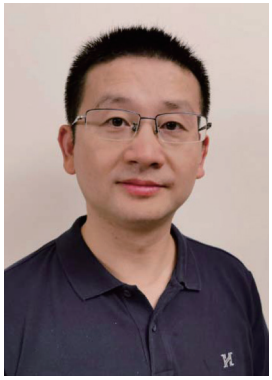
DOI: 10.13376/j.cbls/2021168
文章编号：1004-0374(2021)12-1493-09

· 应用 ·



江会锋，研究员，博士生导师，主要从事代谢合成生物学研究，利用合成生物学原理与技术，解析自然界植物复杂生物代谢途径，研究生物碳缩合与重排的基本规律与生物催化的反应机理，构建生物代谢新酶新途径，创建植物源天然产物等微生物细胞工厂。

合成生物学酶改造设计技术的研究进展

王　千[1#]，白　杰[1,2#]，江会锋[1*]
(1 中国科学院天津工业生物技术研究所，中国科学院系统微生物工程重点实验室，天津 300308；2 天津科技大学生物工程学院，天津 300457)

摘　要：酶是一种优秀的生物催化剂，酶促反应以高效性、专一性、条件温和、绿色环保等优点著称，但天然酶仍存在活性低、稳定性差、专一性不佳等问题。合成生物学与人工智能技术的不断进步为天然酶的催化机制研究、催化活性改造以及新功能酶的设计提出了更高的要求和全新的思路。定向进化、理性及半理性设计和新酶设计等是目前发展和应用较为成熟的酶改造设计技术。同时，机器学习也为拓展酶改造设计技术提供了新的可能。该文对经典的酶改造设计方法和机器学习指导的酶改造设计进行了概括，着眼于酶改造设计技术的发展及其应用，为合成生物学中关键酶的设计改造提供参考和依据。

关键词：酶改造；定向进化；理性设计；机器学习
中图分类号：Q55；Q814　　**文献标志码：**A

Research progress on technologies of enzyme engineering and design in synthetic biology

WANG Qian[1#], BAI Jie[1,2#], JIANG Hui-Feng[1*]
(1 Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China; 2 College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract: As natural catalysts, enzymes have excellent characteristics like high catalytic efficiency, high substrate specificity, mild reaction conditions, environmental friendliness, and so on. However, natural enzymes still suffer the shortages like low catalytic activity, poor stability, and promiscuous substrate specificity. The continuous

收稿日期：2021-11-08
基金项目：天津市杰出青年基金项目(18JCJQJC48300)；天津市杰出人才培养计划
***通信作者：**E-mail: jiang_hf@tib.cas.cn；Tel: +86-22248287-32
[#]**共同第一作者**

advancement of synthetic biology and artificial intelligence put forward higher requirements and provide new insights for understanding the catalytic mechanisms, engineering the catalytic properties, and designing enzymes with novel functions. Currently developed and maturely applied enzyme engineering methods include directed evolution, rational and semi-rational design, new enzyme design, etc. Meanwhile, machine learning also provides new possibilities for enzyme engineering and design. This paper outlines the classical and machine learning-guided enzyme engineering and design methods, focuses on the advancements and applications of enzyme engineering and design technologies, so as to provide reference and basis for the engineering of key enzymes serve in synthetic biology.

Key words: enzyme engineering; directed evolution; rational design; machine learning

合成生物学将来自不同物种的基因所表达的酶重建为一条新的生物制造途径，而酶是构成合成生物学系统的最小单元，但有时酶的天然功能与具体应用的实际要求间存在一定差异，限制了其在工业上的应用。为了解决酶在生产应用中的问题，需要利用分子生物学、生物信息学、结构生物学和计算生物学等手段，对酶进行合理的设计与改造，从而提高其稳定性、催化活性与底物专一性等。随着对酶的了解不断加深，酶改造技术在近些年也经历了快速的发展，根据技术原理的不同，可以分为以定向进化为主的传统酶改造技术；以序列与结构信息为基础的理性及半理性改造技术和新酶设计技术；同时，随着人工智能技术的发展，机器学习指导的酶改造设计也逐渐崭露头角。

1 定向进化

2018 年，诺贝尔化学奖授予弗朗西斯 · 阿诺德 (Frances H. Arnold)、乔治 · 史密斯 (George P. Smith) 和格雷格 · 温特 (Gregory P. Winter)，以表彰三人在定向进化研究领域的开创性贡献。酶的定向进化技术是模仿自然进化过程中基因突变与自然选择的过程，通过向基因中引入突变，并直接或间接对突变体进行筛选，经过多轮迭代，最终得到如酶活性提高、底物特异性改进、酶热稳定性提高等符合预期的突变体 (图 1)。相比于自然进化，定向进化技术可以在短时间内对酶进行多轮突变与筛选，快速搜索序列突变空间，且不受限于蛋白质结构或机理是否明确，是对蛋白质进行改造的有效且高效的手段。定向进化技术需要在每轮迭代中向基因引入突变，引入突变的方法按照原理可以分为随机突变、定点突变、DNA shuffling 等多种类型。

