

文章编号: 1006-3080(2025)03-0291-04

DOI: 10.14135/j.cnki.1006-3080.20250429002

人工智能赋能丝氨酸水解酶的设计进展

张一飞, 丁于敬, 余傲

(北京化工大学化工资源有效利用全国重点实验室, 北京软物质科学与工程高精尖创新中心, 北京 100029)

摘要:近年来, 人工智能技术在蛋白质结构预测与功能设计方面取得了突破性进展, 为人工酶的从头设计提供了全新的方法和工具。丝氨酸水解酶因其催化机制明确, 成为人工酶设计任务的代表性模型。本文综述了近年来开发的重要深度学习算法及其在人工酶设计中的典型应用, 并重点介绍了人工丝氨酸水解酶的最新设计成果。这些突破性进展为创制具有更高催化效率和稳定性的人工酶铺平了道路, 标志着酶设计领域迈入了新的阶段, 有望带来生物制造相关产业的变革。

关键词:人工智能; 计算酶设计; 深度学习; 蛋白质工程; 丝氨酸水解酶

中图分类号: Q81

文献标志码: A

近年来, 人工智能技术在蛋白质研究中的应用引发了蛋白结构与功能预测领域的革命性突破, 也为人工酶的从头设计(*de novo design*)提供了新的方法和工具, 如图 1 所示。2020 年, Madani 等^[1]发表了基于语言生成模型的蛋白生成模型 ProGen, 该模型的训练采用了来自 1.9 万个蛋白家族的总计 2.8 亿条蛋白序列。运用该模型, Madani 等获得了序列有别于天然酶的溶菌酶、马来酸脱氢酶等人工酶^[2]。2021 年, Baker 团队^[3]提出了用于从头设计蛋白质的深度网络幻想方法, 所设计的蛋白质序列中有约 20% 能够在细菌中顺利表达, 且蛋白折叠结构与预测结果相似。2022 年, Baker 团队^[4]开发了“hallucination”和“inpainting”这两种可用于生成包含功能位点的骨架结构的深度学习方法, 展示了其在蛋白设计领域的应用潜力, 但并未对所设计的酶进行活性验证。2023 年, 他们在 RoseTTAFold 基础上结合去噪扩散概率模型, 开发出模型 RFdiffusion^[5]。该模型可从最小活性位点(催化残基主链)构象出发, 通过去噪迭代生成完整的新骨架, 在单体蛋白质、对称寡聚体、蛋白质结合蛋白的设计任务中表现出色。Baker 团队还开发了用于生成给定蛋白骨架的氨基酸序列的逆折叠模型 ProteinMPNN^[6]和 LigandMPNN^[7], 并在金属离子、小分子以及 DNA 结合蛋白的设计中进行了实验验证。这类模型还能对天然蛋白质的氨基酸

序列进行优化, 提高其结构稳定性和可表达性^[8]。在酶的设计方面, Yang 团队^[9]开发了基于综合计算评分蛋白质序列筛选标准 COMPSS(Co-evolutionary and Motif-based Protein Sequence Screening), 它能够从人工智能生成的候选序列中高效识别出具有活性的酶。经过三轮迭代优化, 最终筛选出的苹果酸脱氢酶和铜锌超氧化物歧化酶的人工酶能够在大肠杆菌中表达、纯化, 且具备催化活性。Baker 团队^[10]通过 Rosetta Match 算法将血红素结合位点整合至环形支架中心空腔, 接着利用 Rosetta 组合方法设计血红素结合位点两侧的氨基酸序列, 获得了活性口袋大、可塑性和稳定性高的人工血红素酶。基于图神经网络算法, Baker 团队采用全家族幻觉(Family-wide hallucination)算法和基于软件 Rosetta 的序列设计, 对核转运因子 2(NTF2)家族的蛋白骨架结构进行幻想生成, 并对底物结合的口袋精细调整, 成功构建了一种小型的高热稳定性的荧光素酶, 展示了人工酶设计的巨大潜力^[11]。但是, 由于荧光素酶催化机制简单, 在设计过程中仅需要考虑人工酶的活性中心对目标底物二苯四嗪阴离子的稳定化作用, 并不涉及更复杂的蛋白分子动力学, 因此该方法并不适用于其他复杂酶的设计。对具有更复杂的活性位点、并需要介导多步复杂催化反应的酶, 其设计仍然面临诸多挑战。并且, 上述方法所设计出的蛋白, 其骨架

收稿日期: 2025-04-29

基金项目: 国家自然科学基金(32371325); 北京市自然科学基金(Z240030); 中国石油化工集团公司种子基金(223260)

作者简介: 张一飞(1988—), 男, 江苏省常州人, 教授, 主要研究方向为生物化工。E-mail: yifeizhang@mail.buct.edu.cn

引用本文: 张一飞, 丁于敬, 余傲. 人工智能赋能丝氨酸水解酶的设计进展[J]. 华东理工大学学报(自然科学版), 2025, 51(3): 291-294.

