出高度预组织的Ser-His氢键，而无活性的设计则常常表现出构象异构体的变化，导致相互作用的丧失(图5A, B)。催化周转的设计的Ser(Oy):His(Nε-Cε)键角比无活性设计更为锐利(中位数，所有状态=94 )，而无活性设计(中位数，所有状态 )则更类似于PDB中丝氨酸-组氨酸氢键( ) (图5C)。考虑到反应机制，这种锐利的氢键在化学上是直观的，因为这种几何形状允许组氨酸在不改变构象的情况下参与所有必要的质子转移，涉及丝氨酸、TI1中的离去基团氧和水解水 。这种定位的妥协不仅在我们的活性设计中观察到，也在自然界中发现的许多设计中存在 。

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 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. 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. 5G), increases the median Ser-His H-bond distance (Fig. 5G), and reduces the frequency that the Ser-His and oxyanion hole-acyl group H-bonds form (Fig. 5E). The same retention of the g- rotamer in the AEI is observed in native crystal structures . 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. 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.

在催化循环中，丝氨酸旋转异构体的几何形状与实验结果也有很强的相关性。对于显示酰化或转化的设计，我们发现丝氨酸在无配体状态下主要占据活性g-旋转异构体 。显示转化的设计在形成AEI时保留g-丝氨酸构象，但不可逆酰化的设计在AEI中切换到g+旋转异构体(图5H, I, J)。在这些设计中，g+丝氨酸旋转异构体在催化上无能，因为它导致的酰基构象阻碍了水分子与组氨酸的相互作用(图5G)，增加了丝氨酸-组氨酸氢键的中位距离(图5G)，并减少了丝氨酸-组氨酸和负离子孔-酰基氢键形成的频率(图5E)。在天然晶体结构中也观察到了AEI中g-旋转异构体的相同保留现象 。ChemNet分析还揭示，第二个负离子孔残基的存在有利于活性g-丝氨酸旋转异构体:那些仅有一个负离子孔氢键(来自丝氨酸亲核体的主链酰胺)的设计在酰化后从g-转变为g+，而具有两个负离子孔氢键的设计则主要占据g-丝氨酸旋转异构体(图5J，右)。因此，丝氨酸水解酶中的第二个负离子孔接触不仅稳定了过渡态，还可能有助于将中间体定向到催化有效的构象中。

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. 5K). Although both observed rotamers place T99 Oy within hydrogen bonding distance of the oxyanion, the designed rotamer-oxyanion dihedral angle much more closely matches the angles observed in native serine hydrolases, suggesting it is likely more optimal for selective transition state stabilization (see Methods). 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 win 1 displays a less pronounced rotameric shift, which leads to shorter serine-histidine hydrogen 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. 5K).

差异预组织也可能解释win、win1和win31系列中的活性趋势。对这些设计的晶体结构进行的ChemNet分析显示，在AEI状态下，更活跃的win1和win31样本在56%和60%的预测中采用了设计的T99氧阴离子孔旋转构象，而win从未采用这种旋转构象(图5K)。尽管观察到的两种旋转构象都将T99氧原子置于与氧阴离子形成氢键的距离内，但设计的旋转构象-氧阴离子二面角 与天然丝氨酸水解酶中观察到的角度更为接近，这表明它可能更适合选择性过渡态的稳定 (见方法)。我们还观察到在AEI状态下丝氨酸的旋转构象状态和酰基的预组织存在差异。win和win31在整个AEI集合中都占据了催化不利的g+旋转构象，而win1则表现出不太明显的旋转构象变化，这导致丝氨酸-组氨酸氢键距离更短(win1为2.8 Å，而win和win31为3.1 Å)。总体而言，win1，尤其是win31的酰基显示出明显低于win的构象异质性，这可能增加了组氨酸介导的水攻击的可能性(图5K)。

# Conclusions

# 结论

The substantial catalytic efficiencies of , the complexity of the active site geometry, and the accuracy of sidechain placement considerably surpass previous serine hydrolase computational design efforts despite the testing of a relatively small number of designs and complete omission of laboratory optimization. The folds of the designed catalysts are very different from those of natural serine hydrolases, demonstrating the ability of generative deep learning design methods to find completely new solutions to design challenges that differ from those found by natural evolution. 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 limited activation of the leaving groups and water (Fig. S15) . 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.

的显著催化效率、活性位点几何形状的复杂性以及侧链放置的准确性大大超过了之前的丝氨酸水解酶计算设计工作，尽管测试的设计数量相对较少，并且完全省略了实验室优化。所设计催化剂的折叠与天然丝氨酸水解酶的折叠截然不同，展示了生成性深度学习设计方法在设计挑战中找到与自然进化不同的全新解决方案的能力。之前基于催化三联体的设计努力未能实现多次转化；在某些情况下，例如我们初步的基于NTF2的设计，由于支架的限制，无法实现主链酰胺负离子孔，而在其他基于天然支架的设计中，组氨酸几何形状难以控制，这限制了离去基团和水的活化(图S15) 。从指定的活性位点向外生成主链的de novo生成结合RFdiffusion克服了这些限制，使几乎可以生成任何所需的催化几何形状。

Assessing design compatibility over the entire catalytic cycle has been a longstanding challenge in enzyme design. We show that the deep neural network ChemNet can rapidly generate ensembles for a series of reaction intermediates which directly assess preorganization, and provide structural insights that would otherwise require labor-intensive structural studies. For example, ChemNet revealed pervasive off-target conformational changes in the acyl-enzyme, which could be responsible for the failure to catalyze turnover for many previously designed esterases. The stochastic nature of ChemNet provides ensemble views of the energy landscapes around key reaction intermediates; the agreement we observe between ChemNet preorganization and experimental success rates suggests that such ensemble generation will be broadly useful for enzyme design moving forward.

评估整个催化周期的设计兼容性一直是酶设计中的一个长期挑战。我们展示了深度神经网络ChemNet能够快速生成一系列反应中间体的集合，这些集合直接评估了预先组织，并提供了结构洞察，这些洞察在其他情况下需要耗时的结构研究。例如，ChemNet揭示了酰基酶中普遍存在的非靶向构象变化，这可能是导致许多先前设计的酯酶未能催化转化的原因。ChemNet的随机特性提供了关键反应中间体周围能量景观的集合视图；我们观察到ChemNet预先组织与实验成功率之间的一致性，表明这种集合生成在未来的酶设计中将广泛有用。

While the designed catalysts described here are far more active than previous de novo designed serine hydrolases obtained by direct computation, they are still two to three orders of magnitude less efficient than native serine hydrolases, particularly in terms of turnover number. There are several possible explanations for the remaining activity gap: (1) the oxyanion hole identities and geometries differ slightly from those in native structures, which could reduce selective transition state stabilization, (2) the catalytic aspartate in the designs rarely participates in 2-3 additional hydrogen bonds (like those found in nature) which may limit its modulation of the catalytic histidine’s , and (3) the designed active sites are more buried than those of natural serine proteases, which could inhibit water entry into the active site for acylenzyme hydrolysis. Our de novo buildup approach using RFdiffusion coupled with ChemNet ensemble analysis to ensure design accuracy and preorganization should allow us to test all of these hypotheses by direct construction, which should further complement more traditional approaches based on structural examination and mutation of highly evolved native enzymes.

尽管这里描述的设计催化剂比通过直接计算获得的先前新设计的丝氨酸水解酶活性高得多，但它们的效率仍然比天然丝氨酸水解酶低两个到三个数量级，特别是在周转数方面。剩余活性差距的几个可能解释是:(1) 氧阴离子孔的身份和几何形状与天然结构略有不同，这可能减少选择性过渡态的稳定性，(2) 设计中的催化天冬氨酸很少参与2-3个额外的氢键(如自然界中所发现的)，这可能限制了其对催化组氨酸的调节，(3) 设计的活性位点比天然丝氨酸蛋白酶更为埋藏，这可能抑制水进入活性位点进行酰基酶水解。我们使用RFdiffusion结合ChemNet集成分析的自新构建方法，以确保设计的准确性和预组织，应该能够通过直接构建测试所有这些假设，这将进一步补充基于结构检查和高度进化的天然酶突变的更传统方法。

More generally, 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 in the near future.

更一般地说，我们预期利用RFdiffusion以亚埃级精度精确定位多个催化基团，并通过ChemNet评估复杂反应周期中的活性位点组织，将能够在不久的将来设计出多种新催化剂。

# Acknowledgments

# 致谢

We thank Luki Goldschmidt and Kandise VanWormer for maintaining the computational and wet lab resources at the Institute for Protein Design. We thank Anthony P. Green, Florence J. Hardy, and Donald Hilvert for helpful advice during the development of this project. We thank Florence J. Hardy and Madison A. Kennedy for reading and editing drafts of the manuscript.

