

· 生物元器件智能设计合成 ·

唐双焱 中国科学院微生物研究所研究员，现任微生物生理与代谢工程研究室副主任。主要从事合成生物学、代谢工程和蛋白质工程研究。擅长高通量筛选方法的设计，建立并发展了通过挖掘和改造转录调控因子定制小分子化合物感应/调控元件的技术平台，并成功将元件应用到高通量筛选、动态调控、生物标记物和农残检测以及跨膜转运蛋白筛选等领域。擅长生物合成途径的胞内进化和底盘细胞定向基因组进化技术，快速挖掘并解析与细胞工厂适配性相关的重要基因靶点，为细胞工厂理性设计提供了理论与技术基础。相关成果发表在 *JACS*、*Angewandte Chemie International Edition*、*Nature Communications*、*Metabolic Engineering* 等国际期刊上，同时获批了通过改造获得的高效酶和全细胞催化剂的专利。



基于转录调控蛋白的生物传感器的智能设计构建

梁朝宁，向腊，唐双焱*

中国科学院微生物研究所 微生物多样性与资源创新利用全国重点实验室 微生物生理与代谢工程研究室，北京 100101

梁朝宁，向腊，唐双焱. 基于转录调控蛋白的生物传感器的智能设计构建[J]. 生物工程学报, 2025, 41(3): 1011-1022.

LIANG Chaoning, XIANG La, TANG Shuangyan. Intelligent design of transcription factor-based biosensors[J]. Chinese Journal of Biotechnology, 2025, 41(3): 1011-1022.

摘要：基于转录调控蛋白的生物传感器已在代谢工程、合成生物学、代谢物监测等领域发挥重要作用。该工具具有高度模块化、正交性、易于构建及操作等优良特性，但在实际应用中，天然调控蛋白对目标化合物的响应仍然存在响应程度较低、特异性不符合需求等缺陷。本文综述了近年来在计算机模拟和人工智能技术辅助下，采用蛋白质工程手段，智能设计、改造转录调控蛋白以提升其工作性能所取得的研究进展。主要包括利用蛋白质结构预测、配体结合模拟等优化改造策略快速获取目标突变体，或借助对突变体数据分析并经机器学习等构建数学模型，预测转录调控蛋白突变响应效果等。相较于传统方式，计算机模拟和人工智能辅助技术可实现对生物元件更精准快捷地设计构建，将推动新型生物传感器的创制研发，更好地满足实际应用的需求。

关键词：转录调控蛋白；生物传感器；计算模拟；蛋白质工程；机器学习

资助项目：国家重点研发计划(2021YFC2103901)；中国科学院科研仪器设备研制项目(YJKYYQ20210032)

This work was supported by the National Key Research and Development Program of China (2021YFC2103901) and the Instrument Developing Project of Chinese Academy of Science (YJKYYQ20210032).

*Corresponding author. E-mail: tangsy@im.ac.cn

Received: 2024-07-25; Accepted: 2024-09-06; Published online: 2024-09-09

Intelligent design of transcription factor-based biosensors

LIANG Chaoning, XIANG La, TANG Shuangyan*

Department of Microbial Physiological & Metabolic Engineering, State Key Laboratory of Microbial Diversity and Innovative Utilization, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Abstract: Transcription factor (TF)-based biosensors have been widely applied in metabolic engineering, synthetic biology, metabolites monitoring, etc. These biosensors are praised for the high orthogonality, modularity, and operability. However, most natural TFs with weak responses and low specificity still demand optimization for desired performance in applications. Herein, we comprehensively summarize the recent advances in the engineering and optimization of TF-based biosensors with the assistance of computational simulation and artificial intelligence. This review includes the regulatory protein engineering aided by protein structure prediction and ligand binding simulation and the regulatory protein responses predicted by a mathematical model obtained from machine learning of mutagenesis data. In comparison with conventional tools, computational simulation and artificial intelligence enable more accurate and rapid design and construction of biosensors. Thus, these technologies will greatly promote the development of novel biosensors for applications.

Keywords: transcription factor; biosensor; computational simulation; protein engineering; machine learning

随着合成生物学和代谢工程等生物技术的发展,众多生物元器件在生物制造、医疗监测、环境污染监控等领域得到广泛应用。其中,可识别并响应化学或物理等信号的生物传感器,作为新型元器件应运而生。生物传感器主要由生物识别元件和信号输出元件这 2 部分构成。生物识别元件的类型包括转录调控蛋白、核糖体开关等^[1]。其中,基于转录调控蛋白的生物传感器因为具有高度模块化、正交性、易于构建及操作等优良特性,已在工程菌株优化^[2-3]、动态调控^[4-6]、高通量筛选^[7-8]、环境污染监控^[9]等领域发挥了重要作用。尤其是在大容量工程菌株文库的高通量筛选上,基于转录调控蛋白的生物传感器技术与流式细胞分选、微流控等技术结合,可突破物理、化学等常规分析检测方案的限速瓶颈,推动了“设计-构建-测试-学习”循环的进程^[10-11]。

这一类型生物传感器的核心元件转录调控蛋白广泛存在于自然界中。转录调控蛋白在转录水平调控目标蛋白表达的机制为:转录调控蛋白通过特异性结合配体化合物引发自身的构象变化,使其解离或与特定的启动子 DNA 序列结合,从而促进或干扰 RNA 聚合酶与启动子 DNA 的结合,起到激活或阻遏下游目标基因表达的目的。转录调控蛋白通常由 3 部分构成,分别是 DNA 结合区、配体结合区以及两者间的连接区。转录调控蛋白自身性质决定了生物传感器的调控性能,如配体识别特异性、灵敏度、动态范围及可操作区间等(图 1)。

尽管高度模块化使得转录调控蛋白具有“即插即用”的特性,便于构建应用于不同场景的生物传感器,同时挖掘和表征的天然转录调控蛋白的工作也已取得长足进展,但在实际应用中,天然调控蛋白的各方面特性尚存在不足

