

特约评述

DOI: 10.12211/2096-8280.2022-038

蛋白质稳定性计算设计与定向进化前沿工具

阮青云¹, 黄莘¹, 孟子钧¹, 全舒^{1,2}¹ 华东理工大学生物工程学院, 生物反应器工程国家重点实验室, 上海生物制造技术协同创新中心, 上海 200237;² 上海市细胞代谢光遗传学技术前沿科学研究基地, 上海 200237)

摘要: 天然蛋白质具有临界稳定性的特征, 这种较低稳定性使蛋白质结构具有足够的灵活性, 从而支持其发挥生物学功能。然而, 临界稳定性使得蛋白质遭受胁迫压力后极易发生错误折叠并失去功能, 导致天然蛋白质往往无法满足科学研究与工业应用的需求。此外, 体内蛋白质在错误折叠后产生的聚集沉淀被认为是多种疾病发生发展的原因, 包括阿尔兹海默病、帕金森综合征等。因此, 优化蛋白质的稳定性是科学研究与工程应用领域亟待解决的关键问题。本文从蛋白质的折叠与稳定性机制出发, 聚焦于序列优化与折叠环境优化两种改善蛋白质稳定性的手段, 综述了基于理性设计、计算机辅助设计改善蛋白质稳定性的研究方法, 介绍了用于高通量筛选蛋白质稳定化突变体或折叠相关因子的定向进化技术。通过多项蛋白质序列改良、折叠环境优化的案例介绍, 展示了蛋白质稳定化技术在蛋白质工程与生物医药领域的广阔应用, 包括酶的稳定化设计、疫苗蛋白质的构象控制、分子伴侣与蛋白质聚集抑制剂的筛选、蛋白质稳态药物的开发等。最后, 展望了蛋白质稳定化技术未来的研究方向与前景, 定制化的蛋白质稳定性检测技术将会迎来蓬勃发展。

关键词: 蛋白质折叠; 蛋白质稳定性; 蛋白质稳定化工程; 理性设计; 计算机辅助设计; 定向进化

中图分类号: Q816 **文献标志码:** A

Computational design and directed evolution strategies for optimizing protein stability

RUAN Qingyun¹, HUANG Xin¹, MENG Zijun¹, QUAN Shu^{1,2}

(¹State Key Laboratory of Bioreactor Engineering, School of Biotechnology, East China University of Science and Technology, Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), Shanghai 200237, China; ²Shanghai Frontiers Science Center of Optogenetic Techniques for Cell Metabolism, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Most natural proteins tend to be marginally stable, which allows them to gain flexibility for biological functions. However, marginal stability is often associated with protein misfolding and aggregation under stress conditions, presenting a challenge for protein research and applications such as proteins as biocatalysts and therapeutic

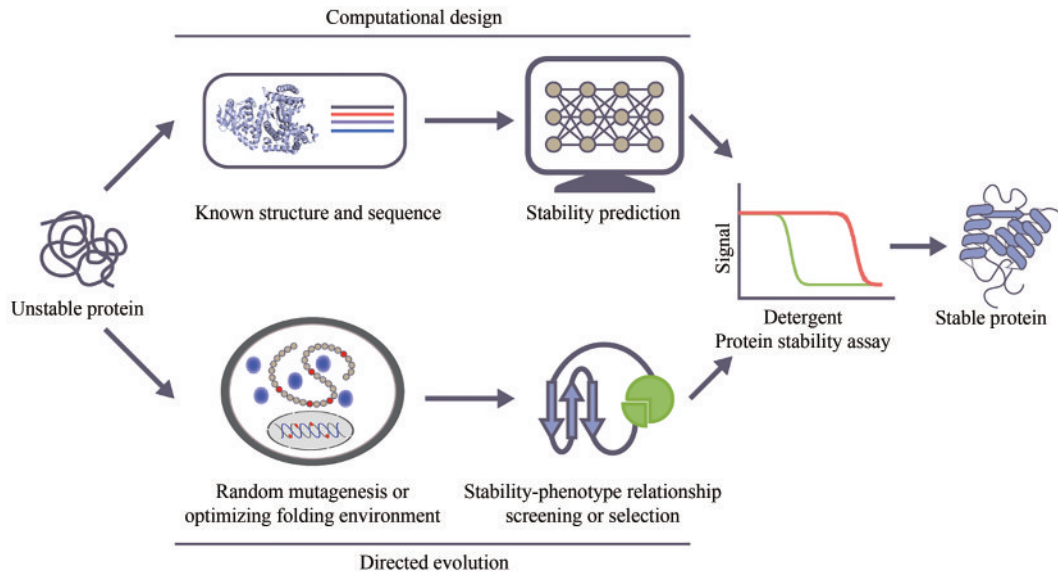
收稿日期: 2022-07-02 修回日期: 2022-07-30

基金项目: 国家自然科学基金面上项目 (31870054, 32171269)

引用本文: 阮青云, 黄莘, 孟子钧, 全舒. 蛋白质稳定性计算设计与定向进化前沿工具[J]. 合成生物学, 2023, 4(1): 5-29

Citation: RUAN Qingyun, HUANG Xin, MENG Zijun, QUAN Shu. Computational design and directed evolution strategies for optimizing protein stability [J]. Synthetic Biology Journal, 2023, 4(1): 5-29

agents. In addition, protein instability has been increasingly recognized as one of the major factors causing human diseases. For example, the formation of toxic protein aggregates is the hallmark of many neurodegenerative diseases, including Alzheimer's and Parkinson's diseases. Therefore, optimizing protein folding and maintaining protein homeostasis in cells are long-standing goals for the scientific community. Confronting these challenges, various methods have been developed to stabilize proteins. In this review, we classify and summarize various techniques for engineering protein stability, with a focus on strategies for optimizing protein sequences or cellular folding environments. We first outline the principles of protein folding, and describe factors that affect protein stability. Then, we describe two main approaches for protein stability engineering, namely, computational design and directed evolution. Computational design can be further classified into structure-based, phylogeny-based, folding energy calculation-based and artificial intelligence-assisted methods. We present the principles of several methods under each category, and also introduce easily accessible web-based tools. For directed evolution approaches, we focus on library-based, high-throughput screening or selection techniques, including cellular or cell-free display and stability biosensors, which link protein stability to easily detectable phenotypes. We not only introduce the applications of these techniques in protein sequence optimization, but also highlight their roles in identifying novel folding factors, including molecular chaperones, chemical chaperones, and inhibitors of protein aggregation. Moreover, we demonstrate the applications of protein stability engineering in biomedicine and pharmacotherapeutics, including identifying small molecules to stabilize disease-related, aggregation-prone proteins, obtaining conformation-fixed and stable antigens for vaccine development, and targeting protein stability as a means to control protein homeostasis. Finally, we look forward to the trends and prospects of protein stabilization technologies, and believe that protein stability engineering will lead to a better understanding of protein folding processes to facilitate the development of precision medicine.



Keywords: protein folding; protein stability; protein stability engineering; rational design; computational design; directed evolution

蛋白质是生物体维持生命活动的基本元件，其功能的发挥依赖于自身折叠形成正确的三维结构。此过程受一系列物理因素驱动，包括氢键、

范德华力、氨基酸骨架角度的偏好性 (backbone angle preferences)、静电相互作用、疏水相互作用和氨基酸链的熵 (chain entropy)^[1]，其中疏水相

相互作用被认为是最主要的驱动因素^[2]。在蛋白质的天然构象中，疏水氨基酸主要被包裹在蛋白质内部，形成一个疏水内核，而蛋白质表面则分布着大量亲水的极性氨基酸。这样的氨基酸分布巩固了蛋白质的内部构架，减少了蛋白质间的非特异相互作用，有效提高了蛋白质稳定性。

蛋白质的折叠是一个复杂且精密的过程。一条100个氨基酸组成的肽链理论上存在超过 10^{30} (2^{100})种可能的构象，其中仅有少数几种天然构象(native state)^[3]。蛋白质在折叠过程中，自由能逐渐降低，同时也伴随着构象多样性的降低，其势能面可被具象化成一张漏斗形折叠能量景观图(图1)：漏斗顶部是未折叠构象(unfolded state)；底部是天然构象；而在漏斗的中间存在许多或深或浅的“山谷”，包括折叠中间态、部分折叠态、错误折叠态等^[4]。由于势能面的崎岖特点，动力学因素在折叠过程中也占据了很大的比重。处于势能面上“山谷”处的部分折叠态需要克服一定的能垒才能重新回到折叠路径上去。若这一能垒过高，则很难回到正确路径上，在生理相关的时间尺度上表现为错误折叠态^[5]。错误折叠态会暴露大量本该包裹在内部的疏水氨基酸残基，使蛋白质间发生非特异性相互作用，具有诱发蛋白质聚集的强烈倾向，形成无定形沉淀。聚集的蛋白质还可以形成稳定的、主要由 β -片层结构规则排布组成的纤维状聚集体(也称淀粉样纤维)，这是很多重大疾病发生发展的原因^[5]。

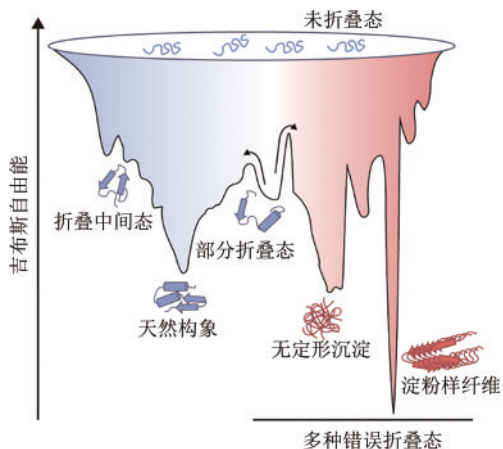


图1 蛋白质折叠能量景观图^[4]

Fig. 1 Energy landscape for protein folding^[4]

Modified with permission

蛋白质的稳定性即是蛋白质维持天然构象的能力^[6]，其与蛋白质的折叠密切相关。在微观上，蛋白质稳定性可分为热力学稳定性与动力学稳定性。热力学稳定性是指蛋白质的天然构象与未折叠构象之间的吉布斯自由能差值(折叠自由能, ΔG_{UN})，其反映了天然构象与未折叠构象之间的平衡(图2)。 ΔG_{UN} 的绝对值越大，蛋白质的热力学稳定性越高。动力学稳定性与蛋白质从天然构象转变为非天然构象过程中所需克服的活化能壁垒(ΔG_u^\ddagger)有关(图2)。该能量壁垒越高，则蛋白质的动力学稳定性越高^[7]。蛋白质的动力学稳定性会反映到宏观的蛋白质稳定性事件中，包括高温下的蛋白质结构与功能、去垢剂压力下的蛋白质变性、蛋白质聚集倾向、蛋白酶降解的难易程度等^[8]，由此衍生出了许多评价蛋白质稳定性指标，具体见表1。

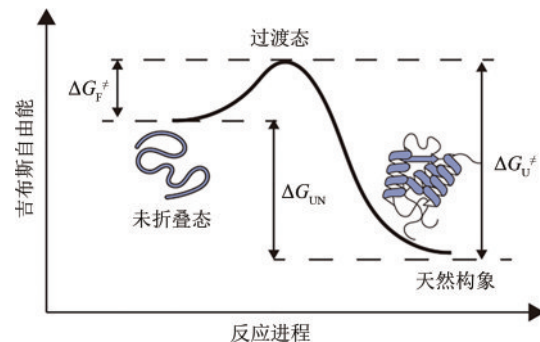


图2 蛋白质热力学稳定性和动力学稳定性

Fig. 2 Protein thermodynamic stability and kinetic stability

生物体内的蛋白质稳定性极为复杂，其会受到包括生物过程、生物大分子、物理化学环境等因素的影响。例如，蛋白质折叠的速率(微秒或毫秒级)远远快于翻译的速率(秒级)，新生肽链未完成翻译便已启动了自发折叠，这意味着蛋白质在较长的时间范围内处于不稳定的未折叠、部分折叠状态；细胞内拥挤的生物大分子环境会增强蛋白质的非特异相互作用，加剧了蛋白质的聚集倾向^[3]。自然进化的蛋白质具有临界稳定性(marginal stability)的特征，其折叠构象与未折叠构象的吉布斯自由能差值较小，在自身突变或受到环境因素扰动后极易发生错误折叠或聚集沉淀^[9]。为保障生物体内蛋白质功能的正确发挥，生物体进化出了一套蛋白质质量控制网络，其成

表1 不同表征蛋白质稳定性参数的定义

Table 1 Definitions of stability parameters

符号	稳定性类型	度量	定义
ΔG_{UN}	热力学	折叠自由能	蛋白质未折叠状态到天然构象的吉布斯自由能变化
ΔG_{NU}	热力学	去折叠自由能	蛋白质天然构象到未折叠状态的吉布斯自由能变化
$\Delta\Delta G$	热力学	折叠自由能变化	蛋白质突变前后折叠自由能的变化
T_m	热力学	熔解温度	使一半的蛋白质解折叠时的温度
$C_{1/2}$	热力学	半变性浓度	使一半的蛋白质解折叠时的变性剂浓度
K_U	热力学	解折叠平衡常数	未折叠状态与天然状态的浓度比值
k_f	动力学	折叠速率常数	蛋白质折叠过程的速率常数
k_u	动力学	去折叠速率常数	蛋白质去折叠过程的速率常数
$k_{d,obs}$	动力学	表观失活速率常数	从天然状态到完全失活(deactivation)的表观速率常数
T_{50}	—	半数失活温度	在一定时间内酶活降至一半时的温度
$t_{1/2}$	动力学	半衰期	酶活降至初始的一半时所需的时间

员包括分子伴侣、辅助分子伴侣及其调控因子(辅助蛋白质的折叠,详细内容读者可参阅西湖大学张鑫课题组撰写的综述^[10])、泛素-蛋白酶体系统与细胞自噬系统(清理错误折叠或聚集沉淀的蛋白质)^[2]。

尽管如此,天然蛋白质的临界稳定性依旧伴随大量的问题。有研究表明,在超过40种的疾病中观察到了蛋白质纤维状聚集体,包括阿尔兹海默病中的 β -淀粉样蛋白^[11]和tau蛋白^[12]、帕金森综合征中的 α -核突触蛋白^[13]、二型糖尿病中的人胰岛淀粉样多肽^[14]等。提升蛋白质稳定性以治疗该类疾病的药物开发正如火如荼,其中转甲状腺素蛋白淀粉样变性心肌病(transthyretin amyloid cardiomyopathy, ATTR-CM)的药物开发最为成功。该疾病中,负责转运甲状腺素和维生素A的转甲状腺素蛋白(transthyretin, TTR)失去四聚体构象后发生错误折叠与堆积,在心脏区域的细胞上沉降并诱发疾病^[15]。美国 Scripps 研究所 Kelly 课题组开发了氯苯唑酸(tafamidis)作为TTR蛋白的稳定剂^[16]。氯苯唑酸作为底物类似物,结合在TTR蛋白结合甲状腺素的口袋中,稳定了该蛋白质的四聚体构象,提升了其动力学稳定性,延缓了疾病发展,并将ATTR-CM的全因死亡率降低了41%。该药物2021年销售额达20.15亿美元,成功跻身全球药品销售额排行榜TOP100中的86位。