我们感谢Luki Goldschmidt和Kandise VanWormer维护蛋白质设计研究所的计算和湿实验室资源。我们感谢Anthony P. Green、Florence J. Hardy和Donald Hilvert在本项目开发过程中提供的有益建议。我们感谢Florence J. Hardy和Madison A. Kennedy对手稿草稿的审阅和编辑。

# Funding

# 资金支持

This work was supported by the Howard Hughes Medical Institute (HHMI) (I.K., A.B, D.B.), the Open Philanthropy Project Improving Protein Design Fund (S.J.P, K.H.S., I.K., E.B., A.B.), the Washington Research Foundation (S.J.P), National Institutes of Health (NIH) and/or the National Institute of General Medical Sciences (NIGMS) Award (T32GM008268) (A.L.), a gift from Microsoft (I.A., D.J., D.B.), the Defense Threat Reduction Agency Grant (HDTRA1-19-1- 0003) (S.J.P., A.L.), and the Audacious Project at the Institute for Protein Design (A.K., E.B., A.B., A.L.). Crystallographic data were collected at the Advanced Light Source (ALS), which is supported by the Director, Office of Science, Office of 20 Basic Energy Sciences, and US Department of Energy under contract number DE-AC02-05CH11231.

本研究得到了霍华德·休斯医学研究所(Howard Hughes Medical Institute, HHMI)(I.K., A.B, D.B.)、开放慈善项目改善蛋白质设计基金(Open Philanthropy Project Improving Protein Design Fund)(S.J.P, K.H.S., I.K., E.B., A.B.)、华盛顿研究基金会(Washington Research Foundation)(S.J.P)、美国国立卫生研究院(National Institutes of Health, NIH)和/或美国国立普通医学科学研究所(National Institute of General Medical Sciences, NIGMS)奖项(T32GM008268)(A.L.)、微软的赠款(I.A., D.J., D.B.)、国防威胁减少局资助(Defense Threat Reduction Agency Grant)(HDTRA1-19-1-0003)(S.J.P., A.L.)以及蛋白质设计研究所的大胆项目(Audacious Project at the Institute for Protein Design)(A.K., E.B., A.B., A.L.)的支持。晶体学数据是在先进光源(Advanced Light Source, ALS)收集的，该设施由科学办公室(Office of Science)、基础能源科学办公室(Office of Basic Energy Sciences)和美国能源部(US Department of Energy)根据合同号DE-AC02-05CH11231提供支持。

# Author contributions

# 作者贡献

Conceived the study: A.L., S.J.P., and D.B; Trained chemnet: I.A. Conceived, implemented, and trained the models comprising all-atom CA RFdiffusion described here: D.J.; Implemented code to support training of all-atom RFdiffusion models: W.A.; Performed DFT calculations used to model the substrate geometry for design calculations: C.J.; Developed motif generation script: I.K.; Computationally designed serine hydrolases: A.L., S.J.P, A.S., A.H., and K.H.S.; Experimentally characterized serine hydrolases: A.L., S.J.P, K.H.S., A.S., A.H.; Prepared samples for crystallography: A.L. and S.J.P.; Performed crystallization and crystal preparation: E.B. and A.K. Performed data collection for crystal structures: A.B. and B.S.; Solved and refined crystal structures: A.L., S.J.P, and A.B. Wrote the manuscript: A.L., S.J.P, I.A., and D.B.; All authors revised and edited the manuscript. Supervision: D.B. and K.H. Methods Computational design of serine hydrolases

构思研究:A.L.、S.J.P. 和 D.B；训练化学网络:I.A. 构思、实施并训练了这里描述的全原子CA RF扩散模型:D.J.；实施代码以支持全原子RF扩散模型的训练:W.A.；执行用于建模基底几何形状的DFT计算以进行设计计算:C.J.；开发了基序生成脚本:I.K.；计算设计丝氨酸水解酶:A.L.、S.J.P、A.S.、A.H. 和 K.H.S.；实验表征丝氨酸水解酶:A.L.、S.J.P、K.H.S.、A.S.、A.H.；准备晶体学样品:A.L. 和 S.J.P.；执行结晶和晶体准备:E.B. 和 A.K.；执行晶体结构的数据收集:A.B. 和 B.S.；解决并精炼晶体结构:A.L.、S.J.P 和 A.B.；撰写手稿:A.L.、S.J.P、I.A. 和 D.B.；所有作者修订和编辑了手稿。监督:D.B. 和 K.H. 方法:丝氨酸水解酶的计算设计

# Motif generation

# 图案生成

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 1Ins, 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.

图案是在迭代过程中构建的。首先，根据参考文献7中的几何形状，将处于过渡态几何形状的底物旋转异构体(4MU-Bu或4MU-Ac)放置在与一个由三个氨基酸残基组成的丝氨酸和来自两个天然丝氨酸水解酶晶体结构(1scn，残基220-222和1Ins，残基347-349)中的局部负离子孔相关的位置上，其中除丝氨酸外的所有残基均被突变为丙氨酸。底物酯基的过渡态几何形状通过DFT几何优化(B3LYP-D3(BJ)/6-31G(d))确定。接下来，在由丙氨酸夹住的三个氨基酸螺旋或链段的突变体上，围绕催化丝氨酸采样组氨酸的位置和旋转异构体，并过滤出那些同时与催化丝氨酸和底物离去基团氧形成氢键的结构。这个过程产生了108个独特的第一轮图案。对于第三轮图案，最初以类似的方式添加了天冬氨酸或谷氨酸残基和第二个负离子孔氢键，使用氢键构象和旋转异构体的几何采样。然而，从这些图案生成的骨架的AF2成功率极低，可能是由于生成了不兼容的骨架构象组合。为了确保剩余的催化残基突变体被放置在可实现的几何形状中，我们使用简单的底物-丝氨酸-组氨酸图案作为输入生成了10,000个骨架，并使用Rosetta搜索这些骨架在次级结构上可以容纳天冬氨酸或谷氨酸三联残基以与组氨酸形成氢键的位置。然后提取这些突变体，在最后一步中，重复相同的过程以生成第二个负离子孔的突变体，考虑所有氢键供体侧链，最终产生了2238个独特的第三轮图案，具有丝氨酸-组氨酸-天冬氨酸/谷氨酸催化三联体，以及丝氨酸/苏氨酸/酪氨酸/组氨酸/色氨酸负离子孔。

# Backbone generation

# 骨架生成

See supplemental methods for a detailed description of CA diffusion, which was employed to generate backbones to scaffold motifs.

有关CA扩散的详细描述，请参见补充方法，该方法用于生成骨架以支撑图案。

# Sequence design

# 序列设计

We performed three cycles of LigandMPNN and Rosetta FastRelax 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 within of the active site. Catalytic residues were kept fixed and Rosetta enzyme constraints 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 for distances and for angles and dihedrals.

我们进行了三轮LigandMPNN 和Rosetta FastRelax 以设计从RFdiffusion生成的骨架的序列。为了促进与催化组氨酸(第一轮图案)和催化天冬氨酸/谷氨酸(第三轮图案)形成氢键接触，LigandMPNN在选择残基时使用的对数概率偏向于所有在活性位点 范围内的极性氨基酸。催化残基保持固定，并在放松步骤中应用Rosetta酶约束 以保持设计周期中的催化几何形状。每个氢键相互作用的约束使用起始图案几何形状定义，距离的容差为 ，角度和二面角的容差为 。

# Filtering

# 过滤

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 passing designs were used as input to for single sequence structure prediction. was run using model 4 with three recycles. Designs were filtered for a global Ca RMSD < 1.5 Å, pLDDT > 75, and catalytic residue Ca RMSD < 1.0 Å. 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 Supplemental Methods) were measured. To analyze native hydrolases with Chemnet, a set of native crystal structures was collected (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 above-described process was applied.

