



mRNA疫苗理性设计: 从经验到人工智能设计

胡宇轩², 濮澄韬¹, 刘博翔³, 张亮^{1*}

1. 中国科学院杭州医学研究所, 杭州 310018;

2. Harvard Medical School, Boston 02115, USA;

3. National University of Singapore, Singapore 119077, Singapore

* 联系人, E-mail: zhangliang@him.cas.cn

2024-05-05 收稿, 2024-06-23 修回, 2024-08-29 接受, 2024-09-02 网络版发表

国家自然科学基金和比尔及梅琳达·盖茨基金会联合资助项目(7231101367/2023YFVA1002)、新加坡教育部学术研究基金(FY2023, 23-0434-A0001, 22-5800-A0001, MOE-T2EP30123-0015)、精准医学转化研究项目核心资金资助项目(NUHSRO/2020/080/MSC/04/PM)和2024年海南省卫生健康科技创新联合项目(WJSK2024MS230)资助

摘要 mRNA疫苗因其免疫原性强、无基因整合风险和低成本制造等天然优势,已在传染病的预防以及癌症治疗等领域展示出巨大的潜力。目前,针对mRNA疫苗面临的稳定性差和蛋白表达效率不高等挑战,研究人员开展了广泛的mRNA设计和优化工作。传统的mRNA疫苗设计依赖于经验和实验结果的反复迭代,而随着人工智能技术的发展,mRNA疫苗设计正向更加理性和高效的方向转变。本文首先系统介绍了各种疫苗的特性,并概述了多款已上市和临床评估阶段的mRNA疫苗。同时,针对mRNA疫苗的优化问题,根据mRNA的五个不同结构区域,分别详细介绍了经验设计和人工智能设计在mRNA疫苗理性设计中的应用和优缺点。最后,讨论了人工智能模型在mRNA疫苗优化中的局限性和未来发展方向。

关键词 mRNA, 疫苗, 人工智能, 核酸药物设计

疫苗是一类能够激活人体免疫系统,从而预防和治疗特定传染病或恶性疾病的生物制剂,在维护人类健康方面发挥着至关重要的作用^[1-3]。作为预防和控制传染病的重要手段,疫苗种类多样,各有优劣。灭活病毒疫苗是由化学处理、加热或辐射方式灭活的病毒或细菌制成的疫苗^[4]。这类疫苗安全性高且稳定,但引发的免疫反应较弱,通常需要多次接种。减毒活病毒疫苗利用减毒病原体引发人体强烈的免疫反应,免疫效果持久,但对免疫系统较弱的人群存在潜在风险,并且需要严格的储存条件^[5]。亚单位疫苗使用病原体的部分成分作为抗原,具有高安全性和特异性强的优点,但免疫原性较弱,需要加入佐剂并进行多次接种^[6]。重组蛋白疫苗通过基因工程技术生产,安全性高且易于大规模

生产,但生产成本较高且免疫原性较弱^[7]。DNA疫苗通过将编码病原体抗原的DNA序列注入机体细胞,具有稳定性好、生产简便的优势,但其主要问题是基因整合的风险,即外源DNA可能插入宿主基因组,导致基因组不稳定,甚至引发癌症等安全问题^[8]。mRNA疫苗是基于信使RNA(messenger RNA, mRNA)构建的疫苗,由一组携带蛋白质合成编码信息的核苷酸组成。基本原理是通过mRNA在人体细胞内翻译成蛋白质抗原,诱导免疫反应,从而防治特定疾病。

1 mRNA疫苗背景

早在1990年,研究就表明将目标基因的mRNA分子注射到小鼠体内可以检测到蛋白质的产生^[9]。然而,

引用格式: 胡宇轩, 濮澄韬, 刘博翔, 等. mRNA疫苗理性设计: 从经验到人工智能设计. 科学通报, 2024, 69: 4805–4812

Hu Y X, Pu C T, Liu B X, et al. The rational design of mRNA vaccine: From empirical method to artificial intelligence-based design (in Chinese). Chin Sci Bull, 2024, 69: 4805–4812, doi: 10.1360/TB-2024-0478