此外,天然蛋白质的临界稳定性更是蛋白质体外应用的巨大掣肘,因而稳定性改良一直是国

内外蛋白质工程领域的热点。相关的研究包括:提高工业酶温度耐受性,拓展使用范围;提高蛋白质在储存与运输过程中的稳定性,降低整体成本;延长药物蛋白的体内半衰期,改善其聚集倾向;优化重组蛋白表达,减少表达过程中的聚集与降解等。例如,神经生长因子(nerve growth factor, NGF)与原肌球蛋白受体激酶A(tropomyosin receptor kinase A, TrkA)的结合是启动和维持疼痛的重要环节,因此筛选针对NGF的稳定化抗体并将其开发成镇痛药物是当下生物医药领域的热点^[17]。英国利兹大学的Radford课题组利用定向进化的方式,优化了前期筛选到的NGF抗体MEDI1912的稳定性,显著改善了其聚集倾向与免疫原性,减少了其潜在的副作用^[18]。

鉴于蛋白质稳定性在人类健康与工业应用上的重要意义,本文将从计算设计和定向进化两方面讨论蛋白质稳定性的改良方案。在计算设计部分,我们将介绍基于进化分析、结构分析、折叠自由能计算和机器学习等优化蛋白质稳定性的方法。在定向进化部分,我们详细介绍了基于文库表面展示和基于蛋白质稳定性检测探针的两种定向进化手段,总结它们的原理与特点,并列以应用实例具体分析。

1 计算设计提升蛋白质稳定性

增强蛋白质稳定性的计算设计是指根据蛋白

质的序列与结构,借助理性(经验)推断或数据库中揭示的蛋白质序列、结构与稳定性之间的关系信息,预测影响蛋白质稳定性的潜在位点,并对这些位点进行定向突变以期提升蛋白质稳定性^[19]。在进行突变设计前,通常需要对蛋白质的序列、三维结构、是否存在寡聚体、功能相关的关键残基以及发挥功能的胞内场所等信息有较为充分的了解,以避免影响蛋白质原有的功能。在进行突变设计后还可结合定向进化的方法对目标蛋白进行进一步的改造,以达到更好的效果。根据底层原理和输入信息的不同,计算设计提升蛋白质稳定性的方法可分为四类:基于进化分析的蛋白质稳定性设计、基于结构分析的蛋白质稳定性设计、基于折叠自由能计算的蛋白质稳定性设计和基于机器学习的蛋白质稳定性设计。

1.1 基于进化分析的蛋白质稳定性设计

基于进化分析的设计通常不依赖于蛋白质的三维结构,而是通过多序列比对获得目标蛋白序列中潜在影响稳定性的关键位点或重构目标蛋白可能的进化路径,进而对蛋白质稳定性进行理性设计。根据策略不同,可以细分为同源序列比对、祖先序列重构和共进化分析。在后基因组时代,二代测序、宏基因组测序等技术高速发展,数据库中海量的蛋白质序列为基于进化分析的稳定性设计提供了充足的素材,因此除从头设计(*de novo design*)得到的蛋白质或没有明显同源序列的孤儿蛋白(*orphan protein*),其他大部分的蛋白质都能够基于进化分析开展稳定性设计。

同源序列比对一方面可以将目标蛋白的序列与嗜热生物来源的同源蛋白(通常具有较高的稳定性)进行序列对比,找出具有明显区别的位点,继而将目标序列中这些位点定向突变为嗜热微生物中对应的氨基酸。如Goihberg等^[20]对具有工业应用价值的贝氏梭状芽孢杆菌的乙醇脱氢酶(*alcohol dehydrogenase, ADH*)进行稳定性改造,通过序列比对,发现嗜热微生物来源的乙醇脱氢酶有多个位点带有脯氨酸,进一步结合三维结构鉴定出一个潜在的突变位点Q100P,后续实验表明仅此一个突变就可将贝氏梭状芽孢杆菌来源的乙

醇脱氢酶的 T_m 值提高10 °C以上。另一方面,同源序列比对也可以采用共识分析(*consensus design*)的方法。共识分析并不局限于使用嗜热生物来源的蛋白质进行多序列比对,而是将目标蛋白序列与大量不同来源的同源蛋白进行序列比对,若某一氨基酸在其他同源蛋白中都较为保守,而在目标蛋白中与这一保守的氨基酸不同,即将该位点突变为较为保守的氨基酸。Sternke等^[21]利用共识分析的方法对6种不同家族的蛋白质进行理性设计,每种蛋白质使用1000条以上的序列进行同源比对,成功使这6种蛋白质的稳定性得到不同程度的提高,且维持了原有的功能。

相较通过同源序列比对找出热点残基进行改造,基于祖先序列重构的理性设计则是对全局序列进行较为显著的改造。由于地球远古时期的恶劣环境,祖先序列通常拥有较强的稳定性,因此重构出目标蛋白的祖先序列即可能实现提升稳定性的目的。祖先序列重构首先通过序列比对和进化树的构建得到已知序列的进化特征,再通过最大似然法(如软件FastML)或贝叶斯推理等方法(如软件MrBayes)重构出祖先蛋白的序列。如Gumulya等^[22]对具有广泛工业应用价值的细胞色素P450 CYP102A1进行祖先序列重构,基于来源于42个物种的138条序列使用FastML重构出祖先蛋白序列CYP3_N1,使其在60 °C的半衰期由野生型的小于10 min提升至10 h以上。此后,作者以CYP3_N1作为定向进化的模板,借助串联黄色荧光蛋白进行筛选,得到更多更稳定的突变体。值得注意的是,祖先序列重构有可能会改变蛋白质原有的功能,如酶的底物特异性、催化反应的类型等,需要后续进一步加以验证。

共进化分析通过多序列比对预测可能存在共进化特征的两个氨基酸残基,若其中一个氨基酸残基发生突变,另一个氨基酸残基有较大概率出现补偿突变。通常存在共进化关系代表在两个氨基酸残基间存在相互作用,因此通过共进化分析可以获得存在相互作用的残基对,再对评分靠前的残基对进行定向进化或理性设计,以提升蛋白质的稳定性。如Wang等^[23]对 α -淀粉酶使用共进化分析得到了10对具有共进化特征的残基对,后续对这些残基对进行饱和突变得到的最佳突变体

对比野生型其半数失活温度 (half-inactivation temperature, T_{50}) 提升了 8 °C。总体而言, 共进化分析是被关注较少的理性设计方法, 各种方法还有待开发。

1.2 基于结构分析的蛋白质稳定性设计

蛋白质的三维结构包含了残基间相互作用模式、潜在的残基动态性、寡聚状态等诸多信息, 因此基于结构分析的蛋白质稳定性设计旨在从蛋白质整体结构或局部结构出发, 预测影响蛋白质稳定性的缺陷位点, 对这些位点进行定点突变以提高蛋白质稳定性。

针对蛋白质整体结构的稳定性设计包括优化蛋白表面电荷和降低整体灵活性等方法。优化蛋白表面电荷是指突变蛋白质表面的氨基酸为带电氨基酸, 在表面残基之间形成更多的静电相互作用, 从而提升蛋白质的稳定性。Strickler 等^[24]对优化蛋白表面电荷的方法进行了系统性的研究, 他们使用遗传算法 (genetic algorithm) 对 5 种不同的蛋白质进行了表面电荷的理性设计, 在每一轮虚拟优化中均将表面氨基酸随机突变为带电氨基酸, 然后计算蛋白质表面静电相互作用的能量, 能量较优的突变体将进入下一轮的优化。经过 3 轮优化, Strickler 等发现最终得到的所有 11 个突变体的稳定性都得到不同程度提升。后续 Chan 等对该策略进行进一步优化^[25], 通过计算折叠前后表面带电氨基酸的 pK_a 值的变化 (ΔpK_a), 选取折叠后 pK_a 值增大的位点进行突变, 根据 ΔpK_a 值计算得到的 $\Delta\Delta G$ 理论值与实验测得的数值相关性高达 0.81。Lawrence 等^[26]对绿色荧光蛋白 (green fluorescence protein, GFP) 进行优化表面电荷的改造, 将蛋白质表面的非保守氨基酸替换成带正电的氨基酸或带负电的氨基酸, 分别得到带有 36 个正电荷或 30 个负电荷的绿色荧光蛋白, 这些带有强电荷的绿色荧光蛋白展现了更强的抗聚集能力, 且维持原有的荧光蛋白功能。

从结构出发降低蛋白质整体灵活性的方法通过对结构中灵活性较高的氨基酸进行定点突变以期提高蛋白质的稳定性。灵活性较高的氨基酸的挑选可以以温度因子 (B-factor) 为标准。如

Reetz 等^[27]对脂肪酶 (lipase A, LipA) 进行温度因子分析, 挑出其中 10 个温度因子最高的位点进行迭代饱和突变, 成功将 LipA 在 55 °C 下的半衰期从小于 2 min 提升至 980 min。Yu 等^[28]更进一步发展了“刚化”柔性 (高灵活性) 位点策略, 他们综合利用蛋白质结构中的温度因子、分子动力学模拟得到的均方根涨落 (root mean square fluctuation, RMSF) 以及其他指标 (如氢氘交换质谱表征的结构灵活性), 寻找出灵活性较高的位点, 进一步对这些位点使用定向进化或设计二硫键、计算 Rosetta 自由能等方法进行“刚化”。使用这一策略, Zhang 等^[29]对 Fab 抗体等一系列蛋白质进行了稳定性改造。

针对蛋白质局部结构的理性设计, 通过在蛋白质局部结构中引入新的相互作用或消除局部的不良影响, 包括设计二硫键、破坏表面大面积疏水片区等方法, 以实现蛋白质整体的稳定化。理性设计二硫键的方法是指在特定位置上引入半胱氨酸, 将蛋白质置于氧化环境使半胱氨酸之间形成二硫键。Fang 等^[30]在可卡因酯酶 (cocaine esterase, CocE) 的二聚体界面上设计出两对全新二硫键, 使 CocE 在 37 °C 下的半衰期从约 12 min 延长至 100 d 以上。破坏表面大面积疏水片区的方法是指将蛋白质发生自聚集的潜在位点处 (大面积疏水片区) 的氨基酸突变为亲水氨基酸, 从而降低蛋白质的自聚集倾向。如 Gil-Garcia 等^[31]基于此原理开发了 AGGRESKAN3D 软件, 利用其对易沉淀的人生殖细胞抗体的重链部分 DP47 进行改造, 仅通过三个表面的突变就显著降低了该抗体在 37 °C 的自聚集倾向。

1.3 基于折叠自由能计算的蛋白质稳定性设计

基于折叠自由能计算的设计首先分别预测突变前后的最佳构象, 然后预测两种构象的折叠自由能 (ΔG_{UN}), 从而得到突变对稳定性的影响 ($\Delta\Delta G$), 经典的方法有 FoldX^[32]和 Rosetta^[33]。

FoldX 将多种相互作用力 (如氢键、范德华力、静电相互作用) 以及折叠过程中的熵变等通过线性相加构建能量方程, 再将能量方程中每一项参数的权重通过来自不同蛋白质的 339 个突变体

的结构和稳定性实验数据进行拟合，建立起结构到能量的计算方程^[34]。在计算折叠自由能时，FoldX首先固定蛋白骨架和周围环境，只对突变位点进行采样得到最佳的构象，再使用能量方程计算出突变对折叠自由能的影响。Buss等^[35]使用FoldX对 ω -转氨酶所有8246个可能的单点突变进行预测，得到了11个打分最佳的候选突变。通过实验验证，其中4个突变稳定性得到了提升，稳定性最佳的突变体G98M在55 °C下的半衰期提升为野生型的3倍以上。

Rosetta能量方程的定义更为复杂和精细。与FoldX类似，Rosetta同样涉及多种相互作用力的计算，但在计算时根据原子是否包埋于蛋白质内部、原子之间的远近等进行分类计算，还添加了非理想键角和键长的相互作用计算、骨架和侧链的扭转、二面角、脯氨酸对骨架的影响等项目。添加的项目如骨架和侧链的扭转使用了结构数据库中的统计数据，再根据玻尔兹曼分布进行能量计算。在采样上，Rosetta允许蛋白骨架和蛋白侧链进行小规模移动，并使用蒙特卡洛法逐步收敛得到最佳的构象，结合能量方程计算得到 $\Delta\Delta G$ 值。Rosetta能量方程和采样方式都较为精细，因此运算速度远低于FoldX。Yu等^[36]对大肠杆菌来源的转酮酶使用Rosetta设计了49个突变体，通过实验验证得到Rosetta的准确率高达65.3%。

在单一软件的基础上，对多种折叠自由能计算软件进行组合或整合其他理性设计的方法，能进一步提升准确率或计算速度。Wijma等^[37]开发了FRESCO (framework for rapid enzyme stabilization by computational libraries) 流程，综合FoldX、Rosetta和二硫键预测软件Disulfide Discovery的结果得到候选突变体，再进行短时间的分子动力学模拟，排除可能大幅影响蛋白质整体结构的突变。作者利用FRESCO对柠檬烯环氧化物水解酶 (limonene epoxide hydrolase, LEH) 进行稳定性设计，得到的17个突变体中有10个突变体的稳定性得到了提升，最终组合得到的突变体的 T_m 值提高了35 °C^[37]。

PROSS (protein repair one stop shop) 方法舍弃了FRESCO流程中计算量要求较大的分子动力学模拟。对于庞大的序列空间，首先通过多重序

列比对，排除每个位点中不常见的突变，并进一步排除潜在影响功能的突变位点，再对剩余的突变使用Rosetta进行稳定性预测并进行组合，最终得到多点突变体。PROSS使用了序列比对对罕见突变进行了排除，从而提高了计算的效率。Goldenzweig等^[38]使用PROSS对人源的乙酰胆碱酯酶 (human acetylcholinesterase, hAChE) 进行改造，仅用2.5 h就完成了整个预测流程，最佳的hAChE突变体在大肠杆菌中的表达量提高了100倍以上，原有酶活也得到了保留。

FireProt^[39]软件同时进行基于折叠自由能的位点预测和基于共识分析的位点预测。前者使用FoldX和Rosetta串联的方法，将FoldX预测得到的潜在稳定突变体 ($\Delta\Delta G_{UN} < -1$ kcal/mol) 再通过Rosetta进行预测。后者使用共识分析得到潜在的突变位点，再使用FoldX进行预测。将两条并行的途径得到的位点进行组合得到最终的突变体。Musil等^[39]使用三种不同的蛋白质对FireProt流程进行了验证，最终得到的突变体 T_m 值提升在15~25 °C之间。