在序列设计之后，设计通过FastRelax后的动机催化几何形状的重现和结合位点与底物的形状互补性进行了筛选。通过筛选的序列被用作 的单序列结构预测输入。使用模型4和三次回收运行 。设计的筛选标准为全局Ca RMSD < 1.5 Å，pLDDT > 75，以及催化残基Ca RMSD < 1.0 Å。通过AF2筛选的设计随后使用ChemNet进行了分析。ChemNet是一个去噪神经网络，经过训练以从PDB中的高分辨率和中等分辨率X射线和电子显微镜结构中重现部分损坏输入结构的正确原子位置，前提是从一开始就知道被建模系统的所有化学信息。ChemNet的预测是针对与活性位点最接近的600个原子的空间裁剪进行的。网络的输入包括裁剪内的蛋白质主链坐标和氨基酸序列，侧链坐标在各自的C-alpha原子周围随机初始化。对于没有晶体结构的蛋白质，使用AF2模型。对于每个设计的蛋白质，我们建模了5种反应状态，这些状态在催化丝氨酸在反应过程中经历的化学修饰上有所不同:(1) apo，(2) 底物结合，(3) 四面体中间体1，(4) 酰基酶中间体，以及(5) 四面体中间体2。我们使用50个不同的种子为每种反应状态(apo、底物结合、TI1、AEI和TI2)生成50个ChemNet模型的集合。然后分别分析这些集合中活性位点氢键模式的保留情况。对于每个集合中的50个预测，测量该步骤中每个氢键相互作用的几何形状(见补充方法)。为了使用ChemNet分析天然水解酶，收集了一组天然晶体结构 (PDB ID: 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)，识别活性位点位置，并应用上述过程。

# In-gel fluorescence screening with activity-based probes

# 基于活性探针的凝胶内荧光筛选

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 of LB supplemented with 50 kanamycin. For expression, of overnight was used to inoculate of media and grown for 1.5 hours at on a Heidolph shaker and then of IPTG was added and cultures were shaken at for an additional 3 hours. Cultures were centrifuged at 4000g for 10 minutes and supernatant removed. Cell pellets were resuspended in mM HEPES (pH 7.4), containing 50 mM NaCl, 0.1 mg/mL lysozyme, and 0.01 mg/mL DNasel. After 15 minutes, lysates were frozen in liquid nitrogen and subsequently thawed. of lysate was incubated with 1 M FP-TAMRA probe ( of stock in lysis buffer) for 1 hour at room temperature before quenching using Laemmli sample buffer. Labeled samples were heated at for 5 minutes and 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

从IDT订购的编码设计蛋白的DNA作为eblocks，并克隆到包含C末端SNAC和六组氨酸标签的载体LM627(addgene)中。将得到的质粒转化到BL21(DE3)细胞中，并在补充50μg/mL卡那霉素的LB培养基中过夜培养。为了表达，过夜培养物的部分用于接种新的培养基，并在Heidolph摇床上以 的温度培养1.5小时，然后添加 IPTG的 ，并在 下继续摇动培养3小时。培养物在4000g下离心10分钟，去除上清液。细胞沉淀重悬于含有50 mM NaCl、0.1 mg/mL溶菌酶和0.01 mg/mL DNasel的 mM HEPES(pH 7.4)中。15分钟后，裂解液在液氮中冷冻，然后解冻。将 的裂解液与1 M FP-TAMRA探针( 的 库存溶液在裂解缓冲液中)在室温下孵育1小时，然后使用 Laemmli样品缓冲液终止反应。标记样品在 下加热5分钟，然后将每个样品的 分离在BioRad AnykD Criterion预制凝胶上，并使用LI-COR Odyssey M成像仪可视化凝胶中的荧光。随后，凝胶用考马斯亮蓝染色，以可视化每个设计的分子量和表达水平。裂解液筛选

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. Cultures were grown overnight at scale in 96-well plates on a Heidolph shaker at and . For expression, of the overnight cultures were used to inoculate of autoinduction media in 96-well round bottom plates and incubated at and 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 of lysis buffer (20 mM HEPES,50 mM NaCl,0.01 mg/mL DNAsel,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 . Plates were centrifuged at for 30 minutes and supernatant collected for screening. For activity screening,4or 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.

从IDT订购的编码设计蛋白的DNA作为eblocks并克隆到包含C末端mScarlet-i3融合和His标签的载体pCOOL1中。在 规模的96孔板上，在Heidolph摇床上以 和 的条件过夜培养。为了表达，使用过夜培养物的 接种 的自诱导培养基，在96孔圆底板中培养，并在 和 下孵育约24小时。培养物以4000g离心10分钟，弃去上清液，然后用缓冲液(20 mM HEPES，50 mM NaCl，pH 7.4)洗涤，并在室温下以1300 rpm在Heidolph摇床上孵育5分钟以重悬。再次以4000g离心10分钟，弃去上清液。为了裂解，将细胞沉淀重悬于 的裂解缓冲液(20 mM HEPES，50 mM NaCl，0.01 mg/mL DNAsel，0.01 mg/mL 溶菌酶，1 mM EDTA，0.1% Triton X-100)中，并在Heidolph摇床上以1300 rpm和 孵育2小时。以 离心30分钟，收集上清液进行筛选。为了活性筛选，将4或 的裂解液分装到微量滴定板中，通过添加含有111.1 µM 4MU-Ac或4MU-Bu、20 mM HEPES、50 mM NaCl、pH 7.4、5% DMSO的缓冲液36或54 µL来启动反应。

# Protein expression and purification

# 蛋白质表达与纯化

Genes encoding the designed proteins were ordered from IDT as eblocks and cloned into vector LM627 (addgene) (ref). Resulting plasmid was transformed into BL21(DE3) cells and grown overnight in of supplemented with kanamycin, after which of overnight was used to inoculate of autoinduction media, which was grown 4-6 hours at and then overnight at . Cultures were spun down at for 15 minutes, and supernatant decanted. Cell pellets were resuspended in of cold wash buffer ( imidazole,500 mM NaCl, 50 mM sodium phosphate, pH 7.4) with 1mg/mL lysozyme and 0.1mg/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 for 30 minutes and the supernatant was applied to 1 of Ni-NTA resin equilibrated with wash buffer. The resin was subsequently washed with 15 of wash buffer 3 times and once with of elution buffer imidazole,500 sodium phosphate, pH 7.4) followed by elution with elution buffer. The eluate was purified by size-exclusion chromatography on a Superdex 75 Increase 10/300 GL with running buffer of HEPES, , . 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 LC-MS.

编码设计蛋白的基因从IDT以eblocks形式订购，并克隆到载体LM627(addgene)(参考文献)。所得质粒转化到BL21(DE3)细胞中，并在补充有 卡那霉素的 的 中过夜培养，随后用过夜培养的 接种 的自诱导培养基，在 下培养4-6小时，然后在 下过夜培养。培养物在 下离心15分钟，弃去上清液。细胞沉淀重悬于含有1mg/mL溶菌酶和0.1mg/mL DNAse I的冷洗涤缓冲液 ( 咪唑，500 mM NaCl，50 mM磷酸钠，pH 7.4)中。细胞悬液在冰上以80%幅度超声处理2.5分钟，10秒开，10秒关。所得裂解液在 下离心30分钟，上清液被应用于用洗涤缓冲液平衡的1 Ni-NTA树脂。随后，树脂用15 洗涤缓冲液洗涤3次，并用 洗脱缓冲液 (咪唑，500 磷酸钠，pH 7.4)洗涤一次，然后用 洗脱缓冲液洗脱。洗脱液通过在Superdex 75 Increase 10/300 GL上进行尺寸排阻色谱进行纯化，运行缓冲液为 HEPES， ， 。样品要么立即用于后续实验，要么在液氮中快速冷冻并储存于-80°C。蛋白质分子量通过LC-MS确认。

# Kinetic analysis

# 动力学分析

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 4MU-PhAc. Kinetic screens were either performed in reaction volumes in 96-well half area plates or reaction volume in 96-well full-area plates. Protein and substrate were prepared in HEPES, DMSO. Either4or 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 was monitored continuously (excitation , emission ). 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 . For those designs that displayed a clear burst phase followed by a slower steady-state rate, straight-line fits of the steady-state velocities were used to determine Michaelis-Menten catalytic parameters.

为了表征从凝胶荧光和裂解液筛选中识别的催化转化命中，我们将纯化的蛋白质样品与荧光底物4MU-Ac、4MU-Bu和4MU-PhAc孵育。动力学筛选在96孔半面积板的 反应体积或96孔全面积板的 反应体积中进行。蛋白质和底物在 HEPES和 DMSO中准备。将酶的4或 加入微量滴定板中，反应通过加入底物(36或54 μL)启动。荧光产物 的生成被持续监测(激发 ，发射 )。使用自定义脚本对结果数据进行分析(见计算方法)。在观察到单次转化活性的情况下，初始速度用于确定 。对于那些显示出明显的突发阶段后跟随较慢的稳态速率的设计，使用稳态速度的直线拟合来确定米哈利斯-门腾催化参数。

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 .

为了确定在测定缓冲液(20 mM HEPES，50 mM NaCl，pH 7.4，5% DMSO)中的非催化反应速率，底物仅在缓冲液中稀释，并在多个底物浓度下测定速率，之后通过拟合[S]与速率的关系，使用形式为速率 的方程确定速率。

# Crystallography

# 晶体学

Proteins for crystallography were prepared as described above, but SEC was done with SNAC tag cleavage buffer . After SEC, protein eluate was incubated with guanidinium hydrochloride and 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 HEPES, as the running buffer. Samples were concentrated and stored at 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 , phased and refined with Phenix , and model building performed with . Coordinates are deposited in the PDB with PDB IDs of 9DED (slap215.8), 9DEE (super), 9DEF (win), 9DEG (win1), and 9DEH (win31).