由于mRNA稳定性不足、非特异性先天免疫反应较强以及体内递送效率低等问题, mRNA疫苗的应用一直受到限制. 随着纳米技术的突破, 脂质纳米颗粒(lipid nanoparticles, LNPs)递送系统可以高效保护mRNA并将其传送到细胞中, 使mRNA疫苗的递送和稳定性得到显著改善, 为其大规模应用奠定了基础^[10]. mRNA疫苗具有免疫原性强, 特异性高, 基因整合风险低等多种天然优势. 截至目前, 已有多款mRNA疫苗被批准上市或处于临床试验阶段. 特别在COVID-19疫情大流行期间, 由Pfizer和Moderna公司分别开发的mRNA疫苗BNT162b2和mRNA-1273均获得美国FDA的紧急授权, 为全球近10亿人提供针对COVID-19的免疫^[11~13]. 同时, 针对流感病毒、狂犬病毒等病毒引起的传染病的mRNA疫苗也在临床研究中^[14]. 例如, Moderna针对季节性流感的mRNA疫苗mRNA-1010已经完成了Ⅲ期临床试验(<https://news.modernatx.com/news/news-details/2023/Moderna-Announces-Interim-Phase-3-Safety-and-Immunogenicity-Results-for-mRNA-1010-a-Seasonal-Influenza-Vaccine-Candidate/default.aspx>), 结果表明其在2种甲流病毒(H3N2和H1N1)上血清转化率更具优势. 此外, 基于mRNA疫苗技术的个性化癌症治疗药物显示出巨大的治疗潜力. 这种个性化癌症疫苗(personalized cancer vaccine, PCV)可以在人体细胞内表达肿瘤抗原, 从而激活特异性抗肿瘤免疫并攻击癌细胞^[15]. 2023年12月15日, Moderna和默沙东公司公布了mRNA癌症疫苗mRNA-4157在157名晚期黑色素瘤患者参与的2b期随机临床试验结果. 个性化mRNA疫苗mRNA-4157通过引导患者体内细胞合成特定肿瘤新抗原, 激发针对这些抗原的免疫反应, 同时使用免疫检查点抑制剂Keytruda阻断PD-1/PD-L1通路, 解除T细胞活性抑制, 增强其对肿瘤细胞的识别和杀伤能力, 最终显著改善黑色素瘤患者的无复发生存期和无远端转移生存期, 并降低复发或死亡风险^[16]. 与此同时, 基于mRNA的蛋白替代疗法在心脏病和一些罕见病中也展示出广泛的应用前景^[17,18]. 例如, 根据Moderna公司近期公布的Ⅰ/Ⅱ期临床研究结果, mRNA-3927作为基于mRNA的蛋白替代疗法, 可以显著降低丙酮酸脱氢酶缺乏症患者的发病风险^[19]. 当然, mRNA疫苗在实际应用中也面临一些挑战, 包括稳定性差、蛋白表达效率不高、免疫系统的非特异性激活等. 针对这些挑战, 研究人员已经开展了许多针对mRNA的设计和 optimization 工作.

2 mRNA疫苗优化

2.1 mRNA疫苗发展里程碑

首先, 核苷酸的化学修饰是优化mRNA疫苗的常用方法之一. 外源的mRNA在进入人体后会被TLRs (Toll-like receptors)识别, 从而引发机体产生固有免疫应答, 进而诱导炎症反应, 破坏外源mRNA并产生药物副作用. 通过使用化学修饰过的核苷酸, 例如假尿苷(Ψ)^[20]和N1-甲基-假尿苷($m^1\Psi$)^[21], 可以显著降低mRNA疫苗的炎症副作用. 2023年诺贝尔生理学或医学奖颁发给了来自匈牙利的生物化学家卡塔琳·卡里科(Katalin Karikó)和来自美国的免疫学家德鲁·韦斯曼(Drew Weissman), 以表彰他们在mRNA疫苗技术和核苷酸修饰方面的开创性贡献. 目前, $m^1\Psi$ 是进入临床使用的治疗性mRNA中最常用的修饰方式. 然而, 最新的研究表明, $m^1\Psi$ 修饰的mRNA在翻译过程中可能发生核糖体移码^[22], 产生非预期的蛋白质, 引发免疫反应, 这一问题需要进一步关注.

2.2 mRNA疫苗理性设计

除了对核苷酸进行化学修饰外, 还可以直接设计和优化mRNA不同结构域的序列. mRNA主要由五个结构域构成(图1), 分别是5'帽子结构、5'端和3'端的非翻译区域(untranslated regions, UTR)、编码区(coding sequence, CDS)和Poly A尾结构. 传统的经验方法通过设计和优化mRNA的不同区域, 可以在一定程度上解决mRNA疫苗的一些固有缺陷. 然而, 这种传统方法依赖于实验结果的反复迭代, 存在研发效率低、实验成本高的问题. 人工智能(artificial intelligence, AI)技术的引入为mRNA设计带来了革命性的变化. 通过大数据分析和机器学习, AI模型能够基于公开或自有的生物实验数据进行学习, 快速高效地优化设计高可成药性mRNA序列. 此外, AI模型还可以根据病人个体基因差异定制个性化mRNA疫苗.

