

· 综述 ·

## mRNA 疫苗翻译效率的研究进展

刘涛, 王升启, 李伍举\*

军事医学研究院生物信息中心, 北京 100850

刘涛, 王升启, 李伍举. mRNA 疫苗翻译效率的研究进展[J]. 生物工程学报, 2024, 40(11): 3930-3950.

LIU Tao, WANG Shengqi, LI Wujun. Research progress in the translation efficiency of mRNA vaccines[J]. Chinese Journal of Biotechnology, 2024, 40(11): 3930-3950.

**摘要:** 面向快速应对新发传染病的需求, 与传统疫苗相比, mRNA 疫苗技术在设计、生产和应用等方面具有优势, 在新型冠状病毒疫情期间率先获得世界卫生组织的紧急使用推荐。mRNA 疫苗设计的关键之一是确保编码蛋白在受种者体内稳定、充分表达。本文综述了近年来在提高 mRNA 疫苗翻译效率方面的相关研究进展和存在的问题, 以期为相关研究提供帮助。

**关键词:** mRNA; 疫苗; 翻译; 表达

## Research progress in the translation efficiency of mRNA vaccines

LIU Tao, WANG Shengqi, LI Wujun\*

Bioinformatics Center of Academy of Military Medical Sciences, Beijing 100850, China

**Abstract:** Compared with conventional vaccines, mRNA vaccines have considerable advantages in design, production, and application, especially in dealing with emerging infectious diseases. Particularly, mRNA vaccines were the first to be recommended by the World Health Organization for emergency use during the COVID-19 pandemic. A key to the design of mRNA vaccines is to ensure the stable and sufficient expression of the encoded protein in the recipient. In recent years, advances have been attained in the experimental and computational research in this area. This review focused on the progress and problems in improving the translation efficiency of mRNA vaccines in recent years, aiming to promote related research.

**Keywords:** mRNA; vaccine; translation; expression

资助项目: 国家自然科学基金(31471244); 国家自然科学基金重大研究计划(91540202)

This work was supported by the National Natural Science Foundation of China (31471244) and the Major Research Plan of National Natural Science Foundation of China (91540202).

\*Corresponding author. E-mail: liwj@nic.bmi.ac.cn

Received: 2024-03-06; Accepted: 2024-06-11; Published online: 2024-06-12

近年来快速发展的信使 RNA (messenger RNA, mRNA)疫苗免疫技术,可将体外转录(*in vitro*-transcription, IVT)的 mRNA 包装并递送进入细胞,这些 mRNA 在细胞内翻译表达成抗原,从而刺激机体产生免疫反应。该技术可以实现更加高效的疫苗研发,成为快速应对新发传染病的新型有效方法之一<sup>[1-2]</sup>。例如,2019 冠状病毒病(coronavirus disease 2019, COVID-19)的致病病毒是严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2),其基因组序列于 2020 年 1 月公布,而据此研发的 mRNA 疫苗在 2020 年 3 月便已率先进入人体临床试验阶段<sup>[3]</sup>,于 2020 年 12 月再次率先获得世界卫生组织(World Health Organization, WHO)的紧急使用推荐<sup>[4]</sup>,响应速度前所未有。根据 WHO 在 2023 年 3 月 30 日公布的统计数据,在处于临床开发阶段的 SARS-CoV-2 疫苗中, mRNA 疫苗占比 24%,仅次于占比 32%的蛋白质亚单位疫苗<sup>[5]</sup>。

针对 SARS-CoV-2 的 mRNA 疫苗的快速研发使用,最核心和关键的因素就是数十年的相关研究积累,例如抗原的快速设计得益于先前对中东呼吸综合征冠状病毒(Middle East respiratory syndrome coronavirus, MERS-CoV)和严重急性呼吸综合征冠状病毒(severe acute respiratory syndrome coronavirus, SARS-CoV)的研究<sup>[6]</sup>,疫苗的快速产品化得益于脂质纳米颗粒封装递送 mRNA 方面的研究<sup>[7]</sup>。当安全有效

的 mRNA 封装递送技术实现通用化和工业化后, mRNA 疫苗研发的瓶颈转变为抗原的快速设计。其中,首先是如何选择最有效的抗原,然后是如何设计编码抗原的 mRNA 序列,使其满足序列稳定和高效翻译表达的要求。抗原的选择受制于新发病毒及其致病机制,只能借助先前对原型病毒的研究积累,或者依赖最新的分子病理研究结果。而在 mRNA 序列设计方面,可以根据以往对 mRNA 稳定性和翻译表达机制的研究,通过总结和针对性深入研究而得到优化设计原则,以实现计算机辅助的快速设计。

典型的真核生物 mRNA 可划分为 5 个部分(图 1),从 5'到 3'依次是:5'帽(5' cap)、5'非翻译区(untranslated region, UTR)、编码蛋白质的开放阅读框(open reading frame, ORF)、3' UTR 和多聚腺苷酸[polyadenylated, poly(A)]尾。研究表明这些核酸序列,或者说它们所形成的结构,对 mRNA 的稳定性和翻译表达具有重要调节意义<sup>[8]</sup>。因此, mRNA 疫苗中指导抗原合成的 mRNA 分子也是按照这一组成模式进行设计的。此外,在设计上还要考虑对用于体外转录合成的 RNA 聚合酶的适应性。由于噬菌体 T7 RNA 聚合酶(T7 RNA polymerase, T7 RNAP)是一种易于获取且具有高度转录活性的单肽 RNA 聚合酶,对 T7 启动子具有严格选择性,只要启动子区具有双链结构,通过 1-3 个核苷酸连接所需合成序列的模板即可,不需要其他蛋白质因子,因此体外合成 mRNA 通常使用 T7 RNAP<sup>[9-10]</sup>。

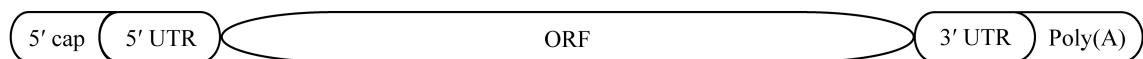


图 1 典型真核 mRNA 结构示意图 可划分为 5'帽、5'非翻译区、编码蛋白质的开放阅读框、3' UTR 和多聚腺苷酸尾

Figure 1 Schematic diagram of typical eukaryotic mRNA structure. It can be divided into five parts: 5' cap, 5' untranslated region (5' UTR), open reading frame (ORF), 3' UTR and polyadenylated (poly(A)) tail.

对应 mRNA 结构中的 5 个部分, 本文对优化设计疫苗 mRNA 序列以实现稳定高效翻译表达方面的研究进展进行综述, 以供相关研究者参考。鉴于蛋白替代疗法也需要 mRNA 在靶细胞组织中稳定高效表达, 本综述也可供该领域相关研究者参考。

## 1 5' cap

通常, 5'帽结构是指 5'末端 N<sup>7</sup>-甲基鸟苷(N<sup>7</sup>-methylguanosine, m<sup>7</sup>G)以 5'-5'三磷酸桥连接到 mRNA 其余部分的一个结构<sup>[11-14]</sup>。该结构普遍存在于真核生物内源性 mRNA 中, 是自体 RNA 区别于外源 RNA 的标志之一, 以避免引起意外的 I 型干扰素反应<sup>[15-17]</sup>; 另外也是 mRNA 稳健翻译所需, 保护 mRNA 免受 5'外切核酸酶的降解<sup>[18]</sup>, 与真核翻译起始因子 4E (eukaryotic translation initiation factor 4E, EIF4E)结合以便于翻译进程的有效启动<sup>[8]</sup>。

由于二核苷聚磷酸(dinucleoside polyphosphates, Np<sub>n</sub>N)是 T7 RNAP 的优良底物, 可在 IVT 体系中用作起始核苷酸<sup>[19]</sup>, 若在 IVT 反应体系中添加合成的 5'帽模拟物, 如 m<sup>7</sup>G(5')p<sub>3</sub>(5')G, 则可以获得 5'末端带有帽模拟物的 mRNA。这种伴随转录添加 5'帽结构的方式简单、高效、经济<sup>[20-21]</sup>, 常用于体外合成含 5'帽结构的 mRNA。

这一加帽方式存在的一个主要问题是, 帽模拟物和三磷酸鸟苷(guanosine triphosphate, GTP)会相互竞争, 导致一些 IVT mRNA 不能被成功加帽而无法获得翻译活性。在实际应用中, 需优化帽模拟物和 GTP 的比例, 以取得 mRNA 产量和加帽效率的平衡<sup>[22]</sup>。

另一个主要问题是, 帽模拟物中 m<sup>7</sup>G 的 3'-OH 也会对转录的第 1 个核苷三磷酸的亲电子 α-磷酸进行亲核攻击, 从而使加到 mRNA 5'端的帽模拟物有 1/3-1/2 是反向的<sup>[23]</sup>, 这些反向连

接的不利于被翻译机器识别并启动翻译<sup>[22]</sup>。2001 年 Stepinski 等<sup>[22]</sup>和 2002 年 Peng 等<sup>[24]</sup>将 3'-OH 修饰为 3'-OCH<sub>3</sub>, Stepinski 等还将 3'-OH 修饰为 3'-H, 均可抑制帽模拟物的反向连接, 也不会削弱翻译机器对它们的识别和翻译的启动, 从而得到了抗反向的帽模拟物(anti-reverse cap analogues, ARCA)。2003 年, Jemielity 等<sup>[25]</sup>发现, 仅对 2'-OH 进行类似的修饰也可达到同样的目的。2004 年 Grudzien 等<sup>[26]</sup>还发现, 采用 m<sup>7</sup>Gp<sub>3</sub>m<sup>7</sup>G 这种对称性修饰的帽模拟物即可避免反向连接的问题。

5'帽结构提供了与 EIF4E 相结合的结构基础, 结合力主要来自后者与 m<sup>7</sup>G 间的疏水性阳离子-π 相互作用和与负电性三磷酸桥间的静电相互作用<sup>[27]</sup>, 翻译的启动需要两者的结合, 而翻译的继续又需要两者的解离, 因此只有适度的结合力才能达成高效的翻译效率。另外, 5'帽结构还是 mRNA 脱帽全酶 DCP1 (decapping mRNA 1)-DCP2 (decapping mRNA 2)的靶向结合对象, DCP2 可靶向切割三磷酸桥 α-和 β-磷酸基团之间的连接, 造成 mRNA 脱帽并进入后续衰变过程<sup>[28]</sup>。因此, 通常需要对 5'帽结构进行进一步的设计, 尤其是 m<sup>7</sup>G 和磷酸桥, 以优化其对 EIF4E 的亲合力并增强脱帽酶的抗性。

2003 年, Jemielity 等<sup>[25]</sup>报道了 2'或 3'抗反向修饰以及不同长度磷酸桥在兔网织红细胞裂解物(rabbit reticulocyte lysate, RRL)中对翻译效率的影响, 发现氧甲基化修饰时, 翻译效率比脱氧修饰时要高; 磷酸桥长度为 4 时, 翻译效率比三磷酸桥时高, 而超过 4 则不再升高; 氧甲基化修饰并叠加四磷酸桥时, 虽然加帽效率不高, 但翻译效率最高, 可达 m<sup>7</sup>Gp<sub>3</sub>G 的 2.4-2.5 倍左右。2004 年, Grudzien 等<sup>[26]</sup>在 RRL 中发现, 对称性修饰可以获得比抗反向修饰更高的翻译效率, 其中 m<sup>7</sup>Gp<sub>4</sub>m<sup>7</sup>G 的最高, 可达 m<sup>7</sup>Gp<sub>3</sub>G