为晶体学准备的蛋白质如上所述，但SEC使用了SNAC标签切割缓冲液 。在SEC之后，蛋白质洗脱液与 氯化胍和 在室温下孵育过夜，以去除C端His标签。SNAC切割反应应用于用洗涤缓冲液平衡的镍柱，以去除任何未切割的产物，得到的洗脱液应用于Superdex 75 Increase 10/300 GL柱，运行缓冲液为 HEPES， 。样品被浓缩并储存于 或立即用于结晶。结晶筛选使用STP Labtech的Mosquito LCP进行，得到的晶体直接从筛选板上收集。每种设计的结晶条件如下:slap215.8(15 mg/mL)在0.1 M Bis-Tris pH 5.5中，25%(w/v)PEG 3350；super(50 mg/mL)在0.2 M氟化钾中，20%(w/v)PEG 3350；win(42 mg/mL)在0.1 M醋酸钠pH 4.6中，8%(w/v)PEG 4000；win1(54 mg/mL)在60% v/v Tacsimate pH 7.0中；win31(60 mg/mL)在0.2 M二铵酒石酸盐和20%(w/v)PEG 3350中。数据使用XDS 处理，使用Phenix 进行相位和精修，模型构建使用 。坐标已存入PDB，PDB ID为9DED(slap215.8)，9DEE(super)，9DEF(win)，9DEG(win1)，和9DEH(win31)。

# Mass spectrometry

# 质谱

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 ) were incubated overnight with substrate in SEC running buffer at room temperature prior to mass spectrometry analysis.

通过在Agilent G6230B TOF上使用AdvanceBio RP-Desalting柱进行脱盐后，采用反相液相色谱/质谱法获得了蛋白质样品的完整质谱。使用总熵算法进行解卷积，采用Bioconfirm软件。在某些情况下，蛋白质样品(1 )在质谱分析之前与底物 在室温下的SEC运行缓冲液中孵育过夜。

# Acyltransferase activity screening

# 酰基转移酶活性筛选

Enzymes were incubated with cognate substrate in assay buffer in the presence of varying concentrations of acyl acceptor, PEA(50,25,12.5,6.3,3.1,1.6,0.8,0mM), 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.

酶 在不同浓度的酰基受体PEA(50,25,12.5,6.3,3.1,1.6,0.8,0mM)存在下，与 相应底物在测定缓冲液中孵育，并监测底物水解反应1小时，如上所述。通过拟合每个进程曲线的开始部分获得初始速度，并将其除以在没有PEA情况下的水解速率，以获得相对水解速率。

Structural similarity search of the PDB and AFDB

PDB和AFDB的结构相似性搜索

To assess the structural novelty of our designed enzymes, we used TMalign to compare our crystal structures against the Protein DataBank (PDB) and AlphaFold database . 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 (version 4) proteins were fetched April, 2024. We report the average TM-score for the top hit.

为了评估我们设计的酶的结构新颖性，我们使用TMalign 将我们的晶体结构与蛋白质数据银行(PDB)和AlphaFold数据库 进行比较。我们在2024年4月4日下载了所有通过X射线晶体学或冷冻电子显微镜解决的PDB中的蛋白质聚合物，并从每个条目中提取了所有蛋白质链。AFDB50 (版本4)蛋白质的模型于2024年4月获取。我们报告了最佳匹配的平均TM-score。

# References

# 参考文献

1. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089-1100 (2023).

1. Watson, J. L. 等. 通过RFdiffusion进行蛋白质结构和功能的全新设计. 自然 620, 1089-1100 (2023).

2. Lovelock, S. L. et al. The road to fully programmable protein catalysis. Nature 606, 49-58 (2022).

2. Lovelock, S. L. 等. 通往完全可编程蛋白质催化的道路. 自然 606, 49-58 (2022).

3. Röthlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. Nature 453, 190-195 (2008).

3. Röthlisberger, D. 等. 通过计算酶设计的Kemp消除催化剂. 自然 453, 190-195 (2008).

4. Jiang, L. et al. De novo computational design of retro-aldol enzymes. Science 319, 1387- 1391 (2008).

4. Jiang, L. 等. 新型计算设计的逆醛醇酶. 科学 319, 1387-1391 (2008).

5. Siegel, J. B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329, 309–313 (2010).

5. Siegel, J. B. 等. 立体选择性双分子Diels-Alder反应的酶催化剂的计算设计. 科学 329, 309–313 (2010).

6. Privett, H. K. et al. Iterative approach to computational enzyme design. Proc. Natl. Acad. Sci. U. S. A. 109, 3790-3795 (2012).

6. Privett, H. K. 等. 计算酶设计的迭代方法. 美国国家科学院院刊 109, 3790-3795 (2012).

7. Richter, F. et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. J. Am. Chem. Soc. 134, 16197-16206 (2012).

7. Richter, F. 等. 用于酯水解的催化二元体和氧阴离子孔的计算设计. 美国化学会志 134, 16197-16206 (2012).

8. Rajagopalan, S. et al. Design of activated serine-containing catalytic triads with atomic level accuracy. Nat. Chem. Biol. 10, 386-391 (2014).

8. Rajagopalan, S. 等. 具有原子级精度的活化丝氨酸催化三元体的设计. 自然化学生物学 10, 386-391 (2014).

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

9. Kiss, G., Röthlisberger, D., Baker, D. 和 Houk, K. N. 酶设计的评估与排名。蛋白质科学 19, 1760-1773 (2010)。

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

10. Fleishman, S. J., Khare, S. D., Koga, N. 和 Baker, D. 原生蛋白及其复合物结构中的侧链塑性受限。蛋白质科学 20, 753-757 (2011)。

11. Bunzel, H. A. et al. Evolution of dynamical networks enhances catalysis in a designer enzyme. Nat. Chem. 13, 1017-1022 (2021).

11. Bunzel, H. A. 等. 动态网络的演化增强了设计酶的催化作用。自然化学 13, 1017-1022 (2021)。

12. Frushicheva, M. P., Cao, J., Chu, Z. T. & Warshel, A. 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).

12. Frushicheva, M. P., Cao, J., Chu, Z. T. 和 Warshel, A. 通过模拟人工凯姆消除酶中的催化反应探索合理酶设计中的挑战。美国国家科学院院刊 107, 16869-16874 (2010).

13. Krishna, R. et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science 384, eadl2528 (2024).

13. Krishna, R. 等. 使用 RoseTTAFold 全原子进行广义生物分子建模和设计。科学 384, eadl2528 (2024).

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

14. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. 和 Woolfson, D. N. 在完全新构建的蛋白质框架中安装水解活性。自然化学 8, 837-844 (2016).

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

15. Bolon, D. N. 和 Mayo, S. L. 通过计算设计获得酶样蛋白质。美国国家科学院院刊 98, 14274-14279 (2001).

16. Burke, A. J. et al. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. Nature 570, 219-223 (2019).

16. Burke, A. J. 等. 具有非典型有机催化机制的酶的设计与演化. 自然 570, 219-223 (2019).

17. Studer, S. et al. Evolution of a highly active and enantiospecific metalloenzyme from short peptides. Science 362, 1285-1288 (2018).

17. Studer, S. 等. 从短肽中演化出一种高活性和对映选择性的金属酶. 科学 362, 1285-1288 (2018).

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

18. Der, B. S., Edwards, D. R. & Kuhlman, B. 通过新型锌介导的蛋白质界面催化: 对自然酶演化和理性酶工程的启示. 生物化学 51, 3933-3940 (2012).

19. Moroz, Y. S. et al. 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).

19. Moroz, Y. S. 等. 老蛋白的新把戏: 非酶蛋白中的单一突变产生多种酶活性. 美国化学会志 137, 14905–14911 (2015).

20. Blow, D. M. Structure and mechanism of chymotrypsin. Acc. Chem. Res. 9, 145-152 (1976).

20. Blow, D. M. 胰蛋白酶的结构与机制. 化学研究进展 9, 145-152 (1976).

21. Carter, P. & Wells, J. A. Dissecting the catalytic triad of a serine protease. Nature 332, 564-568 (1988).

21. 卡特(P. Carter)与威尔斯(J. A. Wells)。剖析丝氨酸蛋白酶的催化三联体。《自然》332, 564-568 (1988)。

22. Carter, P. & Wells, J. A. Functional interaction among catalytic residues in subtilisin BPN’. Proteins 7, 335-342 (1990).