碱基正确的配对是基因准确复制和转录的前提，而碱基错配是基因在体内或体外复制时的低概率事件，这时腺嘌呤 (A) 与鸟嘌呤 (C) 配对，胸腺嘧啶 (T) 与胞嘧啶 (G) 配对，是自然界中基因变异的主要来源之一。受此启发，1989 年，Leung[1] 首次提出通过易错 PCR 对基因进行随机突变的观点。1992 年，Cadwell 和 Joyce[2] 在前人研究的基础上进一步完善，建立了相对成熟的易错 PCR 体系，并沿用至今。1993 年，Frances H. Arnold 团队通过易错 PCR 向酶中引入突变，并进行多轮迭代筛选，最终得到了能在极端环境下高效发挥作用的酶突变体 [3]，此工作被视为蛋白质定向进化技术的开端。

历史上的首次定点突变由 Michael Smith 团队在 1978 年完成 [4]。定点突变利用了 PCR 过程中引物会一直保留在扩增的基因上这一特点，在设计引物时覆盖要突变的序列区域，只要在合成引物时对目标位点进行碱基替换修改，突变就会引入扩增的基因片段之中。NNK 是一种设计引物时的简并策略，N 代表任意 4 种碱基之一，K 表示 G 或 T，这样的设计以 32 (4 × 4 × 2) 种组合覆盖 20 种氨基酸，同时还避免了终止密码子的出现，是构建突变文库最常用的策略。Schwaneberg 课题组使用易错 PCR 与多点饱和突变组合的方法对葡萄糖氧化酶 (glucose oxidase, GOx) 进行定向进化，使酶对氧气的依赖降低约 97%，并提高酶活 5.7 倍 [5]。

自然选择对基因突变的筛选与有性繁殖是生物进化的主要动力，真核生物在有性繁殖的过程中存在的同源重组现象更是其基因多样性的基石。在体外基因扩增过程中同样也存在 DNA 重组现象，20 世纪 90 年代，Marton 等 [6] 与 Stemmer[7] 提出并证明在 PCR 过程中存在 DNA 重组，并以此提出 DNA shuffling 技术。DNA shuffling 技术首先将一组同源序列酶解为小片段，在经历变性退火后，含有同源序列的来自不同基因的片段间发生配对，之后不添加引物进行延伸，使同源序列间发生重组。Stemmer[7] 运用 DNA shuffling 对 β- 内酰胺酶进行了体外分子进化，经过 3 轮定向进化后，获得了一个对头孢霉

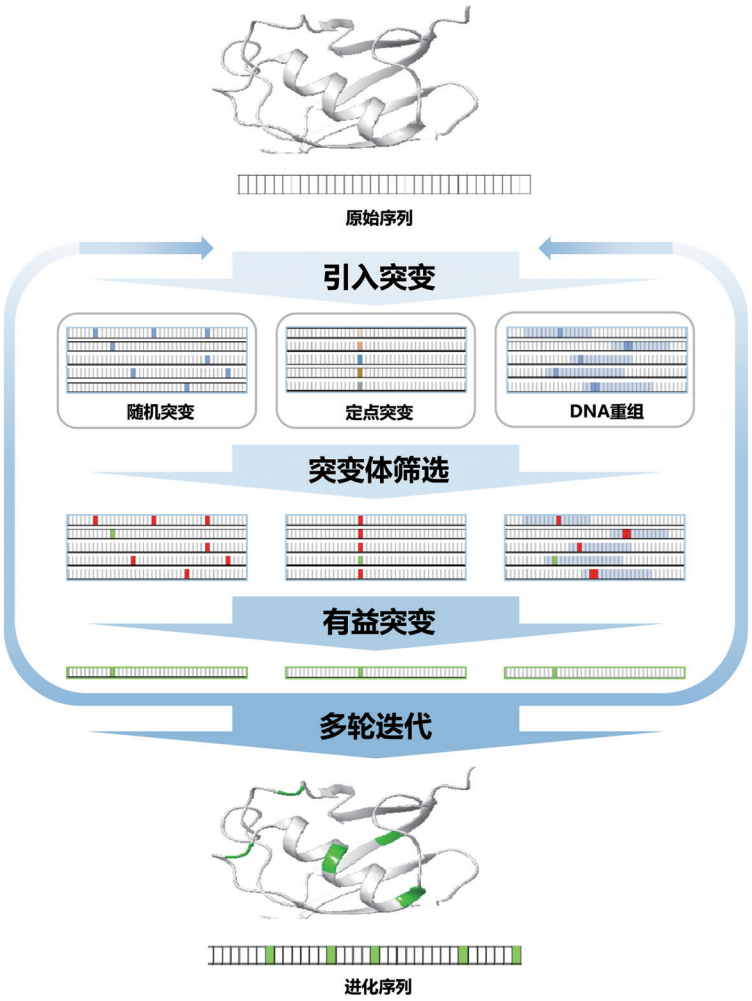


图1　定向进化

素抗性提高 16 000 倍的突变体。

在实际应用中，以上策略通常会被组合使用，以提高突变体库的多样性。如 Shi 等 [8] 采用易错 PCR 与 DNA-shuffling 结合的技术构建突变体库，筛选到 β 琼脂糖酶突变体，其 Tm 值提高 4.6 倍，在 40 ℃时的半数失活时间为 350 min，比野生型提高 18.4 倍。

