

# 使用高分辨 LC/MS 分析 mRNA Poly-A 序列变异体

## 作者

Brian Liou  
安捷伦科技有限公司

## 前言

2020 年 SARS-CoV-2 疫情带来的紧迫形势促使政策制定者和制药公司以前所未有的速度开发和部署 mRNA 疫苗。mRNA 疫苗在临床试验<sup>[1-4]</sup> 中表现出令人惊叹的安全性和有效性，远优于采用其他技术研发的疫苗。由于 mRNA 疫苗被认为是基因治疗产品<sup>[5]</sup>，因此 FDA 指南要求对产品相关杂质进行广泛的表征。其中可能包括序列中存在微小错误的 mRNA 分子群，称为序列变异体。此外，mRNA 疫苗要求 3' 端具有长重复 A 核苷酸序列 (poly-A)，以获得最佳的稳定性和生物活性<sup>[6]</sup>。因此，poly-A 序列的长度和组成都是关键质量属性。

本研究使用 Agilent AdvanceBio 6545XT LC/Q-TOF 分析由大肠杆菌 Poly-A 聚合酶 (PAP) 形成的 poly-A 加尾序列, PAP 是体外转录系统的常见组成部分。研究结果表明, PAP 对 ATP 并不具有完全的选择性, 在标准体外转录条件下, PAP 可以作用于 CTP 和 UTP 前体, 结合大量不需要的 C 和 U 核苷酸。由于可以将这些序列变异体视为产品相关杂质, 因此结果提醒不要使用 PAP, 并展现了 LC/MS 作为灵敏、高效的过程优化和核酸疗法质量控制方法的价值。

本研究中使用的缩写:

- ATP — 三磷酸腺苷
- CTP — 三磷酸胞苷
- UTP — 三磷酸尿苷
- GTP — 三磷酸鸟苷
- A、C、U 和 G 核苷酸 — 一磷酸腺苷、一磷酸胞苷、一磷酸尿苷和一磷酸鸟苷
- Poly-A — 多聚腺苷酸
- PAP — 大肠杆菌 Poly-A 聚合酶
- RNA-seq — RNA 测序

## 实验部分

### mRNA 的体外转录

编码 3822 nt 基因, 两侧为上游 T7 启动子和下游 BGH 终止子序列的 pCMV3 质粒购自 Sino Biological。使用 T7 和 BGH 终止子引物 (Agilent Herculase II Fusion DNA 聚合酶, 部件号 600677) 对 DNA 序列进行 35 个循环的 PCR 扩增。净化 (Agilent StrataPrep PCR 纯化试剂盒, 部件号 400771) 后, 采用 DNA 7500 试剂盒于 Agilent 2100 生物分析仪 (部件号 5067-1506) 上对扩增的 dsDNA 进行分析, 测量其浓度并评估扩增均一度。然后使用 HiScribe T7 ARCA mRNA 试剂盒 (New England Biolabs, 部件号 E2060S) 在体外转录扩增的 dsDNA (约 13 nmol/L), 并按照制造商推荐的方案使用随附的 PAP 酶对其进行加尾, 然后用 LiCl 进行沉淀。采用 RNA 6000 Nano 试剂盒于 2100 生物分析仪 (部件号 5064-1511) 上分析 PAP 加尾前后的转录 mRNA 等分试样, 监测其反应。

在 PAP 选择性研究中, 使用 PAP 酶于 37 °C 下扩展具有 5' 和 3'-OH 的合成 10 mer poly-A 序列 (Integrated DNA Technologies) 30 分钟, 每次反应仅使用一种前体核苷三磷酸 (1 mmol/L 的 ATP、CTP、UTP 或 GTP), 如图 1A 所示。

### 样品前处理

在 37 °C 下, 使用 1000 U RNase T1 将约 20 pmol poly-A 加尾 mRNA 酶解 3 小时, 以释放 poly-A 序列。使用 200 μL oligo-dT 磁珠对每个样品进行五轮纯化, 得到 poly-A 序列<sup>[7]</sup>。在 50 μL 1x IDTE 缓冲液 (Integrated DNA Technologies, 部件号 11-05-01-05) 中洗脱每个 pull-down 序列, 并将其合并为最终体积为 250 μL 的溶液。在进行 LC/MS 分析之前, 使用具有 10 kDa MWCO 的 Vivaspin 500 柱 (Sartorius, 部件号 VS0102) 将合并的洗脱液脱盐到 60 μL 的去离子水中。

### Poly-A 加尾实验

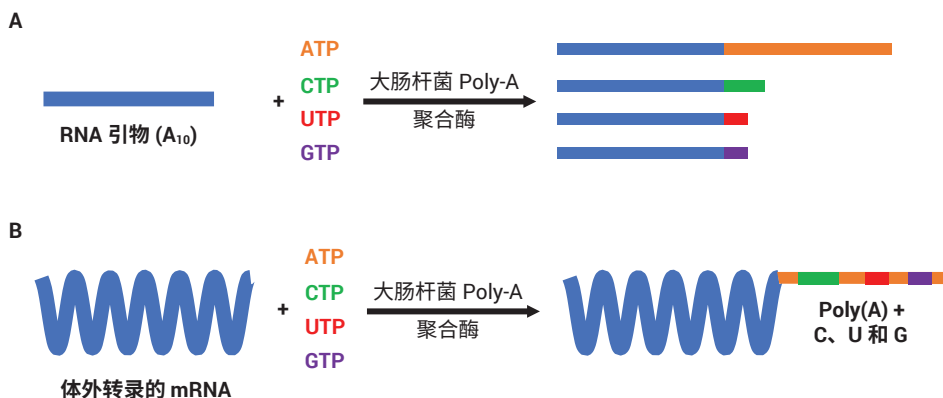


图 1. 本应用简报中进行的加尾反应示意图。(A) 对 RNA 引物进行的反应, 每次反应仅使用一种前体。(B) 在标准条件下 (所有前体) 对体外转录的 mRNA 进行的反应