之处^[12-14],尤其是对于众多的目标化合物来说,仍然缺乏对其进行可特异性识别响应的天然转录调控蛋白。因此,研究者们通过蛋白质工程等策略“量身定制”转录调控蛋白突变体来满足众多不同特异性生物传感器的设计需求。经由人工智能与计算机辅助,包括蛋白质结构预测、配体结合模拟或者借助数学建模的机器学习以及深度学习,可快速精准实现转录调控蛋白性能改造,以满足不同应用需求。本文将对近年来这一领域取得的研究进展进行综述(表 1)。

1 改造转录调控蛋白的配体识别特异性

转录调控蛋白具有高度模块化、可塑性等

特性,其配体识别特异性具备被重塑的潜力。体现转录调控蛋白诱导特性最关键的性能参数即配体识别特异性,也是设计新型生物传感器的主要限制因素。因此,近年来有多项研究着力于采用理性设计、定向进化等技术拓展转录调控蛋白的配体识别特异性。但改变配体识别特异性同时也可能扰乱转录调控蛋白结构的别构效应,影响其调控功能,是设计新型生物传感器的难点。

蛋白结构信息有助于理性或半理性设计,尤其适用于目标化合物与天然可识别配体结构相似的转录调控蛋白改造。结合流式细胞分选等高通量筛选手段,可大概率获得符合预期的突变体^[2,7,12-13,27-29](图 2A)。

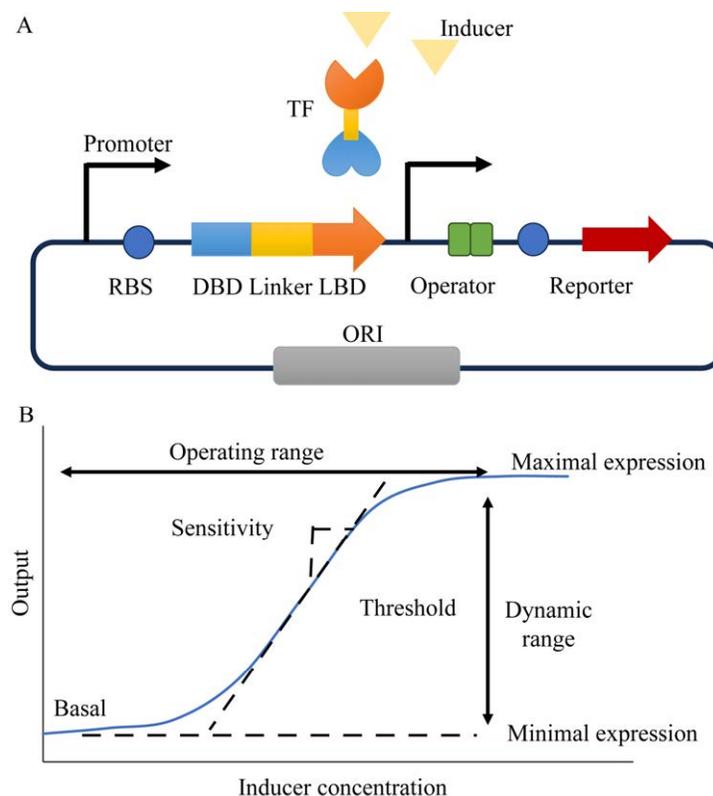


图 1 基于转录调控蛋白生物传感器工作原理(A)及常用性能参数(B)^[10] TF: 转录调控蛋白; RBS: 核糖体结合位点; DBD: DNA 结合区; LBD: 配体结合区; ORI: 复制起始位点。

Figure 1 Working mechanisms (A) and performance parameters (B) of transcription factor-based biosensors^[10]. TF: Transcription factor; RBS: Ribosome binding site; DBD: DNA binding domain; LBD: Ligand binding domain; ORI: Replication origin.

鉴于别构调节的复杂性, 计算机辅助设计调控蛋白在很长时间内存在较大困难。多数情况下, 蛋白设计的准确性依赖于目标蛋白与已知结构蛋白的序列相似性, 而通过比较建模 (comparative modelling) 策略构建蛋白结构模型, 可为序列相似性较低的蛋白建模, 为研究蛋白质-配体化合物结合方式提供参考。在 Jha 等^[15]的研究中, 尽管待改造转录调控蛋白 PobR 与用于建模的红球菌属 (*Rhodococcus*) 来源的模板蛋白序列相似性仅为 36%, 比较建模策略仍实现了可信度较高的蛋白结构模型的构建。将模型与目标配体化合物进行分子对接, 并据此对 PobR 进行改造, 获得的 PobR 突变体不仅可响应新的配体化合物 3,4-二羟基苯甲酸, 对天然配体化合物 4-羟基苯甲酸的响应强度亦有所提升。因此, 在蛋白结构信息欠缺时, 该蛋白

建模策略值得借鉴。

TetR 家族转录调控蛋白 QacR 具有 2 个不同的配体结合区域, 可以结合一系列结构相似性较小的配体化合物。在 de los Santos 等^[16]的研究中, 通过结合或未结合配体化合物时 QacR 蛋白的结构比对, 发现配体化合物进出结合口袋引发的构象变化直接影响 QacR 与 DNA 的结合或解离; 在计算机辅助下预测了 QacR 蛋白与目标配体香草醛的复合物结构, 并对其结合口袋中与香草醛分子距离 15 Å 范围内的氨基酸位点进行了突变, 通过体外筛选系统的筛选, 最终获得了 2 个胞内和胞外皆可响应香草醛的 QacR 突变体。

Church 团队^[17]通过 Rosetta 算法设计, 对乳糖操纵子调控蛋白 LacI 进行定点饱和突变以及随机突变, 获得了一系列突变体, 可分别与海藻糖、乳糖醇或三氯蔗糖结合, 这 3 个化合

表 1 近年来基于转录调控蛋白生物传感器智能设计研究小结

Table 1 The summary of current researches on computer-aided protein engineering of transcription factor-based biosensors