2　理性及半理性改造

尽管定向进化被广泛应用于酶改造并取得了很大的成功，但是这种方法的缺陷也比较明显，因为对于指定蛋白质，其单点饱和突变空间为 20^N (N 为蛋白质序列长度)，多点组合突变的突变空间更是高达 $(20^N)^X$，定向进化所使用的突变库仅占据所有可能的序列的一小部分 [9]。使用更大的库和更多的筛选诚然可以解决该问题，但随之而来的高几个数量级的工作量将会是一个巨大的阻碍。与定向进化相比，酶的理性及半理性改造方法将序列、结构、功能等信息作为先验知识，大大缩小了要考虑的氨基酸范围，降低了实验工作量，增加了有益突变的概率，并且可以在突变过程中了解突变背后酶活性改善的机理 (图 2)。

酶改造区域的选择是酶改造设计过程中首当其冲要面临的一步。正确选择改造位点可以极大提高酶改造的效率。根据与底物的距离，活性位点周围的氨基酸被分为不同的层次：距离底物最近或与底物存在相互作用的氨基酸划分为 first shell，与 first shell 相邻或存在相互作用的氨基酸划分为 second shell，以此类推。其中，相邻的定义是原子间距离 (Å) 等于或小于 4.1 Å (C-C)、3.3 Å (O-O)、3.4 Å (N-N, N-O)、3.8 Å (C-N) 和 3.7 Å (C-O)[10]。对于改善酶的对映选择性、立体选择性，活性中心氨基酸 (first shell 及 second shell) 的突变往往能发挥出令人意想不到的效果；而对于改变酶的活性及稳定性，距离

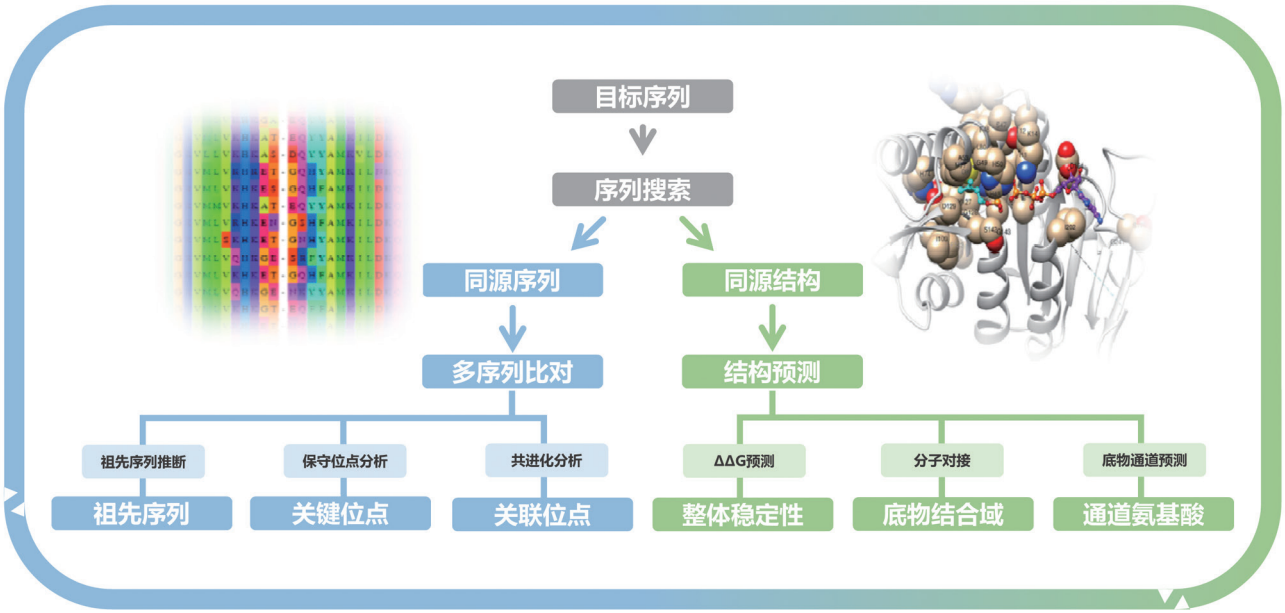


图2 理性与半理性改造设计

活性中心稍远些的位点的效果可能比活性中心的效果更好 (7~10 Å，third 和 fourth shell)[11]。Keasling 实验室在对 γ- 律草烯合成酶 (gamma-humulene synthase) 的研究中，对活性中心 19 个氨基酸进行饱和突变和片段重组，由于选择了底物活性中心的位点，研究人员只用了不到 2 500 个突变体就完成了对酶活性的提升 [12]，证明了正确选择改造位点的重要性。