Citation: ZHANG Yifei, DING Yujing, SHE Ao. Recent Advances in Artificial Intelligence-Enabled Design of Serine Hydrolases[J]. Journal of East China University of Science and Technology, 2025, 51(3): 291-294.

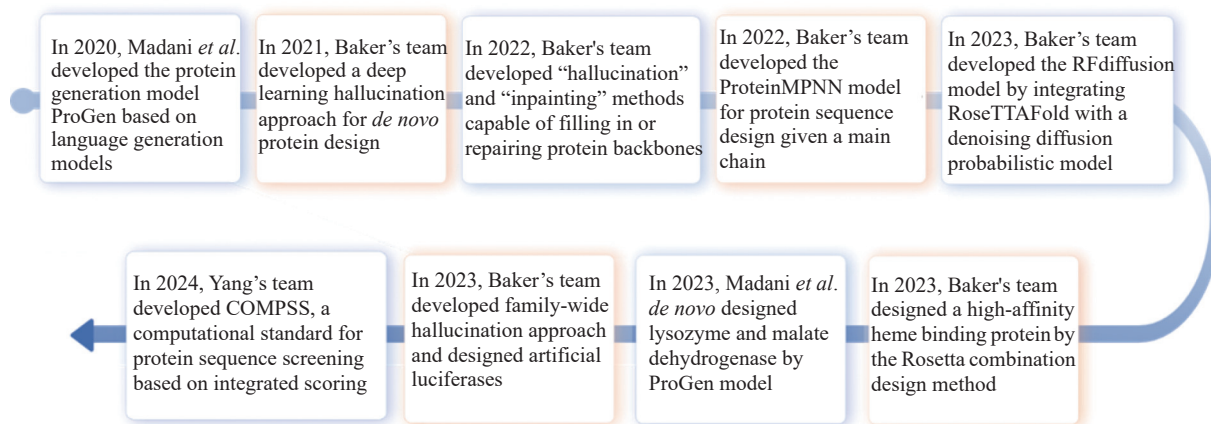


图 1 计算工具开发方面的最新进展及其在酶设计中的应用

Fig. 1 Recent progress in computational tool development and their applications in enzyme design

多样性仍受限于所选择的天然蛋白家族模板。如何突破进化限制,设计具有不同蛋白骨架或折叠形式的人工酶,也是人工酶设计领域的难题。

丝氨酸水解酶(Serine hydrolase)是一类以丝氨酸的羟基为亲核基团,通过进攻羰基来实现酯键、酰胺键水解的酶。丝氨酸水解酶广泛参与包括脂质代谢、蛋白质降解、信号转导、代谢调控等多种生理代谢过程,在疾病治疗方面也发挥着重要作用。例如,乙酰胆碱酯酶调节神经递质乙酰胆碱的代谢,其异常与阿尔茨海默病相关^[12]。脂肪酸酰胺水解酶能够降解内源性大麻素类物质 *N*-花生酰乙醇胺,影响内源性大麻素系统的功能,进而对肿瘤细胞的生长、增殖、凋亡等生理过程产生影响^[13]。丝氨酸水解酶还是一类重要的工业酶。例如,脂肪酶在油脂改性、肉类脱脂、酯类香料合成以及日化洗涤剂等领域中具有广泛的应用。最近发现的可催化聚酯塑料高效降解的角质酶、聚对苯二甲酸乙二醇酯(PET)水解酶等也都属于丝氨酸水解酶^[14-15]。这类酶通过酸碱催化机制加速底物的水解,其活性中心由一个丝氨酸残基、一个酸残基(通常为谷氨酸或天冬氨酸)和一个碱残基(通常为组氨酸)组成,构成一组催化三联体。丝氨酸的羟基作为亲核基团攻击底物的羰基碳原子,形成四面体中间产物,同时酸残基和碱残基协同调节质子转移,促进反应进行。氧负离子穴通过氢键网络稳定过渡态的氧负离子,降低反应活化能,从而加速底物的水解^[16]。

丝氨酸水解酶清晰的催化机制为其从头设计奠定了基础。Guallar 团队^[17]利用天然蛋白骨架,在无催化活性的 Fragaceatoin C 蛋白质纳米孔中构建了 8 组催化三联体,实现了对 PET 纳米颗粒的水解,但对微米级以上的 PET 颗粒无水解活性。这类复杂催化酶的设计需要综合评价关键催化残基在整个反应进程中结合底物的位置变化。为解决这个问题,