Year	TF	Host strain	AI assistance	Target	References
2015	PobR	<i>Escherichia coli</i>	Rossetta, comparative modelling	Ligand specificity	[15]
2015	QacR	Cell free	Protein structure prediction	Ligand specificity	[16]
2016	LacI	<i>E. coli</i>	Protein structure prediction	Ligand specificity	[17]
2017	LasR	–	Visual screening	–	[18]
2020	BenR, XylS	<i>E. coli</i>	Homology modelling, docking	Ligand specificity	[19]
2020	LysG	<i>E. coli</i>	Molecular dynamics simulation	Ligand specificity	[13]
2020	PcaV	<i>E. coli</i>	Design of experiments	Optimization, modelling	[20]
2020	DmpR	<i>E. coli</i>	Machine learning	Dynamic range	[21]
2020	CdaR	<i>E. coli</i>	Machine learning	Dynamic range	[22]
2021	LuxR	<i>E. coli</i>	Molecular dynamics simulation	Sensitivity, stability, etc.	[23]
2021	AR	–	Visual screening		[24]
2022	FapR	<i>Saccharomyces cerevisiae</i>	Machine learning	Dynamic range	[25]
2023	LysR, BenM	<i>Corynebacterium glutamicum</i>	Docking, molecular simulation	Dynamic range, ligand specificity	[26]

物与 LacI 的天然配体 IPTG 分子结构差异较大；通过对突变热点氨基酸的聚类分析，评估其在配体化合物识别中所扮演的角色，如突变位点 I79S 和 I79T 仅针对乳糖醇的识别，而 N246D 和 Y273H 突变与特异性识别三氯蔗糖相关等；该研究认为，定点饱和突变或随机突变策略能够改造转录调控蛋白的识别特异性，但利用计算机辅助手段则更有助于打造可响应与原配体在化学结构上差异较大的新配体的转录调控蛋白突变体。

分子对接广泛应用于预测小分子化合物和蛋白的结合模式。通过分子对接可以表征蛋白质-配体化合物结合强度以及配体化合物与关

键残基相互作用方式。序列相似性高达 60% 的同家族转录调控蛋白 BenR 和 XylS 的诱导特异性差别较大，前者仅可识别苯甲酸盐，作用于 Pm/Pb 这 2 种启动子；后者虽可被多种结构相似芳香族化合物诱导，但仅可识别 Pm 启动子^[19]。有研究^[19]通过对 BenR 中与配体结合相关的 N 端进行同源建模和分子对接，推测位于蛋白表面的空腔可能参与其与天然配体苯甲酸盐的结合，同时推断其结合配体的选择性受到配体化合物分子空间位阻(size exclusion)的制约。基于该推断对 XylS 的配体结合口袋表面非保守氨基酸 His28 和 Val111 进行突变，进一步扩大了 XylS 结合口袋，使其可结合并响应甲基化或卤代芳香族类似物。

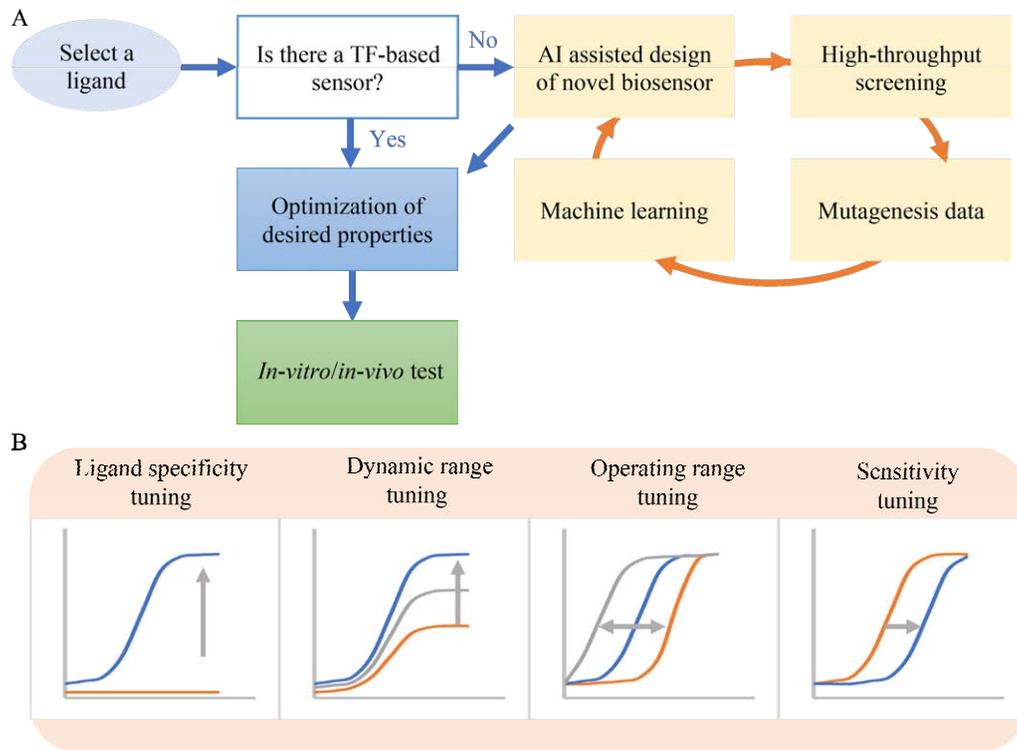


图 2 计算机辅助设计基于转录调控蛋白生物传感器的工作流程图(A)及主要改造性能(B)^[10] 以不同颜色曲线示意诱导特性的优化方向。

Figure 2 Computer-aided design of transcription factor-based biosensor. A: The workflow of the designing-building-testing-learning; B: The options of altering the biosensor response curve through protein engineering^[10]. The colors of the curves indicate the trend of the optimizations.