第三代测序技术的普及使蛋白质序列数据得到进一步的丰富，依靠进化信息指导目的酶的改造是当前比较成熟的一种策略。基于多序列比对 (multi-sequences alignment, MSA) 和系统发育分析定位和识别蛋白质序列中的功能区域，探索蛋白质氨基酸的保守性和祖先进化关系是序列进化信息应用的主要方面。蛋白质的共进化 (coevolution) 是指在自然进化的过程中，一些位点的变异会引起与其相关联的位点的变化，这一概念在 1971 年由 Fitch[13] 提出。在一项对异戊烯基磷酸激酶 (isopentenyl phosphate kinase, IPK) 的改造中，Liu 等 [14] 利用蛋白质序列的共进化信息，设计出 9 个位点并对其进行突变筛选，最终将 IPK 的活性提高 8 倍。祖先序列重建 (ancestral sequence reconstruction, ASR) 是一种常被应用于序列进化分析的方法，该方法认为当前序列是由远古的共同祖先进化而来，并且在进化过程中保留了一些祖先的特性和痕迹，因此可以使用现存序列信息 (DNA、氨基酸序列) 推断当前蛋白质的祖先序列 [15]。多个研究团队以 ASR 作为辅助手段，分别在酶结构稳定性 [16]、活性提高 [17] 与底物特异性 [18] 等方面实现了优化提升。

蛋白质结晶数据的逐步积累与蛋白质建模技术的发展，为结构信息指导的酶改造策略奠定了基础。当前比较常用的半理性策略包括基于同源结构元件交叉原理的 SCOPE (structure-based combinatorial protein engineering) 策略 [19]、基于酶结构功能关系对酶活性中心进行饱和突变的 CASTing (combinatorial active-site saturation test) 策略 [20]、基于硫代核苷酸的 DNA 片段重组方法 PTRec (phosphorothioate-based DNA recombination)[21] 和基于 Rosetta[22] 及 FoldX[23] 自由能计算结果预测潜在稳定性突变点的 FRESCO (framework for rapid enzyme stabilization by computational libraries strategy)[24] 等。在对黑曲霉环氧化酶的改造中，Reetz 等 [20] 使用 CASTing 方法对活性中心位点进行饱和突变，实现了酶对多种底物水解活性不同程度的提升。

此外，越来越多把结构信息和序列信息结合起来运用于半理性设计的方法被开发出来，如 3DM[25] 和 HotSpot Wizard[26]。3DM 首先对目的蛋白质序列进行以结构为基础的多序列比对，然后根据从 PDB、GenBank、PubMed 和 Swiss-Prot 数据库中收集的结构功能关系对影响酶活性的关键位点进行推测，再通过多点饱和突变对关键位点进行验证。HotSpot Wizard 则是整合了众多数据库与进化和结构计算工具，推测出对酶活性、稳定性、底物特异

性等具有影响的潜在位点。例如，在 LinWu 等 [27] 对细菌 I 型硝基还原酶 NfsB 进行活性改造的研究中，6 个潜在活性提高位点中的组合突变 N71S/F124W，使产物 7- 氨基苯并二氮杂卓的产量在有氧环境下提高 11 倍，在无氧环境下提高 6 倍。

酶的理性改造设计是建立在对酶结构与功能的关系及催化机理具有一定了解的基础上，对特定的位点进行突变，从而改变或优化酶分子的性质。近年来，伴随着分子动力学模拟与结构生物学的发展以及对蛋白质折叠机理的研究，理性设计得到了更广泛的应用。Pikkemaat 等 [28] 以分子动力学模拟为研究手段，解析了嗜盐脱卤素酶的解折叠机理，模拟过程表明酶 cap 结构域的 helix-loop-helix 区域具有很高的柔性，根据模拟结果在 201 和 16 位氨基酸之间引入了一个二硫键使 loop 区域的刚性增加，并最终使酶的 Tm 值由 47.5 ℃增加至 52.5 ℃。Kazlauskas 实验室通过构建活性中心周围氨基酸与中间产物过氧基团之间的氢键稳定过渡态，使荧光假单胞菌 (*Pseudomonas fluorescens*) 酯酶 (PFE) 突变体对其过氧化物底物的水解活性提高了 28 倍 [29-30]。同时，把非天然氨基酸 (UAAs) 并入氨基酸突变库以对酶进行理性改造的方法也被开发出来，弥补了天然氨基酸侧链功能基团种类较少和氨基酸生化性质单一的缺点，极大地补充和拓展了当前的酶改造设计方法 [31-34]。在对脂水解和转酰基反应酶的研究过程中，通过引入多种疏水性更强的氟化 UAAs，脂肪酶的稳定性大大增强，使得其在工业上的应用价值有了显著提升 [35-36]。在对 P450 酶的抗氧化研究中，通过对蛋白质的甲硫氨酸 (Met) 进行惰性正亮氨酸 (norleucine) 替换，极大地增强了 P450 酶的抗氧化能力，并显著提高了催化活性 [37-38]。此外，UAAs 引入也对酶新功能的设计有很大帮助。France Arnold 课题组将 P450 酶 HEM 链接的半胱氨酸突变为其他 UAAs，可以使一种 P450 酶获得对多种底物的催化能力 [39]。

3　新酶设计

新酶设计，顾名思义指的是设计出自然界尚未发现的可以催化特定化学反应的酶。在计算机运算能力不断提高的背景下，酶的从头设计已经成为新酶设计的一个重要方向 (图 3)。根据计算过程中使用策略的不同，酶设计可以分为基于能量函数的新酶设计和基于深度学习的新酶设计。