1.4 基于机器学习的蛋白质稳定性设计

随着蛋白质结构和序列数据的不断积累和计算机运算速度的提升，基于机器学习的蛋白质稳定性预测方法迅猛发展。机器学习是能自动改进的计算机算法，其通过使用大量数据进行训练，不断改进内部权重和参数等，建立自变量到因变量的映射关系，最终训练完成的模型能够对全新的自变量预测对应的因变量。基于机器学习的蛋白质稳定性理性设计是通过数据库中的大量数据 (自变量)，如结构、突变位点、稳定性实验数据，对机器学习模型进行训练，使其面对全新的蛋白质突变能够给出合理的稳定性变化 (因变量) 预测。

Shroff等^[40]使用PDB数据库中的大量结构对卷积神经网络进行训练，使其能够判断特定结构变化是否对蛋白质有利，开发了MutCompute模型。Lu等^[41]进一步使用这一模型对聚对苯二甲酸乙二醇酯水解酶 (poly ethylene terephthalate hydrolase, PETase) 进行改造，将PETase的 T_m 值

提升了7 °C,使其能够在50 °C下对PET塑料进行降解。Capriotti等^[42-43]使用ProTherm数据库中1948个突变的 $\Delta\Delta G$ 数据和相应的序列对支持向量机(Support Vector Machine)模型进行训练,得到I-mutant模型,在测试数据集上预测得到的 $\Delta\Delta G$ 与实验数据达到0.71的相关性,近年来也被应用于新冠病毒刺突蛋白的突变稳定性评估中^[44]。华东师范大学张增辉课题组^[45]基于人工校对的来自242个蛋白质的5766个 $\Delta\Delta G$ 数据使用SRP神经网络进行训练,同时将氨基酸残基对在结构上的距离、氨基酸保守性、序列上邻近氨基酸的种类等信息作为特征进行神经网络输入。多方面对训练数据的优化使得DeepDDG预测的准确率得到了提升,在测试数据集中得到高达0.68的相关性。

同时机器学习也可以用于指导定向进化实验。如加利福尼亚理工学院的Arnold课题组^[46]利用定向进化得到的少量突变体数据(242个细胞色素P450突变体的 T_{50} 值)对高斯过程模型进行训练,并用于P450的进一步改造,模型预测得到的最优突变体在定向进化的基础上将 T_{50} 值提升了8.7 °C。

1.5 不同计算设计方法的比较

进化分析、结构分析、折叠自由能计算和机器学习等方法基于不同的理论或经验,虽然发展历程不尽相同,但彼此并不存在优劣,应综合考虑目标蛋白的特点选取相应的方法。

首先,各种计算设计的输入信息截然不同,这会导致对不同目标蛋白的适用性具有明显差异。基于进化分析的方法需要大量的同源蛋白序列,对从头设计得到的蛋白质或没有明显同源序列的孤儿蛋白不适用;基于结构分析、折叠自由能计算以及部分机器学习的方法需要蛋白质三维结构为基础,对未解析三维结构需要先预测三维结构,对于预测结果置信度低的目标蛋白以及内在无序的目标蛋白等不适用。其次,各种计算设计的方法存在明显的算力需求差异,这也导致了各种方法的硬件需求不同。总体而言,基于进化分析和结构分析的算力需求最低;基于折叠自由能计算的计算需求适中,通常需要使用性能较强的工作站;基于机器学习的计算要求最高,通常需要使

用专业的工作站或计算集群。基于各种方法具有明显的适用性和计算需求差异,可以对这些方法进行组合,取长补短,得到如FireProt、PROSS、FRESCO这样的组合型方法(表2)。

2 定向进化提升蛋白质稳定性

定向进化是提升蛋白质稳定性的一种强有力的手段。该方法需生成目标蛋白的突变体文库,并经过特定的筛选手段,获取稳定性提升的突变体。其中,突变体文库构建的方法已十分成熟,常用手段包括易错PCR、DNA洗牌术(DNA shuffling)、高突变率菌株(mutator strain)等。

本文重点讨论不同的蛋白质稳定性筛选方法,包括各类文库展示技术与蛋白质稳定性检测探针。这些稳定性筛选方法通常均具有如下诸多优点:高通量;不需要目标蛋白的结构、功能等先验知识;可反映蛋白质的体内折叠状态;相较于理性设计,筛选结果阳性率高等。

2.1 利用文库展示技术进行定向进化

文库展示技术(library-based display)是指将目标蛋白文库中的突变体固定在生物系统(噬菌体、细胞、细胞器或核酸等)上,以便针对蛋白质的某一属性直接进行扰动并检测。由于展示的蛋白质与外界环境接触,施加的变性压力(温度、pH、去垢剂、蛋白酶等)会直接影响蛋白质的折叠状态或降解程度。在使用特定的方法富集经受压力考验的蛋白质后,便可读取其关联的生物系统上的基因信息,建立基因型-蛋白质表型的对应关系。

根据生物系统的不同,文库展示技术可分为细胞表面展示、噬菌体表面展示、细菌芽孢表面展示、核糖体展示、mRNA展示和cDNA展示等。

2.1.1 细胞表面展示

细胞表面展示技术将目标蛋白与细胞表面的锚定蛋白融合表达,并利用信号肽将融合蛋白经分泌途径转运并固定到细胞表面。展示的目标蛋白可以通过多种方法检测,如利用荧光抗体进行识别并定量。结合流式细胞术,可对具有不同表

表2 理性设计方法的代表性软件^[31-33, 38-40, 43, 45, 47-56]

Table 2 Summary of software for protein stability rational design

类别	软件名称	输入	网站/本地安装	参考文献
基于结构分析				
优化蛋白质表面电荷	TKSA-MC	结构	http://tksamc.df.ibilce.unesp.br	[47]
	PHEPS	结构	http://pheps.orgchm.bas.bg/home.html	[48]
基于温度因子	B-FITTER	结构	Windows 系统本地安装	[49]
设计二硫键	Disulfide By Design	结构	http://cptweb.cpt.wayne.edu/DbD2	[50]
	DISULFIDE	结构	http://disulfind.disi.unitn.it	[51]
破坏表面大面积疏水区域	AGGRESCAN3D	结构/序列	http://biocomp.chem.uw.edu.pl/A3D2	[31]
基于进化分析				
同源序列比对	3DM	序列	https://3dm.bio-product.com	[52]
	Consensus Finder	序列	http://kazlab.umn.edu	[53]
祖先酶重构	Ancestors 1.0	序列	http://ancestors.bioinfo.uqam.ca/ancestorWeb	[54]
	FireProt ASR	序列	https://loschmidt.chemi.muni.cz/fireprotasr	[55]
基于折叠自由能计算				
单纯 $\Delta\Delta G$ 计算	FoldX	结构	全平台本地安装	[32]
	Rosetta	结构	Linux 系统本地安装	[33]
	PoPMuSiC	结构	http://babylone.ulb.ac.be/popmusic	[56]
组合其他策略的 $\Delta\Delta G$ 计算	PROSS	结构	http://pross.weizmann.ac.il/step/pross-terms	[38]
	FireProt	结构/序列	https://loschmidt.chemi.muni.cz/fireprot	[39]
基于机器学习				
支持向量机	I-mutant	结构/序列	http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi	[43]
SRP 神经网络	DeepDDG	结构	http://protein.org.cn/ddg.html	[45]
卷积神经网络	MutCompute	结构	https://mutcompute.com	[40]

面展示量的细胞加以区分。

利用表面展示进行蛋白质稳定性进化的底层逻辑大致可分为两类。其一，利用细胞分泌途径中存在的蛋白质质量控制网络对蛋白质折叠状态进行区分：错误折叠的蛋白突变体会更多地滞留在分泌途径中进而被降解，正确折叠的蛋白质则会顺利到达细胞表面，由此可以通过蛋白的表面展示量表征其稳定性。其二，鉴于表面展示的蛋白质对外界环境的敏感性，对其施加变性压力，通过检测残余蛋白质的量或活力，表征蛋白突变体在压力条件下的耐受性。

细胞表面展示根据细胞类型可分为细菌表面展示、酵母表面展示、哺乳动物细胞表面展示等。其中，酵母表面展示得到了最为广泛的应用，其优势体现在：①与原核系统相比，酵母的蛋白质质量控制网络与翻译后修饰更为接近高等生物，能够为真核来源的蛋白质提供更合适的表达体系；②与哺乳动物细胞相比，酵母的培养成本低廉，

遗传操作成熟，且具有更好的同质性（相同遗传背景细胞在形态、蛋白表达等各方面具有一致性）。

在酿酒酵母中，典型的表面展示利用定位在细胞表面的 α 凝集素（ α -agglutinin, Aga）核心亚基（Aga1p）与其结合亚基（Aga2p）来锚定目标蛋白，其中 Aga2p 与目标蛋白直接进行融合 [图3(a)]。除外源蛋白本身可被抗体识别外，还可以在锚定端和游离端加上标签，供识别检测。

Kowalski 等^[57] 利用酿酒酵母分泌表达牛胰蛋白酶抑制剂（bovine pancreatic trypsin inhibitor, BPTI）及其突变体，并通过蛋白免疫印迹实验，发现 BPTI 分泌量与其热稳定性具有相关性。作者推断，不稳定蛋白分泌量较低是由于其更容易暴露疏水位点，增强了与内质网内分子伴侣，如免疫球蛋白结合蛋白（binding immunoglobulin protein, BiP）的结合，并引发内质网相关降解（ER-associated degradation, ERAD）。后续亦有研

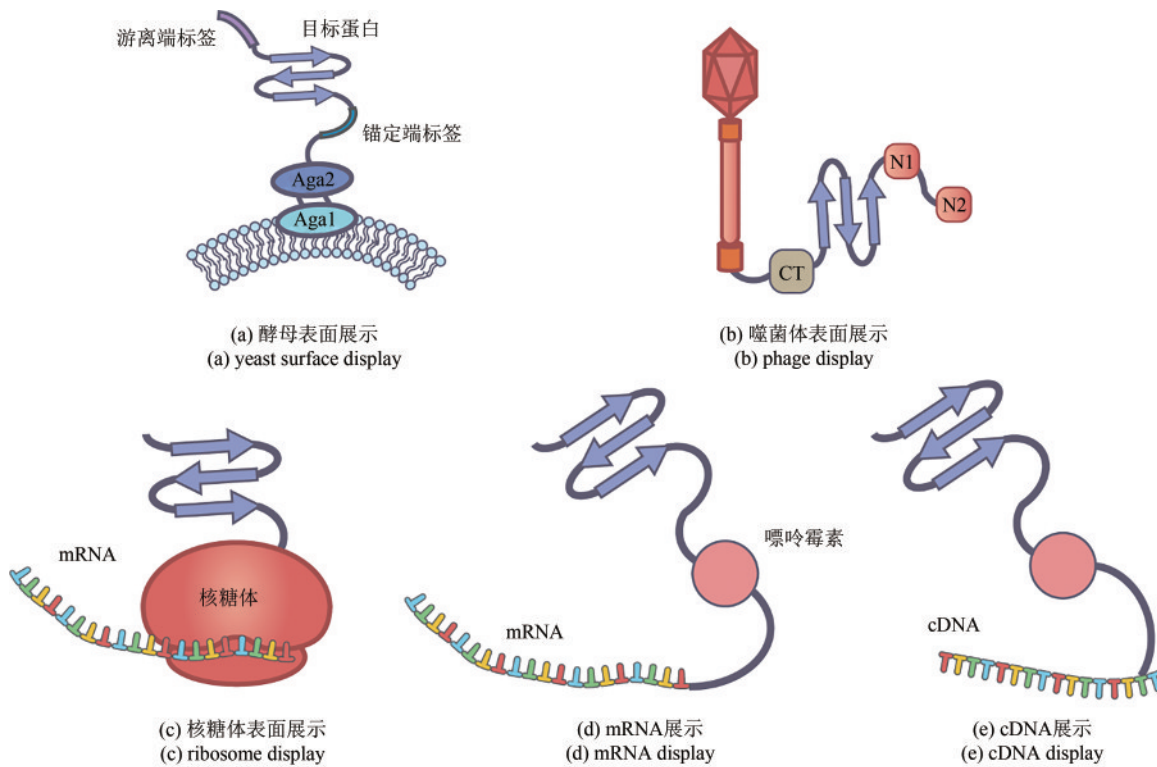


图3 各类文库展示技术

Fig. 3 Various library-based display technologies

究发现，不稳定BPTI突变体会经高尔基体转运至液泡中降解^[58]。这些研究为酵母表面展示鉴定蛋白质稳定性提供了理论与数据基础。

Shusta等^[59-60]构建了单链鼠T细胞受体(single-chain T-cell receptor, scTCR)突变体文库，并通过Aga2p将其展示在酵母表面，利用荧光偶联的scTCR构象特异性抗体结合表面的scTCR，借助流式细胞术筛选高荧光信号细胞，鉴定到分泌效率更高的scTCR突变体。该研究发现，筛选出的突变体具有更好的热稳定性，且在超过酵母最适生长温度的37℃下，依旧具备更高的分泌量。作者建议，结合热压力的分泌量筛选有助于获取更稳定的突变体。

已有研究发现，虽然酵母的存活率随着温度的上升而大幅下降，但其依旧保留了细胞完整性(85℃可孵育10 min)^[61]，因此可用于后续的流式细胞术评估与高通量测序分析。Orr等^[62]在表面展示的基础上额外引入热压力，以鉴定单链抗体稳定性。作者对细胞进行诱导表达后，将其置于不同的温度(20~80℃)下孵育，并测定孵育后细胞表面的蛋白突变体对构象特异性抗体的结合

能力，最终获得了蛋白突变体的热变性曲线。这些研究表明了酵母表面展示可用于高温下的稳定性定向进化。

构象特异性抗体的使用能够提升稳定性筛选的精度。然而，并非每一个蛋白质都具备该类抗体，大多时候需要使用仅能识别一级序列的一般抗体或针对某种融合标签的抗体替代，但这会带来若干问题。首先，若外源蛋白在分泌过程中发生了末端降解，游离端标签信号会丢失，用该标签的抗体检测会得不到任何信号。此外，Hagihara等^[63]在BPTI突变体文库筛选的过程中发现，有一些结构发生明显变化导致功能丢失的突变体会有极高的展示量，因此也不适宜用锚定端的抗体和一般抗体进行检测。

2.1.2 噬菌体表面展示

噬菌体表面展示技术由George Smith于1985年提出^[64]。该技术中，目标蛋白会与一种噬菌体衣壳蛋白(如M13噬菌体中的pIII、pVIII)融合表达，并展示在噬菌体表面。噬菌体表面展示技术拥有较高的文库容量(理论上可达 10^{10})以及较低的生物系统表面结构复杂度，因此被广泛用于抗

原抗体库的建立、蛋白质相互作用的鉴定、细胞信号转导的研究和生物探针的开发等领域^[65]。在蛋白质稳定性的相关研究中，噬菌体展示的蛋白质通常会被施加各类变性压力，以考查蛋白质对该压力的耐受性。此外，噬菌体表面展示根据富集手段的差异，可大致分为侵染能力富集和结合能力富集。