的 3.1 倍左右;  $m^7Gp_3G$  的  $N^7$ -甲基替换为  $N^7$ -苄基( $N^7$ -benzyl,  $b^7$ )即可提高 mRNA 的体外翻译活性<sup>[29]</sup>, Grudzien 等<sup>[26]</sup>还对  $b^7m^{3'-O}Gp_4G$  进行了研究,发现在 RRL 中其对应的翻译效率高于抗反向修饰,甚至略高于  $m^7Gp_3m^7G$ ,可达  $m^7Gp_3G$  的 2.9 倍左右。上述结果总结于表 1。但值得注意的是,这些结果都是在 RRL 中获得的,体内情况未必完全如此。如 Kuhn 等<sup>[30]</sup>在 2010 年的研究发现,在人树突状细胞(human dendritic cells, hDCs)内, $m^7Gp_4m^7G$  相应的翻译效率、稳定性以及蛋白表达量均远低于  $m^7Gp_3G$ 。

在提高对脱帽酶的抗性方面,研究人员主要对磷酸桥进行了 2 种类型的修饰研究,即对桥接氧或非桥接氧进行取代修饰。用于取代桥接氧的有亚甲基<sup>[31-33]</sup>、亚氨基<sup>[33-34]</sup>和二卤代亚甲基<sup>[35]</sup>。用于取代非桥接氧的有硫<sup>[30,36-39]</sup>、硒<sup>[34,40]</sup>和硼<sup>[34,37]</sup>。研究发现,一些修饰可以提升 mRNA 与 EIF4E 的结合能力、对抗 DCP2 水解脱帽的能力、稳定性以及翻译效率,但由于这些研究在所使用的 mRNA 序列、翻译体系以及检测评价指标方面并不一致,因此,对这方面修饰的原则和方向尚未达成一种规律性的认识;但能够明确的是,相同帽模拟物在不同细胞中所能

发挥的作用并不完全相同,会受所在细胞内环境的影响<sup>[30]</sup>。此外,非桥接修饰会产生非对映异构体,因此需要分离纯化,方便性方面不如桥接修饰。

还有一种加帽方式是在转录后通过加帽酶添加 5'帽结构<sup>[41]</sup>,通常采用痘苗加帽酶(vaccinia capping enzyme, VCE)。通过这种二次反应的方式也可以引入含有不同修饰的 5'帽结构,有的修饰也可以提高翻译效率<sup>[42]</sup>,但这种方法无法对磷酸桥进行优化修饰,而且当大部分 5'碱基形成稳定发夹结构时会导致无法加帽,此外,这种方法虽然可以获得 100%的加帽率,但应用于规模化生产的成本较高。Fuchs 等<sup>[41]</sup>为痘苗加帽酶复合物设计了一种优化的表达系统,可直接纯化得到大量的高活性酶,实现了在很短的时间内获得 mg 量级的加帽 RNA。Corbett 等<sup>[6]</sup>、Zhang 等<sup>[43]</sup>和 Ye 等<sup>[44]</sup>在研发针对 COVID-19 的 mRNA 疫苗时都采用了这种加帽方式。

从 5'端开始的第 1 个甚至第 2 个核苷酸在核糖的 2'羟基上被甲基化,分别称为帽 1 (cap 1)和帽 2 (cap 2)<sup>[11-14]</sup>。 $m^7G$  和 cap 1 形成于细胞核中<sup>[45]</sup>,可将自体 RNA 与外源 RNA 区分开来,避免引起意外的 I 型干扰素反应<sup>[15-17]</sup>。cap 2 则

表 1 几种帽模拟物的性能比较

Table 1 Performance of several cap simulants

Cap simulants	Capping rate (%)	Direction accuracy (%)	Capping efficiency (%)	Relative translation efficiency dependent on cap structure in RRL
$m^7Gp_3G$ <sup>[27]</sup>	69±6	58±4	40±6	1.00
$m^72'dGp_3G$ <sup>[26]</sup>	83	100	83	1.17±0.13
$m_2^{7,2'-O}Gp_3G$ <sup>[26]</sup>	62	100	62	2.10±0.15
$m_2^{7,3'-O}Gp_3G$ <sup>[26]</sup>	72	100	72	1.88±0.40
$m^7Gp_3m^7G$ <sup>[27]</sup>	73±2	100	73±2	2.66±0.64
$m^72'dGp_4G$ <sup>[26]</sup>	62	100	62	1.83±0.18
$m_2^{7,2'-O}Gp_4G$ <sup>[26]</sup>	56	100	56	2.56±0.18
$m_2^{7,3'-O}Gp_4G$ <sup>[26]</sup>	57	100	57	2.42±0.22
$m^7Gp_4m^7G$ <sup>[27]</sup>	74±3	100	74±3	3.14±0.24
$b^7m^{3'-O}Gp_4G$ <sup>[27]</sup>	64±4	100	64±4	2.87±0.38

可形成于细胞质中<sup>[45]</sup>。如果是转录后添加 5'帽结构,则需要再进行一步 2'-O-甲基转移酶催化反应,以进行 cap 1 修饰<sup>[46]</sup>。为了在转录时便能够得到 cap 1 结构, TriLink BioTechnologies 开发了系列的 CleanCap 帽类似物,在其中引入了 cap 1 修饰,其加帽率可高达 90%以上<sup>[47-48]</sup>。BioNTech 在开发针对 COVID-19 的 mRNA 疫苗 BNT162b1 和 BNT162b2 时,就采用了 CleanCap 中对应蛋白表达水平较高的一种帽类似物  $m_2^{7,3'-O}Gp_3m^{2'-O}ApG$ <sup>[49-50]</sup>。

综上, 5' cap 影响到 mRNA 稳定性及翻译效率的因素涉及 m7G 修饰、磷酸桥的长度以及修饰、第 1 个核苷酸及其甲基化修饰。研究人员对于这些因素做了许多尝试和比较研究,但通常未考虑它们与 mRNA 其他部分之间可能的交互作用,也未充分考虑这些因素之间可能的交互作用,而且其结果还会受到多种可变因素的综合影响,如所采用的翻译环境、检测方法的不同,不利于结果之间的系统比较,直接阻碍了相关的定量评价或预测模型的开发。此外,这些修饰因素是否可行还受到所选择加帽方式的限制。

## 2 5' UTR

5' UTR 是位于编码区上游的一段不被翻译的序列,充当核糖体启动翻译的结合位点。43S 核糖体加载到真核 mRNA 上后会按照一定的扫描机制沿着 5' UTR 迁移扫描,并利用与 Met-tRNA 反义密码子的互补性连续检查起始密码子 AUG。5' UTR 主要影响翻译的启动过程,通常关注的 5' UTR 特征包括二级结构、序列元件和 5' UTR 长度<sup>[51]</sup>。哺乳动物 5' UTR 的平均长度为 100–200 nt,一般认为,至少需要 20 nt 才能保证核糖体对起始密码子的有效识别。5' UTR 区域往往富含 GC 序列,这会促

进二级结构的形成,从而影响到翻译起始的效率<sup>[52]</sup>。例如,其中的 G 四联体和发夹结构会起到类似路障的作用,阻碍核糖体在 5' UTR 的装载和迁移扫描<sup>[53]</sup>。然而,有些茎环结构通过结合真核翻译起始因子 3 (eukaryotic initiation factor, eIF3)及相关核糖体又可以发挥翻译增强子或激活子的作用<sup>[54]</sup>。此外,高达 10%的人 5' UTR 含有内部的核糖体进入位点(internal ribosome entry site, IRES)<sup>[55]</sup>,可以在不识别 5' cap 的情况下,直接将 40S 核糖体募集装载到 5' UTR,以保障在依赖 5' cap 的翻译受到抑制的时候还能进行翻译<sup>[56]</sup>。大部分含有 IRES 的 mRNA 同时具备 5' cap 结构,可通过 5' cap 机制进行翻译。

大多数真核生物 5' UTR 区域包含 Kozak 共有序列 RNNAUGG,其中,启动翻译的起始密码子 AUG 周围的核苷酸高度保守<sup>[57]</sup>。一项高通量测序研究表明-3 和+4 位置对于确定起始密码子效率至关重要,而-2、-4 和+5 位置起到辅助作用<sup>[58]</sup>。值得注意的是,越来越多的证据表明翻译起始可以发生在与 AUG 不同的密码子处。在非 AUG 起始密码子中,CUG、GUG、UUG 和 ACG 是最常用的。

为了实现合成 mRNA 编码蛋白质的最大化表达,研究人员对各种 5' UTR 进行了比较,其中  $\beta$  珠蛋白的是较为常用的基准比较对象。基于这些比较研究,得到了一些基本的定性认识。例如,为了最大限度地减少扫描过程,一般建议使用较短的 5' UTR,但当然不能短于 20 nt。应避免含有上游起始密码子尤其是 AUG,以消除上游 ORF 的存在。对于起始密码子,应设计具有 Kozak 共有序列上下文的强起始密码子,以防止出现扫描错过的现象。还应避免出现高度稳定的二级结构,尤其是在 5'端附近,这可能会破坏核糖体的加载、扫描和对起始密码子的选择。

即使如此,研究人员对 5' UTR 如何大量参与 mRNA 的翻译和降解仍不完全清楚<sup>[59-60]</sup>。而且,不同细胞类型或组织中的翻译调节因子的表达会有差异,因此还可能需要根据具体的靶向表达环境对 5' UTR 进行适应性调整。更复杂的是,各种序列依赖性调节机制之间存在复杂的相互作用,正是这种复杂的相互作用最终控制着蛋白质的表达。因此需要建立能够综合考虑这些因素的定量模型,以根据 mRNA 序列预测蛋白质表达水平,优化 mRNA 设计。

在定量研究 5' UTR 中诸因素的影响并建立预测模型方面,Castillo-Hair 等<sup>[61]</sup>认为内源转录数据可能并不是用于训练的最佳数据选择。首先,内源转录本包含高度可变的 UTR 和 CDS 序列,因此很难可靠地分辨 mRNA 的特定部分(例如 5' UTR)如何影响翻译。其次,内源数据集的大小从根本上受到人类转录组大小的限制,因此得到的预测模型也会具有相应的局限。最后,具有有害影响的序列在内源数据中代表性不足,可能构成模型的相应盲区。因此,研究人员以多核糖体分析得到的 mRNA 平均核糖体负荷(mean ribosome load, MRL)作为翻译效率的定量衡量标准,通过大规模并行报告基因检测(massively parallel reporter assay, MPRA)这种无偏好实验分析测定大量随机 5' UTR 序列的 MRL,一方面从中获取 5' UTR 序列与翻译表达水平之间关系的生物学见解,以指导序列的从头设计;另一方面,用深度学习技术综合其中复杂的调节因素,开发根据序列预测蛋白质表达的定量模型,以评估设计的 5' UTR 序列<sup>[62]</sup>,当然同时也有助于有效地发现特别的相关序列模式<sup>[62-63]</sup>。结果发现,当主要起始密码子的上游具有与其三联体阅读框架不一致的框外 AUG 时,平均而言,相应 mRNA 的 MRL 值较低。如果是非规范起始密码子(CUG、GUG),也存