基于能量函数的酶设计策略主要包括中国科学技术大学刘海燕课题组开发的 ABACUS[40-41] 及 SCUBA[42] 方法和华盛顿大学 David Baker 课题组开发的 Rosetta 方法 [22]。ABACUS 和 SCUBA 分别基于主链氨基酸和侧链氨基酸采样的统计能量函数，并结合范德华能量项，适用于主链蛋白质序列设计和侧链氨基酸构象采样及设计。ABACUS 以指定蛋白质结构作为框架输入，使用由已知蛋白质结构训练的统计能量函数对蛋白质结构进行计算取样，最终得到最优的氨基酸残基组合。这种基于统计能量的蛋白质设计方法已经在多个酶的设计和改造中得到了应用 [41, 43-44]；Rosetta 方法作为一种经典的基于能量函数的复合采样方法，在蛋白质同源建模 [45]、分子对接 [46]、抗体设计和新酶设计 [47-50] 等方面都有广泛的应用。使用 Rosetta 设计新酶，需要根据目标反应的反应机理，针对过渡态构象和活性中心的几何形状用量子力学原理进行建模，以此为参考在蛋白质数据库 (PDB) 中搜索可以与过渡态模型紧密结合的蛋白质骨架并优化，之后，根据过渡态自由能及骨架与过渡态位置取向对优化后的结果进行排序，实验鉴定功能后再结合定向进化等方法进一

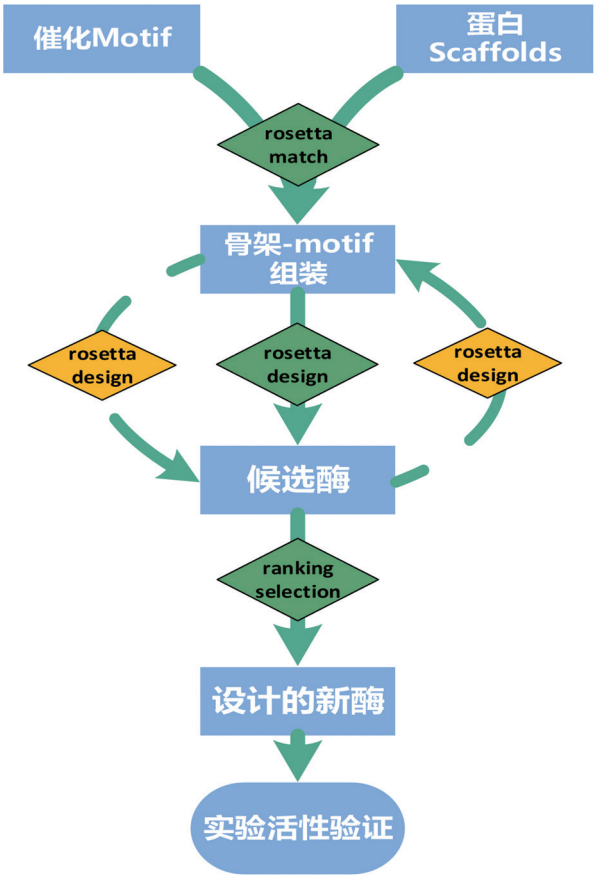


图3　新酶设计

步提高[49]。近年来，新酶设计在新型跨膜纳米孔蛋白[51]、IL2 及 IL5 模拟结合因子[50]、多肽诊疗因子[52]、非天然 β 折叠片[53]等的设计中体现出了巨大优势，使特定蛋白质结构及催化功能的设计成为了可能。

随着人工智能的发展，使用机器学习 (machine learning, ML) 生成具有特定功能的全新蛋白质序列成为新酶设计另一个具有挑战性的领域。通过对大量蛋白质序列的学习，使用神经网络或其他学习模型总结归纳其中的序列 - 结构 - 功能特征，是一种典型的深度学习过程。作为序列 - 结构学习最成功的例子，Alphafold2[54] 和 RoseTTAFold[55] 采用深度学习与结构优化相结合的方法，在 CASP14 大赛中，将蛋白质结构预测的精度提升到近乎晶体结构的水平，为设计特定蛋白质结构的序列提供了可能。通过类似的思路，研究人员成功地设计出了具有多种催化功能的新酶[56-57]。UniRep 模型是目前应用较为广泛的序列设计和生成模型之一，该模型通过对 UniRef50 中 2 400 多万条序列的学习，获得了序列 - 功能特征，在序列 - 功能预测和序列设计上具有很高的应用价值[58]；ProteinGAN[59] 是一种基于自注意生成性对抗网络建立的学习模型，该模型直接从复杂的多维氨基酸序列空间中学习蛋白质序列的进化关系，并创建具有天然物理性质的高度多样的新序列。在对苹果酸脱氢酶 (MDH) 进行序列设计时，显示出了 24% (13/55) 的设计成功性，证实了 ProteinGAN 作为全新序列设计工具的潜力。