Baker 团队^[18]开发了反应路径预组织评估工具 PLACER,利用深度神经网络模拟催化循环中的关键状态,评估各反应步骤的活性位点兼容性和多态预组织程度。基于 RFdiffusion 和 PLACER 设计人工丝氨酸水解酶的工作流程如图 2(a)所示,通过整合 RFdiffusion 和 PLACER,从头设计了具有全新折叠结构和酯催化活性的人工丝氨酸水解酶。以 4-甲基香豆素苯乙酸酯为底物,最佳的人工酶的催化效率(k_{cat}/K_m)达到 $2.2 \times 10^5 \text{ mol}/(\text{L} \cdot \text{s})$ (转化数(k_{cat})为 0.057 s^{-1} ,米氏常数(K_m)为 $0.26 \text{ } \mu\text{mol}/\text{L}$)。这项研究展示了人工智能在蛋白质设计中的巨大潜力。但是,该人工酶与天然丝氨酸水解酶相比,其催化效率仍然较低(例如,乙酰胆碱酯酶对乙酰胆碱的 k_{cat}/K_m 值超过 $10^8 \text{ mol}/(\text{L} \cdot \text{s})$ ^[19],且 k_{cat} 仅为 0.057 s^{-1} ,难以满足实际应用的需要。

除了小分子底物,很多丝氨酸水解酶还能高效地催化蛋白、聚酯等大分子的降解。这类酶的设计不仅要求相关的活性位点满足特定的几何排布和预组织方式,还需要保持活性位点位于蛋白表面。本课题组^[20]展示了以已知的天然酶活性位点为基础,将深度学习算法和常规分子模拟相结合的工作流程,用于设计全新的 PET 水解酶。我们首先从天然的枝叶堆肥角质酶(LCC)中提取了含有催化三联体基序的二级结构片段,然后利用改进的 RoseTTAFold 网络(RF_{joint})生成用于支撑所提取基序的蛋白质序列,并使用 AlphaFold2 对新的序列进行蛋白结构预测。基于 RF_{joint} 和传统计算筛选策略设计人工 PET 水解酶的工作流程如图 2(b)所示。在此基础上,通过评估以下指标筛选具有潜在催化活性的序列:(1)结构预测建模得分(Template Modeling score, TM-score > 0.7);(2)每个残基局部距离差检验的预测值(per-residue Local Distance Difference Test, pLDDT > 70);(3)基序复现精度均方根偏差(Motif