在这种抑制效应,但程度明显较小。这些上游起始密码子周围的环境会影响其的抑制效应。对框外 AUG,当被 Kozak 共有序列,即-3 位嘌呤和+4 位鸟嘌呤包围时,更具抑制性;而对框外 CUG 和 GUG,则仅在被 Kozak 共有序列包围时才会产生在统计学上显著的抑制增强效果。与此相对应的是,上游 ORF 具有相似的抑制效应。这些效应与其对核糖体的竞争占用是相一致的。此外,平均而言,5' UTR 预测自由能较低的 mRNA 具有较低的 MRL,这与稳定的二级结构会干扰核糖体扫描的认识也是一致的。

基于这些数据的训练,研究人员开发了一种名为 Optimus 5-Prime 的卷积神经网络(convolutional neural network, CNN)模型<sup>[62]</sup>,预测 5' UTR 序列的核糖体负载,该模型包含两个带有识别短基序过滤器的卷积层,以及一个最终计算 MRL 预测值的全连接层,预测与实测的决定系数达到 0.93;对 2.5 万条人转录本起始密码子上游 50 nt 的 UTR 序列,预测与实测的决定系数达到 0.82,说明该模型确实能学习到 5' UTR 中一定量的与翻译调控相关的生物学信号。更换 CDS 或对 CDS 进行假鸟苷修饰后,模型分别可以解释 72%–78%、68%–76%的实测 MRL 变化,说明 5' UTR 对翻译效率的作用具有一定程度的相对独立性;在 6 种不同细胞系中,模型可以解释 73%–85%的实测 MRL 变化<sup>[62]</sup>。

由于该模型学到的是位置特异性权重,仅能对 50 nt 的 5' UTR 进行预测,因此,研究人员又建立了新的长度为 25–100 nt 的 5' UTR 库,将模型的预测范围提高到 100 nt,在未参与训练的随机序列和人 5' UTR 序列测试数据集上,其 MRL 预测与实测值的决定系数分别达到 0.84 和 0.75<sup>[62]</sup>。进一步地,研究人员将搜索启发式的遗传算法与该模型相结合进行进化设计,以获取具有目标翻译效率的 5' UTR 序列,

简而言之,就是从一组随机的 50 nt 长的 5' UTR 开始,每次迭代都包含计算机模拟的随机突变和交叉,然后使用 Optimus 5-Prime 进行评分以选择下一轮的最佳序列,发现仅在目标 MRL 不超过 8 时,目标与实测值才会非常一致<sup>[62]</sup>。当目标 MRL 超过 8 时,所设计的序列中出现了训练库中不存在的长聚尿嘧啶序列,说明遗传算法很可能会利用 Optimus 5-Prime 中的盲点来进一步提高 MRL<sup>[62]</sup>。而且,基于搜索启发式的设计还可能具有缓慢且低效的问题,例如,突变和交叉操作一次改变几个核苷酸,并且不能保证每一步都能提高性能。

更有效的方法是通过梯度下降实现激活最大化,其中,性能指标随模型输入变化的梯度用于迭代地优化候选序列<sup>[64-65]</sup>,这样进程总是朝着提高性能的方向发展,从而减少了所需模型评估的次数,提高了设计效率。但激活最大化方法仍有一定局限性。首先,对于每个新生成的序列,需要从头开始运行该算法。此外,优化可能会陷入局部最小值或收敛到序列空间中远离训练数据集的区域,在该区域中模型并不准确。与此相关的一个问题是缺乏一种明确的机制用于约束新生成的序列与旧的序列不同,从而限制了可用于实验测试的序列的多样性,并降低了找到高性能序列的可能性<sup>[61]</sup>。

另外一类不同的设计方法基于深度生成模型,即经过训练的神经网络,用于学习训练数据集的分布,以生成具有相似属性的全新序列<sup>[64,66-67]</sup>。其相对于梯度方法的一个主要优点是速度:训练后,生成新的序列只需要对生成模型进行一次评估,无需任何迭代;然而,这些方法的基本版本并未优化序列性能或明确最大化多样性。

最近开发的深度探索网络(deep exploration networks, DEN)是一种激活最大化深度生成模

型,通过融合 2 种方法以应对这些限制<sup>[68]</sup>;通过最小化成本函数的梯度下降进行训练;成本函数由 2 项组成,一个与由独立的预训练预测模型(例如 Optimus 5-Prime)给出的生成序列的性能相关,另一个根据两个生成序列之间的相似性计算得出。在序列空间的探索上,一方面通过根据任意两个生成序列的相似性进行罚分,明确地最大化了序列多样性;另一方面,通过采用变分自动编码器(variational autoencoder, VAE)<sup>[66]</sup>来维持生成序列与训练数据集的似然比,以避免漂移到预测模型的低置信区域;这样,通过同时最大化性能和最小化相似性, DEN 就可学会生成既具有高性能又具有高度多样化的序列。

以上方法都是基于对定长序列进行预测的模型,应用灵活性不强。为了建立一个可以对任意长度 5' UTR 进行预测的模型,考虑到上游 ORF 所在阅读框架对翻译的影响, Karollus 等<sup>[55]</sup>引入了一种逐个框架池化的神经网络操作,既不影响捕获和量化序列中发现的调节基序,又避免了引入位置特异性权重;与 Optimus 5-Prime 相比较,从同样的 50 nt 训练数据得到的 FramePool 模型不仅保持了对 50 nt 随机序列的预测性能,还大大提高了对 25-100 nt 变长序列的预测性能,如对随机序列的预测与实测的相关系数由 0.743 提升到 0.901,对人 5' UTR 序列翻译效率的预测与实测的相关系数则由 0.700 提升到 0.871;该模型在各种与翻译相关的全基因组数据集上也显示出更好的泛化性能,尽管相关性低至 0.11-0.25,但均为正相关且具有统计显著性,说明通过这种方法,模型还是捕获到了一定的与翻译调控相关的生物学信号。

然而,这种模型架构具有一定的局限性。首先,该架构依赖于卷积和池化,因此无法捕获比感受野长的调节基序,从而会错过相应的调节特征。其次,无法确定所检测到的特征在

序列中出现的顺序,因此无法完全复制泄漏扫描的机制。研究人员试图用循环神经网络如长短期记忆网络来克服这些局限,但没有成功。此外,基于 MPRA 数据训练模型,由于可能序列的空间是随机探测的,若想触及较长基序,所需样本量将大到不可行的地步。而且,具有更细微效果的基序可能会被具有更大效果的基序掩盖,从而导致捕获到生物体中很少出现的基序,而未捕获到通过进化选择的微调复杂基序。例如,高达 10%的人 5' UTR 含有 IRES<sup>[69]</sup>,但该基序并未出现在该模型中<sup>[55]</sup>。

综上所述,单就 5' UTR 而言,如何将其他相关研究得到的先验的相关生物知识,如 IRES 机制、泄漏扫描机制、UTR 长度、与 mRNA 稳定和衰变相关的因素以及细胞或组织特异性表达等与神经网络模型整合起来,以及如何将内源数据与 MPRA 随机数据合理整合起来,以训练构建新的更为通用的模型,成为下一步所要解决的问题。FramePool 模型的构建说明,尽管神经网络常被视为黑盒模型,但通过修改其架构,是可以将先验的生物知识编码到神经网络中的。然而,将特定的生物学模型编码到神经网络中,也会降低其了解偏离该模型的机制的能力。因此,如何进行无偏差的整合也是一个需要解决的问题。

在实际应用中,面对快速应对大流行的压力,BioNTech 开发的 BNT162b2 疫苗直接采用了高表达的人  $\alpha$  珠蛋白基因的 5' UTR 序列,仅对其 Kozak 共有序列做了细微调整,即将 ACCAUG 调整为 GCCACCAUG<sup>[70]</sup>。而 Moderna 开发的 mRNA-1273 疫苗则选用他们已申请的专利序列 V1-UTR 作为 5' UTR 序列。该序列由两部分组成,第一部分是该公司一项专利中一号序列,第二部分是富含 GC 的 CCCCGGCGCC,随后就是 Kozak 共有序列 ACCAUG<sup>[71]</sup>。

### 3 ORF

ORF 是蛋白质的编码部分,也是 mRNA 的重要组成部分。对 ORF 的设计主要包括密码子优化和功能肽编码序列的引入<sup>[72]</sup>,而这同时会影响到 mRNA 的二级结构以及 G+C 含量等,对 mRNA 的稳定性和翻译产生影响<sup>[73-74]</sup>。高表达的 mRNA 常含有更多的常见密码子<sup>[75]</sup>,从而可以使用更常见的 tRNA,减少稀有 tRNA 的使用,避免由稀有密码子引发的翻译停止、阅读框错位或 mRNA 降解<sup>[74,76]</sup>,最终提高蛋白质的产量以及翻译的保真度<sup>[77-78]</sup>。

现有密码子优化工具虽多为原核或真菌表达系统设计,也可用于人源表达序列的优化。早期的优化方法如软件 Codon optimizer<sup>[79]</sup>和网络服务器 JCat<sup>[80]</sup>,均是将待优化序列中的密码子转换为表达宿主中使用频率最高的同义密码子。而高表达 mRNA 所使用的密码子并不完全是最常见密码子,其中同义密码子的存在有其特别的意义,例如,通过二级结构影响 mRNA 的稳定性<sup>[81-82]</sup>,或者通过适时适地降低翻译速度,以确保蛋白质正确折叠,从而确保蛋白质具有正确功能,或者通过 G+C 含量影响 mRNA 抵制核糖核酸内切酶降解的能力<sup>[83]</sup>等。因此,研究人员提出了具有一定灵活性的优化策略。例如,网络服务器 UpGene<sup>[84]</sup>根据表达宿主的同义密码子概率分布,以蒙特卡罗模拟方法对序列进行优化,使其同义密码子概率分布与表达宿主相一致。网络服务器 Gene Design<sup>[85]</sup>可将蛋白质序列回译为最常见密码子序列,而对核酸序列则提供了 4 种不同的优化算法,optimization 算法全部选用最常见密码子,next most optimal 算法选用与原密码子不同的最常见密码子,most different 算法选用与原密码子相似度最低的最常见密码子,random 算法则随



机选用与原密码子不同的同义密码子。网络服务器 Synthetic Gene Designer<sup>[86]</sup>则通过设置一个 0–64 间的比例因子数字,将表达宿主的同义密码子概率分布计算调整为优化的目标概率分布,再据此以蒙特卡罗模拟方法对序列进行优化;比例因子从 0 增长到 64 时,目标概率分布中同义密码子的偏性从最常见密码子逐步过渡到最稀有密码子,从而实现包括全部使用最常见密码子、全部使用最稀有密码子和基于表达宿主概率分布的蒙特卡罗模拟在内的不同优化策略。网络服务器 OPTIMIZER<sup>[87]</sup>在全部最优化和基于表达宿主概率分布的蒙特卡罗模拟之外,提出了 customized one amino acid-one codon 优化策略,对序列中出现的密码子按照它们在表达宿主中的使用概率升序排序,只对低概率的前  $n$  个密码子进行最优化, $n$  由用户指定;如果输入的是蛋白质序列,还可以按照 G+C 或 A+T 最多的方式选用密码子,或者选用 Archetti<sup>[88]</sup>定义的可最大限度减少突变或误译的密码子。