## Poly-A 序列的 LC-DAD/MS 分析

仪器包括:

- 配备二极管阵列检测器的 1290 Infinity II 液相色谱仪 (部件号 G7117B)
- 6545XT AdvanceBio LC/Q-TOF

注意去除流路中的玻璃, 减少碱金属加合效应。将安捷伦 Nalgene 瓶 (部件号 9301-6460) 用作流动相容器, 每个溶剂管线配备一个不锈钢滤芯。使用安捷伦聚丙烯样品瓶 (部件号 5190-2242)。首次使用前, 用 50% MeOH + 0.1% 甲酸将液相色谱系统和色谱柱冲洗过夜, 进一步减少碱金属加合物。如果需要, 在两次实验之间用 50% MeOH + 0.1% 甲酸冲洗 30 分钟即可确保系统清洁<sup>[6]</sup>。

在 PLRP-S 色谱柱 (2.1 × 50 mm, 5 μm, 1000 Å, 部件号 PL1912-1502) 上分离 poly-A 序列。为了获得更高的色谱分离度, 在 PAP 选择性实验中使用 Infinity Poroshell 120 HPH-C18 色谱柱 (2.1 × 50 mm, 1.9 μm, 120 Å, 部件号 699675-702)。流动相和液相色谱梯度如表 1 所示。在负离子模式下运行质谱仪 (设置如表 2 所示), 在 BioConfirm 10.0 (解卷积设置如表 3 所示) 中进行数据分析。

表 1. 流动相和液相色谱梯度

Agilent 1290 Infinity II 液相色谱系统		
色谱柱	InfinityLab Poroshell 120 HPH-C18, 1.9 μm, 2.1 × 50 mm, 120 Å	Agilent PLRP-S, 5 μm, 2.1 × 50 mm, 1000 Å
溶剂 A	15 mmol/L 二丁胺 + 25 mmol/L HFIP 的去离子水溶液	
溶剂 B	15 mmol/L 二丁胺 + 25 mmol/L HFIP 的甲醇溶液	
梯度	0–2 min, 15% B 12 min, 30% B 12.1–13 min, 90% B	0–1 min, 15% B 10.5 min, 45% B 10.6–11.5 min, 90% B
柱温	50 °C	80 °C
流速	0.4 mL/min	
进样量	10–20 μL	

表 2. 质谱仪设置

Agilent 6545XT AdvanceBio LC/Q-TOF		
	LC/MS	LC/MS/MS
采集模式	负离子, 标准 (3200 m/z) 质量数范围, 高灵敏度 (2 Ghz)	
气体温度	350 °C	
气体流速	12 L/min	
雾化器	55 psig	
鞘气温度	275 °C	
鞘气流速	10 L/min	
毛细管电压	4500 V	
喷嘴电压	2000 V	
碎裂电压	250 V	
锥孔电压	65 V	
MS1 范围	400–3200 m/z	
MS1 扫描速率	2 Hz	5 Hz
MS2 范围	N/A	50–3200 m/z
MS2 扫描速率		3 Hz
MS2 分离峰宽		中等 (约 4 amu)
碰撞能量		0、40 和 60 V
MS2 阈值		开启; 重复 3 次, 然后排除 0.2 min
基于母离子丰度的扫描速率		是
目标物 (响应值/质谱图)		25000
使用 MS2 累积时间限		是
纯度		严格性 100%, 截留率 30%
母离子排序		仅按丰度; +3、+2、+1
参比质量	1033.9881	

表 3. 解卷积设置

Agilent MassHunter BioConfirm B10.0 设置		
寡核苷酸长度	≤ 30 nt	≥ 90 nt
解卷积算法	最大熵	
扣除基线	1	
加合物	质子丢失	
质量数范围	3000–10000 Da	30000–60000 Da
质谱精度	0.05 Da	0.05 Da
使用有限的 m/z 范围	1040–3200	800–2500

## 结果与讨论

第一个测试分析了在 1 mmol/L ATP 的条件下, PAP 在由 10 个重复的 A 核苷酸 ( $A_{10}$ ) 组成的合成 RNA 引物上扩展的 poly-A 序列。如图 2 所示, 这导致了 poly-A 序列的双峰分布, 其中一个峰群由

洗脱时间为 2.5–6 分钟的较短寡核苷酸组成, 而另一个峰群由约 10.6 分钟处以宽峰洗脱的较长核苷酸组成。质谱分析结果表明, 较短的寡核苷酸群长度范围从 11 到 22 nt (图 3D), 而较长的核苷酸群长度范围从 108 到 149 nt (35492.89 到 48990.17 Da, 图 4C)。