利用改变配体识别特异性的转录调控蛋白可以在实际应用中实现生物传感器的定制,并被应用于高通量筛选工具或代谢通路中动态调控元件中,在生物制造领域发挥重要作用^[7,26-29],如利用改变配体特异性的 VanR 突变体构建生物传感器,可分别实现对香草醛生物合成中全细胞催化剂的优化以及羟基酪醇生物合成中对限速酶定向进化筛选;利用 HucR 突变体构建的动态调控元件调控合成途径表达,可缓解香草醛生物合成中细胞毒性较高的香草醛对细胞工厂的阻滞;利用 LysR 突变体构建的生物传感器可应用于高产 L-赖氨酸高产菌株的高通量筛选。

2 提升天然转录调控蛋白的诱导特性

评价转录调控蛋白的诱导效果有多项指标,包括诱导剂响应浓度的可操作区间(operating range)、响应动态范围(dynamic range)、最高/最低(本底)输出信号值、灵敏度、诱导倍数、诱导特异性等(图 1B)。针对天然转录调控蛋白存在的诱导性能缺陷,研究者尝试通过设计、改造等方式优化诱导特性,打造更具有应用价值的非天然调控元件(图 2B)。

Li 等^[23]以根瘤农杆菌 (*Agrobacterium tumefaciens*)菌来源的 TraR (PDB ID: 1L3L)结构为模板,对费氏弧菌(*Vibrio fischeri*)来源的同家族群感效应调控蛋白 LuxR 进行同源建模和分子动力学分析;采用 Molecular Mechanics Generalized Born Surface Area (MM-GB/SA)方法计算 LuxR 和配体化合物 3OC6-HSL 的结合自由能,推测位于结合口袋 8 Å 之外的 I46 位点对配体化合物结合起关键作用;利用携带 I46F 突变的 LuxR 突变体构建全细胞生物传感器,其表现出对配体化合物更为灵敏、稳定的响应。

以原核生物中最大的转录调控因子家族——

LysR 型转录调控蛋白(LysR-type transcription regulator)家族的 LysG 为研究对象,研究人员通过分子对接、分子动力学模拟(molecular dynamics simulation)等技术解析配体结合域与 DNA 结合域之间铰链区的调控功能及调控机制^[13];通过对仅包含约 30 个氨基酸的铰链区进行定向进化,获得了响应强度、诱导倍数等诱导性能显著提升的突变体;该策略亦可拓展至同家族另一转录调控蛋白 BenM,同样取得了提升诱导特性的效果;因此,该研究认为,从优化铰链区氨基酸入手改造转录调控蛋白诱导性能,相较于针对建库筛选难度较大的配体结合域的改造,此策略更易快速、便捷获得具有优良性能的突变体。

3 计算机模拟在基于转录调控蛋白生物传感器设计中的应用

3.1 计算机模拟在设计改造基于转录调控蛋白生物传感器中的应用实例

改造转录调控蛋白所获取的数据,凭借现有高通量测序技术和机器学习工具,可推动“设计-构建-测试-学习”的循环,获取丰富的蛋白序列与构效关系数据信息,基于决策系统自动预测配体化合物结构或浓度,为高效设计、构建新型转录调控蛋白元件奠定基础(图 2A)。

研究者用实验设计法(design of experiments, DOE)实现了多维度表征生物传感器元件参数,如转录调控蛋白启动子强度、信号输出启动子强度及信号输出核糖体结合位点强度等^[20];通过打分机制,实现以最少实验次数优化生物传感器参数,如提升最大输出信号强度、扩大动态响应范围、拓宽响应浓度范围、提高灵敏度等,实现生物传感器设计优化的模块化操作,大大提升了高质量生物传感器的设计效率。

另一项研究利用蛋白质工程手段,结合生

物传感器矩阵和机器学习处理海量数据的优势,训练机器学习生成精准测试多种酚类化合物的算法^[21]。针对 DmpR 配体结合域构建随机突变文库,采用流式细胞分选高通量富集被苯酚诱导、荧光表达较高的细胞;经过两轮定向进化获得了多种类型被苯酚诱导的 DmpR 突变体。经矩阵测试野生型 DmpR 及 5 种突变体对系列浓度下 11 种酚类化合物的诱导响应效果,该数据经机器学习后,可用于预测该生物传感器的响应效果,该方法鉴定酚类化合物的阳性率高达 95.3%;相较于基于野生型 DmpR 的生物传感器,高分辨率蛋白质工程技术将酚类化合物检测下限提高 75 倍^[21]。机器学习技术与单一转录因子驱动的进化生物传感器阵列相结合,将为分子指纹识别技术开辟新途径。

利用生物传感器可将细胞内代谢物浓度转化为输出(如报告基因)信号,但代谢物浓度-信号强度剂量关系仍不明确。研究者利用 DNA 可追踪组装法,结合荧光信号细胞分选和测序技术,以丙二酰辅酶 A 生物传感器为例,在酿酒酵母中构建并表征包含 5 184 种组合的生物传感器文库,包含 6 种转录调控因子表达水平、4 种操纵基因位置和 216 种可能的上游增强子(upstream enhancer sequence, UAS)的组合^[25]。数据通过机器学习后,用于预测未知的基因型-表型关系。借助该流程,研究者成功获得了具有最大动态响应范围的丙二酰辅酶 A 生物传感器。此外,该研究通过设计构建、调整和分析生物传感器诱导剂量曲线平台,推进生物元件的理性设计^[25]。

生物传感器系统中,分别控制转录调控蛋白和报告基因翻译水平的核糖体结合位点的组合(定义为 cross-RBS,简称 cRBS),对生物传感器性能影响较大。Deng 研究团队^[22]基于 RBS 大数据文库开发可深度学习模型 Classification

model between cRBSs and average dynamic range of each sub-library (CLM-RDR)。以响应葡萄糖二酸的转录调控蛋白 CdaR 为研究对象,构建采用 81 种不同 cRBS 组合调控 CdaR 和报告基因 sfGFP 表达的生物传感器,结果显示 cRBS 组合对改善生物传感器响应动态范围贡献显著^[22]。该研究显示,RBS 区域不仅影响蛋白质翻译水平,还影响其折叠状态。根据 RBS 设计原则,结合流式细胞分选和二代测序技术,该研究对 7 053 种 cRBS 序列数据通过卷积神经网络模型进行了分析和深度学习,最终构建了预测性能优越的 cRBS 组合和生物传感器响应动态范围相互关联的分类模型(CL M-RDR 模型),实现通过 RBS 对生物传感器动态范围进行智能调控^[22]。