这些策略的优化目标单一局限于密码子使用偏好性,而影响翻译表达的还有其他因素<sup>[89-90]</sup>,因此研究人员又提出了多目标密码子优化算法。例如,在对宿主同义密码子概率分布进行蒙特卡罗模拟基础上,软件 Visual Gene Developer<sup>[91]</sup>增加了局部或全局二级结构优化的目标;软件 Eugene<sup>[92]</sup>增加了对密码子上下文偏好性、G+C 含量、隐藏终止密码子和密码子重复的优化;网络服务器 COOL<sup>[93]</sup>增加了个体密码子使用、密码子对、隐藏终止密码子和 G+C 含量等 4 个优化目标。此外,这些软件或网络服务器还具有添加或去除特定序列的功能。如 UpGene 同时还添加了排除富腺苷酸尿苷酸元件(adenylate uridylate-rich element, ARE)序列、设定固定不优化的序列、设定不可出现的序列

以及添加 Kozak 共有序列等功能。Synthetic Gene Designer 会去除重复片段如 GGGGG/CCCCC 和 AAAAA/TTTTT。OPTIMIZER、Visual Gene Developer 和 Codon Optimization OnLine 还具有特定序列排除功能。

综合来看,这些优化方法虽然已经考虑到模拟表达宿主密码子使用方式的重要性,但也仅仅是停留在对使用频率方面的带有经验、机械和随机性质的模拟。而对于一个特定蛋白质的最优表达,其编码 mRNA 中各种使用频率的密码子的排列组合应该存在某种最优解,其中密码子常见性、结构稳定性以及特定序列模式在整个 mRNA 序列中可能并不是均匀的,而是可能符合某种规范的。对不同的蛋白质,这种不均匀性可能是不同的,它们之间可能既有个体差异,也有共通之处,显然,以上这些优化算法在应对这些问题方面应该是远远不够的。例如,关于 mRNA 二级结构在蛋白质翻译中的作用,Mauger 等<sup>[81]</sup>通过核苷酸修饰进行研究,结果显示高表达的 mRNA 往往含有高度结构化的 CDS。Leppek 等<sup>[94]</sup>通过序列设计改变 CDS 的二级结构,结果显示结构化程度高的 CDS 翻译效率也高。但相反的是,Mao 等<sup>[95]</sup>的研究却显示 m<sup>6</sup>A 修饰可以通过减轻 mRNA 的结构化程度以促进翻译的进行。

Zhang 等<sup>[96]</sup>通过开发新的算法来平衡结构稳定性和密码子最优性,得到了具有优异蛋白质表达的 mRNA 序列。该研究受计算语言学启发,借用确定性有限状态自动机(deterministic finite-state automaton, DFA)点阵表示法对编码区的设计空间进行表示,借用点阵解析对编码区稳定性及密码子使用进行联合优化,将序列设计问题简化为计算语言学中点阵解析的经典问题,最终获得了一种简单的解决算法,称为 LinearDesign。关于该算法,有 2 个有趣的现象

值得一提。(1) 通过该算法虽然仅能得到编码区序列, 但将其与广泛使用的 UTR 序列组成完整 mRNA 时, 并不会过于干扰后者的结构, 仍能改善相应蛋白质的表达和免疫反应, 这应该是得益于所设计的编码区具有较强的自身结构<sup>[94,96]</sup>。(2) 虽然使用化学修饰的核苷酸已被广泛认为对 mRNA 疫苗的成功至关重要<sup>[97-99]</sup>, 但目前 LinearDesign 设计的 mRNA 并未使用化学修饰的核苷酸, 仍然可以表现出高水平的稳定性、翻译效率和免疫原性, 使其兼具较低制造成本的优势<sup>[96]</sup>。由于该算法所采用点阵表示法的通用性, 当与 mRNA 设计相关的其他参数可用时, 该算法容易调整适应, 从而增加对这些新加入参数的联合优化, 并可能取得更好的效果<sup>[96]</sup>。尽管如此, 这种优化算法本质上仍然是基于有限的经验假设的。近年来, 基于大数据的深度学习不断用于生物学领域的研究, 不仅可以抛开经验认识方面的不足和限制, 还能够从数据中学习得到传统方法难以发掘到的各组成部分之间的关联规律, 其中也应该包括数据个性与共性的关联性<sup>[100-104]</sup>, 有望弥补传统优化方法的缺陷。

Fu 等<sup>[105]</sup>注意到核苷酸组成相同的密码子共有 20 组, 恰好与氨基酸数目相同, 而且每一组密码子中不会出现同义密码子, 若将密码子组用于对蛋白质序列进行编码, 通过深度学习方法构建设计高表达 mRNA 的模型时, 相比用密码子编码, 降低了学习复杂度并使模型简化。其结果显示, 对几个不同编码序列的优化表达结果优于商业软件 ThermoFisher ([www.thermofisher.com](http://www.thermofisher.com)) 和 Genewiz ([www.genewiz.com](http://www.genewiz.com))。

多转换双向编码器表示(bidirectional encoder representation from transformers, BERT)<sup>[106]</sup>是近年提出的一个语言表示模型, 在多项自然语言处

理任务中取得了更好的效果。Li 等<sup>[107]</sup>将该模型扩展应用于 mRNA 语言, 将密码子用作输入符号, 通过包括细菌病毒以及人等多物种的千万条 mRNA 序列的训练, 建立了 CodonBERT 模型, 发现可以从头学习到遗传密码和同源关系等生物学知识, 扩展到用于 mRNA 稳定性及翻译等预测任务的监督学习, 在多个数据集上取得了比其他模型更好的结果。值得注意的是, 密码子优化虽然已经被广泛使用, 关于其安全性却仍存在争议<sup>[108]</sup>, 因为它可能会改变蛋白质构象并在体内产生具有未知生物活性的新肽<sup>[109-110]</sup>, 其机制还有待进一步研究。

上述密码子优化工具简要总结于表 2。

另外, 信号肽对于在细胞外发挥功能的蛋白质是必需的<sup>[111]</sup>, 因此已被广泛用于改善 mRNA 疫苗的抗原呈递<sup>[112]</sup>。例如, Kreiter 等<sup>[113]</sup>通过在 ORF 中编码分泌信号和 MHC-I 分子的跨膜胞质结构域来改善蛋白质抗原的运输特性, 这使其在树突状细胞中的抗原呈递增加了约 10 倍, 并提高了 mRNA 疫苗在小鼠中的抗肿瘤功效。

已上市的 2 款新型冠状病毒 mRNA 疫苗的 ORF 均保留了 S 蛋白的信号肽编码序列, 在密码子优化方面, 无明确信息显示这 2 个序列采用了哪种优化软件, 但应该是采取了传统的优化策略。如研究显示 BioNTech 的 BNT162b2 采取了宽松的优化策略, 保留了一定比例的非主流密码子, 而 Moderna 的 mRNA-1273 则采取了严格策略, 偏向优化为最常见密码子。严格优化策略可能会导致最常见密码子对应的 tRNA 耗竭现象, 从而不利于翻译的进行, 这可能是现实应用中 mRNA-1273 的接种剂量高于 BNT162b2 的原因。至于相关联的二级结构优化方面, 这 2 款疫苗可能都未考虑<sup>[71]</sup>。

表 2 密码子优化工具概览

Table 2 Overview of codon optimization tools

Tool	Optimization objectives							Algorithm	Web URL
	Codon usage	Codon context	Codon reuse	Hidden stop codon	Secondary structure	GC/AT content	Motif avoidance		
Codon optimizer <sup>[79]</sup>	Optimal							Select optimal codon	N/A
Jcat <sup>[80]</sup>	Optimal							Select optimal codon	<a href="https://www.jcat.de/">https://www.jcat.de/</a>
UpGene <sup>[84]</sup>	Flexible						Yes	Monte Carlo simulation	<a href="http://www.vectorcore.pitt.edu/upgene.html">http://www.vectorcore.pitt.edu/upgene.html</a>
GeneDesign <sup>[85]</sup>	Flexible							Select optimal codon; Next most optimal: most different: random	<a href="http://genedesign.org/">http://genedesign.org/</a>
Synthetic gene designer <sup>[86]</sup>	Flexible						Yes	Monte Carlo simulation based on scaled probability distribution	<a href="http://www.evolvingcode.net/codon/sgd/">http://www.evolvingcode.net/codon/sgd/</a>
OPTIMIZER <sup>[87]</sup>	Flexible					Yes	Yes	Select optimal codon; Monte Carlo simulation; customized one amino acid-one codon; G+C rich; A+T rich; Select Archetti's codons	<a href="http://genomes.urv.es/OPTIMIZER">http://genomes.urv.es/OPTIMIZER</a>
Visual gene developer <sup>[91]</sup>	Flexible				Yes		Yes	Monte Carlo simulation	<a href="http://www.visualgenedeveloper.net/">http://www.visualgenedeveloper.net/</a>
Eugene <sup>[92]</sup>	Flexible	Yes	Yes	Yes		Yes	Yes	Monte Carlo simulation	<a href="http://bioinformatics.ua.pt/eugene">http://bioinformatics.ua.pt/eugene</a>
COOL <sup>[93]</sup>	Flexible	Yes		Yes		Yes	Yes	Monte Carlo simulation	<a href="http://bioinfo.bti.a-star.edu.sg/COOL/">http://bioinfo.bti.a-star.edu.sg/COOL/</a>
LinearDesign <sup>[96]</sup>	Flexible				Yes			Lattice analysis	N/A
CodonBERT <sup>[107]</sup>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	BERT	N/A

Note: Optimal means always selecting optimal codons; Flexible means not always selecting optimal codons; Empty means not considering that objective.

## 4 3' UTR

3' UTR 是位于编码区下游的一段不被翻译的序列，其长度从几个到数千个核苷酸不等。对于内源 mRNA，3' UTR 的长度主要由聚腺苷酸化位点(polyadenylation site, PAS)决定。前体 mRNA 可含有多个 PAS，通过可选多聚腺苷酸

化(alternative polyadenylation, APA)创建不同长度的 3' UTR。有证据表明，APA 以细胞类型特异性的方式发生<sup>[114-115]</sup>，这提示相同 UTR 在不同类型细胞中的性能有所不同，因此，3' UTR 的设计需要考虑根据靶向细胞类型进行优化的问题。

与 5' UTR 类似, 3' UTR 也包含许多顺式调控元件, 对 mRNA 的稳定性、亚细胞定位和翻译效率具有重大贡献<sup>[116-117]</sup>。3' UTR 对于将 mRNA 靶向到特定的细胞器至关重要, 这对于 mRNA 的定位和后续的蛋白质合成是必需的, 尤其是在高度极化或分化的细胞中<sup>[118]</sup>。其中最常见的是 ARE 元件, 它可与 AU 结合蛋白相互作用, 引发 mRNA 的降解<sup>[119]</sup>。还有许多其他元件, 如富 GU 元件、富 CU 元件、富 CA 元件、富 GC 元件和铁反应元件, 其作用取决于它们所结合的反式作用因子。

3' UTR 中还可能含有 miRNA 结合序列, 其与 miRNA 的结合可沉默 mRNA 的表达<sup>[120-122]</sup>。目前普遍认为, 较短的 3' UTR 中 miRNA 结合序列会相对减少, 从而可使 mRNA 更加稳定。Sandberg 等<sup>[123]</sup>的研究显示, 快速增殖的细胞中会出现较短的 3' UTR 以及减少的 miRNA 结合序列, 这与细胞对稳定的 mRNA 以及相应增强的蛋白质表达的需求是一致的。这一结果表明, 通过对 3' UTR 内的 miRNA 结合序列进行设计, 可以控制 mRNA 在不同细胞类型中的表达活性, 甚至进行细胞特异性的翻译表达<sup>[124]</sup>。