5'帽子结构可以保护mRNA免受核酸外切酶的切割, 调节前体mRNA的剪切, 启动mRNA的翻译并促进mRNA从细胞核输出至细胞质^[23]. 共转录加帽是体外转录mRNA最常用的加帽方法. 目前, CleanCap[®]类似物相比于第一代帽子类似物在加帽效率上有了显著提升^[24]. 此外, 有研究表明, 使用苄基修饰的帽子类似物能够促进mRNA的纯化, 并提高其在体内外的蛋白翻译性能^[25].

5'UTR通常对基因转录的起始和翻译速率有显著

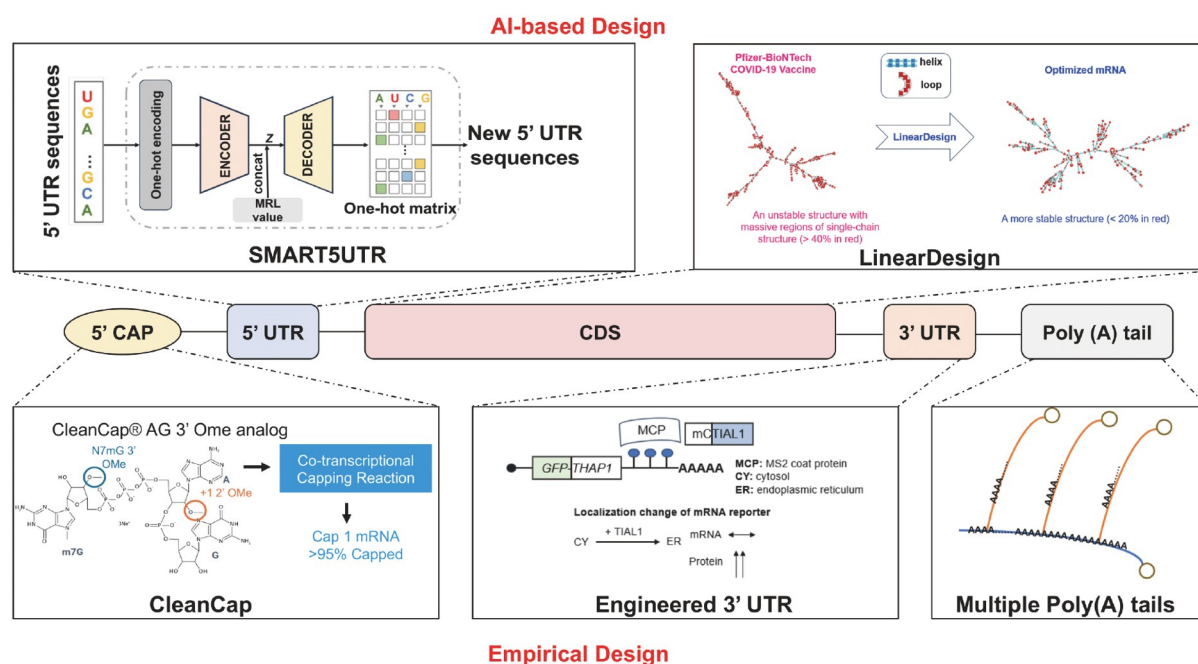


图1 mRNA的5个结构域及代表性优化方法

Figure 1 Five structural domains of mRNA and representative optimization methods

影响。通过调节这些过程, 5'UTR可以最终控制蛋白表达的水平 and mRNA 的稳定性。针对5'UTR的经验设计方法是选取在人体细胞内蛋白表达水平高的基因所对应的5'UTR。例如, COVID-19疫苗BioNTech BNT162b2就是使用含有优化的Kozak序列的人 α -球蛋白RNA的5'UTR^[26]。Kozak序列^[27,28]是位于翻译启动位点附近的特定核苷酸序列, 它对于确保核糖体准确识别并定位到正确的翻译启动密码子至关重要, 进而对蛋白表达水平产生重要影响。目前针对Kozak序列的优化方法主要还是凭借经验而非人工智能方法。同时, Mauger等人^[29]指出, 通过减少在5'前导区(即5'UTR和CDS的前约10个密码子)内的二级结构, 可以提高蛋白质的表达水平。但是, 5'UTR的经验设计方法需要通过不断进行实验尝试来寻找最优的5'UTR序列, 直接导致了研发周期过长且效率较低。AI技术的发展进一步推动了针对5'UTR的设计, 使得研发人员可以精准、快速地获得最优的5'UTR序列。Sample等人^[30]基于28万条大规模平行翻译实验测定数据和卷积神经网络算法开发了预测模型Optimus 5-Prime, 用于预测5'UTR序列对mRNA核糖体负载(ribosome loading)的影响。该模型在测试集上的决定系数(R^2)达到0.93, 显示出良好的预测性能。这一方