4 机器学习指导的酶改造设计

传统的酶改造技术具有突变采样空间巨大、实验成本昂贵以及依赖高通量技术等缺点，这在很大程度上限制了酶改造设计的进程。随着新一代测序技术、高通量筛选方法、蛋白质改造数据库和人工智能的发展，以统计数据为驱动进行酶改造设计正成为解决这些挑战的一个有潜力且行之有效的方案[60]。近些年，机器学习辅助酶改造的方法已经被初步提出并在酶活性改造、立体选择性改造以及热稳定性改造上取得了一些可观的成就[61]。这种对现有酶改造数据加以学习，并辅之以实验验证和数学模型验证的方法，极大提高了酶改造的效率 (图 4)。

作为一种以数据为驱动力的研究方法，机器学习对特定性质的实验数据有着较高的要求，针对特定的蛋白质性质 (对特定底物的活性、选择性及稳定性)，输入特定有效的数据集，通过对数据集的

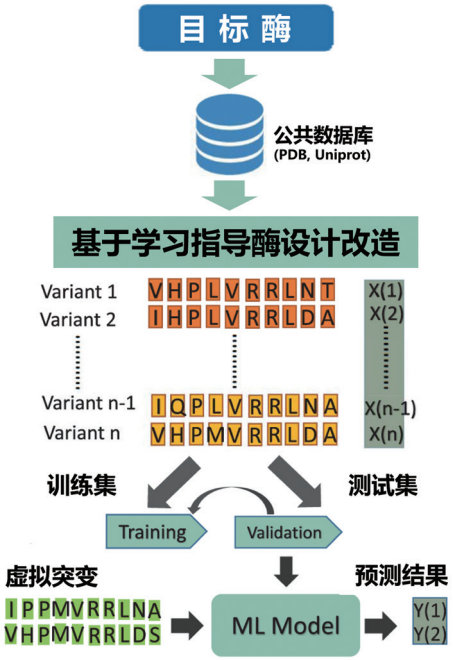


图4 机器学习指导的酶改造设计

学习和预测来指导酶改造和设计。高通量筛选作为一种可以大量产出实验数据的方法，可以与数学建模和数理统计相结合来分析蛋白质序列与功能之间的联系，指导酶学性质的改造。比较经典的案例为 Fox 等[62] 通过高通量筛选及测序获得了全饱和突变对应酶性质变化的数据库，然后使用 ProSAR 方法构建了突变 - 活性关系的数学模型，依据此模型最终通过多点组合突变使卤代醇脱卤酶的活性提高约 4 000 倍。但是这种线性模型应用到酶改造多点组合突变时，有可能出现较强的弱 (或无) 上位显性 (weak/no epistasis)，即多个有效单点突变的组合可能会使酶的活性降低，甚至丧失[63]。当面对这种传统静态函数不能描述的复杂情况时，就需要机器学习来寻找并训练一种函数进行解释[64]。

相对于传统定向进化的单一途径酶活性提升模型，机器学习可以结合现有序列 - 功能数据，在不丢失中性及微阳性数据的情况下，“智能”全面地计算全局活性优化路径，提出最优的序列活性景观 (landscape) 模型，指导酶的进一步活性改造[65-66]。机器学习整体上可以分为有监督 (supervised) 的学习和无监督 (unsupervised) 的学习。当应用于酶的定向进化时，机器学习被认为是一种有监督的学习，即为算法提供一组给定标签的数据 (活性与稳定性等)，通过学习产生一个能预测非标签数据的函数。有监督的学习应用于酶改造及策略研究已经有比较

成熟的案例[60, 65, 67]。相对于有监督的机器学习，半监督的学习使用大部分数据进行无监督的学习，不设立任何已有特征，让机器自动学习序列内部的各项性质，生成一个学习模型后，再使用少部分序列 - 功能数据进行有监督的学习，对前期生成的学习模型进行有偏好性的调节。无监督的学习在整个机器学习过程中都是机器自动学习序列数据内部的特征，自动生成计算模型。前文提到的 UniRep 和 ProteinGAN 就是典型的无监督学习过程，以这两个模型为基础，可以预测或生成具有特定功能特征的新序列，同时在高活性新突变体产生上也展示出了潜在的应用价值；此外，依据进化背景信息的深度学习框架 ECNet[68] 和基于自然语言处理模型的 MT-LSTM[69] 均在蛋白质序列 - 功能关系的构建和预测上具有成功的案例，为机器学习指导高活性新酶的设计提供了实践基础。

5　总结与展望

酶改造设计技术已经成为合成生物学发展必不可少的工具，酶作为生命活动的主要践行者，它的催化活性和酶学性质极大地影响其在合成生物学中的应用。随着蛋白质分子改造技术的不断完善、蛋白质数据库的不断丰富以及计算科学的快速发展，酶改造设计技术经过了一段快速发展的时期，期间涌现出了定向进化、理性及半理性设计、新酶设计等方法，这些方法极大促进了酶改造设计技术的进步，同时也使合成生物学有了长足的发展。

但是，我们也要看到当前酶改造设计技术的限制，例如采样空间还是相对较大、实验验证成本高昂、对高通量筛选技术的依赖性较高、对蛋白质结构信息要求较高、新酶设计的成功率较低且活性较差等。这也为酶改造设计技术在未来的发展提出了更高的要求。值得庆幸的是，伴随着数据科学和人工智能的发展，深度学习方法逐渐被应用到对酶改造设计技术的改进上，相信在不久的将来，与深度学习相结合的酶改造设计技术必将迎来快速的发展，也必将极大地促进合成生物学的进展。