对于深度学习来说,足够容量的数据集至关重要。设计生物传感器,对其进行改造,经流式细胞分选或微流控技术等手段进行高通量筛选,结合数据收集和高通量测序技术,有助于突破连接机器学习和生物传感器数据收集的瓶颈。与传统实验室检测相比,将机器学习构建的函数应用于生物传感器预测及设计,提高其预测的可靠性和准确性,将极大推动该领域的高速发展(图 2)。

3.2 常用计算工具

计算工具在调控蛋白的改造中发挥了重要作用,不仅能够帮助深入理解调控蛋白的结构与功能关系,指导实验设计和改造策略,也为更高效、更精确的蛋白设计提供了可能(图 3)。

3.2.1 分子动力学模拟

分子动力学模拟是一种通过模拟体系原子和分子运动轨迹研究体系微观和宏观性质的计算机模拟方法,是研究分子体系的重要手段。根据经典力学,只要给定分子体系中原子和分子之间的相互作用,就可以利用牛顿运动方程

求解分子或原子的运动轨迹, 然后使用一定的统计方法计算出系统的力学、热力学、动力学性质。分子动力学模拟是研究蛋白在水溶液中的稳定性和动力学特征的重要方法, 有助于理解生物大分子结构-功能关系。调控蛋白的构象转变是其发挥功能的关键。Qing 等^[30]利用分子动力学模拟研究调控蛋白 PhoP 活性和非活性的同源二聚体, 揭示其二聚体相互作用机制和潜在结合区, 为理性设计新型抗菌药物分子提供依据。Kamaraj 等^[31]通过分子动力学模拟, 研究了干扰素调节因子与 DNA 识别的分子机制。Frezza 等^[32]利用分子动力学模拟研究转录调控蛋白 GabR 的不对称形式, 揭示了活性位点之间可能存在的变构干扰。

3.2.2 量子力学/分子力学计算

量子计算作为一种利用量子力学现象的新兴技术, 被广泛应用于生化系统、生化化学、蛋白-配体相互作用、蛋白折叠等方面, 与传统的计算相比, 量子计算速度更快, 广泛用于酶结构和机制的计算建模。Morales 等^[33]使用量子力学/分子力学优化内部潜在能量, 以确定 Mg^{2+} 离子与调控蛋白 PhoP 蛋白结构之间的相互作用。

de Carvalho Matias 等^[34]利用量子化学方法研究了 TtgR 与不同小分子的偶联谱, 量子计算结果显示 TtgR 与不同小分子的收敛半径、总结合能、配体区域的相关性(能量)、最相关的氨基酸残基, 提高了对 TtgR 结合配体机制的理性认识。

3.2.3 分子对接和虚拟筛选

分子对接是广泛应用于预测配体结合模式和基于结构的虚拟筛选的建模工具。通过分子对接, 筛选得到对调控蛋白亲和力高的配体分子, 表征调控蛋白-配体结合强度并识别其与关键残基之间的相互作用。Patel 等^[35]通过分子对接研究了调控蛋白 HbpR 和 PhnR 对不同芳香烃污染物的相互作用, 表明利用这些调控蛋白能够实现芳香烃污染物的监测。虚拟筛选可实现在生物活性筛选实验前, 利用化学结构数据库的计算分析, 为特定的调控蛋白或受体确定可能的候选配体, 从而降低实际筛选化合物的数目。将该技术与分子对接技术相结合, 可降低实验成本、增加筛选量、缩短实验周期, 因此该技术已在药物筛选中被广泛应用。利用这一策略, Kalia 等^[18]从 2 603 种天然化合物中筛

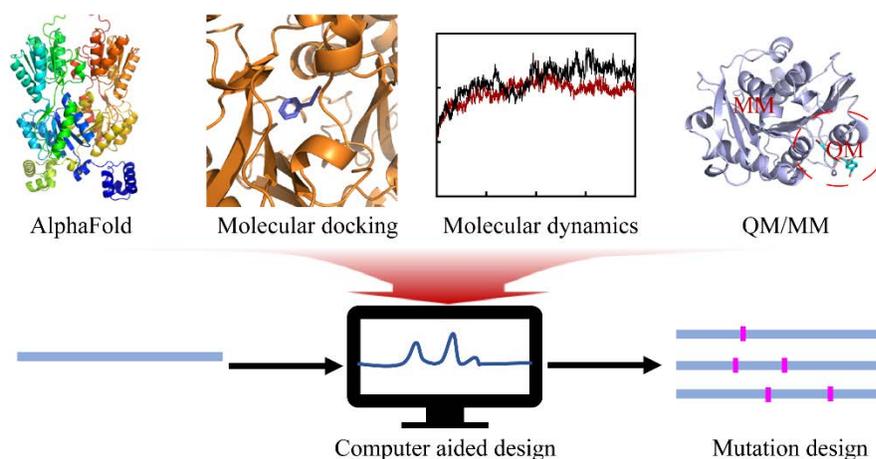


图3 转录调控蛋白设计改造常用工具 QM/MM: 量子力学/分子力学。

Figure 3 Tools for designing transcription factor-based biosensor. QM/MM: Quantum-mechanics/molecular-mechanics.