法可以进一步拓展到5'帽子结构和3'端UTR对mRNA核糖体负载的影响。同时, Cao等人^[31]利用公开的Ribo-seq和RNA-seq数据基于随机森林算法训练了AI模型用于预测翻译效率和mRNA表达, 随后将该预测模型和遗传算法结合构建了一个预计具有高翻译效率的合成5'UTR库。此外, 考虑到化学修饰会影响5'UTR对翻译的调控, Tang等人^[32]建立的Smart5UTR模型则专门用于m¹ Ψ -mRNA的5'UTR设计。近期, Chu等人^[33]将大语言模型应用于5'UTR区域并将其用于mRNA的功能预测任务中。UTR-LM模型基于Transformer算法框架在多物种来源的天然mRNA序列以及二级结构、最小自由能等多模态数据上进行预训练, 然后经过微调后在核糖体负载、翻译效率和mRNA表达等多个任务上展现出优异的性能。同时, 该研究团队设计了211条自然界中不存在的mRNA 5'UTR区域序列, 并测试了这些序列用于mRNA疫苗的潜力。结果显示, 相较于目前广泛应用的传统mRNA疫苗序列, 该团队设计的全新mRNA疫苗蛋白表达水平最高提高了32.5%。

3'UTR与5'UTR类似, 其主要功能是调控蛋白表达水平和维持mRNA的稳定性^[34]。目前采用血红蛋白亚基 α 和血红蛋白亚基 β 等天然高表达基因对应的3'UTR

区域是设计首选^[28]。Holtkamp等人^[35]发现采用两个连续的 β 珠蛋白3'UTR首尾相连的方式可以增强RNA的稳定性和翻译效率。此外,使用随机序列或天然存在的序列片段进行大规模并行报告分析(massively parallel reporter assays, MPRA)^[36,37],进而筛选出所需的特定3'UTR,也是有效的方法之一。来自纪念斯隆-凯特琳癌症中心的最新研究表明^[38],通过工程化3'UTR可以改变mRNA细胞质内定位,从而显著增加蛋白质的丰度。

CDS区域是编码蛋白质的序列,也是mRNA最关键的组成部分。考虑到遗传密码的简并性,对应同一个蛋白的mRNA编码区可能的序列数量众多^[39],而不同的mRNA编码区序列在稳定性,蛋白翻译效率,蛋白表达水平,免疫原性和生物毒性等可成药性的关键指标方面差异极大,突显了CDS区域设计的重要性。由于CDS区域的设计空间极为庞大,所以单纯基于经验的设计方法难以实现优化。现有的方法基本都是基于计算机经典算法或AI算法。针对CDS的设计一般关注两个方面,即基于密码子的优化和基于mRNA结构稳定性的优化。同义密码子的选择会直接影响蛋白质翻译效率,因此选择“优选密码子”是至关重要的^[40]。常用于衡量密码子最优性质的指标包括:CAI(codon adaptation index)^[41],tAI(tRNA adaptation index)^[42],RSCU(relative synonymous codon usage)^[43]等。目前已经开发出了一些密码子优化(codon optimization)工具,其中OPTIMIZER^[44]和JCat^[45]是被广泛应用的两个工具。它们使用基于生物体的密码子使用偏好(codon usage preference)来优化核酸序列。同时,对于mRNA的结构稳定性优化来说,主要的指标为MFE(minimum free energy)和AUP(average unpaired probability)等。Wayment-Steele等人^[46]开发一种名为RiboTree的算法工具,用于对RNA序列进行优化以获得最小的AUP。CodonBERT^[47]则是来自赛诺菲团队开发的针对CDS序列设计的大语言模型,它使用了一个多头注意力Transformer框架在多样化生物体来源的1000万个mRNA编码序列上进行预训练,可以广泛用于针对mRNA属性预测的下游任务中,包括蛋白质表达和mRNA降解预测。2023年在Nature发表的LinearDesign算法^[39]进一步将密码子和稳定性联合优化,能够在数分钟时间内设计出更稳定、更高效的mRNA编码区序列,展示了计算机算法应用于mRNA设计的巨大潜力。目前,LinearDesign算法已成功应用于多种mRNA疫苗序列的设计和优化。首先,在论文中^[39]报道了两款使用LinearDesign算法优化后的