最近的研究显示, 3' UTR 内的非结构化序列不利于 poly(A)尾发挥促翻译作用, 这使得 3' UTR 的结构成为一个新的优化目标<sup>[125]</sup>。Zhao 等<sup>[126]</sup>通过 MPRA 技术对 3' UTR 进行功能注释, 结果鉴定出了 106 个不稳定元件和 44 个稳定元件, 其中 83 个不含有可被已知 miRNA 和 RNA 结合蛋白结合的基序, 为序列的优化设计提供了重要的参考。MPRA 技术也被用于 3' UTR 的优化设计, 通过基于细胞的选择过程, 识别可增强蛋白质表达的 3' UTR 序列<sup>[127]</sup>。

3' UTR 的 5' 末端是终止密码子, 共有 3 种, 它们的终止效率和准确性具有很大差异。一般认为, 高表达的哺乳动物基因更多使用 UAA 终

止密码子。在终止密码子的选用方面, Moderna 的 mRNA-1273 将 3 种不同的终止密码子串联使用(UGAUAAUAG), 而 BioNTech 的 BNT162b2 mRNA 则使用 2 个连续的 UGA 终止密码子。这种设计可能是期望达到一种故障安全的效果, 但实际上鉴于 UGAU 存在被移码通读的可能, 这可能并不是一种最优设计<sup>[71]</sup>。

在 3' UTR 设计方面, 2 款疫苗均选用了已被证实成功有效的现有序列。如 Moderna 的 mRNA-1273 使用了人  $\alpha$  珠蛋白基因 110 nt 的 3' UTR。BioNTech 的 BNT162b2 则将来自 2 个基因的 RNA 片段串联使用, 先是 136 nt 的 *TLE5* 基因的 RNA 片段(含 2 个 C-U 变异), 然后是 139 nt 的 *RMR1* 基因的 RNA 片段, 这 2 个片段在之前的相关研究中被证明优于其他候选序列<sup>[127]</sup>。

## 5 poly(A)

人源 poly(A)尾通常由 10–250 个腺苷酸组成, 最常见的为 50–100 nt, 超过 250 nt 的也有<sup>[128]</sup>, 但很少见<sup>[129]</sup>。poly(A)一方面可与 poly(A)尾结合蛋白[poly(A) binding protein, PABP]结合, PABP 随后结合 eIF4G, 以增强 eIF4F 与 5' 帽的结合, 并促进 mRNA 环化形成“闭环结构”, 有助于将 eIF4F “拴”在 mRNA 附近, 从而提高翻译的循环效率<sup>[8]</sup>; 另一方面可以保护 mRNA 编码区免受降解酶的影响<sup>[129-130]</sup>。poly(A)尾的长度一般是 30 nt 的倍数<sup>[131-133]</sup>。与一些研究认为较长 poly(A)尾会导致较高稳定性和翻译效率<sup>[130, 134-135]</sup>不一致的是, 也有研究显示许多高度翻译、稳定的 mRNA 的 poly(A)尾在稳定状态下相对较短, 在 30 nt 左右, 也有少于 20 nt 的, 如  $\beta$ -actin<sup>[136-137]</sup>, 而翻译不佳的 mRNA 则具有相对较长的 poly(A)尾。另有研究显示, 稳态 poly(A)尾的长度和 mRNA 半衰期的相关性较差或呈负相关<sup>[138]</sup>, 稳定、高度翻译的转录本更

能抵抗脱腺苷化<sup>[129]</sup>。可见, poly(A)尾的长度在 mRNA 的稳定性和翻译效率中扮演重要角色, 但其机制可能涉及各种复杂的交互作用, 而这些目前还不完全清楚<sup>[139]</sup>。

最新的全长 poly(A)和 mRNA 测序[full-length poly(A) and mRNA sequencing, FLAM-seq]技术可准确地对包括 poly(A)尾在内的 mRNA 转录本进行测序, 结果表明 poly(A)中除了腺苷酸外还掺入了其他核苷酸, 主要是胞苷酸<sup>[128]</sup>。2022 年, Li 等<sup>[140]</sup>报道称, 因为去腺苷酸酶去除胞苷酸的效率较低, 在 poly(A)尾部掺入胞苷可保护 mRNA 免受去腺苷酸酶的去腺苷作用, 从而可增强 mRNA 的稳定性和翻译产出。

BioNTech BNT162b2 的 poly(A)长 110 nt, 由 30 个腺苷残基之后加 GCAUAUGACU 接头序列和另外 70 个腺苷残基组成。插入接头序列可能是出于生产工艺考虑<sup>[141-142]</sup>, 而其中掺入的胞苷酸也可能对增强稳定性和翻译产出有所贡献。

## 6 修饰的核苷

为减少合成 mRNA 引起的先天免疫刺激, 研究人员还广泛采用了掺入修饰核苷的方法, 但这对 mRNA 整体翻译的影响并不确定。例如有报道显示掺入 m<sup>6</sup>A 修饰可能会扰乱 tRNA 在密码子处的装载从而阻碍翻译的进行<sup>[143]</sup>。掺入假尿苷会改变 tRNA 的结合从而阻碍翻译的延伸和增加氨基酸取代的发生<sup>[144]</sup>。但也有报道编码区的 m<sup>6</sup>A 可以通过减轻 mRNA 二级结构而促进翻译进行<sup>[95]</sup>。这些研究结果提示, 修饰的核苷因其本身以及所在位置的不同, 可能会从多个不同方面对翻译的进行产生不同的影响, 所以需要针对 mRNA 序列本身具体情况进行具体设计, 但这种过于具体的设计结果, 在目前的体外转录生产过程中应该是难以实现的。Moderna

和 BioNTech 均直接选择将疫苗 mRNA 所有 U 修饰为 N1-methylpseudouridine ( $\Psi$ ), 以避免先天免疫刺激对疫苗 mRNA 的攻击破坏, 而未顾及这种修饰可能引起的翻译误读。

## 7 各因素的系统组合研究

mRNA 分子作为一个系统有机的整体, 其稳定性和表达性能应该是由分子的整体以及所处环境决定的, 为系统探究高稳定性和高表达性的 mRNA 的设计原则, Leppek 等<sup>[94]</sup>开发了一种新的 MPRA 方法, 即 mRNA 溶液稳定性、细胞内稳定性和翻译的汇总评估 RNA 测序(pooled evaluation of mRNA in-solution stability, and in-cell stability and translation RNA-seq, PERSIST-seq), RNA 测序技术可以对大量 mRNA 序列的溶液内稳定性、细胞内稳定性以及细胞内 MRL 进行大规模并行测定; 通过对 UTR 和编码区进行系统的设计组合, 得到 233 条不同的 mRNA 序列, 测定了其溶液内稳定性、细胞内稳定性以及 MRL, 并分析了这些性能与序列的关系。发现结构化的 CDS 以及细胞和病毒来源的各种 5' UTR 元件可以维持甚至提高 MRL; CDS 和 5' UTR 引起的细胞内稳定性的变化较大; MRL 在适度范围与 mRNA 稳定性之间存在正相关关系, 在较高范围则转为负相关<sup>[94]</sup>。将 MRL 和细胞内稳定性整合到一个简单的定量模型中, 以理解和预测它们对蛋白表达的相对影响, 结果显示, 在较长时间时, 细胞内稳定性是驱动蛋白质产出的更为重要的因素, 而在较短时间内, MRL 是驱动蛋白质产出的更为重要的因素。考虑到样本量偏少, 该研究通过构建随机 5' UTR 库的方式探究 5' UTR 影响表达的具体因素, 结果显示, 高表达的序列明显缺失与主 ORF 不在同一三联体阅读框架的上游 AUG 起始密码子, 而且一些短茎环基序的富集对翻译具有促进作

用。总之，这一研究显示优化的主要焦点应该为 UTR 的优化和 CDS 的结构优化，而假尿苷修饰则几乎是必备操作。但由于该研究的样本量有限，而且所研究的翻译环境仅局限于 HEK293T 细胞，因此相关结果的普遍性还有待验证。

## 8 展望

为了快速应对新型冠状病毒感染大流行，上市的两款 mRNA 疫苗序列的设计主要还是遵循传统经典策略，对于一些潜在的风险因素未能加以规避。而未来出现类似传染大流行的风险依旧存在，因此需要继续推进相关研究，找到规避相关风险的设计和工艺方法，以实现快速设计出更加安全的 mRNA 疫苗的目标。

mRNA 翻译是一个由 mRNA 自身和所处细胞环境共同决定的过程，是一个多因素交互的系统化过程，厘清这些系统关系是成功进行 mRNA 设计的最佳保证。因此，需要针对这一问题进行系统的设计和实验，而 PERSIST-seq 这种 MPRA 方法已经是当前一种比较理想的实验研究方法，如果能够突破采样空间的限制，再结合深度学习技术，应该会得到一个考虑因素较为全面的理想模型，以用于任意目标的 mRNA 设计。

## REFERENCES

- [1] RAPPUOLI R, PIZZA M, del GIUDICE G, de GREGORIO E. Vaccines, new opportunities for a new society[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(34): 12288-12293.
- [2] MASCOLA JR, FAUCI AS. Novel vaccine technologies for the 21st century[J]. *Nature Reviews Immunology*, 2020, 20(2): 87-88.
- [3] THANH LE T, ANDREADAKIS Z, KUMAR A, GÓMEZ ROMÁN R, TOLLEFSEN S, SAVILLE M, MAYHEW S. The COVID-19 vaccine development landscape[J]. *Nature Reviews Drug Discovery*, 2020, 19(5): 305-306.
- [4] COVID-19 Vaccines with WHO Emergency Use Listing[EB/OL]. [2022-07-29]. <https://extranet.who.int/pqweb/vaccines/vaccinescovid-19-vaccine-eul-issued>.
- [5] COVID-19 vaccine tracker and landscape[EB/OL]. [2024-04-09]. <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>.
- [6] CORBETT KS, EDWARDS DK, LEIST SR, ABIONA OM, BOYOGLU-BARNUM S, GILLESPIE RA, HIMANSU S, SCHÄFER A, ZIWAWO CT, DIPIAZZA AT, DINNON KH, ELBASHIR SM, SHAW CA, WOODS A, FRITCH EJ, MARTINEZ DR, BOCK KW, MINAI M, NAGATA BM, HUTCHINSON GB, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness[J]. *Nature*, 2020, 586(7830): 567-571.
- [7] HASSETT KJ, BENENATO KE, JACQUINET E, LEE AS, WOODS A, YUZHAKOV O, HIMANSU S, DETERLING J, GEILICH BM, KETOVA T, MIHAI C, LYNN A, McFADYEN I, MOORE MJ, SENN JJ, STANTON MG, ALMARSSON Ö, CIARAMELLA G, BRITO LA. Optimization of lipid nanoparticles for intramuscular administration of mRNA vaccines[J]. *Molecular Therapy Nucleic Acids*, 2019, 15: 1-11.
- [8] JACKSON RJ, HELLEN CUT, PESTOVA TV. The mechanism of eukaryotic translation initiation and principles of its regulation[J]. *Nature Reviews Molecular Cell Biology*, 2010, 11(2): 113-127.
- [9] STUDIER FW, MOFFATT BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes[J]. *Journal of Molecular Biology*, 1986, 189(1): 113-130.
- [10] MILLIGAN JF, GROEBE DR, WITHERELL GW, UHLENBECK OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates[J]. *Nucleic Acids Research*, 1987, 15(21): 8783-8798.
- [11] ROTTMAN F, SHATKIN AJ, PERRY RP. Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing[J]. *Cell*, 1974, 3(3): 197-199.
- [12] WEI CM, GERSHOWITZ A, MOSS B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA[J]. *Cell*, 1975, 4(4): 379-386.
- [13] ADAMS JM, CORY S. Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA[J]. *Nature*, 1975, 255(5503): 28-33.