Kristensen 等与 Sieber 等^[66-67]在 1998 年开发了一种基于噬菌体展示的蛋白质稳定性定向进化技术 (protein stability increased by directed evolution, Proside)。作者选择丝状噬菌体 fd (Filamentous Phage fd) 作为目标蛋白的表达宿主，并将目标蛋白与 fd 用于侵染细菌的 pⅢ 蛋白融合表达。fd 的 pⅢ 蛋白包括 CT 结构域 (负责锚定 pⅢ 至噬菌体表面) 和 N 结构域 (负责侵染细菌，包括 N1 和 N2) [图 3(b)]。作者在 CT 域与 N 域间插入目标蛋白，这会导致目标蛋白的状态 (折叠状态或降解程度) 间接关联 N 域的结构或保留情况，进而影响噬菌体侵染能力。逐渐增大变性压力后，富集仍能完成侵染过程并大量扩增的噬菌体，即实现了稳定的蛋白突变体的筛选。随后几年，多家团队利用该技术对多种蛋白质进行了稳定化，包括嗜热菌枯草芽孢杆菌中的冷激蛋白 (cold shock proteins from *Bacillus subtilis*, CspB) (一种改良作物的潜在基因资源)^[68]、 β -内酰胺酶 (β -lactamase, β la) (一种常见的抗性基因)^[69] 等。

在部分研究中，作者利用结合能力富集稳定化的蛋白质。他们将携带亲和和标签的目标蛋白置于 pⅢ 蛋白的游离端，并利用固定化的亲和配体吸附变性压力处理后仍保留亲和和标签的噬菌体^[70]。这种设计思路与 Proside 类似，期望目标蛋白的状态会影响到亲和和标签的状态。然而，亲和和标签一般较短且不具备二级结构，其生物可及性不容易受到目标蛋白折叠状态的影响。为提升此系统的敏感度，研究者将固定化的亲和配体替换为目标蛋白的构象特异性抗体，以便直接检测蛋白突变体的状态。尽管该方法的准确性远远高于 Proside，但对于不具备构象特异性抗体的蛋白质没有普适性。

噬菌体超高的生长速率为蛋白质的快速进化提供了可能，但实际过程却有着诸多限制，这些

限制包括：①一次构建的文库需多轮筛选才能获得可靠的结果；②满足实际应用的蛋白突变体需迭代多轮定向进化；③一轮文库构建需要数天的实验周期。针对此现状，哈佛大学和麻省理工学院 Broad 研究所刘如谦课题组于 2011 年开发了一种噬菌体辅助连续进化技术 (phage-assisted continuous evolution, PACE)^[71]，其便捷性体现在目标蛋白突变体的生成与富集同时发生，故可在短时间内完成多轮筛选。该技术富集特定蛋白质属性的核心原理如下：①将目标蛋白突变体的性质与 pⅢ 蛋白的表达相关联，从而影响噬菌体的侵染能力，进而影响其增殖速度。②在含有 M13 噬菌体的溶液内不断流加大肠杆菌培养液，噬菌体侵染大肠杆菌完成增殖的同时也不断被稀释，因此只有增殖速率超过稀释速率的噬菌体得以富集。

在后续研究中，刘如谦课题组将 PACE 技术应用于蛋白质稳定性定向进化，开发了可溶表达-噬菌体辅助连续进化技术 (soluble expression phage-assisted continuous evolution, SE-PACE)^[72]。作者在初始的噬菌体中导入融合表达 T7 RNA 聚合酶 N 端片段与目标蛋白的噬菌体载体 (selection phage vector, SP)；在流加的大肠杆菌中表达 T7 RNA 聚合酶 C 端片段，并导入利用 T7 启动子表达 pⅢ 蛋白的辅助质粒 (accessory plasmid, AP) 与增加目标蛋白突变频率的突变质粒 (mutagenesis plasmid, MP)。当噬菌体侵染大肠杆菌后，T7 RNA 聚合酶 N 端与 C 端的重新结合受目标蛋白突变体稳定性的影响，重构的 T7 RNA 聚合酶又决定了下游 pⅢ 蛋白的表达效率，使噬菌体表现出不同的侵染能力与增殖速度。作者利用此方法成功对麦芽糖结合蛋白 (maltose-binding protein, MBP) 与一种真核胞苷脱氨酶 (apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1, APOBEC1) 进行了稳定性定向进化。

PACE 技术需要使用一套自动流加培养设备以完成目标蛋白的定向进化。该设备价格昂贵，且使用过程中试剂消耗量大，并不利于 PACE 技术的推广。中国科学院深圳先进技术研究院刘陈立课题组^[73]针对上述问题，开发了一种可视的连续空间定向进化方法，在琼脂平板上完成目标蛋白的定向进化，降低了 PACE 技术的使用门槛。该方法

原理如下：大肠杆菌接种于琼脂平板的中央，且在距其 1 cm 处的三角上接种噬菌体；大肠杆菌由中央向边缘生长，且由于初代噬菌体侵染能力较低，大肠杆菌可迅速覆盖全部平板；初代噬菌体低效率侵染大肠杆菌形成噬菌斑，且噬菌斑不断向边缘移动，同时目标蛋白发生突变，并使得子代噬菌体侵染效率差异化；侵染效率高的子代噬菌体会导致其生成的噬菌斑更远离平板中央；从远离平板中央的噬菌斑上挑取噬菌体，供新一轮定向进化使用。

2.1.3 细菌芽孢展示技术

细菌芽孢展示技术将目标蛋白与细菌芽孢衣壳蛋白融合表达，并使其展示在芽孢表面^[74]。细菌芽孢是芽孢杆菌纲（Bacilli）、梭菌纲（Clostridia）的细菌应对营养限制环境而产生的一种休眠型细胞形态。其通过多层膜结构，包括外衣壳、内衣壳、肽聚糖层等，将基因组包裹在芽孢内部，并使基因组免受化学环境、高温、干燥、强电离辐射等因素的破坏。细菌芽孢展示技术被广泛应用于蛋白质进化、口服疫苗开发、抗体生产、生物催化、生物修复和诊断工具体制备等领域。该技术中最常使用的是枯草芽孢杆菌（*Bacillus subtilis*）芽孢。该细菌芽孢的内外层衣壳上拥有超过 70 种的衣壳蛋白，其中使用较多的衣壳蛋白有 CotB、CotC、CotE、CotG 等。与细胞相比，芽孢结构完整性与生物活性均不易被变性压力破坏，故更适用于超高稳定性进化^[75]。

CotA 是一种漆酶，也是枯草芽孢杆菌芽孢衣壳的一种组分，被广泛用于造纸、果酒澄清、染料脱色等领域。由于工业需求，CotA 需要在各类有机溶剂和低 pH 环境下依旧发挥可观的催化能力。新泽西理工学院 Farinas 课题组^[76]对敲除 *CotA* 基因的枯草芽孢杆菌中导入 CotA 突变体库，利用其氧化联胺 2,2'-联氮-双-3-乙基苯并噻唑啉-6-磺酸（diammonium 2,2'-azino-bis-3-ethylbenzothiazoline-6-sulfonate, ABTS）的显色反应进行筛选，先后得到了有机溶剂耐受型突变体 T480A 和中度酸性环境耐受型突变体 E498G（pH 4 失活半衰期提升 25 倍）。二者组合后的突变体更是将 42 h 产物生成提升 5.3 倍。

2.1.4 核糖体展示技术, mRNA 展示技术与 cDNA 展示技术

前文所述的细胞表面展示、噬菌体表面展示和细菌芽孢表面展示等展示技术均需将体外生成的突变体 DNA 文库导入宿主并完成转录和翻译，这会导致这些展示技术的文库容量受转化效率所限制。无细胞文库展示技术包括核糖体展示技术、mRNA 展示技术、cDNA 展示技术等，其利用纯化后的转录、翻译系统，直接在体外生成蛋白质突变体文库。这些技术的文库容量不受转化效率限制，理论上可达 $10^{12} \sim 10^{14}$ ^[65]。

核糖体展示技术将 mRNA 与其翻译得到的蛋白质共同结合在核糖体上 [图 3(c)]，并通过后续的蛋白质属性筛选，关联蛋白突变体的基因型与表型。在该技术中，无终止密码子的 mRNA 会在核糖体协助下进行体外翻译。翻译过程抵达 mRNA 的物理末端后，由于缺失终止密码子，mRNA 与新生肽链会滞留在核糖体上，形成核糖体-mRNA-蛋白质三元复合物^[77]。Jermutus 等^[78]利用该技术实现了抗 HA 标签单链抗体的定向进化。作者通过逐渐提高翻译过程中的二硫苏糖醇浓度，以破坏单链抗体中的二硫键，促使其错误聚集，并通过固定化的抗原富集未聚集的单链抗体突变体。富集的三元复合物中的 mRNA 经逆转录生成游离的 cDNA，以提供下一轮筛选的模板。经 5 轮筛选，单链抗体稳定性得到了大幅提升（ $\Delta\Delta G_{UN} = -30 \text{ kJ/mol}$ ）。

mRNA 展示技术在核糖体展示技术的基础上进行了简化 [图 3(d)]。该技术的关键步骤是在体外转录生成的 mRNA 3' 端通过一段 DNA 序列连接嘌呤霉素（puromycin）。当 mRNA 翻译完成时，嘌呤霉素模拟酪氨酸-tRNA 末端的氨酰基，进入核糖体 A 位点，并与新生肽链羧基端共价交联，生成稳定的酰胺键，终止翻译的同时形成 mRNA-蛋白质的二元复合物^[79]。该技术后续筛选方法与核糖体展示技术类似。

核糖体展示技术与 mRNA 展示技术的主要缺陷在于，mRNA 的不稳定性令其难以在变性压力筛选后完整保留，不利于后续基因信息的读取与新一轮的筛选。针对此缺陷，后续研究者开发了 cDNA 展示技术^[80]。该技术中，嘌呤霉素结合在

一段与 mRNA 互补配对的 DNA 引物序列的氨基支链上 [图 3(e)]。当 mRNA 翻译完成时, 新生肽链与 DNA 引物上的嘌呤霉素共价交联。此后立即启动 mRNA 的逆转录, 并基于 DNA 引物进行延伸, 生成 cDNA-蛋白质二元复合物。cDNA 的半衰期超 mRNA 4 倍, 更高的稳定性保证了其可承受高变性压力下的筛选^[81]。

2.1.5 不同展示技术的比较

细胞表面展示、噬菌体表面展示、细菌芽孢表面展示、核糖体展示、mRNA 展示和 cDNA 展示等技术是几类平行发展的稳定性筛选方法, 具有各自独特的优势, 应当综合考量目标蛋白的特征和应用场景的需求选择合适的筛选方法。

细胞表面展示、噬菌体表面展示、细菌芽孢表面展示中的蛋白质合成是在生物体内发生的, 可归类为体内展示技术, 这与核糖体展示等体外展示技术截然不同。这主要是由于体内合成的蛋白质受到蛋白质质量控制网络调控, 一方面, 泛素-蛋白酶体系统与细胞自噬系统等蛋白质降解系统会加强对错误折叠或聚集沉淀蛋白质的降解^[2], 使得稳定与不稳定的蛋白质在展示效率上更具差异; 另一方面, 体内合成的蛋白质又被分子伴侣等辅助折叠的因子保护^[10], 这使得一些极端不稳定的蛋白质及其突变体能够在体内完成折叠, 并接受后续的稳定性筛查, 而它们在体外则可能发生错误折叠或聚集沉淀, 无法进行后续筛选。同时, 体内表达过程的优化也能更好地指导蛋白质的大规模生产, 这些因素均赋予了体内展示技术广阔的应用前景。然而, 体内展示技术的库容量有限, 与体外展示技术相比低 $10^2 \sim 10^4$, 在探索蛋白质全部的序列空间上存在瓶颈。此外, 体内展示技术中细胞等生物体的表面结构可能影响蛋白质与外界环境的直接接触, 在模拟蛋白质的应用环境(如酶的催化环境)上不如体外展示技术。

细胞表面展示、噬菌体表面展示、细菌芽孢表面展示中, 前二者无论在机制研究、方法搭建以及应用案例上均较后者更加完善与丰富。尽管细菌芽孢的超高压耐受性赋予了其筛选极高稳定性蛋白质的潜能^[75], 但若在筛选过程中遇到障碍, 可供参考的案例较少。细胞表面展示与噬菌体表面展示相比, 其劣势在于细胞的表面结构复

杂度较噬菌体更高, 可能对展示的目标蛋白形成一定的包埋, 掩盖蛋白质本身的属性。同时, 细胞表面展示的筛选速度较慢, 而噬菌体表面展示则可在短时间内迭代进行多轮定向进化。然而, 细胞(这里主要指酵母或哺乳动物细胞)中的真核蛋白质质量控制网络与翻译后修饰为高等生物来源的蛋白质提供了更加合适的表达系统, 这也是噬菌体表面展示所欠缺的。

2.2 利用蛋白质稳定性检测探针进行定向进化

蛋白质稳定性检测探针的核心元件是表型报告蛋白, 其折叠状态受与其融合的目标蛋白的影响, 因此其表型能够反映目标蛋白的稳定性。鉴于报告蛋白的多样性, 人们可以定制开发基于荧光、抗生素抗性等多种表型的蛋白质稳定性检测探针, 并可实现多种探针的联用。

2.2.1 首尾相连型探针

首尾相连型蛋白质稳定性检测探针通常将报告蛋白连接在目标蛋白的 C 末端 [图 4(a)]。目标蛋白在翻译过程中、翻译后的折叠状态与聚集倾向均会直接影响报告蛋白的状态。

Los Alamos 国家实验室 Waldo 课题组^[82]在 1999 年基于绿色荧光蛋白构建了首尾相连型探针。作者发现, 若目标蛋白折叠良好, 与其融合的 GFP 也能良好折叠, 产生绿色荧光。此后, Kawasaki 等^[83]利用该技术筛选得到鼠源鸟嘌呤核苷酸转化因子家族成员 Vav 蛋白的可溶性结构域; Berg 等^[84]利用该技术提升了烟草蚀纹病毒蛋白酶 (tobacco etch virus protease, TEV protease) 的可溶性表达产量。

除了荧光筛选外, 报告蛋白也可以替换为细胞生长相关蛋白, 利用生长表型检测蛋白质稳定性。相关技术包括氯霉素乙酰转移酶探针 (chloramphenicol acetyltransferase, CAT)^[85]、卡那霉素核苷酸转移酶探针 (kanamycin nucleotidyl transferase, Kan)^[86]、二氢叶酸还原酶探针 (dihydrofolate reductase, DHFR)^[87]等。