mRNA疫苗实验结果,即COVID-19 mRNA疫苗和带状疱疹(Varicella-Zoster virus, VZV)mRNA疫苗。结果显示,LinearDesign算法设计的序列在稳定性、蛋白表达水平、抗体水平等各项指标上均相比其他方法设计的基准序列有显著提升,其设计的新冠mRNA疫苗在抗体反应水平(anti-spike IgG antibody titres)上最高提升了128倍。目前基于该算法原理开发的新冠mRNA疫苗,已获得老挝卫生部的紧急使用授权,用于预防COVID-19,适用于18岁及以上人群^[48],该疫苗也于2023年第四季度完成临床Ⅲ期试验^[49]。Moderna最新的研究^[50]则进一步探讨了核糖体负载和mRNA蛋白表达之间的关系,发现最稳定且表达最高的mRNA具有适度的启动/延伸速率和核糖体负载。

位于mRNA末尾的Poly A区域与mRNA的稳定性和翻译效率密切相关^[51],合适的Poly A设计可以提高mRNA的翻译效率和稳定性^[52]。Trepotec等人^[53]的研究表明,将Poly A序列间隔开来可以在不影响体外转录生成mRNA的翻译效率和半衰期的前提下,显著减少质粒DNA扩增时的重组事件并维持尾巴长度。此外,Li等人^[54]证明了在Poly A尾中的胞苷C取代可以进一步延长mRNA的半衰期。近期来自麻省理工学院的Chen等人^[55]通过设计具有多个Poly A分支的mRNA,进一步增强mRNA的稳定性,提升蛋白表达水平。

3 讨论

表1列举了针对mRNA疫苗设计的部分代表性方法。从中可以看出,除了“经验性”设计方法,基于AI算法的mRNA设计模型已被广泛使用。特别是针对5'UTR和CDS区域的设计。当然,AI模型在应用于mRNA技术设计中也面临着一些挑战。首先,AI模型需要大规模且高质量的实验数据来进行有效训练,但目前可用的实验数据数量有限,质量参差不齐且分散。其次,AI模型的可解释性较差,即它们通常被称为“黑箱”模型,这使得解释其决策过程变得困难。尽管如此,AI算法在设计具有优异临床应用潜力的mRNA药物方面展示出了巨大的潜力,特别是在传染病治疗、肿瘤疫苗开发以及蛋白替代疗法方面。随着AI技术的不断发展和生物数据的积累,AI模型的预测能力将不断提升。更可预见的是,例如AlphaFold3^[56]等基于大语言模型的针对包括RNA、蛋白质等生物大分子的基座模型将在药物研发领域发挥出更大的优势,AI模型将能够更全面地分析RNA与DNA、蛋白质等生物分子的相互作用模式,从

表1 mRNA性质预测和序列设计方法总结

Table 1 Summary of methods for mRNA property prediction and sequence design

区域	参考文献	优化策略	方法类型	第一单位
5' Cap	[24]	采用CleanCap®帽类似物	经验设计	TriLink
	[25]	采用苄基修饰的帽类似物	经验设计	华沙大学
	[29]	减少5'UTR的二级结构	经验设计	莫德纳(Moderna)
	[30] (Optimus 5-Prime)	构建5'UTR序列与平均核糖体负载相关的AI预测模型	人工智能	华盛顿大学西雅图分校
5'UTR	[31]	构建5'UTR序列同转录效率、mRNA表达相关的AI预测模型	人工智能	麻省理工学院
	[32] (SMART5UTR)	基于深度生成模型设计优质的m ¹ Ψ-5'UTR	人工智能	四川大学
	[33] (UTR-LM)	基于多物种的5'UTR预训练的语言模型, 可用于预测平均核糖体负载、翻译效率等	人工智能	普林斯顿大学
	[44] (OPTIMIZER)	优化序列密码子	算法设计	西班牙罗维拉-威尔吉利大学
CDS	[45] (Jcat)	优化序列密码子	算法设计	-
	[46] (RiboTree)	最小化mRNA的AUP	人工智能	斯坦福大学
	[39] (LinearDesign)	联合优化序列密码子和稳定性	人工智能	百度美国研究院
	[47] (CodonBert)	基于多物种的编码区序列的语言模型, 可用于预测蛋白质表达、mRNA降解等	人工智能	赛诺菲(Sanofi)
3'UTR	[35]	两次串联添加3'UTRs序列	经验设计	德国美茵茨大学
	[38]	工程化3'UTR序列	经验设计	纪念斯隆·凯特琳癌症中心
polyA尾	[54]	利用胞苷C取代改造Poly A尾	经验设计	香港大学
	[55]	构建多个Poly A尾的mRNA	经验设计	麻省理工学院