选出 6 种可抑制铜绿假单胞菌(*Pseudomonas aeruginosa*)菌群感效应调控蛋白 LasR 的潜在配体。Pang 等^[24]利用基于分子对接的虚拟筛选策略, 筛选获得新型雄性激素受体(androgen receptor, AR)的拮抗剂, 能够阻断 AR 的 DNA 结合结构域与 DNA 序列结合, 并对 AR 转录活性表现出良好抑制效果。

3.2.4 蛋白结构预测工具

AlphaFold 是 DeepMind 公司基于人工智能技术开发的一款用于预测蛋白质三维结构的程序^[36]。于 2020 年问世的 AlphaFold 极大地提高了蛋白质结构模拟的准确性。此后升级的 AlphaFold 2 能根据蛋白质的氨基酸序列预测其三维结构, 为突变设计提供重要的结构信息。结合分子动力学模拟等计算, 对调控蛋白的分子机理研究、蛋白设计等提供了重要的辅助。随后的 AlphaFold-Multimer 推动了对蛋白质-蛋白质复合物的预测。而最新发布的 AlphaFold 3 能够准确预测蛋白质、DNA、RNA、小分子配体结构以及它们相互作用模式, 对结构预测的准确性有了更进一步的提升^[37]。AlphaFold 3 能预测蛋白质与其他蛋白质、核酸、小分子、离子、修饰蛋白质残基的复合物, 以及抗体-抗原的相互作用, 准确性显著超过当前的预测工具。RoseTTAFold 是 David Baker 团队开发的蛋白质预测工具, 不仅能预测单个蛋白质的结构, 还能预测蛋白质复合物^[38]。Uni-Fold 是北京深势科技有限公司推出的蛋白质结构预测工具^[39], 它在保持与 AlphaFold 2 相当的高预测精度的同时, 实现了计算效率显著提升。作为一款完全开源的软件, Uni-Fold 为研究人员提供了自由改造模型的空间。ESM-Fold 是另一种基于蛋白质语言模型的蛋白结构预测工具^[40], 它能够直接从蛋白质的单个序列进行高精度的三维结构预测。ESMFold 效率比 AlphaFold 2 高 1 个数量级。

2022 年 David Baker 团队推出蛋白质设计工具 RoseTTAFold Diffusion (RFDiffusion)^[41], 利用 RFDiffusion 可以设计出与真实蛋白质类似, 但是在自然界从未出现过的蛋白质。RFDiffusion 能够执行多种蛋白质设计任务, 如单体设计、寡聚体设计、结合位点设计等, 并能测试不同结构元素的设计组合, 从头开始产生全新蛋白质。

AlphaFold 和 RoseTTAFold 等都是蛋白质结构预测领域的突破性工具, 在纯蛋白质系统能高精度地预测蛋白质的结构, 但是在处理复杂生物环境中蛋白质时还存在一些局限。在自然界中, 蛋白质所处的环境是复杂的, 很少单独发挥作用, 蛋白质与其他蛋白质、核酸、小分子等相互作用, 这些相互作用会影响蛋白质的构象与功能。目前这些结构预测工具忽略了许多影响蛋白质结构的因素。另外, AlphaFold 3 尚未开源, 这限制了研究人员对其的改进和应用。RFDiffusion 在设计更大蛋白质方面仍面临挑战, 需要进一步地优化算法。

尽管目前存在这些局限, 这些工具在预测蛋白结构和蛋白设计中的应用前景仍然非常广阔, 通过精确预测转录调控蛋白与 DNA、配体结合的方式, 将推动精准设计新型转录调控蛋白、创建生物传感器的创新发展。

4 结语与展望

基于转录调控蛋白的生物传感器已在代谢工程、代谢物或污染物监测等领域取得广泛应用。针对天然转录调控蛋白存在的问题, 结合计算机辅助设计, 经由蛋白结构模拟、分子对接等技术, 辅以深度学习或机器学习建模等先进人工智能工具, 优化生物传感器性能已取得丰硕成果。相较于传统改造方法数据量少、通用性差、改造周期长等问题, 人工智能辅助技术可基于海量数据构建数学模型, 实现对生物

元件更精准、更快速地预测,必将推动大量性能优良的新型生物传感器的创制,丰富合成生物学元器件的知识储备。

最新推出的 AlphaFold 3 对结构预测的准确性有更进一步的提升;然而,对于蛋白质结构预测任务而言,静态的结构预测仍未解决生物体系中蛋白结构动态变化等问题,例如调控蛋白行使功能时所呈现的别构效应,亟需被精确模拟解析。目前,在 AlphaFold 3 中,上述问题依然存在。另外,在设计与特定配体结合的蛋白构象时,AlphaFold 3 也未必能提供符合预期的结果。

无论如何,利用计算机辅助模拟,智能化设计转录调控蛋白,构建新型生物传感器的时代已经到来。随着大数据信息的累积、人工智能技术的长足发展,基于蛋白结构的理性认知,结合现有机器学习、深度学习等方式,将深入探索转录调控蛋白的工作机制,为快捷便利获得性能优异的生物传感器元件奠定基础。