首尾相连型探针也存在一些系统性缺陷: 如部分报告蛋白在目标蛋白聚集的情况下依旧能产生表型; 报告蛋白分子量太大(尤其在目标蛋白

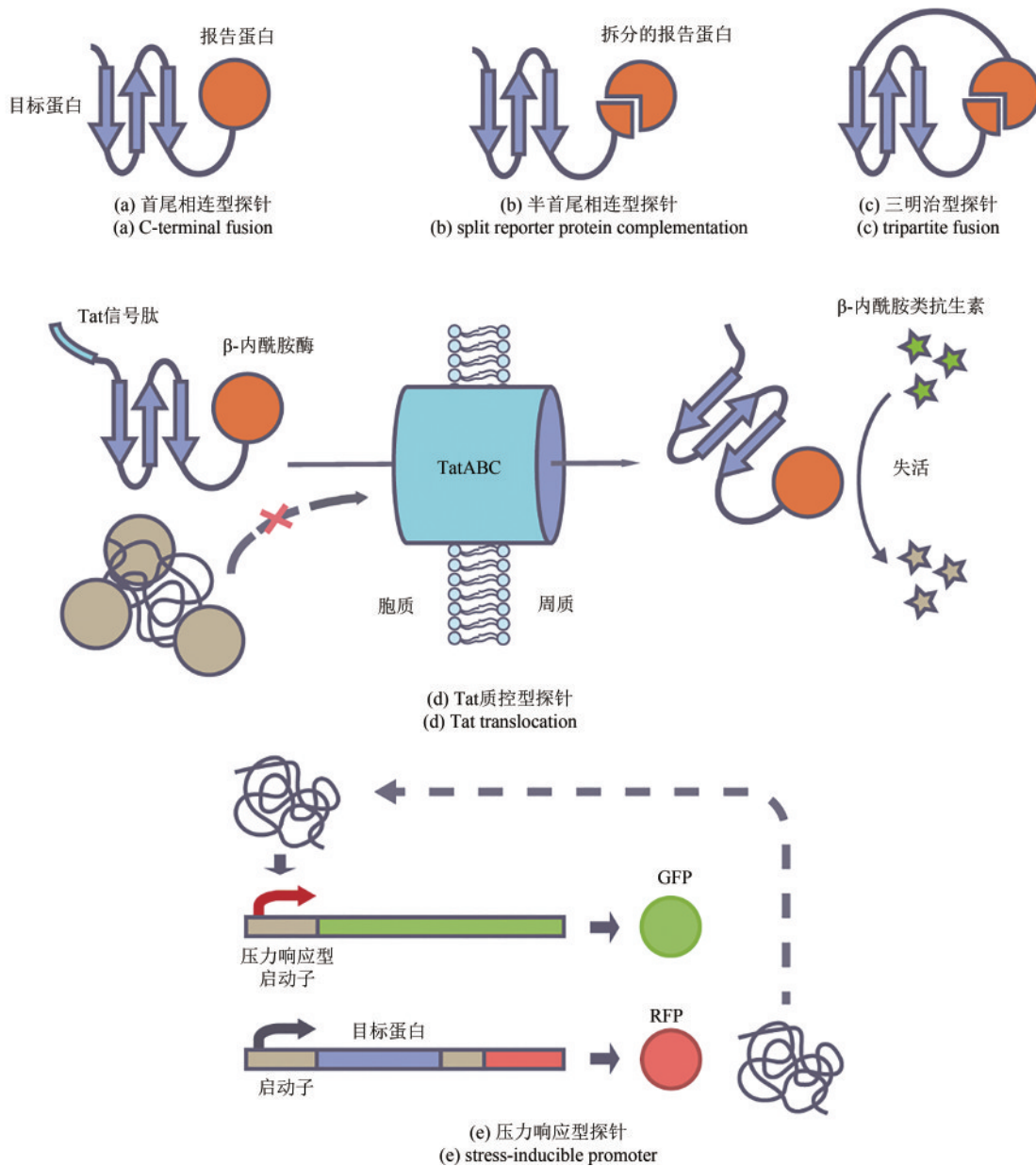


图4 各类蛋白质稳定检测探针

Fig. 4 Various protein stability biosensors

分子量远小于报告蛋白时)可能对目标蛋白的折叠并不敏感;目标蛋白降解导致游离的报告蛋白产生表型,造成假阳性结果。

2.2.2 半首尾相连型探针

Waldo 课题组^[88]于2005年在首尾相连型的基础上,进一步开发了GFP半首尾相连型探针。他们将完整的GFP蛋白拆成两部分:小片段与目标蛋白融合表达;大片段单独表达[图4(b)]。目标蛋白的良好折叠可保证小片段的天然结构与生物可及性,并使得两个GFP片段能互补形成完整的

GFP,显示出绿色荧光的表型。该设计旨在减弱报告蛋白对目标蛋白折叠过程、物理化学属性的影响。Rodríguez-Banqueri等^[89]利用该系统对枯草芽孢杆菌L-丝氨酸/L-苏氨酸转运体(L-serine/L-threonine exchanger, SteT)进行了稳定性定向进化。筛选得到的I134V/A377T稳定性突变体较野生型有全方位的提升,包括其在大肠杆菌中的表达量提升一倍,热孵育后的可溶性比例提高了60%,且活力提升50%。

Wigley等^[90]于2001年开发了基于 β -半乳糖苷

酶 (β -galactosidase, β -gal) 的半首尾相连型探针。与前文 GFP 探针类似, 作者将 β -gal 的小片段 (α) 与目标蛋白融合表达, 大片段 (ω) 则单独表达。可溶的目标蛋白使 β -gal 得以重构, 并发挥分解 5-溴-4-氯-3-吲哚- β -D-半乳糖苷 (5-bromo-4-chloro-3-indolyl β -D-galactoside, X-gal) 的功能, 使菌落显示蓝色。作者用该探针检测了阿尔兹海默病中形成纤维状聚集体的多肽 $A\beta_{42}$ 的聚集, 为该类蛋白质折叠与聚集的研究提供了新方法。然而, β -gal 探针使用的 β -半乳糖苷酶分子质量大于 100 kDa, 具有一定的细胞毒性, 造成信噪比下降、转化效率降低 (影响实际的库容量) 等问题。此外, 该类系统拆分的报告蛋白互补依赖于两片段的相互作用, 目标蛋白降解形成的报告蛋白游离片段仍可能通过该相互作用与另一部分互补, 造成假阳性结果。

2.2.3 三明治型探针

三明治型探针同样需要对报告蛋白进行拆分, 不同点在于目标蛋白会插入到报告蛋白的拆分位点之间, 并作为一个整体融合表达 [图 4(c)]。当被夹在中间的目标蛋白折叠不好或被降解时, 报告蛋白的重构过程将会受阻; 仅当目标蛋白折叠良好时, 报告蛋白才能发挥功能, 并表现出某种表型。由于报告蛋白的两拆分片段存在物理连接, 因此其重构不依赖于两片段的相互作用。故仅需在探针的开发过程中, 规避该互作, 便可保证探针在报告蛋白降解后不输出任何信号。

2005 年 Waldo 课题组 [91] 在先前开发的 GFP 首尾相连型探针的使用中发现, 目标蛋白的缓慢聚集并没有阻碍 GFP 功能的发挥, 探针依旧显示绿色荧光信号。因此, 他们在原有探针的基础上开发了三明治型探针, 其中 GFP 被拆分为氨基酸序列 1~214 和 214~230 的两个片段 [88]。2018 年, Mo 等 [92] 利用该探针, 对酯酶 Aaeol 的稳定性进行了定向进化, 并通过后续的酶活筛选, 同时将表达量与活力提升至野生型的两倍。

2009 年 Foit 等 [93] 开发了 β -内酰胺酶三明治型探针以检测周质蛋白质的稳定性。在该探针中, 目标蛋白的折叠状态或聚集倾向会关联 β la 的结构重构, 进而影响其分解 β -内酰胺类抗生素的能力, 使宿主细胞在抗生素环境下表现出差异化的生长

状况。作者使用该探针对免疫蛋白 7 (immunity protein 7, Im7)、BPTI 等蛋白质与其不同稳定性的突变体进行了测试, 发现目标蛋白突变体的稳定性与探针表型存在良好的相关性。 β la 探针的优势在于: 单体化; 分子质量较小 (29 kDa); 细胞毒性低; 周质表达, 适用于检测含二硫键的蛋白质。Radford 课题组 [18] 利用该探针对单链抗体药物 MEDI1912 进行了稳定性定向进化, 揭示了提升该单链抗体稳定性的热点区域。此外, 该探针在筛选分子伴侣、疾病相关蛋白聚集抑制剂等方面得到了较为广泛的应用, 这部分内容将在后文展开叙述。

此外, 关联细胞生长的三明治型蛋白质稳定性检测探针还包括: DsbA 探针 (氯化镉抗性) [94]、葡糖苷-3'-磷酸转移酶 (aminoglycoside-3'-phosphotransferase, APH) 探针 (卡那霉素、遗传霉素抗性) [95-96]、链霉素/壮观霉素腺苷转移酶 (streptomycin/spectinomycin adenylyltransferase, ANT) 探针 (链霉素、壮观霉素抗性) [96]、诺尔丝菌素乙酰转移酶 (nourseothricin acetyltransferase, NAT) 探针 (诺尔丝菌素抗性) [96]、DHFR 探针 (参与嘌呤及嘧啶合成并影响细胞生长) [97] 等。其中, 不同的关联抗生素抗性的探针可根据宿主对特定抗生素的敏感性、待测目标蛋白发挥功能的胞内场所选择使用。

本课题组 [98] 于 2021 年基于报告蛋白 CysG^A 开发了适用于大肠杆菌胞质的三明治型蛋白质稳定性检测探针。在该探针中, 目标蛋白被插入大肠杆菌尿卟啉原 III 甲基转移酶 (uroporphyrinogen III methyltransferase, CysG^A) 中。折叠良好的目标蛋白能使 CysG^A 发挥催化作用, 将内源性底物尿卟啉原 III (uroporphyrinogen III) 转化成三甲基咕啉 (trimethylpyrrocorphin) 和西罗双氢叶绿三酸 (sirohydrochlorin), 令细胞在紫外光下呈现出明亮的红色荧光信号。CysG^A 探针的优势在于: 基于酶催化的放大效应可提升灵敏度; 在红外检测波长下细胞内源干扰信号少; 荧光产物在无氧的条件下也能生成, 因此可在厌氧乏氧条件下工作, 显著弥补 GFP 等荧光探针的不足 (GFP 荧光团的成熟需要氧气)。利用该探针, 针对多种蛋白质及病理性淀粉样多肽筛选到了稳定性提升的突变体;

与高通量测序结合, 对人源重要表观遗传调控因子——组蛋白H3K4甲基转移酶MLL3的催化结构域进行了深度突变扫描, 获得了其单个氨基酸残基分辨率水平的全局稳定性景观图, 系统揭示了对维持其稳定性起关键作用的残基。

2.2.4 质控型探针

质控型探针的设计思路是在首尾相连型探针的基础上, 利用分泌途径蛋白质质量控制网络, 进一步强化目标蛋白的折叠状态与报告蛋白功能发挥之间的关联, 即只有折叠良好的融合蛋白才能被转运至特定的细胞区室, 从而使报告蛋白发挥功能。2006年康奈尔大学DeLisa课题组^[99]开发了一种基于双精氨酸转运系统(twin-arginine translocation, Tat)的质控型探针。在该探针中, 目标蛋白在N端融合Tat信号肽, 在C端融合 β -内酰胺酶。只有折叠良好的目标蛋白才会被Tat转位酶(translocase)转运到大肠杆菌的周质空间, 并赋予细胞抗性表型[图4(d)]。

在之后的十多年间, Tat探针在蛋白质稳定性优化方面得到了大量应用。Boock等^[100]利用该探针将内切葡聚糖酶(glycoside hydrolase family 5, Cel5A)在大肠杆菌中的产量较野生型提高了30倍; Kim等^[101]利用该探针, 筛选人胚系重链抗体库(human germ-line VH library)中的可溶性突变体, 并对筛选到的突变体进行了定点饱和突变, 揭示了疏水氨基酸堆积(hydrophobic packing)在维持蛋白质稳定性方面的重要作用。Tat探针还可用于鉴定蛋白质复合物的稳定性。对应的设计方案为: 在构成复合物的两个蛋白质上分别串联Tat信号肽与抗性蛋白; 若二者能形成稳定的复合物, 抗性蛋白可通过“搭车”的形式被Tat信号肽转运至周质空间并发挥功能^[102]。2021年DeLisa课题组^[103]对Tat转位酶进行了定向进化, 将该酶的转运效率提升了3倍以上, 大幅提高了探针的信噪比。

2.2.5 压力响应型探针

压力响应型探针可用来检测蛋白质的聚集, 其核心组件是可响应蛋白质错误折叠信号的启动子, 利用其来起始报告蛋白的表达。环境胁迫(如高温、有机溶剂、极端pH等)会造成细胞内大量蛋白质错误折叠或聚集, 细胞为应对该胁迫压力, 会启动一系列的级联应答过程, 这一过程

称为热休克应答(heat shock response, HSR)^[104]。大肠杆菌中的热休克应答被瞬时激活的转录调控因子 σ 32(也称 σ H)所介导, 其引导RNA聚合酶与 σ 32依赖型启动子结合, 并上调大量热休克基因的表达。这些基因的表达产物大多是蛋白质质量控制网络的成员, 如分子伴侣、蛋白酶等, 它们通过辅助蛋白质折叠、降解已聚集蛋白质等方式改善胞内蛋白稳态, 并提高细胞在胁迫环境下的生存能力^[105]。利用 σ 32依赖型启动子表达报告蛋白, 即可输出目标蛋白聚集的信号。

Zutz等^[106]在大肠杆菌中设计了基于*ibpA*启动子(受 σ 32调控)的压力响应型蛋白质稳定性检测探针。作者将GFP置于*ibpA*启动子下游, 以输出目标蛋白聚集的信号[图4(e)]。由于“目标蛋白不聚集”和“蛋白质不表达(GFP和目标蛋白)”这两类事件均不会产生绿色荧光表型, 因此作者将目标蛋白与mCherry(红色荧光)转录融合, 利用红色荧光信号筛选正常表达蛋白质的细胞。目标蛋白折叠良好, 则菌株会产生较高红色荧光和较低绿色荧光。利用该系统, 作者区分了胰凝乳蛋白酶抑制剂(chymotrypsin inhibitor 2, CI2)的6个突变体的稳定性。此外, 作者还针对人聚ADP核糖聚合酶1(human poly[ADP-ribose] polymerase 1, PARP1)的BRCT结构域, 比较了该探针的筛选结果和FoldX的预测结果, 发现二者在鉴定蛋白质的去稳定化位点上存在较好的一致性。

2021年Romero-Suarez等^[107]在酿酒酵母中构建了基于*SSA1*启动子的压力响应型探针。作者首先进行了热休克启动子的挖掘。他们在细胞中分别表达具有不同溶解度的蛋白质, 通过差异表达蛋白质组分析, 筛选到了31个潜在的热休克蛋白, 其中18个为分子伴侣或分子伴侣辅因子。随后, 作者以这些基因对应的启动子为系统的感应元件, 以GFP为报告蛋白, 通过绿色荧光表型, 报告目标蛋白的不稳定性信号。经报告信号强度比较, 最终选取了热休克蛋白70(heat shock protein, HSP70)家族成员Ssa1的启动子为最优的感应元件, 并通过增加热休克启动子上游激活序列的拷贝数以增强系统的响应效果。值得一提的是, 上述两项工作均选择了不稳定的GFP作为报告蛋白,