而进一步提升mRNA设计的效果。综上所述, 基于AI模型的个性化和精准化mRNA设计可能会在未来推动生

物医学领域实现革命性的发展, 为治疗方案提供更多前所未有的可能性。

参考文献

- Gebre M S, Brito L A, Tostanoski L H, et al. Novel approaches for vaccine development. *Cell*, 2021, 184: 1589–1603
- Orenstein W A, Ahmed R. Simply put: Vaccination saves lives. *Proc Natl Acad Sci USA*, 2017, 114: 4031–4033
- Tan S, Zhao J, Hu X, et al. Preclinical evaluation of RQ3013, a broad-spectrum mRNA vaccine against SARS-CoV-2 variants. *Sci Bull*, 2023, 68: 3192–3206
- Sanders B, Koldijk M, Schuitemaker H. Inactivated viral vaccines. In: Nunnally B K, Turula Vincent E, Sitrin R D, eds. *Vaccine Analysis: Strategies, Principles, and Control*. Berlin: Springer, 2015. 45–80
- Trimpert J, Adler J M, Eschke K, et al. Live attenuated virus vaccine protects against SARS-CoV-2 variants of concern B.1.1.7 (Alpha) and B.1.351 (Beta). *Sci Adv*, 2021, 7: eabk0172
- Heidary M, Kaviar V H, Shirani M, et al. A comprehensive review of the protein subunit vaccines against COVID-19. *Front Microbiol*, 2022, 13: 927306
- Pollet J, Chen W H, Strych U. Recombinant protein vaccines, a proven approach against coronavirus pandemics. *Adv Drug Deliver Rev*, 2021, 170: 71–82
- Pagliari S, Dema B, Sanchez-Martinez A, et al. DNA vaccines: History, molecular mechanisms and future perspectives. *J Mol Biol*, 2023, 435: 168297
- Wolff J A, Malone R W, Williams P, et al. Direct gene transfer into mouse muscle *in vivo*. *Science*, 1990, 247: 1465–1468
- Pardi N, Hogan M J, Porter F W, et al. mRNA vaccines — A new era in vaccinology. *Nat Rev Drug Discov*, 2018, 17: 261–279