图 3 显示了从较短的寡核苷酸群中选择的三个峰的提取质谱图和解卷积质谱图。质谱图主要由质子丢失产生的双电荷和三电荷离子以及少量的钠加合物组成。根据各单一同位素峰, 对同位素解析的解卷积质谱图分配了  $A_{20}$ 、 $A_{21}$  和  $A_{22}$  标识 (图 3B–3D), 误差小于 5 ppm。

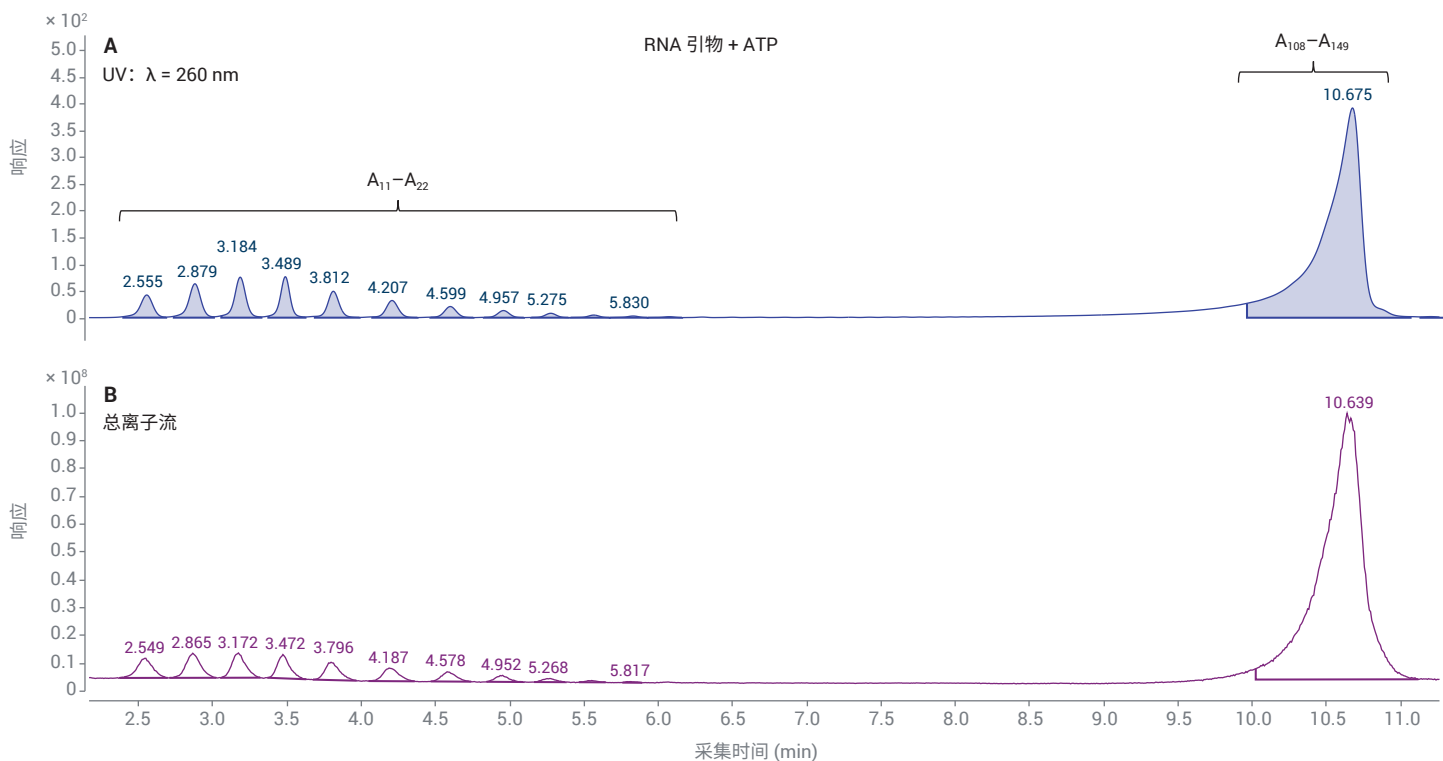
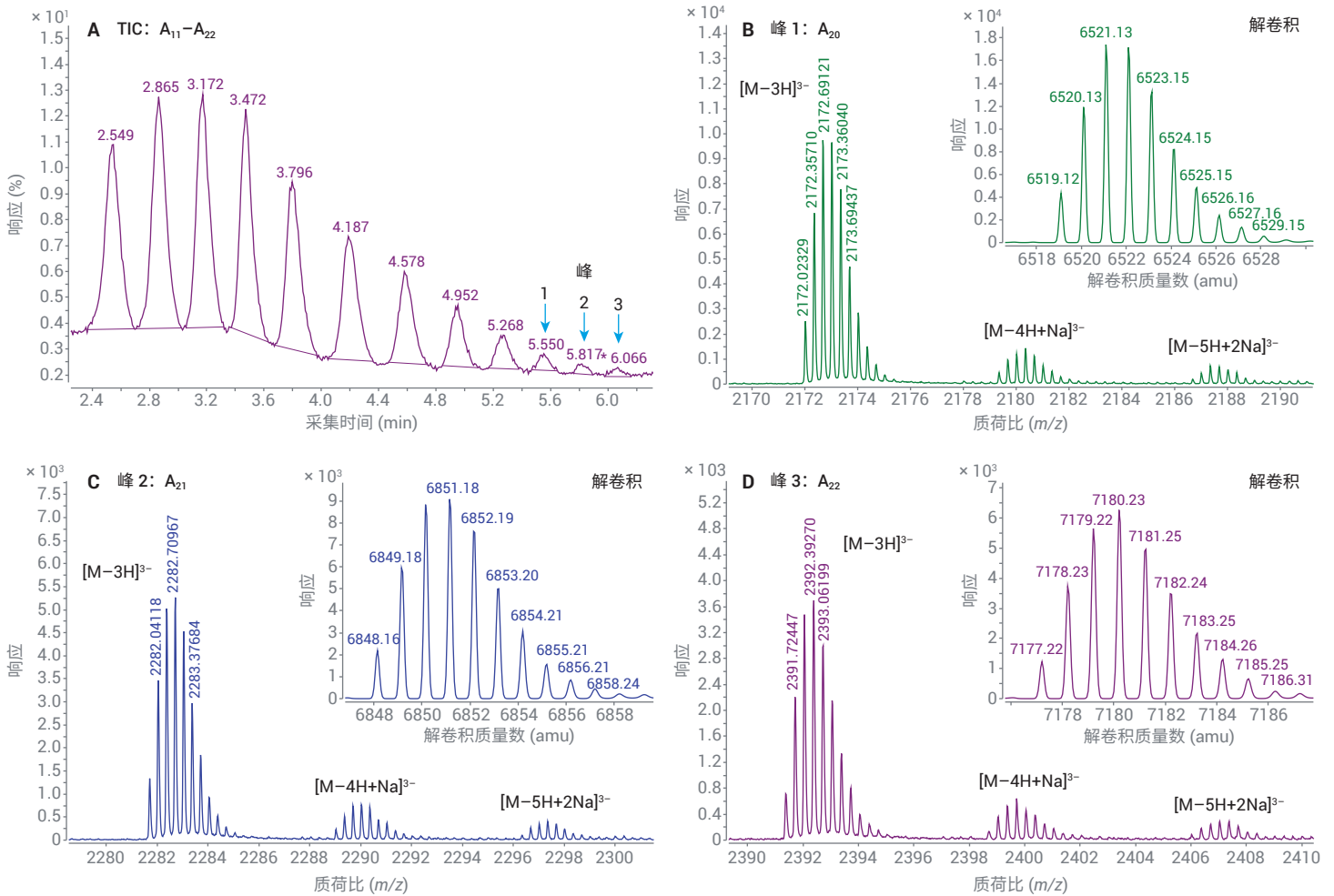