- [14] SHATKIN AJ. Capping of eucaryotic mRNAs[J]. *Cell*, 1976, 9(4 PT 2): 645-653.
- [15] DAFFIS S, SZRETTER KJ, SCHRIEWER J, LI JQ, YOUN S, ERRETT J, LIN TY, SCHNELLER S, ZUST R, DONG HP, THIEL V, SEN GC, FENSTERL V, KLIMSTRA WB, PIERSON TC, BULLER RM, GALE JR M, SHI PY, DIAMOND MS. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members[J]. *Nature*, 2010, 468(7322): 452-456.
- [16] DECROLY E, FERRON F, LESCAR J, CANARD B. Conventional and unconventional mechanisms for capping viral mRNA[J]. *Nature Reviews Microbiology*, 2012, 10(1): 51-65.
- [17] DEVARKAR SC, WANG C, MILLER MT, RAMANATHAN A, JIANG FG, KHAN AG, PATEL SS, MARCOTRIGIANO J. Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(3): 596-601.
- [18] SHIMOTOHNO K, KODAMA Y, HASHIMOTO J, MIURA KI. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74(7): 2734-2738.
- [19] HUDEČEK O, BENONI R, REYES-GUTIERREZ PE, CULKA M, ŠANDEROVÁ H, HUBÁLEK M, RULÍŠEK L, CVAČKA J, KRÁSNÝ L, CAHOVÁ H. Dinucleoside polyphosphates act as 5'-RNA caps in bacteria[J]. *Nature Communications*, 2020, 11(1): 1052.
- [20] MALONE RW, FELGNER PL, VERMA IM. Cationic liposome-mediated RNA transfection[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(16): 6077-6081.
- [21] YISRAELI JK, MELTON DA. *Methods in Enzymology*[M]. Salt Lake City: Academic Press, 1989: 42-50.
- [22] STEPINSKI J, WADDELL C, STOLARSKI R, DARZYNKIEWICZ E, RHOADS RE. Synthesis and properties of mRNAs containing the novel "anti-reverse" cap analogs 7-methyl(3'-O-methyl)GpppG and 7-methyl (3'-deoxy)GpppG[J]. *RNA*, 2001, 7(10): 1486-1495.
- [23] PASQUINELLI AE, DAHLBERG JE, LUND E. Reverse 5' caps in RNAs made in vitro by phage RNA polymerases[J]. *RNA*, 1995, 1(9): 957-967.
- [24] PENG ZH, SHARMA V, SINGLETON SF, GERSHON PD. Synthesis and application of a chain-terminating dinucleotide mRNA cap analog[J]. *Organic Letters*, 2002, 4(2): 161-164.
- [25] JEMIELITY J, FOWLER T, ZUBEREK J, STEPINSKI J, LEWDOROWICZ M, NIEDZWIECKA A, STOLARSKI R, DARZYNKIEWICZ E, RHOADS RE. Novel "anti-reverse" cap analogs with superior translational properties[J]. *RNA*, 2003, 9(9): 1108-1122.
- [26] GRUDZIEN E, STEPINSKI J, JANKOWSKA-ANYSZKA M, STOLARSKI R, DARZYNKIEWICZ E, RHOADS RE. Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency[J]. *RNA*, 2004, 10(9): 1479-1487.
- [27] NIEDZWIECKA A, MARCOTRIGIANO J, STEPINSKI J, JANKOWSKA-ANYSZKA M, WYSLOUCH-CIESZYNSKA A, DADLEZ M, GINGRAS AC, MAK P, DARZYNKIEWICZ E, SONENBERG N, BURLEY SK, STOLARSKI R. Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins[J]. *Journal of Molecular Biology*, 2002, 319(3): 615-635.
- [28] CHARENTON C, GRAILLE M. mRNA decapping: finding the right structures[J]. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 2018, 373(1762): 20180164.
- [29] DARZYNKIEWICZ E, STEPINSKI J, EKIEL I, GOYER C, SONENBERG N, TEMERIUZ A, JIN Y, SIJUWADE T, HABER D, TAHARA SM. Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogues of mRNA 5'-cap: changes in N7 substituent affect analogue activity[J]. *Biochemistry*, 1989, 28(11): 4771-4778.
- [30] KUHN AN, DIKEN M, KREITER S, SELMI A, KOWALSKA J, JEMIELITY J, DARZYNKIEWICZ E, HUBER C, TÜRECI O, SAHIN U. Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses *in vivo*[J]. *Gene Therapy*, 2010, 17(8): 961-971.
- [31] GRUDZIEN E, KALEK M, JEMIELITY J, DARZYNKIEWICZ E, RHOADS RE. Differential inhibition of mRNA degradation pathways by novel cap analogs[J]. *The Journal of Biological Chemistry*, 2006, 281(4): 1857-1867.
- [32] RYDZIK AM, LUKASZEWICZ M, ZUBEREK J,

- KOWALSKA J, DARZYNKIEWICZ ZM, DARZYNKIEWICZ E, JEMIELITY J. Synthetic dinucleotide mRNA cap analogs with tetraphosphate 5',5' bridge containing methylenebis(phosphonate) modification[J]. *Organic & Biomolecular Chemistry*, 2009, 7(22): 4763-4776.
- [33] RYDZIK AM, KULIS M, LUKASZEWICZ M, KOWALSKA J, ZUBEREK J, DARZYNKIEWICZ ZM, DARZYNKIEWICZ E, JEMIELITY J. Synthesis and properties of mRNA cap analogs containing imidodiphosphate moiety-fairly mimicking natural cap structure, yet resistant to enzymatic hydrolysis[J]. *Bioorganic & Medicinal Chemistry*, 2012, 20(5): 1699-1710.
- [34] SU W, SLEPENKOV S, GRUDZIEN-NOGALSKA E, KOWALSKA J, KULIS M, ZUBEREK J, LUKASZEWICZ M, DARZYNKIEWICZ E, JEMIELITY J, RHOADS RE. Translation, stability, and resistance to decapping of mRNAs containing caps substituted in the triphosphate chain with BH<sub>3</sub>, Se, and NH[J]. *RNA*, 2011, 17(5): 978-988.
- [35] RYDZIK AM, WARMINSKI M, SIKORSKI PJ, BARANOWSKI MR, WALCZAK S, KOWALSKA J, ZUBEREK J, LUKASZEWICZ M, NOWAK E, W. CLARIDGE TD, DARZYNKIEWICZ E, NOWOTNY M, JEMIELITY J. mRNA cap analogues substituted in the tetraphosphate chain with CX<sub>2</sub>: identification of O-to-CCl<sub>2</sub> as the first bridging modification that confers resistance to decapping without impairing translation[J]. *Nucleic Acids Research*, 2017, 45(15): 8661-8675.
- [36] GRUDZIEN-NOGALSKA E, JEMIELITY J, KOWALSKA J, DARZYNKIEWICZ E, RHOADS RE. Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells[J]. *RNA*, 2007, 13(10): 1745-1755.
- [37] KOWALSKA J, WYPIJEWSKA del NOGAL A, DARZYNKIEWICZ ZM, BUCK J, NICOLA C, KUHN AN, LUKASZEWICZ M, ZUBEREK J, STRENKOWSKA M, ZIEMNIAK M, MACIEJCZYK M, BOJARSKA E, RHOADS RE, DARZYNKIEWICZ E, SAHIN U, JEMIELITY J. Synthesis, properties, and biological activity of boranophosphate analogs of the mRNA cap: versatile tools for manipulation of therapeutically relevant cap-dependent processes[J]. *Nucleic Acids Research*, 2014, 42(16): 10245-10264.
- [38] STRENKOWSKA M, GRZELA R, MAJEWSKI M, WNEK K, KOWALSKA J, LUKASZEWICZ M, ZUBEREK J, DARZYNKIEWICZ E, KUHN AN, SAHIN U, JEMIELITY J. Cap analogs modified with 1,2-dithiodiphosphate moiety protect mRNA from decapping and enhance its translational potential[J]. *Nucleic Acids Research*, 2016, 44(20): 9578-9590.
- [39] STRENKOWSKA M, KOWALSKA J, LUKASZEWICZ M, ZUBEREK J, SU W, RHOADS RE, DARZYNKIEWICZ E, JEMIELITY J. Towards mRNA with superior translational activity: synthesis and properties of ARCA tetraphosphates with single phosphorothioate modifications[J]. *New Journal of Chemistry*, 2010, 34(5): 993-1007.
- [40] KOWALSKA J, LUKASZEWICZ M, ZUBEREK J, DARZYNKIEWICZ E, JEMIELITY J. Phosphoroselenoate dinucleotides for modification of mRNA 5' end[J]. *ChemBioChem*, 2009, 10(15): 2469-2473.
- [41] FUCHS AL, NEU A, SPRANGERS R. A general method for rapid and cost-efficient large-scale production of 5' capped RNA[J]. *RNA*, 2016, 22(9): 1454-1466.
- [42] OHNO H, AKAMINE S, MOCHIZUKI M, HAYASHI K, AKICHIKA S, SUZUKI T, SAITO H. Versatile strategy using vaccinia virus-capping enzyme to synthesize functional 5' cap-modified mRNAs[J]. *Nucleic Acids Research*, 2023, 51(6): e34.
- [43] ZHANG NN, LI XF, DENG YQ, ZHAO H, HUANG YJ, YANG G, HUANG WJ, GAO P, ZHOU C, ZHANG RR, GUO Y, SUN SH, FAN H, ZU SL, CHEN Q, HE Q, CAO TS, HUANG XY, QIU HY, NIE JH, et al. A ahermostable mRNA vaccine against COVID-19[J]. *Cell*, 2020, 182(5): 1271-1283.e16.
- [44] YE Q, WU M, ZHOU C, LU XS, HUANG BY, ZHANG N, ZHAO H, CHI H, ZHANG XJ, LING DD, ZHANG RR, LI ZF, LUO D, HUANG YJ, QIU HY, SONG HF, TAN WJ, XU K, YING B, QIN CF. Rational development of a combined mRNA vaccine against COVID-19 and influenza[J]. *NPJ Vaccines*, 2022, 7(1): 84.
- [45] LANGBERG SR, MOSS B. Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells[J]. *The Journal of Biological Chemistry*, 1981, 256(19): 10054-10060.
- [46] RAMANATHAN A, ROBB GB, CHAN SH. mRNA capping: biological functions and applications[J]. *Nucleic Acids Research*, 2016, 44(16): 7511-7526.