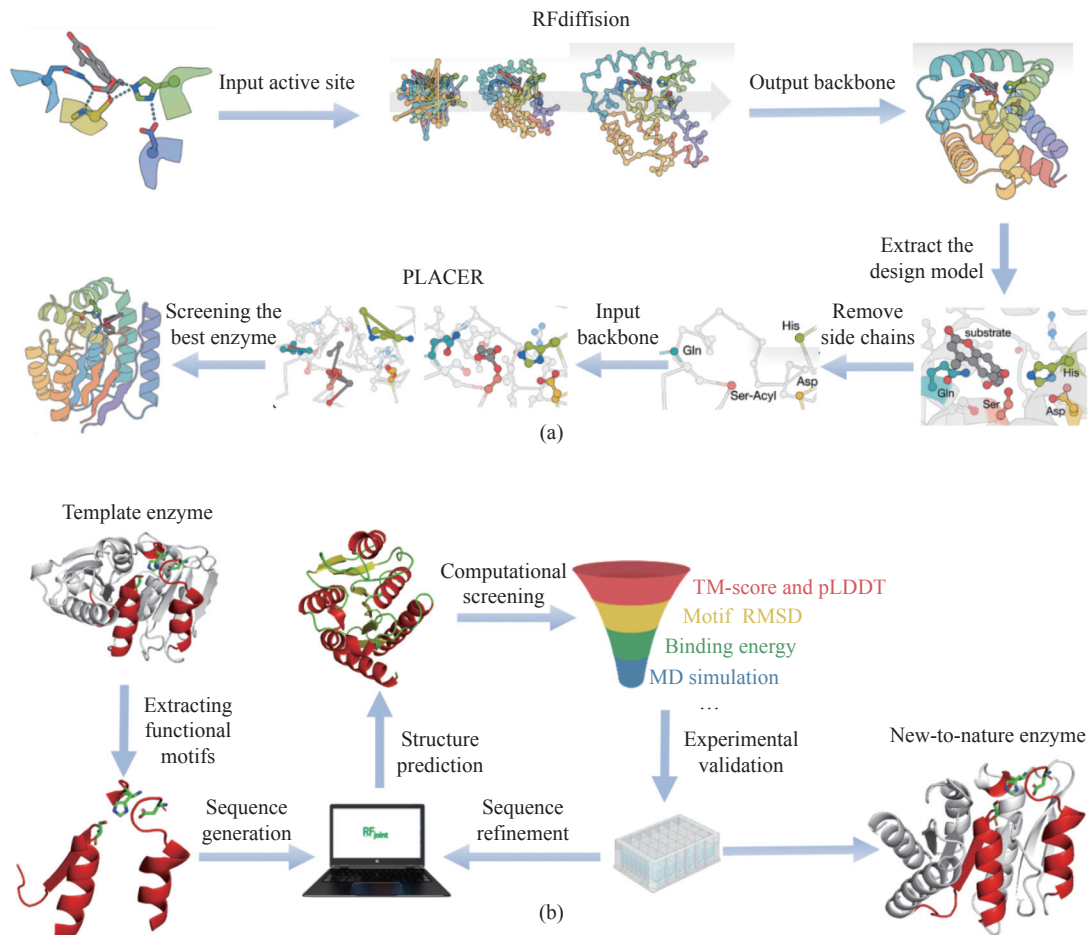


图2 基于模型 RFdiffusion 和工具 PLACER 设计人工丝氨酸水解酶的工作流程(a); 基于 RF_{joint} 和传统计算筛选策略设计人工 PET 水解酶的工作流程(b)

Fig. 2 Workflow for designing artificial serine hydrolase based on RFdiffusion model and PLACER tool (a); Workflow for designing artificial PET hydrolase based on RF_{joint} and conventional computational screening strategies (b)

RMSD < 2.5 Å, 1 Å = 1×10⁻¹⁰ m); (4)模式底物 2-羟乙基-(对苯二甲酸单羟乙基)₃的结合能(Binding energy, -16.7 kJ/mol~0); (5)催化距离约束(丝氨酸残基上的氧原子与底物羰基碳间距小于 4 Å); (6)分子动力学模拟(MD simulation)稳定性(20 ns 分子动力学模拟中骨架 α 碳原子的 RMSD 小于 5 Å)。最后,通过实验验证和序列优化,成功地开发出了比 LCC 序列长度短 30% 的人工酶,且能复制 PET 水解活性。其中,人工酶 RsPETase1 表现突出,其 k_{cat}/K_m 与 LCC 相当,且具备良好的热稳定性(熔融温度为 56 °C),但与 LCC 的序列相似性仅为 34%。该策略的成功实施,证明可以在保留酶的核心功能结构的前提下,尽可能去除冗余序列,获得尺寸更小、折叠更紧凑的新酶。该方法有望应用于更多种类的丝氨酸水解酶的设计中,帮助人们获得更多的“微型酶”,提高酶的结构稳定性和催化效率,降低酶的生产成本。

人工智能技术在丝氨酸水解酶设计方面的突破,为具有复杂催化机制的酶的设计奠定了基础。近年来,各类新的深度学习算法不断涌现,在新催化

中心设计、蛋白质骨架构建和氨基酸序列生成等方面展现出强大的能力,显著提升了人工酶的设计成功率。这些人工酶的构建不仅有效地扩展了酶的序列和结构多样性,也为理解蛋白质序列与功能之间的映射关系、蛋白动力学和催化活性之间的关系提供了新的研究途径。随着人工酶在稳定性、专一性及催化效率等方面不断接近甚至超越天然酶,高性能人工酶有望替代天然酶,在医药合成、食品加工、生物能源、环境治理等关键领域实现广泛应用,推动相关产业技术的深度变革与创新。

参考文献:

- [1] MADANI A, MCCANN B, NAIK N, *et al.* ProGen: Language modeling for protein generation[EB/OL]. (2020-03-07) [2020-05-29]. <https://www.biorxiv.org/content/10.1101/2020.03.07.982272v2>.
- [2] MADANI A, KRAUSE B, GREENE E R, *et al.* Large language models generate functional protein sequences across diverse families[J]. *Nature Biotechnology*, 2023,