[参　考　文　献]

[1] Leung DW. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique, 1989, 1: 11-5

[2] Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. Genome Res, 1992, 2: 28-33

[3] Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA, 1993, 90: 5618-22

[4] Hutchison CA, Phillips S, Edgell MH, et al. Mutagenesis at a specific position in a DNA sequence. J Biol Chem, 1978, 253: 6551-60

[5] Gutierrez EA, Mundhada H, Meier T, et al. Reengineered glucose oxidase for amperometric glucose determination in diabetes analytics. Biosens Bioelectron, 2013, 50: 84-90

[6] Marton A, Delbecchi L, Bourgaux P. DNA nicking favors PCR recombination. Nucleic Acids Res, 1991, 19: 2423-6

[7] Stemmer WP. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 1994, 370: 389-91

[8] Shi C, Lu X, Ma C, et al. Enhancing the thermostability of a novel β-agarase AgaB through directed evolution. Applied Biochem Biotechnol, 2008, 151: 51-9

[9] Wong TS, Zhurina D, Schwaneberg U. The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen, 2006, 9: 271-88

[10] Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci USA, 2000, 97: 10701-5

[11] Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? Trends Biotechnol, 2005, 23: 231-7

[12] Yoshikuni Y, Ferrin TE, Keasling JD. Designed divergent evolution of enzyme function. Nature, 2006, 440: 1078-82

[13] Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. System Biol, 1971, 20: 406-16

[14] Liu Y, Yan Z, Lu X, et al. Improving the catalytic activity of isopentenyl phosphate kinase through protein coevolution analysis. Sci Rep, 2016, 6: 24117

[15] Gumulya Y, Gillam EM. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. Biochem J, 2017, 474: 1-19

[16] Gumulya Y, Baek JM, Wun SJ, et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. Nat Catal, 2018, 1: 878-88

[17] Alcolombri U, Elias M, Tawfik DS. Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. J Mol Biol, 2011, 411: 837-53

[18] Howard CJ, Hanson-Smith V, Kennedy KJ, et al. Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity. Elife, 2014, 3: e04126

[19] O'Maille PE, Bakhtina M, Tsai MD. Structure-based combinatorial protein engineering (SCOPE). J Mol Biol, 2002, 321: 677-91

[20] Reetz MT, Bocola M, Carballeira JD, et al. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. Angewe Chem Int Ed Engl, 2005, 44: 4192-6

[21] Marienhagen J, Dennig A, Schwaneberg U. Phosphorothioate-based DNA recombination: an enzyme-free method for the