其较野生型更短的半衰期会减弱背景荧光，显著提升信噪比。

压力响应型蛋白质稳定性检测探针的优势在于，目标蛋白不需要与报告蛋白融合，其属性不会被报告蛋白干扰。然而，该探针仅对目标蛋白的聚集有响应，富集低信号细胞的筛选方式并不十分适合蛋白质稳定性的定向进化，强行对低信号细胞（表达稳定的蛋白突变体）筛选会获得大量的假阳性结果^[106]。

2.2.6 不同蛋白质稳定性检测探针的比较

首尾相连型、半首尾相连型、三明治型蛋白质稳定性检测探针在设计逻辑上存在递进关系，这使得三明治型蛋白质稳定性检测探针在检测的准确性上更具优势，但这并不意味着前二者不再具备使用价值。应当综合考量目标蛋白与探针的适配性，选择合适的探针进行稳定性定向进化。一些分子量较大或结构复杂的蛋白质在插入到三明治型探针后，报告蛋白的两个拆分片段可能由于空间距离过大或相互作用位点被包埋等因素无法重构，探针不能对该蛋白质及其突变体进行稳定性筛选。半首尾相连型探针的两拆分片段则比较灵活，可以无视上述空间距离的问题；而首尾相连型探针不存在拆分片段重构这一要求。这使得首尾相连型、半首尾相连型探针在某些蛋白质的适用性上更具优势。

不同的蛋白质稳定性检测探针在物种的可选择性上存在较大差异。质控型探针与压力响应型探针的物种选择较为固定，如Tat质控型探针仅适用于大肠杆菌周质^[99]，压力响应型探针在大肠杆菌^[106]与酿酒酵母^[107]中的开发需要筛选各自可响应蛋白质错误折叠信号的启动子。而首尾相连型、半首尾相连型、三明治型探针在物种选择上依赖于报告蛋白能否在目标物种中发挥作用。不同荧光报告蛋白的探针在使用上极为灵活，在大多数物种中均具有开发潜力；不同的关联抗生素抗性的探针可根据宿主对特定抗生素的敏感性选择使用，如APH探针与NAT探针在大肠杆菌与酿酒酵母中均被成功开发^[95-96]；不同的关联细胞通路的探针可根据该通路在目标物种中的有无，判断能否开发使用，如DHFR探针^[97]（在嘌呤、嘧啶的合成通路中发挥催化作用）、CysG^a探针^[98]（尿嘧啶

原Ⅲ相关通路中发挥催化功能，如血红素、维生素B₁₂等的合成），由于其对应通路的保守性，均具备在动物细胞中进行蛋白质稳定性检测的潜力。

2.3 直接检测目标蛋白可溶性的高通量方法

在文库表面展示技术和蛋白质稳定性检测探针中，目标蛋白通常会与其他分子融合，包括核酸、锚定蛋白或报告蛋白等。这种物理连接在一定程度上可能影响目标蛋白本身的属性，导致目标蛋白的稳定化效应是具有报告蛋白依赖性的。尽管压力响应型探针不需要将目标蛋白与报告蛋白融合，实现了独立表达的目标蛋白的检测，但其仅能对聚集蛋白产生响应的性质使其难以满足目标蛋白稳定性定向进化的需求。

Cornvik等^[108]开发了一种菌落过滤印迹法（colony filtration blot, CoFi blot）可对独立表达的可溶目标蛋白突变体进行高通量筛选。该方法的关键在于使用特定孔径的滤膜，选择性透过可溶的目标蛋白突变体。具体流程如下：使用滤膜黏附表达不同突变体的菌落，并将干净的一面置于硝酸纤维素膜上；裂解细胞，使内部蛋白质释放；聚集沉淀的蛋白质被滤膜阻拦，可溶蛋白则穿过滤膜抵达硝酸纤维素膜上；利用蛋白免疫印迹法检测硝酸纤维素膜上的蛋白质，并借鉴影印平板法的思路将其定位到原始琼脂平板上的菌落，即可鉴定到可溶表达水平明显提升的目标蛋白突变体。Asial等^[109]使用此方法成功提高了多种蛋白质的稳定性，包括人核苷二磷酸连接部分X型基元18（nucleoside diphosphate linked moiety X type motif 18, NUDT18）的水解酶结构域、抗生长因子受体结合蛋白2（growth factor receptor-bound protein 2, Grb2）单链抗体、白细胞介素-1受体拮抗剂（interleukin-1 receptor antagonist, IL1RA）、TEV蛋白酶等。

3 蛋白质稳定性改良应用实例

蛋白质的折叠状态由其序列与所处环境共同决定，因此序列改良与折叠环境优化提升蛋白质稳定性在科学研究与工程应用领域具有重要价值。

由于本文在“计算设计提升蛋白质稳定性”及“定向进化提升蛋白质稳定性”两部分内容中已对序列改良做了充分探讨，此处仅对该内容略作补充，着重阐述折叠环境优化相关案例。

3.1 序列改良与疫苗开发

近两年来肆虐的新冠疫情加速了各类新概念疫苗的推广。新冠病毒（SARS-CoV-2）侵染细胞时，其刺突蛋白（spike）上的受体结合域（receptor-binding domain, RBD）与人血管收缩素转化酶2（angiotensin converting enzyme 2, ACE2）结合，诱导spike从不稳定的融合前构象（prefusion state）转变为稳定的融合后构象（postfusion state），这一过程介导病毒与宿主的膜融合，使病毒基因组进入到宿主细胞^[110]。由于针对融合前构象的免疫应答更加有效，因此基于刺突蛋白开发疫苗时，需要将刺突蛋白维持在融合前构象以提升其免疫原性^[111-113]。

Pallesen等^[114]在中东呼吸综合征冠状病毒（Middle East respiratory syndrome coronavirus, MERS-CoV）的spike蛋白的中心螺旋上引入一对连续的脯氨酸，限制spike氨基酸骨架的扭转，以约束spike的构象转化。基于类似设计原理，美国国家卫生研究院与莫德纳公司联合开发了针对新冠病毒的mRNA疫苗mRNA-1273，其编码脯氨酸改造后的spike^[115]。与其他几款主流疫苗的比对发现，mRNA-1273具有最佳预防感染效果^[116]。融合前构象的RBD存在开放（up）与闭合（down）两种构象，仅开放构象的RBD会与ACE2结合并触发spike构象转化，因此通过结构上的设计，将RBD锁定在闭合构象也是一种稳定化融合前构象的策略。Xiong等^[117]与Henderson等^[118]在spike的RBD与S2亚基之间设计了一对二硫键，限制RBD的摆动，并将其全部锁定在了down构象。该spike突变体稳定性大幅提升且不再具备结合ACE2的能力。

3.2 折叠环境优化提升蛋白质稳定性

折叠环境对蛋白质的稳定性起到关键作用，因此挖掘影响蛋白质折叠的相关因子与生物系统、

利用这些因子与系统改良蛋白质稳定性或维持蛋白质稳态是生物医药、工业生产等领域的重要研究内容。

3.2.1 利用蛋白质稳定性检测探针挖掘分子伴侣与蛋白质聚集抑制剂

分子伴侣是细胞中蛋白质质量控制网络的重要成员，通常分子伴侣的共表达是提升目标蛋白可溶性产量的一种重要方式。提升分子伴侣活性、筛选作用于特定目标蛋白的分子伴侣以及挖掘未知的分子伴侣对生命科学研究及蛋白质生产具有重要意义^[119]。

利用蛋白质稳定性检测探针筛选分子伴侣的代表性工作来自于密歇根大学Bardwell课题组^[120-121]。Quan等^[94]针对不稳定蛋白突变体Im7 L53A I54A，利用 β -内酰胺酶（青霉素抗性）和DsbA（氯化镉抗性）两种三明治型蛋白质稳定性检测探针，筛选位于周质空间的新型折叠辅助因子。作者将两种探针同时导入大肠杆菌后，通过化学诱变的方法对大肠杆菌的基因组进行随机突变，针对青霉素和镉离子抗性进行筛选，分离出能够使Im7 L53A I54A的可溶性含量增加700多倍的菌株。这些菌株中均过量表达了周质蛋白Spy（spheroplast protein Y），经鉴定其具有分子伴侣活性。后续利用 β la-Im7 L53A I54A探针，筛选Spy随机突变文库，获得了分子伴侣活性进一步提升的Spy突变体。

化学分子伴侣（chemical chaperone）是一类低分子量、可提升蛋白质稳定性的化合物，如甘油、山梨醇、L-脯氨酸等。其可在不破坏蛋白质共价键的情况下，影响蛋白质的折叠过程。化学分子伴侣潜在的作用机制包括：与蛋白质表面的疏水氨基酸结合，屏蔽蛋白质间的非特异性相互作用并阻止蛋白质聚集沉淀；排阻水分子并给蛋白质提供疏水环境，以此提升了蛋白质的折叠自由能，提高蛋白质的稳定性^[122]。Hailu等^[123]探讨了利用蛋白质稳定性检测探针筛选化学分子伴侣的潜力。在该工作中，作者利用 β -内酰胺酶探针对十余种已知的化学分子伴侣稳定Im7 I22V突变体的效果做了比较，发现探针不仅能响应渗透进细胞的化学分子伴侣，且探针信号与化学分子伴侣的浓度线性相关，实现了针对特定蛋白质的化学

分子伴侣快速筛选。

蛋白质错误折叠可能形成不定形沉淀或纤维状聚集体。据预测在人类孟德尔遗传病数据库 (Online Mendelian Inheritance in Man, OMIM) 中, 25% 以上的氨基酸突变会对蛋白质的稳定性造成显著影响^[124], 这表明蛋白质折叠与疾病的关联可能比我们目前已知的更为广泛。因此, 从折叠角度出发, 增强蛋白质体内稳定性、恢复蛋白质结构与功能已成为现代药物开发的一个重要思路, 例如筛选针对淀粉样蛋白聚集的抑制剂。目前也有基于此原理的药物成功上市, 如前文提到的针对转甲状腺素蛋白淀粉样变性心肌病的氯苯唑酸 (tafamidis)。

Radford 课题组^[125]利用 β -内酰胺酶蛋白质稳定性检测探针对 II 型糖尿病相关的人类胰岛淀粉样多肽 (human islet amyloid polypeptide, hIAPP) 进行了小分子抑制剂的快速筛选。作者将转化 β la-hIAPP 后的大肠杆菌置于含有氨基青霉素的培养基中生长, 且在生长过程中加入待筛选的小分子。若小分子具有稳定 hIAPP 的功能, 可使 β la 结构更好地重构, 使大肠杆菌获得更强的抗性。因此即可通过大肠杆菌的生长初步判定小分子的抑制聚集效果。作者利用此方法对 109 种小分子做了快速鉴定, 发现该系统可以有效表征已知具有抑制效果的小分子, 如姜黄素 (curcumin), 且筛选到了新型 hIAPP 聚集抑制剂多巴胺 (dopamine)。

3.2.2 折叠环境改良与药物开发

随着对疾病机制研究的深入, 蛋白质稳定性在生物医药领域中的重要性日益凸显。明星抑癌基因 *TP53* 是细胞响应 DNA 损伤、进行信号传导并启动修复的关键基因。癌症病例中发现该基因的突变类型超过 1000 种。这些突变大致可分为使蛋白质折叠受影响的结构型突变和使 p53 与 DNA 结合受影响的功能型突变。由于致病机制的差异, 需要针对 p53 突变体的不同类型进行药物设计。上海交通大学卢敏课题组与牛津大学卢欣课题组发现, p53 内部存在的一个空腔造成整体结构的不稳定, 且这种不稳定性会被结构型突变所加剧^[126]。而三氧化二砷中的砷原子可以有效填补这个空腔, 并与邻近的三个半胱氨酸发生共价相互作用, 提升 p53 的稳定性, 维持其抑癌功能。

体内蛋白质稳态的维持对于生命健康具有重要意义。近年来提出的靶向蛋白降解技术成为了该领域的佼佼者。靶向蛋白降解利用一种能同时靶向目标蛋白与细胞蛋白降解系统的多功能分子, 实现目标蛋白的清除。该技术包括适用于胞内蛋白、靶向泛素-蛋白酶体的蛋白降解靶向嵌合体 (proteolysis targeting chimeras, PROTACs)^[127], 以及适用于膜蛋白与胞外蛋白、靶向溶酶体的溶酶体靶向嵌合体系统 (lysosome targeting chimeras, LYTACs)^[128] 等。中国药科大学的尤启冬课题组^[129]开发了针对阿尔兹海默病中异常堆积的 Tau 蛋白的 PROTAC 分子。该 PROTAC 分子由两段分别靶向 Tau 与 E3 连接酶 KEAP1 的短肽组成, 具有较好的细胞通透性。珞诺生物 (Ranok Therapeutics) 公司近期开发中的分子伴侣介导的靶向蛋白质降解技术 (chaperone-mediated protein degraders, CHAMPs) 旨在实现对白血病细胞的靶向清除。该技术使用一个同时靶向转录因子 BRD4 与 HSP90 的双功能分子, 借用 HSP90 复合物招募泛素蛋白酶体机器降解 BRD4 并杀伤细胞。由于 HSP90 复合物在肿瘤细胞内高度活跃, 故该分子具有肿瘤细胞选择性。

与靶向蛋白质并使其降解这一概念相对的便是靶向蛋白质并使其稳定化。加州大学伯克利分校的 Nomura 课题组^[130]开发了一种去泛素化酶靶向嵌合体 (deubiquitinase-targeting chimeras, DUBTACs)。该双功能分子结合目标蛋白的同时, 募集去泛素化酶, 移除目标蛋白上的泛素化标签, 由此达成蛋白质的稳定化。作者利用该平台, 成功实现了囊性纤维化跨膜传导调节因子 (cystic fibrosis transmembrane conductance regulator, CFTR) 突变体以及肿瘤抑制激酶 WEE1 的稳定化。

4 展望

本文详细综述了蛋白质稳定性的内涵以及两大类优化蛋白质稳定性的方法, 包括理性或计算机辅助设计和基于文库展示或蛋白质稳定性检测探针的定向进化方法。随着生命科学与生物工程理念、技术上的进步, 蛋白质稳定化工程也将迎来自我革新。

由于蛋白质折叠状态受自身序列与所处环境协同影响,蛋白质稳定化工程的一个重要目标是提升蛋白质在其天然环境或应用场景中的稳定性。这对于蛋白质的体内稳定化及工业酶的体外应用而言具有重要意义。因此,针对环境定制目标蛋白的稳定化条件是当下蛋白质稳定化工程亟需发展与完善的方向。