- 41(8): 1099-1106.
- [3] ANISHCHENKO I, PELLOCK S J, CHIDYAUSIKU T M, *et al.* *De novo* protein design by deep network hallucination[J]. *Nature*, 2021, 600(7889): 547-552.
- [4] WANG J, LISANZA S, JUERGENS D, *et al.* Scaffolding protein functional sites using deep learning[J]. *Science*, 2022, 377: 387-394.
- [5] WATSON J L, JUERGENS D, BENNETT N R, *et al.* *De novo* design of protein structure and function with RFdiffusion[J]. *Nature*, 2023, 620(7976): 1089-1100.
- [6] DAUPARAS J, ANISHCHENKO I, BENNETT N, *et al.* Robust deep learning-based protein sequence design using ProteinMPNN[J]. *Science*, 2022, 378(6615): 49-56.
- [7] DAUPARAS J, LEE G R, PECORARO R, *et al.* Atomic context-conditioned protein sequence design using Ligand-MPNN[J]. *Nature Methods*, 2025, 22: 717-723.
- [8] SUMIDA K H, NÚÑEZ-FRANCO R, KALVET I, *et al.* Improving protein expression, stability, and function with ProteinMPNN[J]. *Journal of the American Chemical Society*, 2024, 146(3): 2054-2061.
- [9] JOHNSON S R, FU X, VIKNANDER S, *et al.* Computational scoring and experimental evaluation of enzymes generated by neural networks[J]. *Nature Biotechnology*, 2025, 43: 396-405.
- [10] KALVET I, ORTMAYER M, ZHAO J, *et al.* Design of heme enzymes with a tunable substrate binding pocket adjacent to an open metal coordination site[J]. *Journal of the American Chemical Society*, 2023, 145(26): 14307-14315.
- [11] YEH A H W, NORN C, KIPNIS Y, *et al.* *De novo* design of luciferases using deep learning[J]. *Nature*, 2023, 614(7949): 774-780.
- [12] REMYA R S, RAMALAKSHMI N, NALINI C N, *et al.* Design synthesis and *in vitro* evaluation of tacrine-flavone hybrids as multifunctional cholinesterase inhibitors for Alzheimer's disease[J]. *Current Computer-Aided Drug Design*, 2022, 18(4): 271-292.
- [13] TIAN M, TIAN Z, YAO D, *et al.* A NIR fluorescent probe for fatty acid amide hydrolase bioimaging and its application in development of inhibitors[J]. *Journal of Materials Chemistry B*, 2021, 9(32): 6460-6465.
- [14] HAN X, LIU W, HUANG J W, *et al.* Structural insight into catalytic mechanism of PET hydrolase[J]. *Nature Communication*, 2017, 8: 2106.
- [15] SULAIMAN S, YOU D J, KANAYA E, *et al.* Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase[J]. *Biochemistry*, 2014, 53(11): 1858-1869.
- [16] DU S, KRETSCH R C, PARRIS-GOLD J, *et al.* Conformational ensembles reveal the origins of serine protease catalysis[J]. *Science*, 2025, 387(6735): eado5068.
- [17] ROBLES-MARTÍN A, AMIGOT-SÁNCHEZ R, FERNANDEZ-LOPEZ L, *et al.* Sub-micro- and nano-sized polyethylene terephthalate deconstruction with engineered protein nanopores[J]. *Nature Catalysis*, 2023, 6(12): 1174-1185.
- [18] LAUKO A, PELLOCK S J, SUMIDA K H, *et al.* Computational design of serine hydrolases[J]. *Science*, 2025, 388(5744): eadu2454.
- [19] BAR-EVEN A, NOOR E, SAVIR Y, *et al.* The moderately efficient enzyme: Evolutionary and physico-chemical trends shaping enzyme parameters[J]. *Biochemistry*, 2011, 50(21): 4402-4410.
- [20] DING Y, ZHANG S, KONG X, *et al.* Replicating PET hydrolytic activity by positioning active sites with smaller synthetic protein scaffolds[J]. *Advanced Science*, 2025, 12(18): 2500859.

Recent Advances in Artificial Intelligence-Enabled Design of Serine Hydrolases

ZHANG Yifei, DING Yujing, SHE Ao

(State Key Laboratory of Chemical Resources Engineering, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: Recent advances in artificial intelligence have revolutionized the structure prediction and functional design of proteins, providing novel methods and tools for the *de novo* design of artificial enzymes. Serine hydrolases have emerged as a representative model in the task of artificial enzyme design due to their well-defined catalytic mechanism. This review summarizes key deep learning algorithms recently developed for artificial enzyme engineering and highlights their typical applications, with a focus on the latest successes in designing artificial serine hydrolases. These breakthroughs pave the way for the development of artificial enzymes with enhanced catalytic efficiency and stability, marking a new stage in enzyme design that could transform biomanufacturing and related industries.

Key words: artificial intelligence; computational enzyme design; deep learning; protein engineering; serine hydrolase

(责任编辑: 李娟)