- 11 Polack F P, Thomas S J, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*, 2020, 383: 2603–2615
- 12 Baden L R, El Sahly H M, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med*, 2021, 384: 403–416
- 13 Our World in Data. CanSino. <https://ourworldindata.org/>
- 14 Kumar A, Blum J, Thanh Le T, et al. The mRNA vaccine development landscape for infectious diseases. *Nat Rev Drug Discov*, 2022, 21: 333–334
- 15 Blass E, Ott P A. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat Rev Clin Oncol*, 2021, 18: 215–229
- 16 Weber J S, Carlino M S, Khattak A, et al. Individualised neoantigen therapy mRNA-4157 (V940) plus pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): A randomised, phase 2b study. *Lancet*, 2024, 403: 632–644
- 17 Rohner E, Yang R, Foo K S, et al. Unlocking the promise of mRNA therapeutics. *Nat Biotechnol*, 2022, 40: 1586–1600
- 18 Magadam A, Kaur K, Zangi L. mRNA-based protein replacement therapy for the heart. *Mol Ther*, 2019, 27: 785–793
- 19 Koerberl D, Schulze A, Sondheimer N, et al. Interim analyses of a first-in-human phase 1/2 mRNA trial for propionic acidemia. *Nature*, 2024, 628: 872–877
- 20 Karikó K, Buckstein M, Ni H, et al. Suppression of RNA recognition by Toll-like receptors: The Impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, 2005, 23: 165–175
- 21 Andries O, Mc Cafferty S, De Smedt S C, et al. N¹-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *J Control Release*, 2015, 217: 337–344
- 22 Mulrone T E, Pöry T, Yam-Puc J C, et al. N¹-methylpseudouridylation of mRNA causes +1 ribosomal frameshifting. *Nature*, 2024, 625: 189–194
- 23 Ramanathan A, Robb G B, Chan S H. mRNA capping: Biological functions and applications. *Nucleic Acids Res*, 2016, 44: 7511–7526
- 24 Henderson J M, Ujita A, Hill E, et al. Cap 1 messenger RNA synthesis with co-transcriptional CleanCap[®] analog by *in vitro* transcription. *Curr Protocols*, 2021, 1: 39
- 25 Warminski M, Trepkowska E, Smietanski M, et al. Trinucleotide mRNA cap analogue N⁶-benzylated at the site of posttranscriptional m⁶Am mark facilitates mrna purification and confers superior translational properties *in vitro* and *in vivo*. *J Am Chem Soc*, 2024, 146: 8149–8163
- 26 Fang E, Liu X, Li M, et al. Advances in COVID-19 mRNA vaccine development. *Sig Transduct Target Ther*, 2022, 7: 94
- 27 Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 1986, 44: 283–292
- 28 Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*, 1987, 15: 8125–8148
- 29 Mauger D M, Cabral B J, Presnyak V, et al. mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci USA*, 2019, 116: 24075–24083
- 30 Sample P J, Wang B, Reid D W, et al. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol*, 2019, 37: 803–809
- 31 Cao J, Novoa E M, Zhang Z, et al. High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies. *Nat Commun*, 2021, 12: 4138
- 32 Tang X, Huo M, Chen Y, et al. A novel deep generative model for mRNA vaccine development: Designing 5' UTRs with N1-methyl-pseudouridine modification. *Acta Pharm Sin B*, 2024, 14: 1814–1826
- 33 Chu Y, Yu D, Li Y, et al. A 5' UTR language model for decoding untranslated regions of mRNA and function predictions. *Nat Mach Intell*, 2024, 6: 449–460
- 34 Mayr C. Regulation by 3'-untranslated regions. *Annu Rev Genet*, 2017, 51: 171–194
- 35 Holtkamp S, Kreiter S, Selmi A, et al. Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood*, 2006, 108: 4009–4017
- 36 Orlandini von Niessen A G, Poleganov M A, Rechner C, et al. Improving mRNA-based therapeutic gene delivery by expression-augmenting 3' UTRs identified by cellular library screening. *Mol Ther*, 2019, 27: 824–836
- 37 Litterman A J, Kageyama R, Le Tonqueze O, et al. A massively parallel 3' UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. *Genome Res*, 2019, 29: 896–906
- 38 Horste E L, Fansler M M, Cai T, et al. Subcytoplasmic location of translation controls protein output. *Mol Cell*, 2023, 83: 4509–4523.e11
- 39 Zhang H, Zhang L, Lin A, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature*, 2023, 621: 396–403
- 40 Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol*, 2018, 19: 20–30
- 41 Sharp P M, Li W H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 1987, 15: 1281–1295
- 42 Sabi R, Volvovitch Daniel R, Tuller T, et al. stAICalc: tRNA adaptation index calculator based on species-specific weights. *Bioinformatics*, 2017, 33: 589–591
- 43 Sharp P M, Tuohy T M F, Mosurski K R. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic*