图 2. 在仅存在 ATP 的情况下, 用 PAP 延伸的 RNA 引物在 260 nm 处的 UV 吸光度 (A: 参比波长 = 360 nm) 和总离子流色谱图 (B)。在 PLRP-S 色谱柱上的分离



**图 3.** PAP 形成的 A<sub>11</sub>-A<sub>22</sub> 寡核苷酸。(A) 显示三个选定峰的总离子流色谱图，提取质谱图如 (B-D) 所示。A<sub>20</sub> ( $M_{\text{obs}} = 6519.12$  Da,  $M_{\text{theo}} = 6519.09$  Da)、A<sub>21</sub> ( $M_{\text{obs}} = 6848.16$  Da,  $M_{\text{theo}} = 6848.15$  Da) 和 A<sub>22</sub> ( $M_{\text{obs}} = 7177.22$  Da,  $M_{\text{theo}} = 7177.20$  Da) 的解卷积质谱图以内插图的形式显示。 $M_{\text{obs}}$ : 实测单一同位素质量数;  $M_{\text{theo}}$ : 理论单一同位素质量数

取一部分较长的寡核苷酸进行解卷积 (图 4A)。10-10.3 分钟范围内的提取质谱图中的电荷分布包迹主要落在 800-2500  $m/z$  之间 (图 4B)，然后将其解卷积至 30-60 kDa 的目标质量数范围。解卷积质谱图 (图 4C) 清楚显示了 34-50 kDa 范围内的样品峰的异质群体，它们的平均间隔为  $329.2 \pm 1$  Da (图 4D)。这些质量数增量与 A 核苷酸的单次添加一致，

使理论平均质量数增加了 329.209 Da。如表 4 所示，图 4D 中的质量峰被可靠标注为 A<sub>121</sub>-A<sub>138</sub>，理论和实测质量数相差  $\leq 1.16$  Da。

为了评估 PAP 对 ATP 的选择性，进行了重复实验，其中，仅在存在 1 mmol/L CTP、UTP 或 GTP 的情况下将 PAP 添加到 RNA 引物中。虽然未观察到长聚合物链的延伸，但色谱分离向 RNA 引物中添

加最多两个 C 核苷酸单体 (图 5A) 或一个 U 核苷酸 (图 5B)，表明 PAP 对 ATP 并不具有完全的选择性。本实验未观察到一磷酸鸟苷的添加 (图 5C)，但不能排除在更长的反应时间或更高的 GTP 浓度下大量添加的可能性。总体而言，PAP 在 ATP 中显示出最高的活性，然后依次是 CTP、UTP 和 GTP。