combinatorial assembly of multiple DNA fragments. Biotechniques, 2012, 52: 1-6
[22] Richter F, Leaver-Fay A, Khare SD, et al. *De novo* enzyme design using Rosetta3. PLoS One, 2011, 6: e19230
[23] Schymkowitz J, Borg J, Stricher F, et al. The FoldX web server: an online force field. Nucleic Acids Res, 2005, 33: W382-8
[24] Wijma HJ, Floor RJ, Jekel PA, et al. Computationally designed libraries for rapid enzyme stabilization. Protein Eng Des Sel, 2014, 27: 49-58
[25] Kuipers RK, Joosten HJ, van Berkel WJ, et al. 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. Proteins, 2010, 78: 2101-13
[26] Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic Acids Res, 2009, 37: W376-83
[27] LinWu SW, Wu CA, Peng FC, et al. Structure-based development of bacterial nitroreductase against nitro-benzodiazepine-induced hypnosis. Biochem Pharmacol, 2012, 83: 1690-9
[28] Pikkemaat MG, Linssen AB, Berendsen HJ, et al. Molecular dynamics simulations as a tool for improving protein stability. Protein Eng, 2002, 15: 185-92
[29] Bernhardt P, Hult K, Kazlauskas RJ. Molecular basis of perhydrolase activity in serine hydrolases. Angew Chem Int Ed Engl, 2005, 117: 2802-6
[30] Yin DL, Bernhardt P, Morley KL, et al. Switching catalysis from hydrolysis to perhydrolysis in *Pseudomonas fluorescens* esterase. Biochemistry, 2010, 49: 1931-42
[31] Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem, 2010, 79: 413-44
[32] Chin JW. Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem, 2014, 83: 379-408
[33] Agostini F, Völler JS, Koksch B, et al. Biocatalysis with unnatural amino acids: enzymology meets xenobiology. Angew Chem Int Ed Engl, 2017, 56: 9680-703
[34] Mukai T, Lajoie MJ, Englert M, et al. Rewriting the genetic code. Annu Rev Microbiol, 2017, 71: 557-77
[35] Hoesl MG, Acevedo-Rocha CG, Nehring S, et al. Lipase congeners designed by genetic code engineering. ChemCatChem, 2011, 3: 213-21
[36] Budisa N, Wenger W, Wiltschi B. Residue-specific global fluorination of *Candida antarctica* lipase B in *Pichia pastoris*. Mol BioSyst, 2010, 6: 1630-9
[37] Gilles A, Marliere P, Rose T, et al. Conservative replacement of methionine by norleucine in *Escherichia coli* adenylate kinase. J Biol Chem, 1988, 263: 8204-9
[38] Cirino PC, Tang Y, Takahashi K, et al. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. Biotechnol Bioeng, 2003, 83: 729-34
[39] Key HM, Dydio P, Clark DS, et al. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. Nature, 2016, 534: 534-7
[40] Xiong P, Wang M, Zhou X, et al. Protein design with a comprehensive statistical energy function and boosted by experimental selection for foldability. Nat Commun, 2014, 5: 5330
[41] Xiong P, Hu X, Huang B, et al. Increasing the efficiency and accuracy of the ABACUS protein sequence design method. Bioinformatics, 2020, 36: 136-44
[42] 操帆, 陈耀晞, 缪阳洋, 等. 蛋白质计算设计: 方法和应用展望. 合成生物学, 2021, 2: 15
[43] Cui Y, Chen Y, Liu X, et al. Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. ACS Catal, 2021, 11: 1340-50
[44] Zhou X, Xiong P, Wang M, et al. Proteins of well-defined structures can be designed without backbone readjustment by a statistical model. J Struct Biol, 2016, 196: 350-7
[45] Song Y, DiMaio F, Wang RY-R, et al. High-resolution comparative modeling with RosettaCM. Structure, 2013, 21: 1735-42
[46] Combs SA, DeLuca SL, DeLuca SH, et al. Small-molecule ligand docking into comparative models with Rosetta. Nat Protoc, 2013, 8: 1277-98
[47] Siegel JB, Zanghellini A, Lovick HM, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science, 2010, 329: 309-13
[48] Röthlisberger D, Khersonsky O, Wollacott AM, et al. Kemp elimination catalysts by computational enzyme design. Nature, 2008, 453: 190-5
[49] Jiang L, Althoff EA, Clemente FR, et al. *De novo* computational design of retro-aldol enzymes. Science, 2008, 319: 1387-91
[50] Silva DA, Yu S, Ulge UY, et al. *De novo* design of potent and selective mimics of IL-2 and IL-15. Nature, 2019, 565: 186-91
[51] Xu C, Lu P, El-Din T, et al. Computational design of transmembrane pores. Nature, 2020, 585: 129-34
[52] Mulligan VK, Workman S, Sun T, et al. Computationally designed peptide macrocycle inhibitors of New Delhi metallo-β-lactamase 1. Proc Natl Acad Sci USA, 2021, 118: e2012800118
[53] Marcos E, Chidyausiku TM, McShan AC, et al. *De novo* design of a non-local β-sheet protein with high stability and accuracy. Nat Struct Mol Biol, 2018, 25: 1028-34
[54] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583-9
[55] Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a 3-track network. Science, 2021, 373: 871-6
[56] Aziz R, Verma C, Srivastava N. Artificial neural network classification of high dimensional data with novel optimization approach of dimension reduction. Ann Data Sci, 2018, 5: 615-35
[57] Basanta B, Bick MJ, Bera AK, et al. An enumerative algorithm for *de novo* design of proteins with diverse pocket structures. Proc Natl Acad Sci USA, 2020, 117: 22135-45
[58] Alley EC, Khimulya G, Biswas S, et al. Unified rational

protein engineering with sequence-based deep representation learning. Nat Methods, 2019, 16: 1315-22
[59] Repecka D, Jauniskis V, Karpus L, et al. Expanding functional protein sequence spaces using generative adversarial networks. Nat Mach Intell, 2021, 3: 324-33
[60] Siedhoff NE, Schwaneberg U, Davari M. Machine learning-assisted enzyme engineering. Methods Enzymol, 2020, 643: 281-315
[61] Chaparro-Riggers JF, Polizzi KM, Bommarius AS. Better library design: data-driven protein engineering. Biotechnol J, 2007, 2: 180-91
[62] Fox RJ, Davis SC, Mundorff EC, et al. Improving catalytic function by ProSAR-driven enzyme evolution. Nat Biotechnol, 2007, 25: 338-44
[63] Sarkisyan KS, Bolotin DA, Meer MV, et al. Local fitness landscape of the green fluorescent protein. Nature, 2016, 533: 397-401
[64] Yang G, Anderson DW, Baier F, et al. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. Nat Chem Biol, 2019, 15: 1120-8
[65] Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. Nat Methods, 2019, 16: 687-94
[66] Wu Z, Kan SBJ, Lewis RD, et al. Machine learning-assisted directed protein evolution with combinatorial libraries. Proc Natl Acad Sci USA, 2019, 116: 8852-8
[67] Li G, Dong Y, Reetz M. Can machine learning revolutionize directed evolution of selective enzymes? Adv Synth Catal, 2019, 361: 2377-86
[68] Luo Y, Jiang G, Yu T, et al. ECNet is an evolutionary context-integrated deep learning framework for protein engineering. Nat Commun, 2021, 12: 5743
[69] Bepler T, Berger B. Learning the protein language: evolution, structure, and function. Cell Syst, 2021, 12: 654-69.e3