22. 卡特(P. Carter)与威尔斯(J. A. Wells)。亚硫酸酶BPN’中催化残基的功能相互作用。《蛋白质》7, 335-342 (1990)。

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

23. 波尔加尔(L. Polgár)。丝氨酸肽酶的催化三联体。《细胞与分子生命科学》62, 2161-2172 (2005)。

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

24. 布莱恩(P. Bryan)、潘托利亚诺(M. W. Pantoliano)、奎尔(S. G. Quill)、肖(H. Y. Hsiao)与普洛斯(T. Poulos)。定点突变及其在亚硫酸酶中的氧阴离子孔的作用。《美国国家科学院院刊》83, 3743-3745 (1986)。

25. Corey, D. R. & Craik, C. S. 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).

25. Corey, D. R. & Craik, C. S. 通过突变胰蛋白酶的催化三联体研究肽水解的最低要求。美国化学学会杂志 114, 1784-1790 (1992)。

26. Zerner, B., Bond, R. P. M. & Bender, M. L. Kinetic Evidence for the Formation of Acyl-Enzyme Intermediates in the α-Chymotrypsin-Catalyzed Hydrolyses of Specific Substrates. J. Am. Chem. Soc. 86, 3674-3679 (1964).

26. Zerner, B., Bond, R. P. M. & Bender, M. L. 关于特定底物在α-胰蛋白酶催化水解中形成酰基-酶中间体的动力学证据。美国化学学会杂志 86, 3674-3679 (1964)。

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

27. Kraut, J. 丝氨酸蛋白酶:结构与催化机制。生物化学年鉴 46, 331-358 (1977)。

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

28. Frey, P. A. & Hegeman, A. D. 酶促反应机制。(牛津大学出版社，2007)。

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

29. Walsh, C. 酶促反应机制。(W. H. Freeman，1979)。

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

30. Jencks, W. P. 化学与酶学中的催化作用。(Courier Corporation, 1987).

31. Hedstrom, L. Serine protease mechanism and specificity. Chem. Rev. 102, 4501-4524 (2002).

31. Hedstrom, L. 丙氨酸蛋白酶的机制与特异性。化学评论 102, 4501-4524 (2002).

32. Smith, A. J. T. et al. 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).

32. Smith, A. J. T. 等. 酶活性位点的结构重组与预组装:多步丙氨酸酯酶反应周期中实验与理论理想活性位点几何形状的比较。美国化学学会杂志 130, 15361–15373 (2008).

33. Du, S. et al. Conformational Ensembles Reveal the Origins of Serine Protease Catalysis. bioRxiv 2024.02.28.582624 (2024) doi:10.1101/2024.02.28.582624.

33. Du, S. 等. 构象集合揭示丙氨酸蛋白酶催化的起源。bioRxiv 2024.02.28.582624 (2024) doi:10.1101/2024.02.28.582624.

34. Radisky, E. S., Lee, J. M., Lu, C.-J. K. & Koshland, D. E., Jr. Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates.

34. Radisky, E. S., Lee, J. M., Lu, C.-J. K. & Koshland, D. E., Jr. 从胰蛋白酶反应中间体的原子分辨率结构中获得的丙氨酸蛋白酶机制的见解。

675

676

678

679

681

682 Proc. Natl. Acad. Sci. U. S. A. 103, 6835-6840 (2006).

682 美国国家科学院院刊 103, 6835-6840 (2006).

35. Yeh, A. H.-W. et al. De novo design of luciferases using deep learning. Nature 614, 774- 780 (2023).

35. Yeh, A. H.-W. 等. 使用深度学习进行荧光素酶的全新设计. 自然 614, 774-780 (2023).

36. Dauparas, J. et al. Atomic context-conditioned protein sequence design using LigandMPNN. bioRxiv 2023.12.22.573103 (2023) doi:10.1101/2023.12.22.573103.

36. Dauparas, J. 等. 使用LigandMPNN的原子上下文条件蛋白序列设计. bioRxiv 2023.12.22.573103 (2023) doi:10.1101/2023.12.22.573103.

37. Das, R. & Baker, D. Macromolecular modeling with rosetta. Annu. Rev. Biochem. 77, 363- 382 (2008).

37. Das, R. & Baker, D. 使用Rosetta的宏观分子建模. 生物化学年鉴 77, 363-382 (2008).

38. Bennett, N. R. et al. Improving de novo protein binder design with deep learning. Nat. Commun. 14, 2625 (2023).

38. Bennett, N. R. 等. 通过深度学习改善全新蛋白结合体设计. 自然通讯 14, 2625 (2023).

39. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021).

39. Jumper, J. 等. 使用AlphaFold进行高精度蛋白质结构预测. 自然 596, 583-589 (2021).

40. Müller, H. et al. Sequence based prediction of promiscuous acyltransferase activity in hydrolases. Angew. Chem. Weinheim Bergstr. Ger. 132, 11704-11709 (2020).

40. Müller, H. 等. 基于序列的水解酶中多重酰基转移酶活性的预测. 应用化学. 维尔海姆 贝格斯特. 德国. 132, 11704-11709 (2020).

41. Buller, A. R. & Townsend, C. A. 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).

41. Buller, A. R. & Townsend, C. A. 对蛋白酶结构、酶酰化和催化三联体的内在进化约束。美国国家科学院院刊 110, E653-61 (2013).

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

42. Zakharova, E., Horvath, M. P. & Goldenberg, D. P. 准备重新合成肽键的丝氨酸蛋白酶的结构。美国国家科学院院刊 106, 11034-11039 (2009).

43. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. Trends Biochem. Sci. 23, 347-352 (1998).

43. Dodson, G. & Wlodawer, A. 催化三联体及其相关物。生物化学科学趋势 23, 347-352 (1998).

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

44. Simón, L. & Goodman, J. M. 氢键稳定在氧阴离子孔中的作用:从二维到三维的飞跃。有机生物化学 10, 1905-1913 (2012).

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

45. Simón, L. & Goodman, J. M. 通过氢键的酶催化:过氧阴离子孔中过渡态结合与底物结合之间的平衡。J. Org. Chem. 75, 1831-1840 (2010)。

46. Tyka, M. D. et al. Alternate states of proteins revealed by detailed energy landscape bioRxiv preprint doi: https://doi.org/10.1101/2024.08.29.610411; this version posted August 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

46. Tyka, M. D. 等。通过详细的能量景观揭示蛋白质的替代状态。bioRxiv预印本 doi: https://doi.org/10.1101/2024.08.29.610411；此版本于2024年8月30日发布。该预印本的版权持有者(未经过同行评审)为作者/资助者，已授予bioRxiv永久展示该预印本的许可。该文档根据CC-BY-ND 4.0国际许可协议提供。

708 mapping. J. Mol. Biol. 405, 607-618 (2011).

708映射。J. Mol. Biol. 405, 607-618 (2011)。

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

47. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. 使用Rosetta3进行新型酶设计。PLoS One 6, e19230 (2011)。

48. Zanghellini, A. et al. New algorithms and an in silico benchmark for computational enzyme design. Protein Sci. 15, 2785-2794 (2006).

48. Zanghellini, A. 等. 新算法和计算酶设计的计算基准. 蛋白质科学. 15, 2785-2794 (2006).

49. Dang, B. et al. SNAC-tag for sequence-specific chemical protein cleavage. Nat. Methods 16, 319–322 (2019).

49. Dang, B. 等. SNAC标签用于序列特异性的化学蛋白裂解. 自然. 方法 16, 319–322 (2019).

50. Kabsch, W. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125-132 (2010).

50. Kabsch, W. XDS. 晶体学学报 D 生物晶体学. 66, 125-132 (2010).

51. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75, 861-877 (2019).