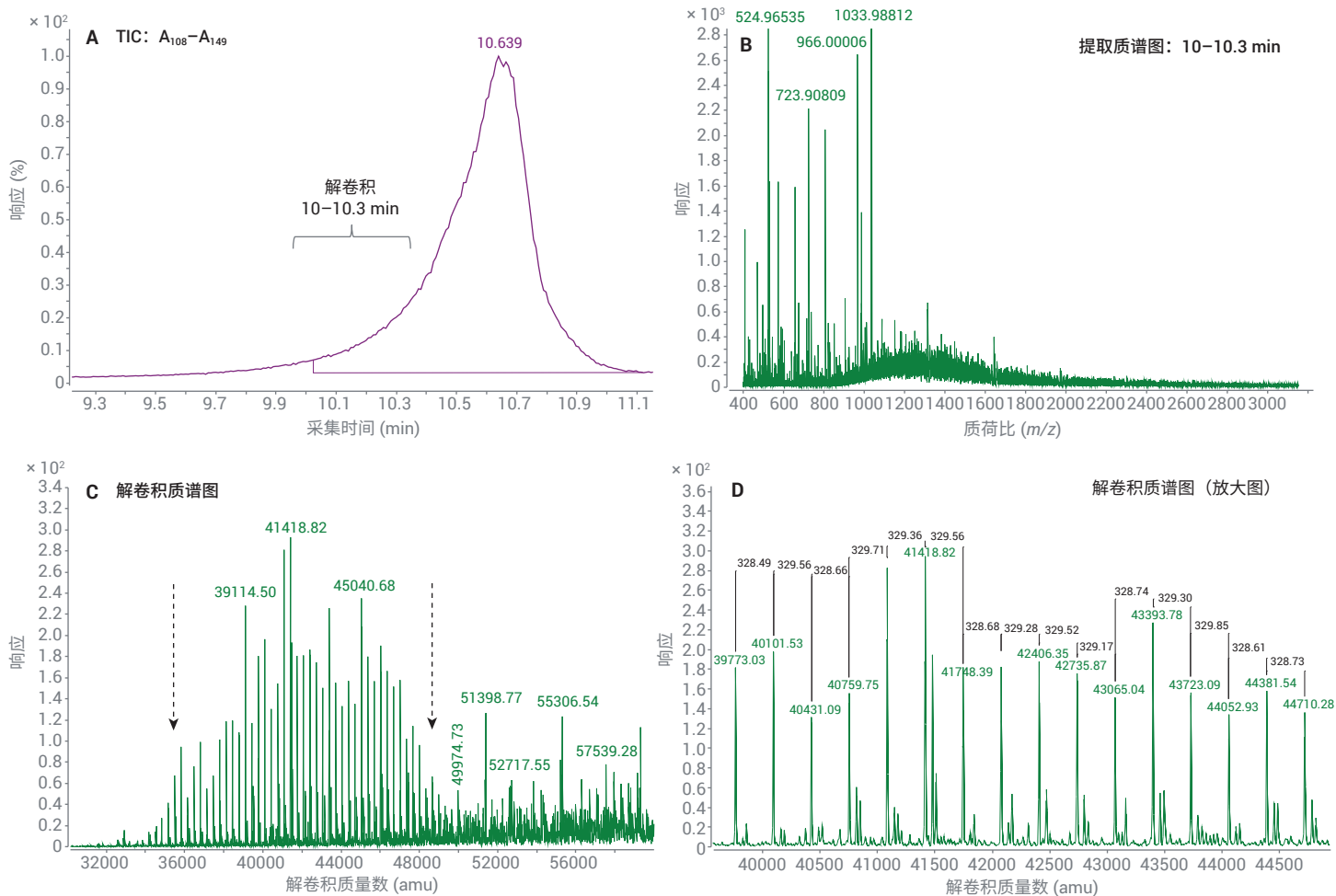


图 4. PAP 形成的 A<sub>108</sub>-A<sub>149</sub> 寡核苷酸。(A) 总离子流色谱图, 显示 10-10.3 分钟的解卷积采样区域。(B) 电荷分布包迹和 (C) 采样区域的解卷积质谱图。虚线箭头 (左=35492.89 Da, 右=48990.17 Da) 表示可以可靠分配 A<sub>108</sub>-A<sub>149</sub> 标识的质量峰范围。(D) 解卷积质谱图的放大图, 显示峰之间的规律性间隔为 39773.03-44710.28 Da

表 4. 图 4D 中标注的质量峰

寡核苷酸	实测质量数 (Da)	理论质量数 (Da)	质量数差异 (Da)
A <sub>121</sub>	39773.03	39772.08	0.95
A <sub>122</sub>	40101.53	40101.28	0.25
A <sub>123</sub>	40431.09	40430.49	0.6
A <sub>124</sub>	40759.75	40759.70	0.05
A <sub>125</sub>	41089.46	41088.90	0.56
A <sub>126</sub>	41418.82	41418.11	0.71
A <sub>127</sub>	41748.39	41747.32	1.07
A <sub>128</sub>	42077.06	42076.52	0.54
A <sub>129</sub>	42406.35	42405.73	0.62
A <sub>130</sub>	42735.87	42734.94	0.93
A <sub>131</sub>	43065.04	43064.15	0.89
A <sub>132</sub>	43393.78	43393.35	0.43
A <sub>133</sub>	43723.09	43722.56	0.53
A <sub>134</sub>	44052.93	44051.77	1.16
A <sub>135</sub>	44381.54	44380.97	0.57
A <sub>136</sub>	44710.28	44710.18	0.1