作者贡献声明

梁朝宁:方案设计、初稿写作、稿件润色修改;向腊:初稿写作、稿件润色修改;唐双焱:经费支持、监督指导、稿件润色修改。

作者利益冲突公开声明

作者声明没有任何可能会影响本文所报告工作的已知经济利益或个人关系。

REFERENCES

- [1] WANG QZ, TANG SY, YANG S. Genetic biosensors for small-molecule products: design and applications in high-throughput screening[J]. *Frontiers of Chemical Science and Engineering*, 2017, 11(1): 15-26.
- [2] LI H, LIANG CN, CHEN W, JIN JM, TANG SY, TAO Y. Monitoring *in vivo* metabolic flux with a designed whole-cell metabolite biosensor of shikimic acid[J]. *Biosensors and Bioelectronics*, 2017, 98: 457-465.
- [3] ROGERS JK, CHURCH GM. Genetically encoded sensors enable real-time observation of metabolite production[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(9): 2388-2393.
- [4] FARMER WR, LIAO JC. Improving lycopene production in *Escherichia coli* by engineering metabolic control[J]. *Nature Biotechnology*, 2000, 18(5): 533-537.
- [5] LIANG CN, ZHANG XX, WU JY, MU SS, WU Z, JIN JM, TANG SY. Dynamic control of toxic natural product biosynthesis by an artificial regulatory circuit[J]. *Metabolic Engineering*, 2020, 57: 239-246.
- [6] ZHANG FZ, CAROTHERS JM, KEASLING JD. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids[J]. *Nature Biotechnology*, 2012, 30(4): 354-359.
- [7] LI H, CHEN W, JIN RN, JIN JM, TANG SY. Biosensor-aided high-throughput screening of hyper-producing cells for malonyl-CoA-derived products[J]. *Microbial Cell Factories*, 2017, 16(1): 187.
- [8] ZHOU SH, LYU YB, LI HZ, KOFFAS MAG, ZHOU JW. Fine-tuning the (2S)-naringenin synthetic pathway using an iterative high-throughput balancing strategy[J]. *Biotechnology and Bioengineering*, 2019, 116(6): 1392-1404.
- [9] ÖZYURT C, ÜSTÜKARCI H, EVRAN S, TELEFONCU A. MerR-fluorescent protein chimera biosensor for fast and sensitive detection of Hg²⁺ in drinking water[J]. *Biotechnology and Applied Biochemistry*, 2019, 66(5): 731-737.
- [10] PHAM C, STOGIOS PJ, SAVCHENKO A, MAHADEVAN R. Advances in engineering and optimization of transcription factor-based biosensors for plug-and-play small molecule detection[J]. *Current Opinion in Biotechnology*, 2022, 76: 102753.
- [11] ROGERS JK, TAYLOR ND, CHURCH GM. Biosensor-based engineering of biosynthetic pathways[J]. *Current Opinion in Biotechnology*, 2016, 42: 84-91.
- [12] DELLA CORTE D, van BEEK HL, SYBERG F, SCHALLMEY M, TOBOLA F, CORMANN KU, SCHLICKER C, BAUMANN PT, KRUMBACH K, SOKOLOWSKY S, MORRIS CJ, GRÜNBERGER A, HOFMANN E, SCHRÖDER GF, MARIENHAGEN J. Engineering and application of a biosensor with focused ligand specificity[J]. *Nature Communications*, 2020, 11(1): 4851.
- [13] PU W, CHEN JZ, LIU P, SHEN J, CAI NY, LIU BY, LEI Y, WANG LX, NI XM, ZHANG J, LIU J, ZHOU YY, ZHOU WJ, MA HW, WANG Y, ZHENG P, SUN JB. Directed evolution of linker helix as an efficient strategy for engineering LysR-type transcriptional regulators as whole-cell biosensors[J]. *Biosensors and*

- Bioelectronics, 2023, 222: 115004.
- [14] XIONG DD, LU SK, WU JY, LIANG CN, WANG W, WANG WZ, JIN JM, TANG SY. Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor[J]. *Metabolic Engineering*, 2017, 40: 115-123.
- [15] JHA RK, CHAKRABORTI S, KERN TL, FOX DT, STRAUSS CEM. Rosetta comparative modeling for library design: engineering alternative inducer specificity in a transcription factor[J]. *Proteins*, 2015, 83(7): 1327-1340.
- [16] de LOS SANTOS EL, MEYEROWITZ JT, MAYO SL, MURRAY RM. Engineering transcriptional regulator effector specificity using computational design and *in vitro* rapid prototyping: developing a vanillin sensor[J]. *ACS Synthetic Biology*, 2016, 5(4): 287-295.
- [17] TAYLOR ND, GARRUSS AS, MORETTI R, CHAN S, ARBING MA, CASCIO D, ROGERS JK, ISAACS FJ, KOSURI S, BAKER D, FIELDS S, CHURCH GM, RAMAN S. Engineering an allosteric transcription factor to respond to new ligands[J]. *Nature Methods*, 2016, 13(2): 177-183.
- [18] KALIA M, SINGH PK, YADAV VK, YADAV BS, SHARMA D, NARVI SS, MANI A, AGARWAL V. Structure based virtual screening for identification of potential quorum sensing inhibitors against LasR master regulator in *Pseudomonas aeruginosa*[J]. *Microbial Pathogenesis*, 2017, 107: 136-143.
- [19] MONTEIRO LMO, ARRUDA LCM, SANCHES-MEDEIROS A, MARTINS-SANTANA L, ALVES LF, DEFELIPE L, TURJANSKI AG, GUAZZARONI MAE, de LORENZO V₁C, SILVA-ROCHA R. Reverse engineering of an aspirin-responsive transcriptional regulator in *Escherichia coli*[J]. *ACS Synthetic Biology*, 2019, 8(8): 1890-1900.
- [20] BEREPIKI A, KENT R, MACHADO LFM, DIXON N. Development of high-performance whole cell biosensors aided by statistical modeling[J]. *ACS Synthetic Biology*, 2020, 9(3): 576-589.
- [21] KIM H, SEONG W, RHA E, LEE H, KIM SK, KWON KK, PARK KH, LEE DH, LEE SG. Machine learning linked evolutionary biosensor array for highly sensitive and specific molecular identification[J]. *Biosensors and Bioelectronics*, 2020, 170: 112670.
- [22] DING NN, YUAN ZQ, ZHANG XJ, CHEN J, ZHOU SH, DENG Y. Programmable cross-ribosome-binding sites to fine-tune the dynamic range of transcription factor-based biosensor[J]. *Nucleic Acids Research*, 2020, 48(18): 10602-10613.
- [23] LI JY, LIU RC, CHEN YL, LIU SX, CHEN C, LIU TY, YANG S, ZHUANG YT, YANG RF, CUI YJ, SONG YJ, WANG T, TENG Y. Computer-aided rational engineering of signal sensitivity of quorum sensing protein LuxR in a whole-cell biosensor[J]. *Frontiers in Molecular Biosciences*, 2021, 8: 729350.
- [24] PANG JP, SHEN C, ZHOU WF, WANG YX, SHAN LH, CHAI X, SHAO Y, HU XP, ZHU F, ZHU DY, XIAO L, XU L, XU XH, LI D, HOU TJ. Discovery of novel antagonists targeting the DNA binding domain of androgen receptor by integrated docking-based virtual screening and bioassays[J]. *Acta Pharmacologica Sinica*, 2022, 43(1): 229-239.
- [25] ZHOU YK, YUAN YM, WU YN, LI L, JAMEEL A, XING XH, ZHANG C. Encoding genetic circuits with DNA barcodes paves the way for machine learning-assisted metabolite biosensor response curve profiling in yeast[J]. *ACS Synthetic Biology*, 2022, 11(2): 977-989.
- [26] WU JY, JIANG PX, CHEN W, XIONG DD, HUANG LL, JIA JY, CHEN YY, JIN JM, TANG SY. Design and application of a lactulose biosensor[J]. *Scientific Reports*, 2017, 7: 45994.
- [27] ZHANG XX, HE Y, WU Z, LIU GX, TAO Y, JIN JM, CHEN W, TANG SY. Whole-cell biosensors aid exploration of vanillin transmembrane transport[J]. *Journal of Agricultural and Food Chemistry*, 2021, 69(10): 3114-3123.
- [28] YAO J, HE Y, SU NN, BHARATH SR, TAO Y, JIN JM, CHEN W, SONG HW, TANG SY. Developing a highly efficient hydroxytyrosol whole-cell catalyst by de-bottlenecking rate-limiting steps[J]. *Nature Communications*, 2020, 11(1): 1515.
- [29] BAHLS MO, PLATZ L, MORGADO G, SCHMIDT GW, PANKE S. Directed evolution of biofuel-responsive biosensors for automated optimization of branched-chain alcohol biosynthesis[J]. *Metabolic Engineering*, 2022, 69: 98-111.
- [30] QING XY, STEENACKERS H, VENKEN T, de MAEYER M, VOET A. Computational studies of the active and inactive regulatory domains of response regulator PhoP using molecular dynamics simulations[J]. *Molecular Informatics*, 2017, 36(11): 1700031.
- [31] KAMARAJ B, AL-SUBAIE AM, AHMAD F, SURAPANENI KM, ALSAMMAN K. Effect of novel leukemia mutations (K75E & E222K) on interferon regulatory factor 1 and its interaction with DNA: insights from molecular dynamics simulations and docking studies[J]. *Journal of Biomolecular Structure & Dynamics*, 2021, 39(14): 5235-5247.
- [32] FREZZINI M, NARZI D, SCIOLARI AM, GUIDONI L, PASCARELLA S. Molecular dynamics of an asymmetric form of GabR, a bacterial transcriptional regulator[J]. *Biophysical Chemistry*, 2020, 262: 106380.
- [33] MORAES G, AZEVEDO V, COSTA M, MIYOSHI A, SILVA A, Da SILVA V, de OLIVEIRA D, TEIXEIRA MF, LAMEIRA J, ALVES CN. Homology modeling, molecular dynamics and QM/MM study of the