51. Liebschner, D. 等. 使用X射线、中子和电子进行大分子结构测定:Phenix的最新进展. 晶体学学报 D 结构生物学 75, 861-877 (2019).

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

52. Emsley, P. & Cowtan, K. Coot: 分子图形的模型构建工具. 晶体学学报 D 生物晶体学. 60, 2126-2132 (2004).

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

53. Zhang, Y. 和 Skolnick, J. TM-align:一种基于TM-score的蛋白质结构比对算法。核酸研究。33, 2302-2309 (2005)。

54. Varadi, M. et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 52, D368-D375 (2024).

54. Varadi, M. 等。2024年AlphaFold蛋白质结构数据库:为超过2.14亿个蛋白质序列提供结构覆盖。核酸研究。52, D368-D375 (2024)。

55. van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. 42, 243-246 (2024). bioRxiv preprint doi: https://doi.org/10.1101/2024.08.29.610411; this version posted August 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license. 724

55. van Kempen, M. 等。使用Foldseek进行快速准确的蛋白质结构搜索。自然生物技术。42, 243-246 (2024)。bioRxiv预印本doi:https://doi.org/10.1101/2024.08.29.610411；此版本于2024年8月30日发布。该预印本的版权持有者(未经过同行评审的版本)为作者/资助者，已授予bioRxiv永久展示该预印本的许可。根据CC-BY-ND 4.0国际许可协议提供。724

740

743

744 745 available under aCC-BY-ND 4.0 International license. 746

744 745 根据CC-BY-ND 4.0国际许可协议提供。746

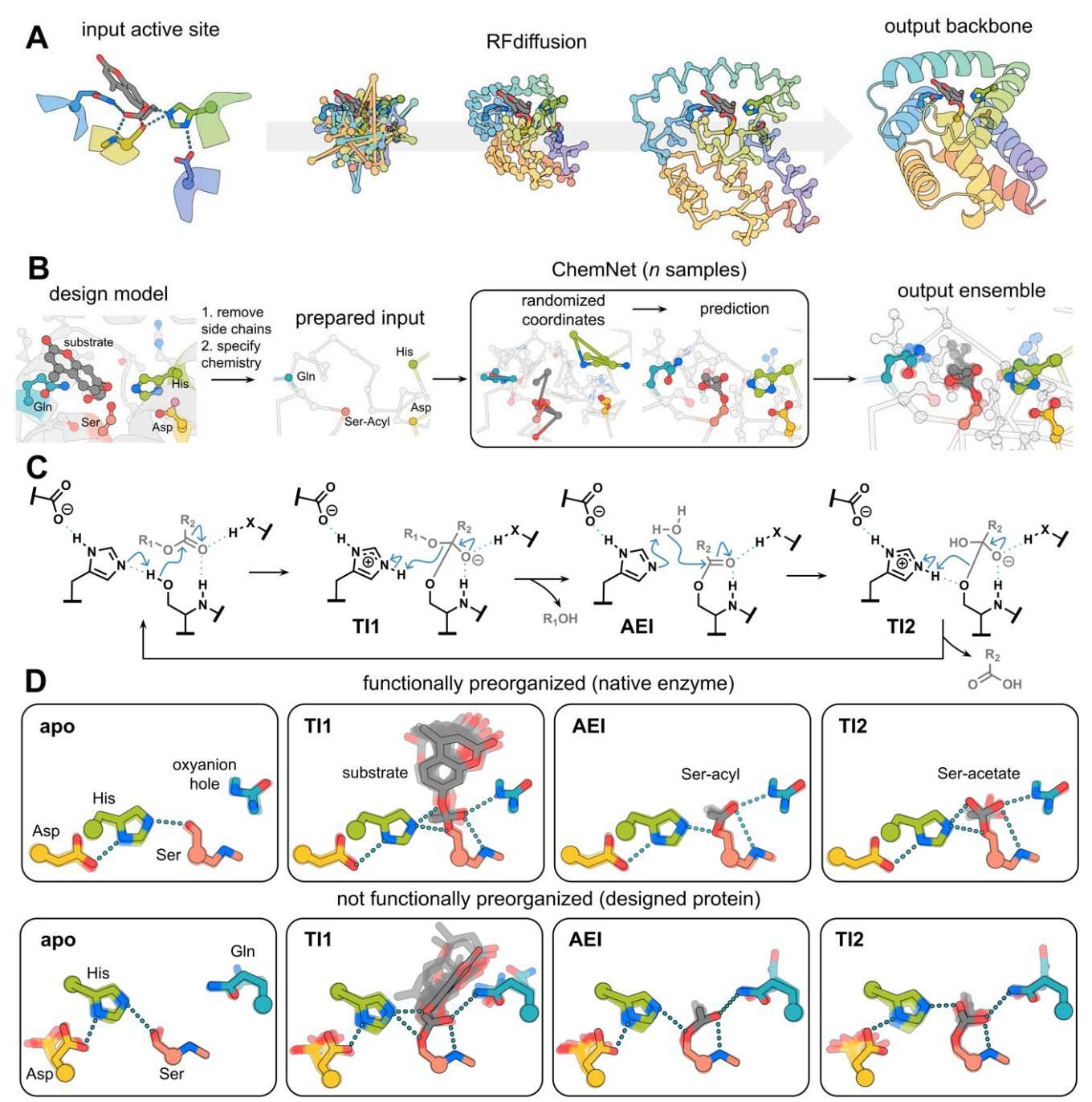


Figure 1. Design methods. (A) Active site specific backbone generation with RFdiffusion. Given the geometry of a possible active site configuration, RFdiffusion denoising trajectories generate backbone coordinates which scaffold the site. (B) Generation of active site ensembles with ChemNet. The coordinates 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).

图1. 设计方法。(A) 使用RFdiffusion生成特定活性位点的骨架。根据可能的活性位点配置的几何形状，RFdiffusion去噪轨迹生成支撑该位点的骨架坐标。(B) 使用ChemNet生成活性位点集合。围绕活性位点的侧链坐标以及反应步骤中任何结合的小分子的坐标被随机化，并进行n次采样以生成预测集合。(C) 丙氨酸水解酶的酯水解机制。(D) ChemNet集合在4MU-Ac水解反应坐标上不同状态的表现，针对一种天然的丙氨酸水解酶(上，PDB: 1IVY)和一种设计的丙氨酸水解酶(下，josie)。

bioRxiv preprint doi: https://doi.org/10.1101/2024.08.29.610411; this version posted August 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made 747 748 available under aCC-BY-ND 4.0 International license. 749

bioRxiv 预印本 doi: https://doi.org/10.1101/2024.08.29.610411; 此版本于2024年8月30日发布。该预印本的版权持有者(未经过同行评审认证)为作者/资助者，已授予bioRxiv永久展示该预印本的许可。该文档根据CC-BY-ND 4.0国际许可协议提供。

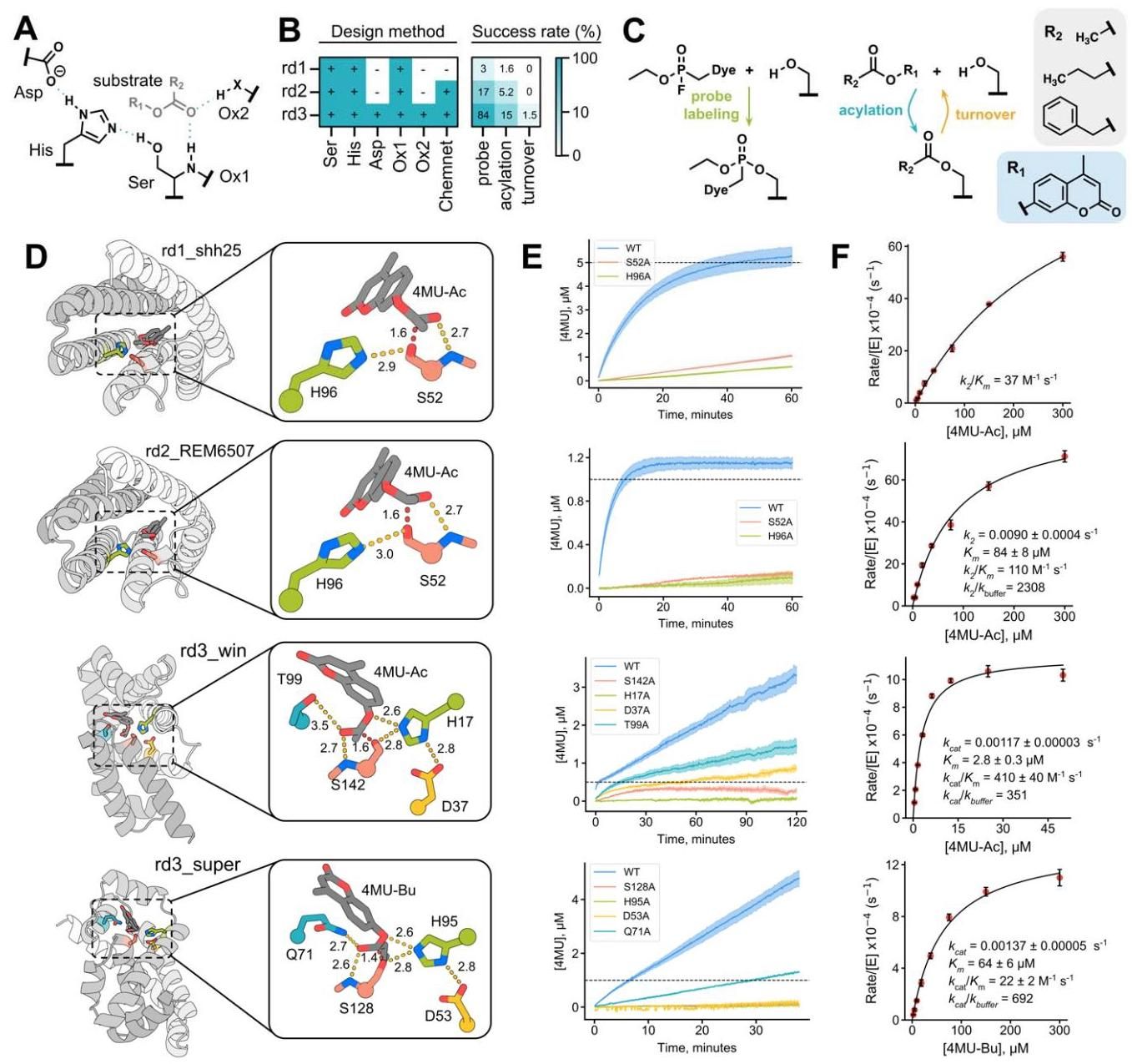


Figure 2. Functional characterization of designed serine hydrolases. (A) Chemical schematic of a serine hydrolase active site. (B) Summary of design method and experimental success rate for probe labeling, single turnover acylation, and catalytic turnover for each design round. (C) Chemical schematic 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. (F) Michaelis-Menten plots derived from initial (rd1, rd2) or steady state velocities (rd3).