- [47] HENDERSON JM, UJITA A, HILL E, YOUSIF-ROSALES S, SMITH C, KO N, McREYNOLDS T, CABRAL CR, ESCAMILLA-POWERS JR, HOUSTON ME. Cap 1 messenger RNA sSynthesis with eCo-transcriptional CleanCap® analog by *in vitro* transcription[J]. *Current Protocols*, 2021, 1(2): e39.
- [48] LINARES-FERNÁNDEZ S, LACROIX C, EXPOSITO JY, VERRIER B. Tailoring mRNA vaccine to balance innate/adaptive immune response[J]. *Trends in Molecular Medicine*, 2020, 26(3): 311-323.
- [49] A prefusion SARS-CoV-2 spike RNA vaccine is highly immunogenic and prevents lung infection in non-human primates[EB/OL]. [2023-09-28]. <https://www.biorxiv.org/content/10.1101/2020.09.08.280818v1>.
- [50] SAHIN U, MUIK A, DERHOVANESEAN E, VOGLER I, KRANZ LM, VORMEHR M, BAUM A, PASCAL K, QUANDT J, MAURUS D, BRACHTENDORF S, LÖRKS V, SIKORSKI J, HILKER R, BECKER D, ELLER AK, GRÜTZNER J, BOESLER C, ROSENBAUM C, KÜHNLE MC, et al. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses[J]. *Nature*, 2020, 586(7830): 594-599.
- [51] LEPPEK K, DAS R, BARNA M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them[J]. *Nature Reviews Molecular Cell Biology*, 2018, 19(3): 158-174.
- [52] ARAUJO PR, YOON K, KO D, SMITH AD, QIAO M, SURESH U, BURNS SC, PENALVA LOF. Before it gets started: regulating translation at the 5' UTR[J]. *Comparative and Functional Genomics*, 2012, 2012: 475731.
- [53] JIA LF, MAO YH, JI QQ, DERSH D, YEWDELL JW, QIAN SB. Decoding mRNA translatability and stability from the 5' UTR[J]. *Nature Structural & Molecular Biology*, 2020, 27(9): 814-821.
- [54] LEE ASY, KRANZUSCH PJ, CATE JHD. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression[J]. *Nature*, 2015, 522(7554): 111-114.
- [55] KAROLLUS A, AVSEC Ž, GAGNEUR J. Predicting mean ribosome load for 5'UTR of any length using deep learning[J]. *PLoS Computational Biology*, 2021, 17(5): e1008982.
- [56] YANG Y, WANG ZF. IRES-mediated cap-independent translation, a path leading to hidden proteome[J]. *Journal of Molecular Cell Biology*, 2019, 11(10): 911-919.
- [57] KOZAK M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes[J]. *Cell*, 1986, 44(2): 283-292.
- [58] NODERER WL, FLOCKHART RJ, BHADURI A, DIAZ de ARCE AJ, ZHANG JJ, KHAVARI PA, WANG CL. Quantitative analysis of mammalian translation initiation sites by FACS-seq[J]. *Molecular Systems Biology*, 2014, 10(8): 748.
- [59] MIGNONE F, GISSI C, LIUNI S, PESOLE G. Untranslated regions of mRNAs[J]. *Genome Biology*, 2002, 3(3): REVIEWS0004.
- [60] GEBRE MS, RAUCH S, ROTH N, YU JY, CHANDRASHEKAR A, MERCADO NB, HE X, LIU JY, McMAHAN K, MARTINOT A, MARTINEZ DR, GIFFIN V, HOPE D, PATEL S, SELLERS D, SANBORN O, BARRETT J, LIU XW, COLE AC, PESSAINT L, et al. Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine[J]. *Nature*, 2022, 601(7893): 410-414.
- [61] CASTILLO-HAIR SM, SEELIG G. Machine learning for designing next-generation mRNA therapeutics[J]. *Accounts of Chemical Research*, 2022, 55(1): 24-34.
- [62] SAMPLE PJ, WANG B, REID DW, PRESNYAK V, McFADYEN IJ, MORRIS DR, SEELIG G. Human 5' UTR design and variant effect prediction from a massively parallel translation assay[J]. *Nature Biotechnology*, 2019, 37(7): 803-809.
- [63] CAO JC, NOVOA EM, ZHANG ZZ, CHEN WCW, LIU DB, CHOI GCG, WONG ASL, WEHRSPAUN C, KELLIS M, LU TK. High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies[J]. *Nature Communications*, 2021, 12(1): 4138.
- [64] Generating and designing DNA with deep generative models[EB/OL]. [2023-09-05]. <http://arxiv.org/abs/1712.06148>.
- [65] LINDER J, SEELIG G. Fast activation maximization for molecular sequence design[J]. *BMC Bioinformatics*, 2021, 22(1): 510.
- [66] Auto-Encoding Variational Bayes[EB/OL]. [2023-09-05]. <http://arxiv.org/abs/1312.6114>.
- [67] Generative Adversarial Networks[EB/OL]. [2023-09-05]. <http://arxiv.org/abs/1406.2661>.
- [68] LINDER J, BOGARD N, ROSENBERG AB, SEELIG G. A generative neural network for maximizing fitness and diversity of synthetic DNA and protein

- sequences[J]. *Cell Systems*, 2020, 11(1): 49-62.e16.
- [69] WEINGARTEN-GABBAY S, ELIAS-KIRMA S, NIR R, GRITSENKO AA, STERN-GINOSSAR N, YAKHINI Z, WEINBERGER A, SEGAL E. Systematic discovery of cap-independent translation sequences in human and viral genomes[J]. *Science*, 2016, 351(6270): aad4939.
- [70] Messenger RNA encoding the full-length SARS-CoV-2 spike glycoprotein[EB/OL]. [2024-04-11]. <https://web.archive.org/web/20210105162941/https://mednet-communities.net/inn/db/media/docs/11889.doc>.
- [71] XIA XH. Detailed dissection and critical evaluation of the pfizer/BioNTech and moderna mRNA vaccines[J]. *Vaccines*, 2021, 9(7): 734.
- [72] ALEXAKI A, HETTIARACHCHI GK, ATHEY JC, KATNENI UK, SIMHADRI V, HAMASAKI-KATAGIRI N, NANA VATY P, LIN B, TAKEDA K, FREEDBERG D, MONROE D, MCGILL JR, PETERS R, KAMES JM, HOLCOMB DD, HUNT RC, SAUNA ZE, GELINAS A, JANJIC N, DiCUCCIO M, et al. Effects of codon optimization on coagulation factor IX translation and structure: implications for protein and gene therapies[J]. *Scientific Reports*, 2019, 9: 15449.
- [73] CANNAROZZI G, SCHRAUDOLPH NN, FATY M, von ROHR P, FRIBERG MT, ROTH AC, GONNET P, GONNET G, BARRAL Y. A role for codon order in translation dynamics[J]. *Cell*, 2010, 141(2): 355-367.
- [74] PRESNYAK V, ALHUSAINI N, CHEN YH, MARTIN S, MORRIS N, KLINE N, OLSON S, WEINBERG D, BAKER KE, GRAVELEY BR, COLLER J. Codon optimality is a major determinant of mRNA stability[J]. *Cell*, 2015, 160(6): 1111-1124.
- [75] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression[J]. *Trends in Biotechnology*, 2004, 22(7): 346-353.
- [76] SIMMS CL, YAN LL, ZAHER HS. Ribosome collision is critical for quality control during No-go decay[J]. *Molecular Cell*, 2017, 68(2): 361-373.e5.
- [77] HANSON G, COLLER J. *Codon* optimality, bias and usage in translation and mRNA decay[J]. *Nature Reviews Molecular Cell Biology*, 2018, 19: 20-30.
- [78] NARULA A, ELLIS J, TALIAFERRO JM, RISSLAND OS. Coding regions affect mRNA stability in human cells[J]. *RNA*, 2019, 25(12): 1751-1764.
- [79] FUGLSANG A. *Codon* optimizer: a freeware tool for *Codon* optimization[J]. *Protein Expression and Purification*, 2003, 31(2): 247-249.
- [80] GROTE A, HILLER K, SCHEER M, MÜNCH R, NÖRTEMANN B, HEMPEL DC, JAHN D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host[J]. *Nucleic Acids Research*, 2005, 33(suppl\_2): W526-W531.
- [81] MAUGER DM, CABRAL BJ, PRESNYAK V, SU SV, REID DW, GOODMAN B, LINK K, KHATWANI N, REYNDERS J, MOORE MJ, McFADYEN IJ. mRNA structure regulates protein expression through changes in functional half-life[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(48): 24075-24083.
- [82] HIA F, YANG SF, SHICHINO Y, YOSHINAGA M, MURAKAWA Y, VANDENBON A, FUKAO A, FUJIWARA T, LANDTHALER M, NATSUME T, ADACHI S, IWASAKI S, TAKEUCHI O. *Codon* bias confers stability to human mRNAs[J]. *EMBO Reports*, 2019, 20(11): e48220.
- [83] KUDLA G, LIPINSKI L, CAFFIN F, HELWAK A, ZYLICZ M. High guanine and cytosine content increases mRNA levels in mammalian cells[J]. *PLoS Biology*, 2006, 4(6): e180.
- [84] GAO WT, RZEWSKI A, SUN HJ, ROBBINS PD, GAMBOTTO A. UpGene: application of a web-based DNA codon optimization algorithm[J]. *Biotechnology Progress*, 2004, 20(2): 443-448.
- [85] RICHARDSON SM, WHEELAN SJ, YARRINGTON RM, BOEKE JD. GeneDesign: rapid, automated design of multikilobase synthetic genes[J]. *Genome Research*, 2006, 16(4): 550-556.
- [86] WU G, BASHIR-BELLO N, FREELAND SJ. The Synthetic Gene Designer: a flexible web platform to explore sequence manipulation for heterologous expression[J]. *Protein Expression and Purification*, 2006, 47(2): 441-445.
- [87] PUIGBÒ P, GUZMÁN E, ROMEU A, GARCIA-VALLVÉ S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences[J]. *Nucleic Acids Research*, 2007, 35(suppl\_2): W126-W131.
- [88] ARCHETTI M. Selection on codon usage for error minimization at the protein level[J]. *Journal of Molecular Evolution*, 2004, 59(3): 400-415.
- [89] SELIGMANN H, POLLOCK DD. The ambush hypothesis: hidden stop codons prevent off-frame gene reading[J]. *DNA and Cell Biology*, 2004, 23(10): 701-705.
- [90] CHUNG BKS, YUSUFI FNK, Mariati, YANG YS, LEE DY. Enhanced expression of codon optimized interferon gamma in CHO cells[J]. *Journal of*