- regulatory protein PhoP from *Corynebacterium pseudotuberculosis*[J]. *Journal of Molecular Modeling*, 2012, 18(3): 1219-1227.
- [34] de CARVALHO MATIAS EG, BEZERRA KS, COSTA AHL, CLEMENTE JUNIOR WS, OLIVEIRA JIN, RIBEIRO JUNIOR LA, GALVÃO DS, FULCO UL. Quantum biochemical analysis of the TtgR regulator and effectors[J]. *Scientific Reports*, 2024, 14(1): 8519.
- [35] PATEL R, CHUDASAMA R, SOLANKI R, PATEL P, PARMAR K, MUNSHI NS. Structure prediction and molecular docking studies of aromatic hydrocarbon sensing proteins TbuT, HbpR and PhnR to detect priority pollutants[J]. *Journal of Environmental Science and Health Part A, Toxic/Hazardous Substances & Environmental Engineering*, 2020, 55(2): 126-141.
- [36] JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL SAA, BALLARD AJ, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [37] ABRAMSON J, ADLER J, DUNGER J, EVANS R, GREEN T, PRITZEL A, RONNEBERGER O, WILLMORE L, BALLARD AJ, BAMBRICK J, BODENSTEIN SW, EVANS DA, HUNG CC, O'NEILL M, REIMAN D, TUNYASUVUNAKOOL K, WU Z, ŽEMGULYTĖ A, ARVANITI E, BEATTIE C, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. *Nature*, 2024, 630(8016): 493-500.
- [38] BAEK M, DiMAIO F, ANISHCHENKO I, DAUPARAS J, OVCHINNIKOV S, LEE GR, WANG J, CONG Q, KINCH LN, SCHAEFFER RD, MILLÁN C, PARK H, ADAMS C, GLASSMAN CR, DeGIOVANNI A, PEREIRA JH, RODRIGUES AV, van DIJK AA, EBRECHT AC, OPPERMAN DJ, et al. Accurate prediction of protein structures and interactions using a three-track neural network[J]. *Science*, 2021, 373(6557): 871-876.
- [39] LI ZY, LIU XY, CHEN WJ, SHEN F, BI HR, KE GL, ZHANG LF. Uni-Fold: an open-source platform for developing protein folding models beyond AlphaFold[J]. *bioRxiv*, 2022, 2022-08: 502811.
- [40] LIN ZM, AKIN H, RAO R, HIE B, ZHU ZK, LU WT, SMETANIN N, VERKUIL R, KABELI O, SHMUELI Y, dos SANTOS COSTA A, FAZEL-ZARANDI M, SERCU T, CANDIDO S, RIVES A. Evolutionary-scale prediction of atomic-level protein structure with a language model[J]. *Science*, 2023, 379(6637): 1123-1130.
- [41] WATSON JL, JUERGENS D, BENNETT NR, TRIPPE BL, YIM J, EISENACH HE, AHERN W, BORST AJ, RAGOTTE RJ, MILLES LF, WICKY BIM, HANIKEL N, PELLOCK SJ, COURBET A, SHEFFLER W, WANG J, VENKATESH P, SAPPINGTON I, TORRES SV, LAUKO A, et al. *De novo* design of protein structure and function with RFdiffusion[J]. *Nature*, 2023, 620(7976): 1089-1100.