图2. 设计的丝氨酸水解酶的功能特征。 (A) 丝氨酸水解酶活性位点的化学示意图。 (B) 探针标记、单次转化和每轮设计的催化转化的设计方法和实验成功率总结。 (C) 描述探针标记、酰化和催化转化的化学示意图。 (D) 丝氨酸水解酶设计模型的折叠(左)和活性位点(右)。 (E) 父设计和催化残基敲除的反应进程曲线。虚线表示酶浓度。 (F) 从初始(rd1, rd2)或稳态速度(rd3)得出的米氏-门捷列夫图。

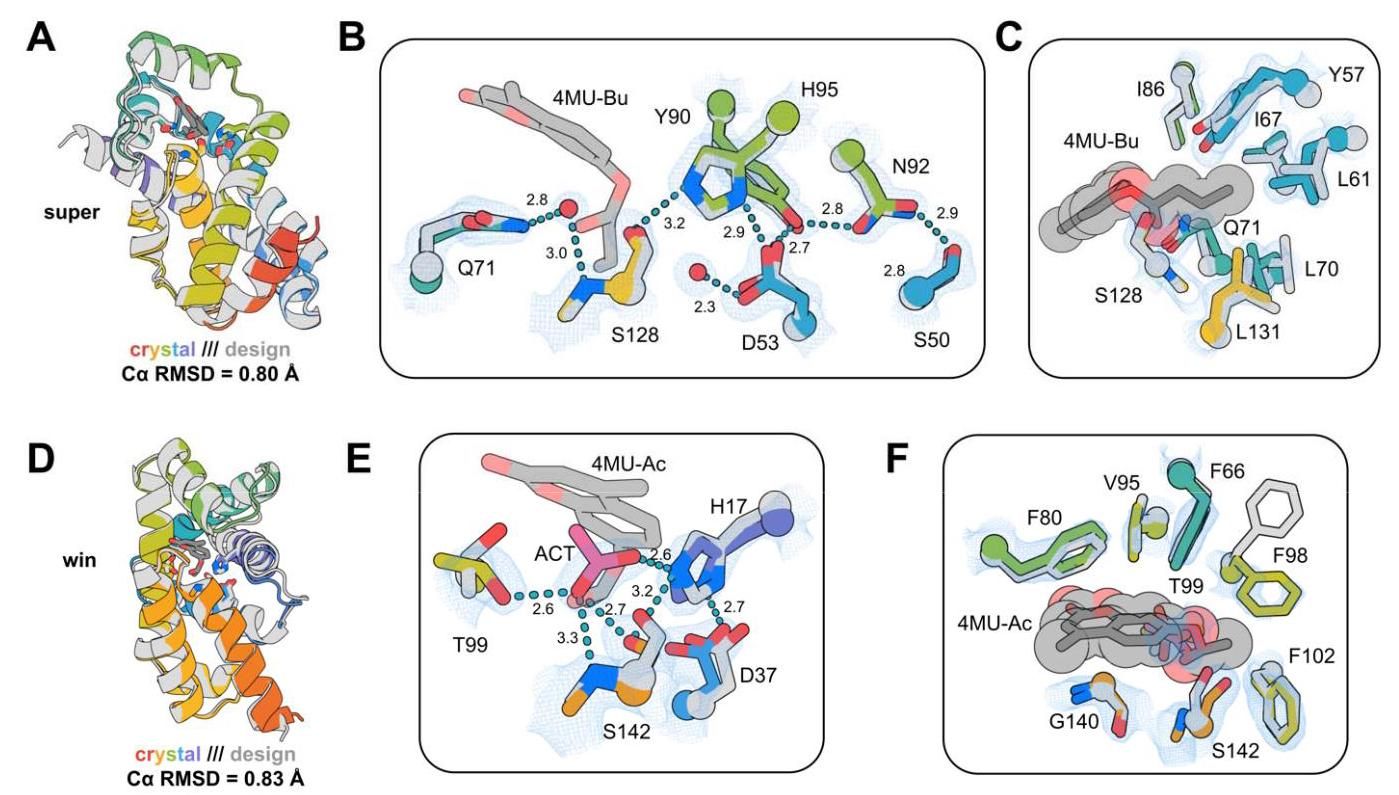


Figure 3. Structural characterization of designed serine hydrolases. (A, D) Structural superposition of design models (gray) and crystal structures (rainbow) for super (A) and win (D). (B, E) Active site overlays of design models (gray) and crystal structures (rainbow) of super (B) and win (E) with 2Fo-Fc 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 (blue mesh). Distances shown are in Å.

图3. 设计的丝氨酸水解酶的结构表征。(A, D) 设计模型(灰色)与超晶体结构(彩虹色)的结构叠加，分别为super(A)和win(D)。(B, E) 设计模型(灰色)与超晶体结构(彩虹色)的活性位点叠加，分别为super(B)和win(E)，2Fo-Fc图在1σ下显示(蓝色网格)。(C, F) 设计模型(灰色)与超晶体结构(彩虹色)的底物结合位点叠加，2Fo-Fc图在 下显示(蓝色网格)。显示的距离单位为Å。

bioRxiv preprint doi: https://doi.org/10.1101/2024.08.29.610411; this version posted August 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license. 750 751 752 753 available under aCC-BY-ND 4.0 International license.

bioRxiv 预印本 doi: https://doi.org/10.1101/2024.08.29.610411; 本版本于2024年8月30日发布。该预印本的版权持有者(未经过同行评审认证)为作者/资助方，已授予 bioRxiv 永久展示该预印本的许可。该文档根据 aCC-BY-ND 4.0 国际许可协议提供。