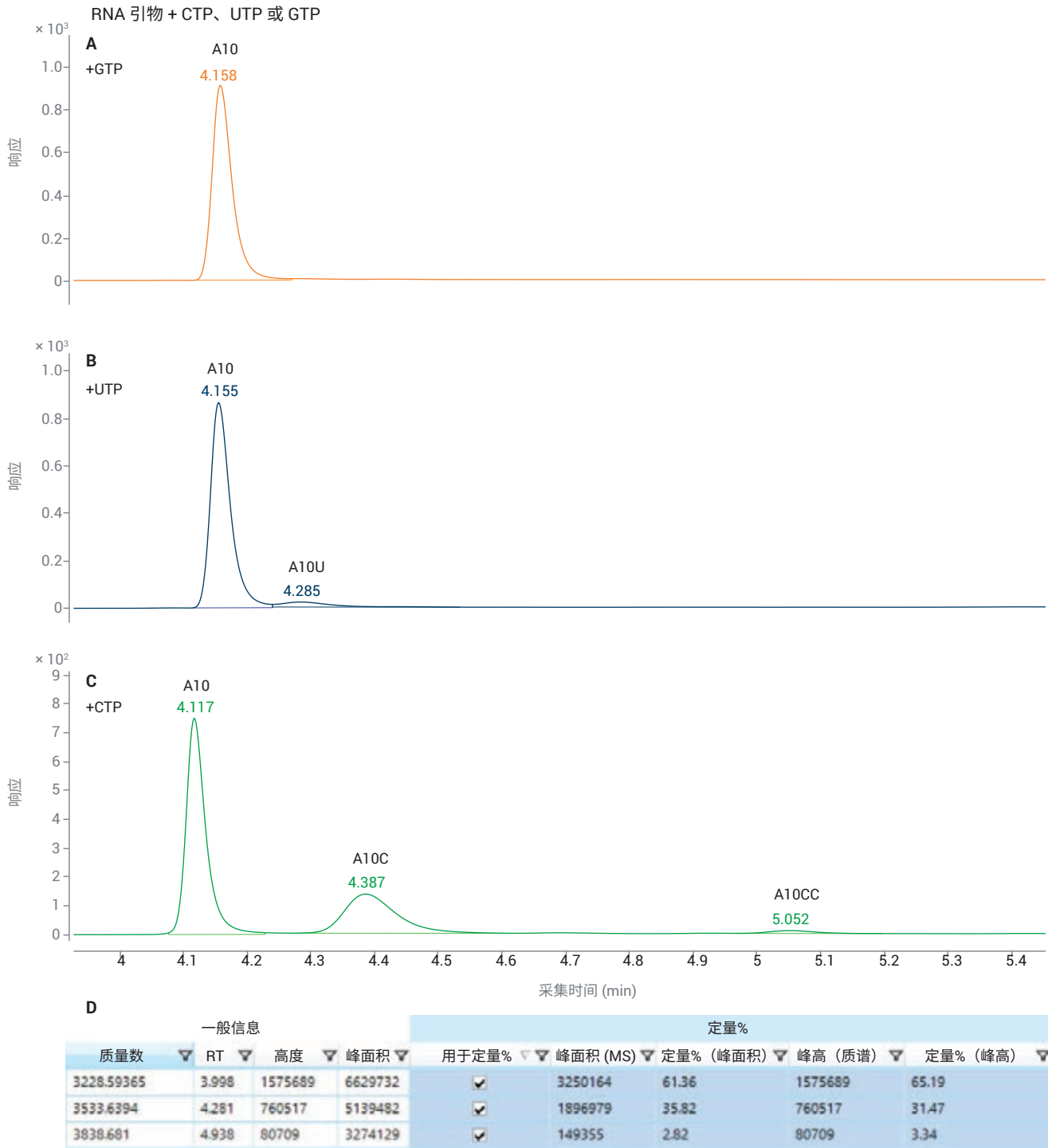


图 5. 紫外吸收 (吸收波长 = 260 nm, 参比波长 = 360 nm) 色谱图, 显示了 PAP 对 (A) GTP、(B) UTP 和 (C) CTP 的不同活性。未检测到一磷酸鸟苷的添加。(D) 图 (C) 中所示的 A<sub>10</sub>、A<sub>10</sub>C 和 A<sub>10</sub>CC 的相对定量结果。在 Agilent Poroshell 120 HPH-C18 色谱柱上的分离

未修饰的 RNA 引物及其以 C 或 U 核苷酸延伸的序列的解卷积质谱图如图 6 所示。根据各单一同位素峰，对同位素解析的解卷积质谱图分配了  $A_{10}$ 、 $A_{10}C$ 、 $A_{10}CC$  和  $A_{10}U$  标识，误差小于 13 ppm。MS/MS 实验表明，在 RNA 引物的 3' 端