目前,计算设计提升蛋白质稳定性的方法尽管能最大限度地探索蛋白质的序列空间,但不能对复杂的环境(包括胞质环境、亚细胞器环境、体外应用场景环境等)进行模拟。细胞、噬菌体或芽孢表面展示技术会将目标蛋白限定在其展示的生物系统上,也很难根据目标蛋白的需求定制折叠环境。核糖体展示技术、mRNA展示技术与cDNA展示技术等体外展示方法在保证文库容量的同时,能够针对目标蛋白的体外应用场景改变筛选环境,十分适用于工业酶的改造。然而,这种体外表达系统忽视了蛋白质的体内表达过程,其潜在的风险是蛋白质可能在生产过程中低表达,提升了应用成本。此外,这类体外展示方法也不能重现细胞内部乃至亚细胞器中的复杂微环境,这对于体内蛋白质的稳定化而言无疑是一种缺陷。

蛋白质稳定性检测探针技术能兼具各类蛋白质稳定化场景的定制需求,其具备典型的合成生物学特征,可通过基本功能元件的组装,如不同的信号肽、锚定蛋白等,将目标蛋白分泌至胞外或运输至特定的亚细胞环境中,供稳定性检测。然而,当前开发的蛋白质稳定性检测探针仅涉及细菌胞质与周质、真核细胞胞质,因此新型探针的开发迫在眉睫。

一些新兴技术为蛋白质稳定性检测探针的设计提供新策略,如近年来开发并完善的微流控技术为分泌蛋白的稳定性优化提供了“硬件”支持^[131]。分泌表达具有对细胞毒性低、分离纯化简单、蛋白质质控严格等特点,逐渐成为了药物蛋白、工业酶的首选表达方案。在利用流式细胞术进行筛选的过程中,要将蛋白质稳定性检测探针限定在细胞内或锚定在细胞表面,目标蛋白并未被完全释放到培养基中,因此其所处环境与分泌的目标蛋白并不完全一致。若要严格模拟目标蛋白的分泌,则可采用微流控技术。其通过“油包

水”的形式,生成大量包裹单细胞的液滴,且每一个液滴可以作为一个独立的培养与检测单元。蛋白质稳定性检测探针与微流控技术联用后,携带目标蛋白的探针在分泌后可保留在液滴内以供检测。

不同亚细胞器的信号肽、膜蛋白等元件为蛋白质稳定性检测探针在特定亚细胞环境中的使用提供了“软件”支持。然而,亚细胞器蛋白质稳定性检测探针的开发需要解决多个关键技术问题。首先,要尽可能减少探针针对目标蛋白的影响,包括目标蛋白的结构、功能、细胞器递送效率等。其次,探针尽量只在进入指定的细胞器后才输出信号,滞留在细胞质等其他区间的探针不能输出干扰信号,这使得泛用的荧光蛋白不满足报告蛋白的要求。最后,探针的输出信号须放大至足够强度以供检测,这主要是由于亚细胞器中的蛋白质较胞质而言表达量较低,加之探针被多重膜结构包埋,输出信号难以摆脱背景。

要实现蛋白质体内稳定性检测的全覆盖,蛋白质稳定性检测探针还需攻克大量的技术难点。生物体中的蛋白质存在着多种呈现形式,如可溶蛋白、磷脂结构中的膜蛋白、功能性的不可溶蛋白(如肌动蛋白)等。目前,蛋白质稳定性检测探针的设计仅能针对可溶蛋白,膜蛋白等探针尚有待开发。此外,蛋白质形成同源或异源多聚体并发挥功能是一种普遍的生物机制,针对该类蛋白质开发相应的探针更可能是一种“case by case”的方式。综上,蛋白质稳定性检测探针的开发还存在着广阔的前景。

随着生命科学研究的推进与认知的深入,蛋白质稳定性将逐渐褪去其理论化的外衣,为生物工程、生物医药带来颠覆性的突破。

参 考 文 献

- [1] DILL K A, MACCALLUM J L. The protein-folding problem, 50 years on[J]. *Science*, 2012, 338(6110): 1042-1046.
- [2] KIM Y E, HIPPE M S, BRACHER A, et al. Molecular chaperone functions in protein folding and proteostasis[J]. *Annual Review of Biochemistry*, 2013, 82: 323-355.
- [3] BALCHIN D, HAYER-HARTL M, HARTL F U. *In vivo* aspects of protein folding and quality control[J]. *Science*, 2016,

- 353(6294): aac4354.
- [4] HARTL F U, BRACHER A, HAYER-HARTL M. Molecular chaperones in protein folding and proteostasis[J]. *Nature*, 2011, 475(7356): 324-332.
- [5] HARTL F U, HAYER-HARTL M. Converging concepts of protein folding *in vitro* and *in vivo*[J]. *Nature Structural & Molecular Biology*, 2009, 16(6): 574-581.
- [6] TRAXLMAYR M W, OBINGER C. Directed evolution of proteins for increased stability and expression using yeast display[J]. *Archives of Biochemistry and Biophysics*, 2012, 526(2): 174-180.
- [7] SANCHEZ-RUIZ J M. Protein kinetic stability[J]. *Biophysical Chemistry*, 2010, 148(1/2/3): 1-15.
- [8] POLIZZI K M, BOMMARIUS A S, BROERING J M, et al. Stability of biocatalysts[J]. *Current Opinion in Chemical Biology*, 2007, 11(2): 220-225.
- [9] TAVERNA D M, GOLDSTEIN R A. Why are proteins marginally stable? [J]. *Proteins: Structure, Function, and Genetics*, 2002, 46(1): 105-109.
- [10] 唐宇琦, 叶松涛, 刘嘉, 等. 分子伴侣作用下的蛋白质稳定与进化[J]. *合成生物学*, 2022, 3(3): 445-464.
TANG Y Q, YE S T, LIU J, et al. Molecular chaperones promote protein stability and evolution[J]. *Synthetic Biology Journal*, 2022, 3(3): 445-464.
- [11] BENILOVA I, KARRAN E, DE STROOPER B. The toxic A β oligomer and Alzheimer's disease: an emperor in need of clothes[J]. *Nature Neuroscience*, 2012, 15(3): 349-357.
- [12] SPILLANTINI M G, GOEDERT M. Tau pathology and neurodegeneration[J]. *The Lancet Neurology*, 2013, 12(6): 609-622.
- [13] SINGLETON A B, FARRER M, JOHNSON J, et al. α -Synuclein locus triplication causes Parkinson's disease[J]. *Science*, 2003, 302(5646): 841.
- [14] ZRAIKA S, HULL R L, VERCHERE C B, et al. Toxic oligomers and islet beta cell death: guilty by association or convicted by circumstantial evidence? [J]. *Diabetologia*, 2010, 53(6): 1046-1056.
- [15] COELHO T, MERLINI G, BULAWA C E, et al. Mechanism of action and clinical application of tafamidis in hereditary transthyretin amyloidosis[J]. *Neurology and Therapy*, 2016, 5(1): 1-25.
- [16] BULAWA C E, CONNELLY S, DEVIT M, et al. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(24): 9629-9634.
- [17] HIROSE M, KURODA Y, MURATA E. NGF/TrkA signaling as a therapeutic target for pain[J]. *Pain Practice*, 2016, 16(2): 175-182.
- [18] EBO J S, SAUNDERS J C, DEVINE P W A, et al. An *in vivo* platform to select and evolve aggregation-resistant proteins[J]. *Nature Communications*, 2020, 11: 1816.
- [19] SETIAWAN D, BRENDER J, ZHANG Y. Recent advances in automated protein design and its future challenges[J]. *Expert Opinion on Drug Discovery*, 2018, 13(7): 587-604.
- [20] GOIHERBERG E, DYM O, TEL-OR S, et al. A single proline substitution is critical for the thermostabilization of *Clostridium beijerinckii* alcohol dehydrogenase[J]. *Proteins: Structure, Function, and Bioinformatics*, 2007, 66(1): 196-204.
- [21] STERNKE M, TRIPP K W, BARRICK D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(23): 11275-11284.
- [22] GUMULYA Y, BAEK J M, WUN S J, et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction[J]. *Nature Catalysis*, 2018, 1(11): 878-888.
- [23] WANG C H, HUANG R B, HE B F, et al. Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis[J]. *BMC Bioinformatics*, 2012, 13: 263.
- [24] STRICKLER S S, GRIBENKO A V, GRIBENKO A V, et al. Protein stability and surface electrostatics: a charged relationship[J]. *Biochemistry*, 2006, 45(9): 2761-2766.
- [25] CHAN C H, WILBANKS C C, MAKHATADZE G I, et al. Electrostatic contribution of surface charge residues to the stability of a thermophilic protein: benchmarking experimental and predicted pK_a values[J]. *PLoS One*, 2012, 7(1): e30296.
- [26] LAWRENCE M S, PHILLIPS K J, LIU D R. Supercharging proteins can impart unusual resilience[J]. *Journal of the American Chemical Society*, 2007, 129(33): 10110-10112.
- [27] REETZ M T, CARBALLEIRA J D, VOGEL A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability[J]. *Angewandte Chemie International Edition*, 2006, 45(46): 7745-7751.
- [28] YU H R, HUANG H. Engineering proteins for thermostability through rigidifying flexible sites[J]. *Biotechnology Advances*, 2014, 32(2): 308-315.
- [29] ZHANG C, SAMAD M, YU H R, et al. Computational design to reduce conformational flexibility and aggregation rates of an antibody fab fragment[J]. *Molecular Pharmaceutics*, 2018, 15(8): 3079-3092.
- [30] FANG L, CHOW K M, HOU S R, et al. Rational design, preparation, and characterization of a therapeutic enzyme mutant with improved stability and function for cocaine detoxification[J]. *ACS Chemical Biology*, 2014, 9(8): 1764-1772.
- [31] GIL-GARCIA M, BAÑÓ-POLO M, VAREJÃO N, et al. Combining structural aggregation propensity and stability predictions to redesign protein solubility[J]. *Molecular Pharmaceutics*, 2018, 15(9): 3846-3859.

- [32] DELGADO J, RADUSKY L G, CIANFERONI D, et al. FoldX 5.0: working with RNA, small molecules and a new graphical interface[J]. *Bioinformatics*, 2019, 35(20): 4168-4169.
- [33] SIMONS K T, BONNEAU R, RUCZINSKI I, et al. Ab initio protein structure prediction of CASP III targets using ROSETTA[J]. *Proteins: Structure, Function, and Genetics*, 1999(S3): 171-176.
- [34] GUEROIS R, NIELSEN J E, SERRANO L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations[J]. *Journal of Molecular Biology*, 2002, 320(2): 369-387.
- [35] BUß O, MULLER D, JAGER S, et al. Improvement in the thermostability of a β -amino acid converting ω -transaminase by using FoldX[J]. *ChemBioChem*, 2018, 19(4): 379-387.
- [36] YU H R, HERNÁNDEZ LÓPEZ R I, STEADMAN D, et al. Engineering transketolase to accept both unnatural donor and acceptor substrates and produce α -hydroxyketones[J]. *The FEBS Journal*, 2020, 287(9): 1758-1776.
- [37] WIJMA H J, FLOOR R J, JEKEL P A, et al. Computationally designed libraries for rapid enzyme stabilization[J]. *Protein Engineering, Design and Selection*, 2014, 27(2): 49-58.
- [38] GOLDENZWEIG A, GOLDSMITH M, HILL S E, et al. Automated structure- and sequence-based design of proteins for high bacterial expression and stability[J]. *Molecular Cell*, 2016, 63(2): 337-346.
- [39] MUSIL M, STOURAC J, BENDL J, et al. FireProt: web server for automated design of thermostable proteins[J]. *Nucleic Acids Research*, 2017, 45(W1): W393-W399.
- [40] SHROFF R, COLE A W, DIAZ D J, et al. Discovery of novel gain-of-function mutations guided by structure-based deep learning[J]. *ACS Synthetic Biology*, 2020, 9(11): 2927-2935.
- [41] LU H Y, DIAZ D J, CZARNECKI N J, et al. Machine learning-aided engineering of hydrolases for PET depolymerization[J]. *Nature*, 2022, 604(7907): 662-667.
- [42] BAVA K A, GROMIHA M M, UEDAIRA H, et al. ProTherm, version 4.0: thermodynamic database for proteins and mutants[J]. *Nucleic Acids Research*, 2004, 32(suppl_1): D120-D121.
- [43] CAPRIOTTI E, FARISELLI P, CASADIO R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure[J]. *Nucleic Acids Research*, 2005, 33(suppl_2): W306-W310.
- [44] RASHID P M A, SALIH G F. Molecular and computational analysis of spike protein of newly emerged omicron variant in comparison to the delta variant of SARS-CoV-2 in Iraq[J]. *Molecular Biology Reports*, 2022, 49(8): 7437-7445.
- [45] CAO H L, WANG J X, HE L P, et al. DeepDDG: predicting the stability change of protein point mutations using neural networks[J]. *Journal of Chemical Information and Modeling*, 2019, 59(4): 1508-1514.
- [46] ROMERO P A, KRAUSE A, ARNOLD F H. Navigating the protein fitness landscape with Gaussian processes[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(3): E193-E201.
- [47] CONTESSOTO V G, DE OLIVEIRA V M, FERNANDES B R, et al. TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions[J]. *Proteins: Structure, Function, and Bioinformatics*, 2018, 86(11): 1184-1188.
- [48] KANTARDJIEV A A, ATANASOV B P. PHEPS: web-based pH-dependent protein electrostatics server[J]. *Nucleic Acids Research*, 2006, 34(suppl_2): W43-W47.
- [49] REETZ M T, SONI P, FERNÁNDEZ L, et al. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method[J]. *Chemical Communications*, 2010, 46(45): 8657-8658.
- [50] CRAIG D B, DOMBKOWSKI A A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins[J]. *BMC Bioinformatics*, 2013, 14: 346.
- [51] CERONI A, PASSERINI A, VULLO A, et al. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server[J]. *Nucleic Acids Research*, 2006, 34(suppl_2): W177-W181.
- [52] KUIPERS R K, JOOSTEN H J, VAN BERKEL W J H, et al. 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities[J]. *Proteins: Structure, Function, and Bioinformatics*, 2010, 78(9): 2101-2113.
- [53] JONES B J, KAN C N E, LUO C, et al. Consensus Finder web tool to predict stabilizing substitutions in proteins[J]. *Methods in Enzymology*, 2020, 643: 129-148.
- [54] DIALLO A B, MAKARENKO V, BLANCHETTE M. Ancestors 1.0: a web server for ancestral sequence reconstruction[J]. *Bioinformatics*, 2010, 26(1): 130-131.
- [55] MUSIL M, KHAN R T, BEIER A, et al. FireProtASR: a web server for fully automated ancestral sequence reconstruction[J]. *Briefings in Bioinformatics*, 2021, 22(4): bbaa337.
- [56] DEHOUCK Y, KWASIGROCH J M, GILIS D, et al. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality[J]. *BMC Bioinformatics*, 2011, 12: 151.
- [57] KOWALSKI J M, PAREKH R N, MAO J, et al. Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control[J]. *Journal of Biological Chemistry*, 1998, 273(31): 19453-19458.
- [58] COUGHLAN C M, WALKER J L, COCHRAN J C, et al. Degradation of mutated bovine pancreatic trypsin inhibitor in the yeast vacuole suggests post-endoplasmic reticulum protein quality control[J]. *Journal of Biological Chemistry*, 2004,