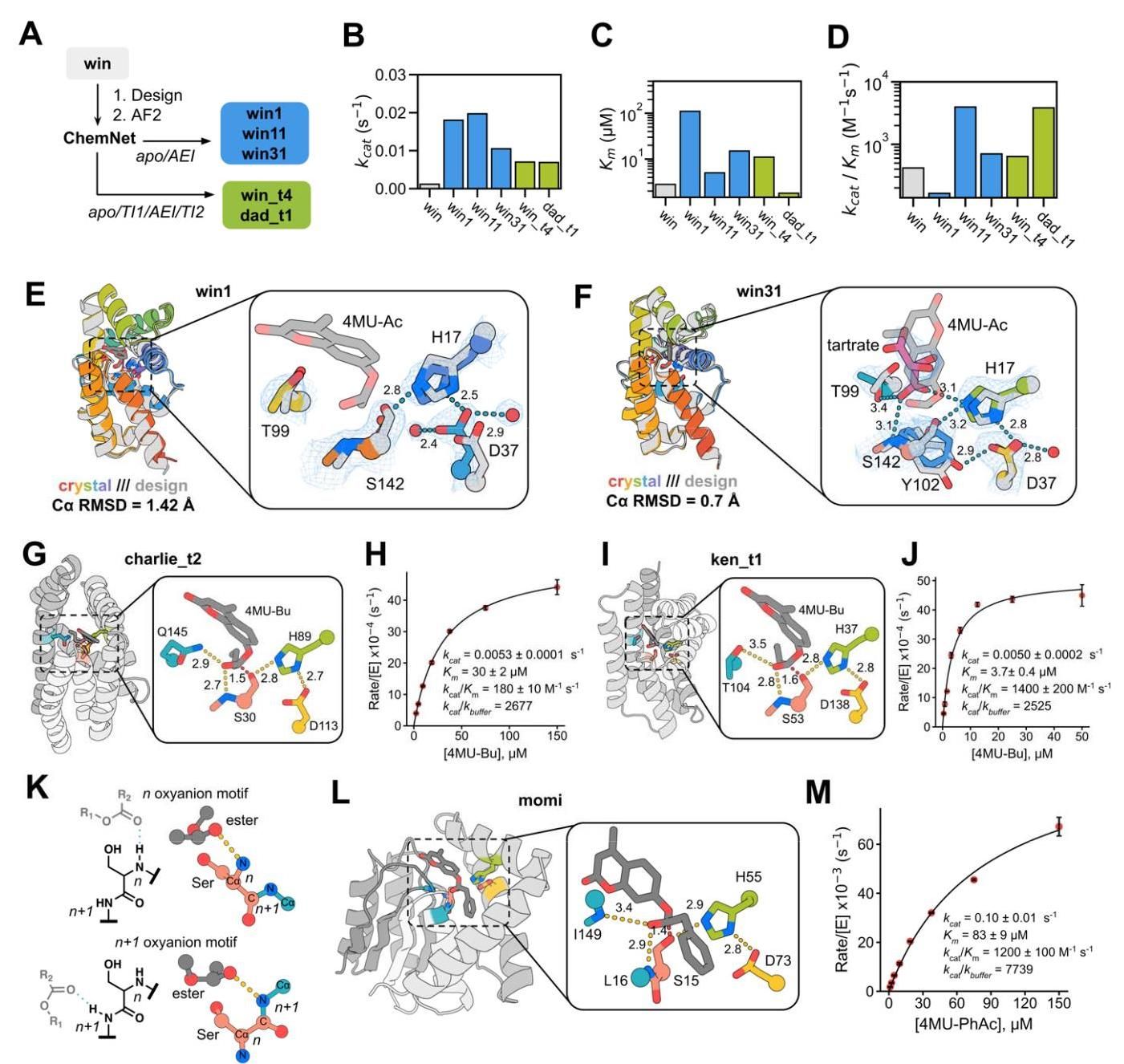


Figure 4. Computational redesign and more complex folds improve catalysis. (A) Computational pipeline for redesign of rd3\_win. (B, C, D) , and of parent rd3\_win compared to computational redesigns. (E, F) Structural superposition of win1 (E) and win31 (F) design and crystal structure. (G, H, I, J) 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 and 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.

图4. 计算重设计和更复杂的折叠改善催化作用。(A) rd3\_win的重设计计算流程。(B, C, D) 与计算重设计相比的母体rd3\_win的 和 。(E, F) win1 (E)和win31 (F)设计与晶体结构的结构叠加。(G, H, I, J) 具有不同折叠和活性位点结构的活性设计的设计模型(G, I)和米哈利斯-门腾图(H, J)。(K) 和 氧阴离子孔特征的化学和结构比较。(L) 利用两个主链酰胺氧阴离子孔接触的活性设计模型，其中一个来自n+1主链酰胺。(M) 活性设计momi的米哈利斯-门腾图。

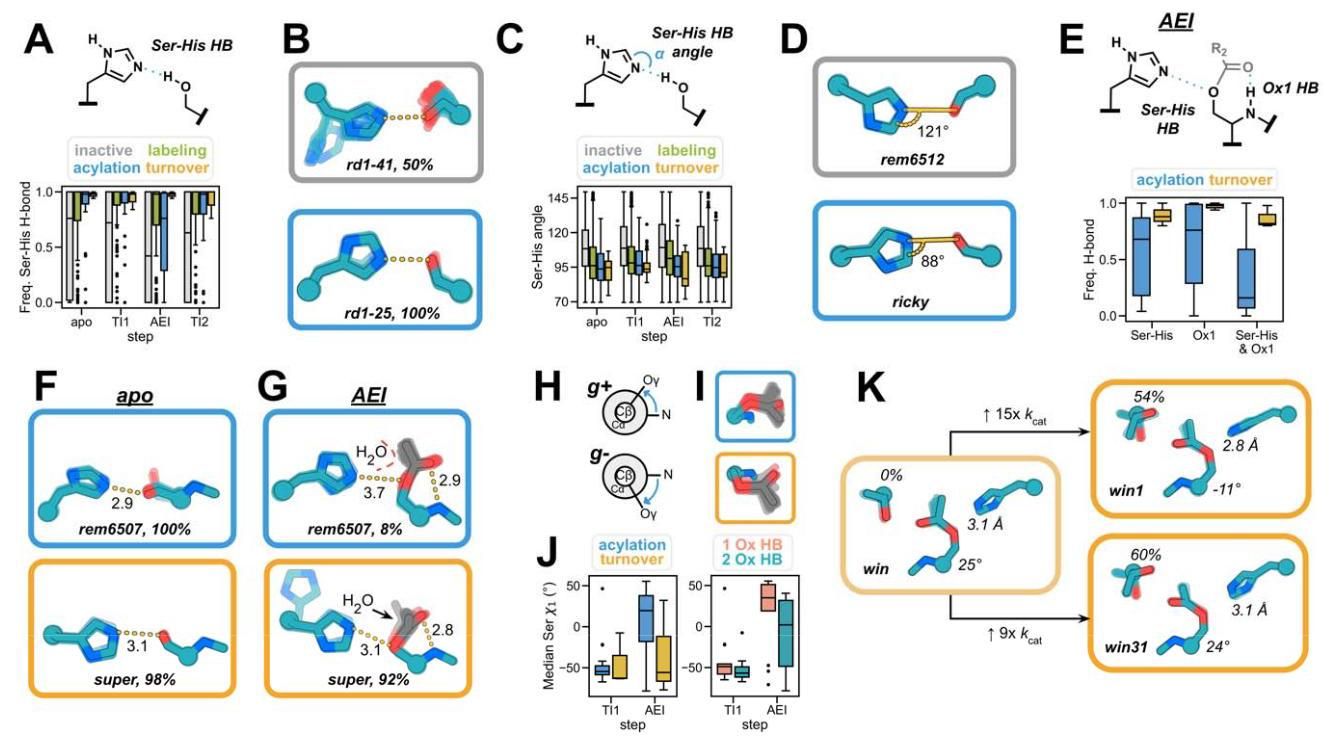


Figure 5. ChemNet ensembles reveal geometric determinants of catalysis. (A) Frequencies of catalytic Ser-His H-bond formation in ChemNet ensembles of each reaction intermediate, grouped by experimental outcome. (B) Apo ChemNet ensembles of representative inactive (top) and acylating (bottom) designs. (C) Median angle ( ) between serine Oy, 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 . (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 and - rotameric states (left). (I) Chemnet ensembles of an acylating design (top) and a design that catalyzes turnover (bottom). (J) Median serine angle across TI1 and AEI state Chemnet ensembles for designs that catalyze acylation or turnover (left). Median serine 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 angle, and catalytic Ser-His H-hbond distance shown.

图5. ChemNet 集合揭示了催化的几何决定因素。(A) 按实验结果分组的每个反应中间体的催化 Ser-His 氢键形成频率。(B) 代表性非活性(上)和酰化(下)设计的 Apo ChemNet 集合。(C) 非活性和酰化设计的 ChemNet 集合中，丝氨酸 Oy、组氨酸 Nϵ 和 Cϵ 之间的中位角( )。(D) 代表性非活性(上)和酰化(下)设计的 Apo ChemNet 集合，角度表示中位数 。(E) 进行酰化或完全转化的设计的 AEI ChemNet 集合氢键频率。(F) 酰化(上)和多次转化设计(下)的 Apo 状态 ChemNet 集合。(G) 进行酰化(上)和催化转化(下)的代表性设计的 AEI 状态 ChemNet 集合。所示测量代表每个集合中关键氢键的中位距离(Å)，百分比表示所有 ChemNet 轨迹中氢键形成的频率。(H) 丝氨酸 和 - 旋转异构体状态的 Newman 投影(左)。(I) 酰化设计(上)和催化转化设计(下)的 ChemNet 集合。(J) 催化酰化或转化的设计在 TI1 和 AEI 状态 ChemNet 集合中的丝氨酸 中位角(左)。按氧阴离子孔氢键数量分组的相同设计在 TI1 和 AEI 状态 ChemNet 集合中的丝氨酸 中位角。(K) win、win1 和 win31 的 AEI 状态 ChemNet 集合，显示具有正确氧阴离子孔旋转异构体的帧百分比、丝氨酸 角度和催化 Ser-His 氢键距离。

bioRxiv preprint doi: https://doi.org/10.1101/2024.08.29.610411; this version posted August 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license. 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768

bioRxiv 预印本 doi: https://doi.org/10.1101/2024.08.29.610411; 本版本于2024年8月30日发布。该预印本的版权持有者(未经过同行评审认证)是作者/资助者，他们已授予bioRxiv永久展示该预印本的许可。该文档根据CC-BY-ND 4.0国际许可协议提供。