添加了 C 和 U 核苷酸 (图 7)，从而形成了特征双电荷  $y$  离子 1601.271  $m/z$  和 1601.758  $m/z$ 。相比之下，未修饰的 RNA 引物以 3' A 核苷酸终止，在碎裂时产生双电荷  $y$  离子 1448.749  $m/z$ 。

然后，在配备 RNA 6000 Nano 试剂盒的生物分析仪上通过 LC/MS 分析体外转录的全长 mRNA。在加尾之前，转录的 mRNA 显示预期长度约为 3800 nt，与 PAP 反应后长度增加到约 4200 nt (图 8A)，表明成功完成 poly-A 加尾。

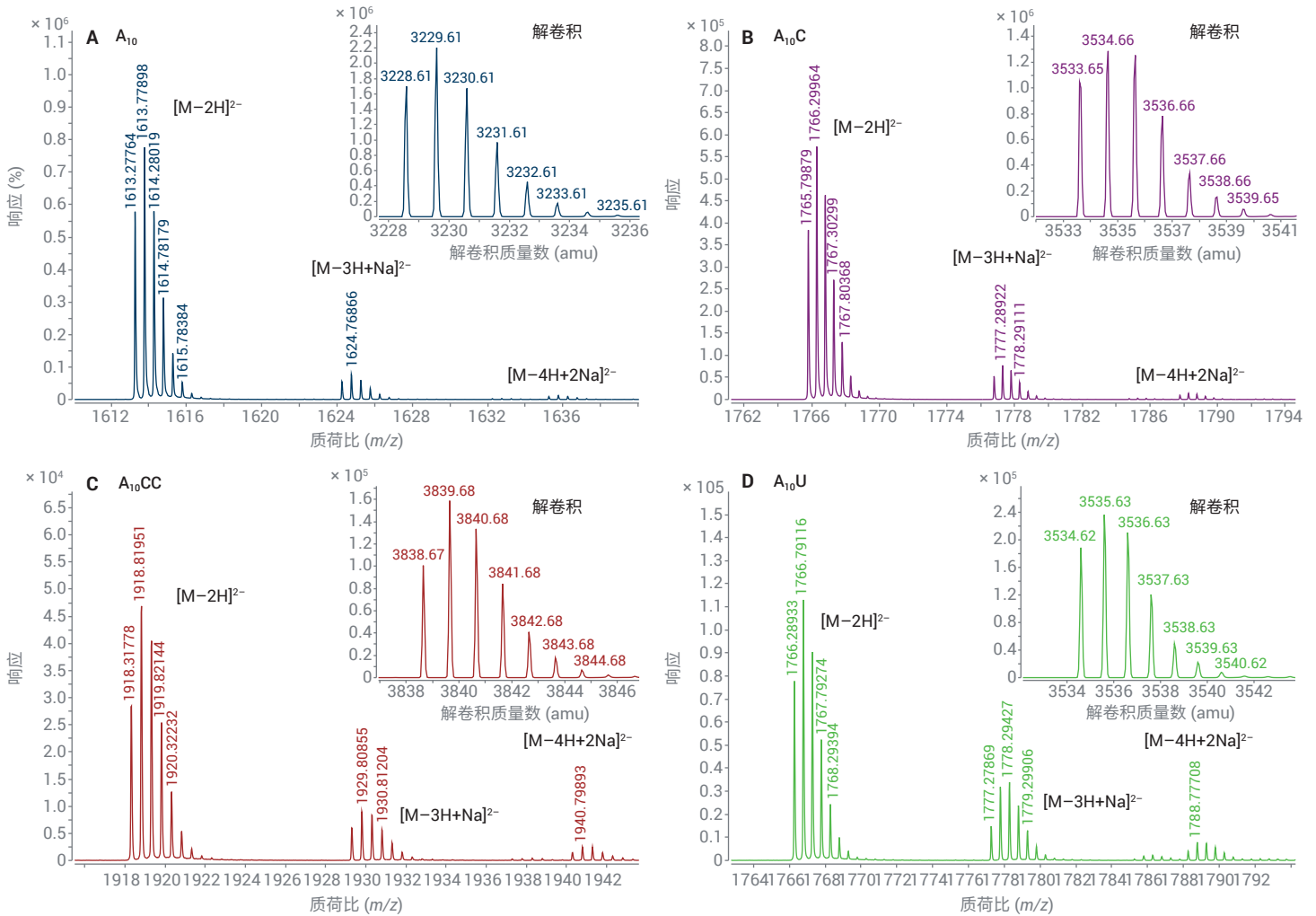


图 6. (A) 未修饰的  $A_{10}$  RNA 引物 ( $M_{obs} = 3228.61$  Da,  $M_{theo} = 3228.57$  Da)，(B) 延伸一个 C 核苷酸 ( $M_{obs} = 3533.65$  Da,  $M_{theo} = 3533.61$  Da)，(C) 延伸两个 C 核苷酸 ( $M_{obs} = 3838.67$  Da,  $M_{theo} = 3838.65$  Da)，(D) 延伸两个 U 核苷酸 ( $M_{obs} = 3534.62$  Da,  $M_{theo} = 3534.59$  Da) 的提取质谱图和解卷积质谱图。 $M_{obs}$ : 实测单一同位素质量数;  $M_{theo}$ : 理论单一同位素质量数