- 279(15): 15289-15297.
- [59] SHUSTA E V, KIEKE M C, PARKE E, et al. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency[J]. *Journal of Molecular Biology*, 1999, 292(5): 949-956.
- [60] KIEKE M C, SHUSTA E V, BODER E T, et al. Selection of functional T cell receptor mutants from a yeast surface-display library[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(10): 5651-5656.
- [61] TRAXLMAYR M W, FAISSNER M, STADLMAYR G, et al. Directed evolution of stabilized IgG1-Fc scaffolds by application of strong heat shock to libraries displayed on yeast[J]. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 2012, 1824(4): 542-549.
- [62] ORR B A, CARR L M, WITTRUP K D, et al. Rapid method for measuring ScFv thermal stability by yeast surface display[J]. *Biotechnology Progress*, 2003, 19(2): 631-638.
- [63] HAGIHARA Y, KIM P S. Toward development of a screen to identify randomly encoded, foldable sequences[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(10): 6619-6624.
- [64] SMITH G P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface[J]. *Science*, 1985, 228(4705): 1315-1317.
- [65] PACKER M S, LIU D R. Methods for the directed evolution of proteins[J]. *Nature Reviews Genetics*, 2015, 16(7): 379-394.
- [66] KRISTENSEN P, WINTER G. Proteolytic selection for protein folding using filamentous bacteriophages[J]. *Folding and Design*, 1998, 3(5): 321-328.
- [67] SIEBER V, PLÜCKTHUN A, SCHMID F X. Selecting proteins with improved stability by a phage-based method[J]. *Nature Biotechnology*, 1998, 16(10): 955-960.
- [68] MARTIN A, SCHMID F X, SIEBER V. Proside: A phage-based method for selecting thermostable proteins[M]//Directed enzyme evolution. New Jersey: Humana Press, 2003: 57-70.
- [69] KATHER I, JAKOB R P, DOBBEK H, et al. Increased folding stability of TEM-1 beta-lactamase by *in vitro* selection[J]. *Journal of Molecular Biology*, 2008, 383(1): 238-251.
- [70] PERSHAD K, KAY B K. Generating thermal stable variants of protein domains through phage display[J]. *Methods*, 2013, 60(1): 38-45.
- [71] ESVELT K M, CARLSON J C, LIU D R. A system for the continuous directed evolution of biomolecules[J]. *Nature*, 2011, 472(7344): 499-503.
- [72] WANG T N, BADRAN A H, HUANG T P, et al. Continuous directed evolution of proteins with improved soluble expression[J]. *Nature Chemical Biology*, 2018, 14(10): 972-980.
- [73] 刘陈立, 赖旺生, 陈茜, 等. 一种可视的连续空间定向进化方法: CN109971693A[P]. 2019-07-05.
- LIU C L, LAI W S, CHEN Q, et al. Visual continuous space directional evolution method: CN109971693A[P]. 2019-07-05.
- [74] PAN J G, KIM E J, YUN C H. *Bacillus* spore display[J]. *Trends in Biotechnology*, 2012, 30(12): 610-612.
- [75] ZHANG G Y, AN Y F, ZABED H, et al. *Bacillus subtilis* spore surface display technology: a review of its development and applications[J]. *Journal of Microbiology and Biotechnology*, 2019, 29(2): 179-190.
- [76] SHENG S L, JIA H, TOPIOL S, et al. Engineering CotA lacase for acidic pH stability using *Bacillus subtilis* spore display[J]. *Journal of Microbiology and Biotechnology*, 2017, 27(3): 507-513.
- [77] REN C, WEN X, MENCIOUS J, et al. Selection and screening strategies in directed evolution to improve protein stability[J]. *Bioresources and Bioprocessing*, 2019, 6: 53.
- [78] JERMUTUS L, HONEGGER A, SCHWESINGER F, et al. Tailoring *in vitro* evolution for protein affinity or stability[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(1): 75-80.
- [79] NEMOTO N, MIYAMOTO-SATO E, HUSIMI Y, et al. *In vitro* virus bonding of mRNA bearing puromycin at the 3' terminal [J]. *Federation of European Biochemical Societies*, 1997, 414(2): 405-408.
- [80] UENO S, NEMOTO N. cDNA display: rapid stabilization of mRNA display[J]. *Methods in Molecular Biology*, 2012, 805: 113-135.
- [81] KURZ M, GU K, AL-GAWARI A, et al. cDNA-protein fusions: covalent protein-gene conjugates for the *in vitro* selection of peptides and proteins[J]. *Chembiochem: a European Journal of Chemical Biology*, 2001, 2(9): 666-672.
- [82] WALDO G S, STANDISH B M, BERENDZEN J, et al. Rapid protein-folding assay using green fluorescent protein[J]. *Nature Biotechnology*, 1999, 17(7): 691-695.
- [83] KAWASAKI M, INAGAKI F. Random PCR-based screening for soluble domains using green fluorescent protein[J]. *Biochemical and Biophysical Research Communications*, 2001, 280(3): 842-844.
- [84] VAN DEN BERG S, LOFDAHL P A, HARD T, et al. Improved solubility of TEV protease by directed evolution[J]. *Journal of Biotechnology*, 2006, 121(3): 291-298.
- [85] BÉHAR G, SOLÉ V, DEFONTAINE A, et al. Evolution of interleukin-15 for higher *E. coli* expression and solubility[J]. *Protein Engineering, Design and Selection*, 2011, 24(3): 283-290.
- [86] CHAUTARD H, BLAS-GALINDO E, MENGUY T, et al. An activity-independent selection system of thermostable protein variants[J]. *Nature Methods*, 2007, 4(11): 919-921.
- [87] LIU J W, BOUCHER Y, STOKES H W, et al. Improving protein solubility: the use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter[J]. *Protein Expression and Pu-*

- rification, 2006, 47(1): 258-263.
- [88] CABANTOUS S, TERWILLIGER T C, WALDO G S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein[J]. *Nature Biotechnology*, 2005, 23(1): 102-107.
- [89] RODRÍGUEZ-BANQUERI A, ERRASTI-MURUGARREN E, BARTOCCIONI P, et al. Stabilization of a prokaryotic LAT transporter by random mutagenesis[J]. *Journal of General Physiology*, 2016, 147(4): 353-368.
- [90] WIGLEY W C, STIDHAM R D, SMITH N M, et al. Protein solubility and folding monitored *in vivo* by structural complementation of a genetic marker protein[J]. *Nature Biotechnology*, 2001, 19(2): 131-136.
- [91] CABANTOUS S, PÉDELACQ J D, MARK B L, et al. Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*[J]. *Journal of Structural and Functional Genomics*, 2005, 6(2/3): 113-119.
- [92] NAGUMO Y, FUJIWARA K, HORISAWA K, et al. PURE mRNA display for *in vitro* selection of single-chain antibodies[J]. *The Journal of Biochemistry*, 2016, 159(5): 519-526.
- [93] FOIT L, MORGAN G J, KERN M J, et al. Optimizing protein stability *in vivo*[J]. *Molecular Cell*, 2009, 36(5): 861-871.
- [94] QUAN S, KOLDEWEY P, TAPLEY T, et al. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy[J]. *Nature Structural & Molecular Biology*, 2011, 18(3): 262-269.
- [95] SACHSENHAUSER V, DENG X X, KIM H H, et al. Yeast tripartite biosensors sensitive to protein stability and aggregation propensity[J]. *ACS Chemical Biology*, 2020, 15(4): 1078-1088.
- [96] MALIK A, MUELLER-SCHICKERT A, BARDWELL J C A. Cytosolic selection systems to study protein stability[J]. *Journal of Bacteriology*, 2014, 196(24): 4333-4343.
- [97] PITTMAN A M C, LAGE M D, POLTORATSKY V, et al. Rapid profiling of disease alleles using a tunable reporter of protein misfolding[J]. *Genetics*, 2012, 192(3): 831-842.
- [98] REN C, WEN X, MENCIOUS J, et al. An enzyme-based biosensor for monitoring and engineering protein stability *in vivo*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(13): e2101618118.
- [99] FISHER A C, KIM W, DELISA M P. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway[J]. *Protein Science*, 2006, 15(3): 449-458.
- [100] BOOCK J T, KING B C, TAW M N, et al. Repurposing a bacterial quality control mechanism to enhance enzyme production in living cells[J]. *Journal of Molecular Biology*, 2015, 427(6): 1451-1463.
- [101] KIM D S, SONG H N, NAM H J, et al. Directed evolution of human heavy chain variable domain (V_H) using *in vivo* protein fitness filter[J]. *PLoS One*, 2014, 9(6): e98178.
- [102] WARAHO D, DELISA M P. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(10): 3692-3697.
- [103] TAW M N, LI M J, KIM D, et al. Engineering a supersecreting strain of *Escherichia coli* by directed coevolution of the multi-protein Tat translocation machinery[J]. *ACS Synthetic Biology*, 2021, 10(11): 2947-2958.
- [104] RICHTER K, HASLBECK M, BUCHNER J, et al. The heat shock response: life on the verge of death[J]. *Molecular Cell*, 2010, 40(2): 253-266.
- [105] LIM B, MIYAZAKI R, NEHER S, et al. Heat shock transcription factor σ_{32} co-opts the signal recognition particle to regulate protein homeostasis in *E. coli*[J]. *PLoS Biology*, 2013, 11(12): e1001735.
- [106] ZUTZ A, HAMBORG L, PEDERSEN L E, et al. A dual-reporter system for investigating and optimizing protein translation and folding in *E. coli*[J]. *Nature Communications*, 2021, 12: 6093.
- [107] ROMERO-SUAREZ D, WULFF T, RONG Y X, et al. A reporter system for cytosolic protein aggregates in yeast[J]. *ACS Synthetic Biology*, 2021, 10(3): 466-477.
- [108] CORNVIK T, DAHLROTH S L, MAGNUSDOTTIR A, et al. Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*[J]. *Nature Methods*, 2005, 2(7): 507-509.
- [109] ASIAL I, CHENG Y X, ENGMAN H, et al. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell[J]. *Nature Communications*, 2013, 4: 2901.
- [110] RILEY T P, CHOU H T, HU R Z, et al. Enhancing the prefusion conformational stability of SARS-CoV-2 spike protein through structure-guided design[J]. *Frontiers in Immunology*, 2021, 12: 660198.
- [111] SANDERS R W, VESANEN M, SCHUELKE N, et al. Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1[J]. *Journal of Virology*, 2002, 76(17): 8875-8889.
- [112] KIRCHDOERFER R N, WANG N S, PALLESEN J, et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis[J]. *Scientific Reports*, 2018, 8: 15701.
- [113] GRAHAM B S, GILMAN M S A, MCLELLAN J S. Structure-based vaccine antigen design[J]. *Annual Review of Medicine*, 2019, 70: 91-104.
- [114] PALLESEN J, WANG N, CORBETT K S, et al. Immunogenic

- ity and structures of a rationally designed prefusion MERS-CoV spike antigen[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(35): E7348-E7357.
- [115] CORBETT K S, EDWARDS D K, LEIST S R, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness[J]. Nature, 2020, 586(7830): 567-571.
- [116] KHOURY D S, CROMER D, REYNALDI A, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection[J]. Nature Medicine, 2021, 27(7): 1205-1211.
- [117] XIONG X L, QU K, CIAZYNSKA K A, et al. A thermostable, closed SARS-CoV-2 spike protein trimer[J]. Nature Structural & Molecular Biology, 2020, 27(10): 934-941.
- [118] HENDERSON R, EDWARDS R J, MANSOURI K, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation[J]. Nature Structural & Molecular Biology, 2020, 27(10): 925-933.
- [119] QUAN S, BARDWELL J C A. Chaperone discovery[J]. BioEssays, 2012, 34(11): 973-981.
- [120] HE W, ZHANG J Y, SACHSENHAUSER V, et al. Increased surface charge in the protein chaperone Spy enhances its anti-aggregation activity[J]. Journal of Biological Chemistry, 2020, 295(42): 14488-14500.
- [121] QUAN S, WANG L L, PETROTCHENKO E V, et al. Super Spy variants implicate flexibility in chaperone action[J]. eLife, 2014, 3: e01584.
- [122] CORTEZ L, SIM V. The therapeutic potential of chemical chaperones in protein folding diseases[J]. Prion, 2014, 8(2): 197-202.
- [123] HAILU T T, FOIT L, BARDWELL J C A, et al. *In vivo* detection and quantification of chemicals that enhance protein stability[J]. Analytical Biochemistry, 2013, 434(1): 181-186.
- [124] WANG Z, MOULT J. SNPs, protein structure, and disease[J]. Human Mutation, 2001, 17(4): 263-270.
- [125] SAUNDERS J C, YOUNG L M, MAHOOD R A, et al. An *in vivo* platform for identifying inhibitors of protein aggregation[J]. Nature Chemical Biology, 2016, 12(2): 94-101.
- [126] CHEN S, WU J L, LIANG Y, et al. Arsenic trioxide rescues structural p53 mutations through a cryptic allosteric site[J]. Cancer Cell, 2021, 39(2): 225-239.e8.
- [127] QI S M, DONG J Y, XU Z Y, et al. PROTAC: an effective targeted protein degradation strategy for cancer therapy[J]. Frontiers in Pharmacology, 2021, 12: 692574.
- [128] RAMADAS B, KUMAR PAIN P, MANNA D. LYTACs: an emerging tool for the degradation of non-cytosolic proteins[J]. ChemMedChem, 2021, 16(19): 2951-2953.
- [129] LU M C, LIU T, JIAO Q, et al. Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway[J]. European Journal of Medicinal Chemistry, 2018, 146: 251-259.
- [130] HENNING N J, BOIKE L, SPRADLIN J N, et al. Deubiquitinase-targeting chimeras for targeted protein stabilization[J]. Nature Chemical Biology, 2022, 18(4): 412-421.
- [131] TEH S Y, LIN R, HUNG L H, et al. Droplet microfluidics [J]. Lab on a Chip, 2008, 8(2): 198-220.



通讯作者：全舒(1982—)，女，教授，博士生导师。全舒课题组聚焦于蛋白质折叠领域的工具开发与机制解析，发展了一系列蛋白质体内稳定性检测探针，建立了以分子伴侣活力改造为基础的蛋白质折叠调控策略，为基础研究及蛋白质在生物催化、生物医药等领域的应用提供了支撑

E-mail: shuquan@ecust.edu.cn



第一作者：阮青云(1996—)，男，博士研究生。研究方向为蛋白质稳定性检测探针的开发与应用。

E-mail: alessandruan@mail.ecust.edu.cn