- [Acids Res](#), 1986, 14: 5125–5143
- 44 Puigbo P, Guzman E, Romeu A, et al. OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. [Nucleic Acids Res](#), 2007, 35: W126–W131
 - 45 Grote A, Hiller K, Scheer M, et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. [Nucleic Acids Res](#), 2005, 33: W526–W531
 - 46 Wayment-Steele H K, Kim D S, Choe C A, et al. Theoretical basis for stabilizing messenger RNA through secondary structure design. [Nucleic Acids Res](#), 2021, 49: 10604–10617
 - 47 Li S, Moayedpour S, Li R, et al. CodonBERT: Large language models for mRNA design and optimization. [bioRxiv](#) 2023.09.09.556981
 - 48 Dolgin E. ‘Remarkable’ AI tool designs mRNA vaccines that are more potent and stable. [Nature](#), 2023, doi: [10.1038/d41586-023-01487-y](#)
 - 49 Fang Y, Li J X, Duangdany D, et al. Safety, immunogenicity, and efficacy of a modified COVID-19 mRNA vaccine, SW-BIC-213, in healthy people aged 18 years and above: A phase 3 double-blinded, randomized, parallel controlled clinical trial in Lao PDR (Laos). [eClinicalMedicine](#), 2024, 67: 102372
 - 50 Bicknell A A, Reid D W, Licata M C, et al. Attenuating ribosome load improves protein output from mRNA by limiting translation-dependent mRNA decay. [Cell Rep](#), 2024, 43: 114098
 - 51 Gallie D R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. [Genes Dev](#), 1991, 5: 2108–2116
 - 52 Yu S, Kim V N. A tale of non-canonical tails: Gene regulation by post-transcriptional RNA tailing. [Nat Rev Mol Cell Biol](#), 2020, 21: 542–556
 - 53 Trepotec Z, Geiger J, Plank C, et al. Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. [RNA](#), 2019, 25: 507–518
 - 54 Li C Y, Liang Z, Hu Y, et al. Cytidine-containing tails robustly enhance and prolong protein production of synthetic mRNA in cell and *in vivo*. [Mol Ther Nucleic Acids](#), 2022, 30: 300–310
 - 55 Chen H, Liu D, Guo J, et al. Branched chemically modified poly(A) tails enhance the translation capacity of mRNA. [Nat Biotechnol](#), 2024, ,
 - 56 Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. [Nature](#), 2024, 630: 493–500

Summary for “mRNA疫苗理性设计: 从经验到人工智能设计”

The rational design of mRNA vaccine: From empirical method to artificial intelligence-based design

Yuxuan Hu², Chengtao Pu¹, Boxiang Liu³ & Liang Zhang^{1*}

¹ Hangzhou Institute of Medicine, Chinese Academy of Sciences, Hangzhou 310018, China;

² Harvard Medical School, Boston 02115, USA;

³ National University of Singapore, Singapore 119077, Singapore

* Corresponding author, E-mail: zhangliang@him.cas.cn

mRNA vaccines, recognized for their strong immunogenicity, low risk of gene integration, and cost-effective manufacturing, hold significant promise in preventing and treating infectious diseases and cancers. Indeed, mRNA vaccines were among the earliest vaccine platforms developed in response to the COVID-19 pandemic, demonstrating robust immunogenicity and playing a crucial role in protecting countless lives from COVID-19 infection. However, they also face practical challenges such as poor stability and suboptimal protein expression efficiency. To overcome these challenges, extensive research has focused on the design and optimization of mRNA vaccines. Traditionally, this design process has relied on iterative empirical refinements. With advances in artificial intelligence (AI) and bioinformatics, mRNA vaccine design is evolving toward more rational and efficient approaches. Thus, this review systematically introduces the characteristics of various vaccines and outlines the principles of mRNA vaccine design. We then focus on the optimization of mRNA vaccines, examining both empirical and AI-based design methods across five key structural domains of mRNA, including the 5' cap, 5' untranslated region (UTR), 3' UTR, coding sequence region, and poly-A tail. Traditional empirical methods involve designing and optimizing these regions to address some of the inherent deficiencies in mRNA vaccines. However, these methods often require repetitive experimental iterations, resulting in low development efficiency and high costs. Currently, the evolution of vaccines is rapidly being revolutionized using advanced AI-based technologies. AI models can rapidly and efficiently optimize and generate highly druggable mRNA sequences by learning from publicly available or proprietary biological data. By comparing empirical and AI-based design approaches, we highlight the advantages of AI in mRNA vaccine design while also discussing its limitations and future potential. Certainly, AI models also face certain challenges when applied to mRNA design. Firstly, AI models require large-scale and high-quality experimental data for training, but currently, the available experimental data is limited in quantity, varies in quality, and is also scattered. Secondly, the interpretability of AI models is relatively poor, as they are often referred to as “black box” models, making it difficult to explain the decision-making processes of AI models. With the development of AI technology and the accumulation of biological data, the predictive capabilities of AI models will keep improving. It is even more foreseeable that base models based on large language models, such as AlphaFold3, targeting biomolecules like RNA and proteins, will play a significant role in drug development. AI models will be able to more comprehensively analyze the interaction patterns between RNA and other biomolecules like DNA and proteins, thereby further enhancing the effectiveness of mRNA design. In conclusion, we argue that personalized and precise mRNA design driven by AI could revolutionize the biomedical field, offering unprecedented therapeutic possibilities for patients, enhancing the vaccine development process, and providing new strategies to address future challenges.

mRNA, vaccine, artificial intelligence, nucleic acid drug design

doi: [10.1360/TB-2024-0478](https://doi.org/10.1360/TB-2024-0478)