- Biotechnology, 2013, 167(3): 326-333.
- [91] JUNG SK, McDONALD K. Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization[J]. BMC Bioinformatics, 2011, 12: 340.
- [92] GASPAR P, OLIVEIRA JL, FROMMLET J, SANTOS MAS, MOURA G. EuGene: maximizing synthetic gene design for heterologous expression[J]. Bioinformatics, 2012, 28(20): 2683-2684.
- [93] CHIN JX, CHUNG BKS, LEE DY. Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design[J]. Bioinformatics, 2014, 30(15): 2210-2212.
- [94] LEPPEK K, BYEON GW, KLADWANG W, WAYMENT-STEELE HK, KERR CH, XU AF, KIM DS, TOPKAR VV, CHOE C, ROTHSCHILD D, TIU GC, WELLINGTON-OGURI R, FUJII K, SHARMA E, WATKINS AM, NICOL JJ, ROMANO J, TUNGUZ B, DIAZ F, CAI H, et al. Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics[J]. Nature Communications, 2022, 13: 1536.
- [95] MAO YH, DONG LM, LIU XM, GUO JY, MA HH, SHEN B, QIAN SB. m6A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2[J]. Nature Communications, 2019, 10: 5332.
- [96] ZHANG H, ZHANG L, LIN A, XU CC, LI ZY, LIU KB, LIU BX, MA XP, ZHAO FF, JIANG HL, CHEN CX, SHEN HF, LI HW, MATHEWS DH, ZHANG YJ, HUANG L. Algorithm for optimized mRNA design improves stability and immunogenicity[J]. Nature, 2023, 621: 396-403.
- [97] KARIKÓ K, MURAMATSU H, WELSH FA, LUDWIG J, KATO H, AKIRA S, WEISSMAN D. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability[J]. Molecular Therapy, 2008, 16(11): 1833-1840.
- [98] PARDI N, HOGAN MJ, PORTER FW, WEISSMAN D. mRNA vaccines: a new era in vaccinology[J]. Nature Reviews Drug Discovery, 2018, 17: 261-279.
- [99] POLACK FP, THOMAS SJ, KITCHIN N, ABSALON J, GURTMAN A, LOCKHART S, PEREZ JL, MARC GP, MOREIRA ED, ZERBINI C, BAILEY R, SWANSON KA, ROYCHOUDHURY S, KOURY K, LI P, KALINA WV, COOPER D, FRENCK RW Jr, HAMMITT LL, TÜRECI Ö, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine[J]. The New England Journal of Medicine, 2020, 383(27): 2603-2615.
- [100] RIVES A, MEIER J, SERCU T, GOYAL S, LIN ZM, LIU J, GUO DM, OTT M, ZITNICK CL, MA J, FERGUS R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(15): e2016239118.
- [101] JI YR, ZHOU ZH, LIU H, DAVULURI RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome[J]. Bioinformatics, 2021, 37(15): 2112-2120.
- [102] Interpretable RNA Foundation Model from Unannotated Data for Highly Accurate RNA Structure and Function Prediction[EB/OL]. [2024-02-02]. <https://europepmc.org/article/PPR/PPR529644>.
- [103] AKIYAMA M, SAKAKIBARA Y. Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning[J]. NAR Genomics and Bioinformatics, 2022, 4(1): lqac012.
- [104] BEPLER T, BERGER B. Learning the protein language: evolution, structure, and function[J]. Cell Systems, 2021, 12(6): 654-669.e3.
- [105] FU HG, LIANG YB, ZHONG XQ, PAN ZL, HUANG L, ZHANG HL, XU Y, ZHOU W, LIU Z. Codon optimization with deep learning to enhance protein expression[J]. Scientific Reports, 2020, 10: 17617.
- [106] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. [2024-02-02]. <http://arxiv.org/abs/1810.04805>.
- [107] CodonBERT: Large Language Models for mRNA design and optimization[EB/OL]. [2023-12-12]. <https://www.biorxiv.org/content/10.1101/2023.09.09.556981v1>.
- [108] KOWALSKI PS, RUDRA A, MIAO L, ANDERSON DG. Delivering the messenger: advances in technologies for therapeutic mRNA delivery[J]. Molecular Therapy, 2019, 27(4): 710-728.
- [109] MAURO VP, CHAPPELL SA. A critical analysis of codon optimization in human therapeutics[J]. Trends in Molecular Medicine, 2014, 20(11): 604-613.
- [110] WEISSMAN D. mRNA transcript therapy[J]. Expert Review of Vaccines, 2015, 14(2): 265-281.
- [111] SCHLAKE T, THRAN M, FIEDLER K, HEIDENREICH R, PETSCH B, FOTIN-MLECZEK M.

- mRNA: a novel avenue to antibody therapy?[J]. *Molecular Therapy*, 2019, 27(4): 773-784.
- [112] KHOURY HJ, COLLINS RH Jr, BLUM W, STIFF PS, ELIAS L, LEBKOWSKI JS, REDDY A, NISHIMOTO KP, SEN D, WIRTH ED 3rd, CASE CC, DiPERSIO JF. Immune responses and long-term disease recurrence status after telomerase-based dendritic cell immunotherapy in patients with acute myeloid leukemia[J]. *Cancer*, 2017, 123(16): 3061-3072.
- [113] KREITER S, SELMI A, DIKEN M, SEBASTIAN M, OSTERLOH P, SCHILD H, HUBER C, TÜRECI Ö, SAHIN U. Increased antigen presentation efficiency by coupling antigens to MHC class I trafficking signals[J]. *Journal of Immunology*, 2008, 180(1): 309-318.
- [114] ZHANG HB, LEE JY, TIAN B. Biased alternative polyadenylation in human tissues[J]. *Genome Biology*, 2005, 6(12): R100.
- [115] JI Z, TIAN B. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types[J]. *PLoS One*, 2009, 4(12): e8419.
- [116] KISLAUSKIS EH, SINGER RH. Determinants of mRNA localization[J]. *Current Opinion in Cell Biology*, 1992, 4(6): 975-978.
- [117] MAYR C. What are 3' UTRs doing?[J]. *Cold Spring Harbor Perspectives in Biology*, 2019, 11(10): a034728.
- [118] ANDREASSI C, RICCIO A. To localize or not to localize: mRNA fate is in 3' UTR ends[J]. *Trends in Cell Biology*, 2009, 19(9): 465-474.
- [119] OTSUKA H, FUKAO A, FUNAKAMI Y, DUNCAN KE, FUJIWARA T. Emerging evidence of translational control by AU-rich element-binding proteins[J]. *Frontiers in Genetics*, 2019, 10: 332.
- [120] BROWN BD, GENTNER B, CANTORE A, COLLEONI S, AMENDOLA M, ZINGALE A, BACCARINI A, LAZZARI G, GALLI C, NALDINI L. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state[J]. *Nature Biotechnology*, 2007, 25: 1457-1467.
- [121] LYTLE JR, YARIO TA, STEITZ JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(23): 9667-9672.
- [122] FRIEDMAN RC, FARH KKH, BURGE CB, BARTEL DP. Most mammalian mRNAs are conserved targets of microRNAs[J]. *Genome Research*, 2009, 19(1): 92-105.
- [123] SANDBERG R, NEILSON JR, SARMA A, SHARP PA, BURGE CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites[J]. *Science*, 2008, 320(5883): 1643-1647.
- [124] GILLIES JK, LORIMER IAJ. Regulation of p27Kip1 by miRNA 221/222 in glioblastoma[J]. *Cell Cycle*, 2007, 6(16): 2005-2009.
- [125] LAI WJ C, ZHU MY, BELINITE M, BALLARD G, MATHEWS DH, ERMOLENKO DN. Intrinsically unstructured sequences in the mRNA 3' UTR reduce the ability of poly(A) tail to enhance translation[J]. *Journal of Molecular Biology*, 2022, 434(24): 167877.
- [126] ZHAO WX, POLLACK JL, BLAGEV DP, ZAITLEN N, McMANUS MT, ERLE DJ. Massively parallel functional annotation of 3' untranslated regions[J]. *Nature Biotechnology*, 2014, 32: 387-391.
- [127] ORLANDINI von NIESSEN AG, POLEGANOV MA, RECHNER C, PLASCHKE A, KRANZ LM, FESSER S, DIKEN M, LÖWER M, VALLAZZA B, BEISSERT T, BUKUR V, KUHN AN, TÜRECI Ö, SAHIN U. Improving mRNA-based therapeutic gene delivery by expression-augmenting 3' UTRs identified by cellular library screening[J]. *Molecular Therapy*, 2019, 27(4): 824-836.
- [128] LEGNINI I, ALLES J, KARAIKOS N, AYOUB S, RAJEWSKY N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control[J]. *Nature Methods*, 2019, 16: 879-886.
- [129] PASSMORE LA, COLLIER J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression[J]. *Nature Reviews Molecular Cell Biology*, 2022, 23: 93-106.
- [130] MOCKEY M, GONÇALVES C, DUPUY FP, LEMOINE FM, PICHON C, MIDOUX P. mRNA transfection of dendritic cells: synergistic effect of ARCA mRNA capping with poly(A) chains in cis and in trans for a high protein expression level[J]. *Biochemical and Biophysical Research Communications*, 2006, 340(4): 1062-1068.
- [131] LIMA SA, CHIPMAN LB, NICHOLSON AL, CHEN YH, YEE BA, YEO GW, COLLIER J, PASQUINELLI AE. Short poly(A) tails are a conserved feature of highly expressed genes[J]. *Nature Structural & Molecular Biology*, 2017, 24: 1057-1063.
- [132] YI H, PARK J, HA MJ, LIM J, CHANG H, KIM VN.

- PABP cooperates with the CCR4-NOT complex to promote mRNA deadenylation and block precocious decay[J]. *Molecular Cell*, 2018, 70(6): 1081-1088.e5.
- [133] WEBSTER MW, CHEN YH, STOWELL JAW, ALHUSAINI N, SWEET T, GRAVELEY BR, COLLIER J, PASSMORE LA. mRNA deadenylation is coupled to translation rates by the differential activities of Ccr4-not nucleases[J]. *Molecular Cell*, 2018, 70(6): 1089-1100.e8.
- [134] HOLTkamp S, KREITER S, SELMI A, SIMON P, KOSLOWSKI M, HUBER C, TÜRECI O, SAHIN U. Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells[J]. *Blood*, 2006, 108(13): 4009-4017.
- [135] GRIER AE, BURLEIGH S, SAHNI J, CLOUGH CA, CARDOT V, CHOE DC, KRUTEIN MC, RAWLINGS DJ, JENSEN MC, SCHARENBERG AM, JACOBY K. pEVL: a linear plasmid for generating mRNA IVT templates with extended encoded poly(A) sequences[J]. *Molecular Therapy Nucleic Acids*, 2016, 5(4): e306.
- [136] CHOI YH, HAGEDORN CH. Purifying mRNAs with a high-affinity eIF4E mutant identifies the short 3' poly(A) end phenotype[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(12): 7033-7038.
- [137] MEIJER HA, BUSHELL M, HILL K, GANT TW, WILLIS AE, JONES P, de MOOR CH. A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells[J]. *Nucleic Acids Research*, 2007, 35(19): e132.
- [138] RISSLAND OS, SUBTELNY AO, WANG M, LUGOWSKI A, NICHOLSON B, LAVER JD, SIDHU SS, SMIBERT CA, LIPSHITZ HD, BARTEL DP. The influence of microRNAs and poly(A) tail length on endogenous mRNA-protein complexes[J]. *Genome Biology*, 2017, 18(1): 211.
- [139] NICHOLSON AL, PASQUINELLI AE. Tales of detailed poly(A) tails[J]. *Trends in Cell Biology*, 2019, 29(3): 191-200.
- [140] LI CY, LIANG ZH, HU YX, ZHANG HX, SETIASABDA KD, LI JW, MA SH, XIA XJ, KUANG Y. Cytidine-containing tails robustly enhance and prolong protein production of synthetic mRNA in cell and *in vivo*[J]. *Molecular Therapy Nucleic Acids*, 2022, 30: 300-310.
- [141] TREPOTEC Z, GEIGER J, PLANK C, ANEJA MK, RUDOLPH C. Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life[J]. *RNA*, 2019, 25(4): 507-518.
- [142] Reverse Engineering the source code of the BioNTech/Pfizer SARS-CoV-2 Vaccine[EB/OL]. [2024-04-12]. <https://berthub.eu/articles/posts/reverse-engineering-source-code-of-the-biontech-pfizer-vaccine/>.
- [143] CHOI J, IEONG KW, DEMIRCI H, CHEN J, PETROV A, PRABHAKAR A, O'LEARY SE, DOMINISSINI D, RECHAVI G, SOLTIS SM, EHRENBERG M, PUGLISI JD. N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics[J]. *Nature Structural & Molecular Biology*, 2016, 23: 110-115.
- [144] EYLER DE, FRANCO MK, BATOOL Z, WU MZ, DUBUKE ML, DOBOSZ-BARTOSZEK M, JONES JD, POLIKANOV YS, ROY B, KOUTMOU KS. Pseudouridylation of mRNA coding sequences alters translation[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, 116(46): 23068-23074.

(本文责编 陈宏宇)