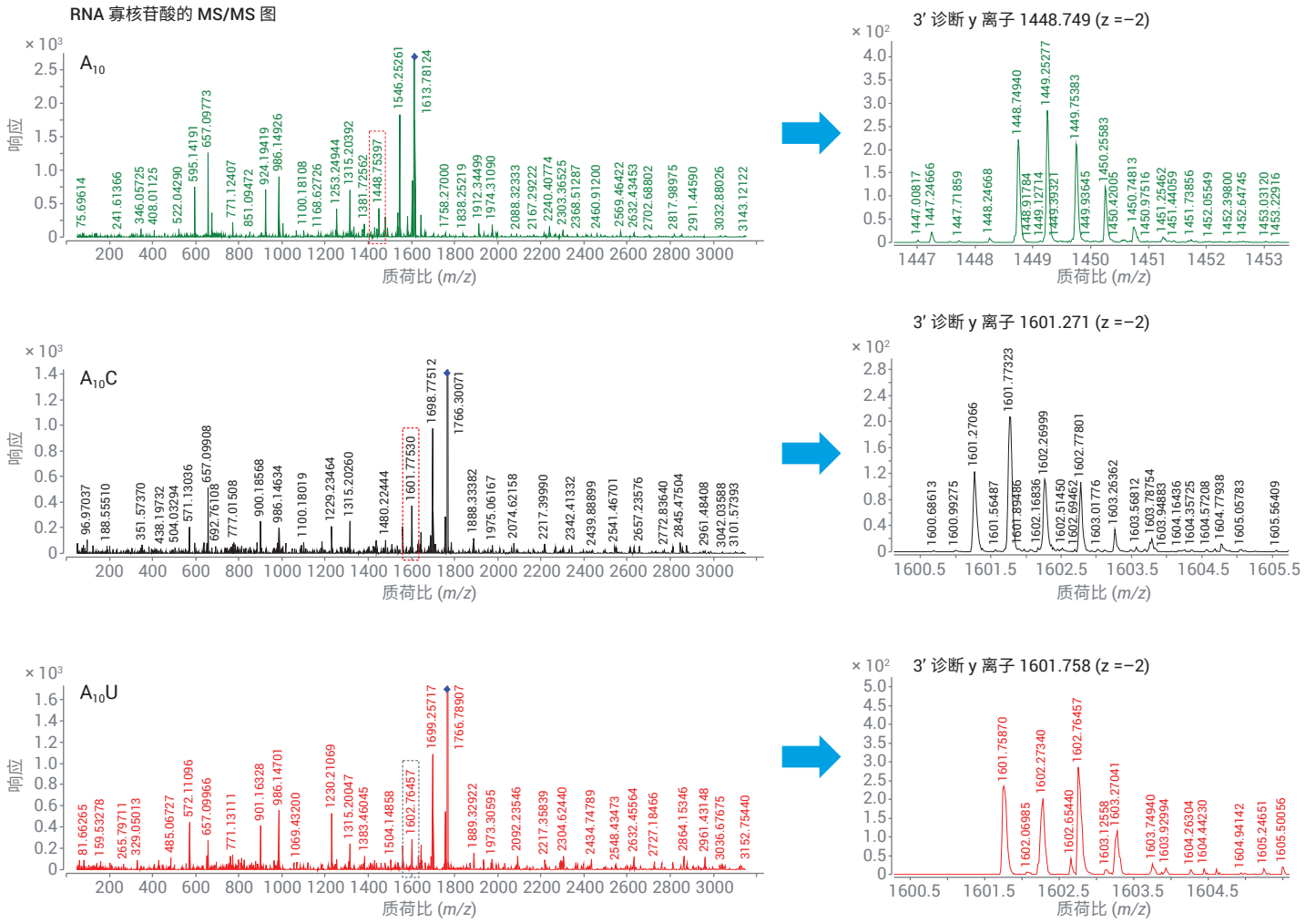


图 7. 所选寡核苷酸的 MS/MS 图，显示其不同 3' 端的诊断离子特征

用 RNase T1 酶解全长 mRNA 样品，然后用 oligo-dT 磁珠反复 pull-down 生成纯化的加尾序列。与 PAP 延伸的 RNA 引物一样，来自体外转录 mRNA 的加尾序列由洗脱时间为 3.7–7.5 分钟的较短寡核苷酸群以及在约 10.6 分钟处洗脱的较长核苷酸群组成（图 8B）。较短寡核苷酸群

中所选峰的提取质谱图和解卷积质谱图显示 poly-A 序列的长度为 16–27 nt，每个序列都包含一个错误结合的 U 核苷酸（图 9）。虽然在此数据集中未观察到这一现象，但在其他实验中也观察到错误结合的 C 核苷酸。

正如 M. Beverly 等<sup>[7]</sup>所指出的那样，与基因模板化 poly-A 序列相比，PAP 形成的加尾序列在长度方面的异质性更高。结果表明，当在四种前体核苷三磷酸均存在的标准条件下发生加尾反应时，错误结合不同数量的 C 和 U 核苷酸会加剧这种异质性，使得较长加尾序列质谱图的解卷积非常具有挑战性。

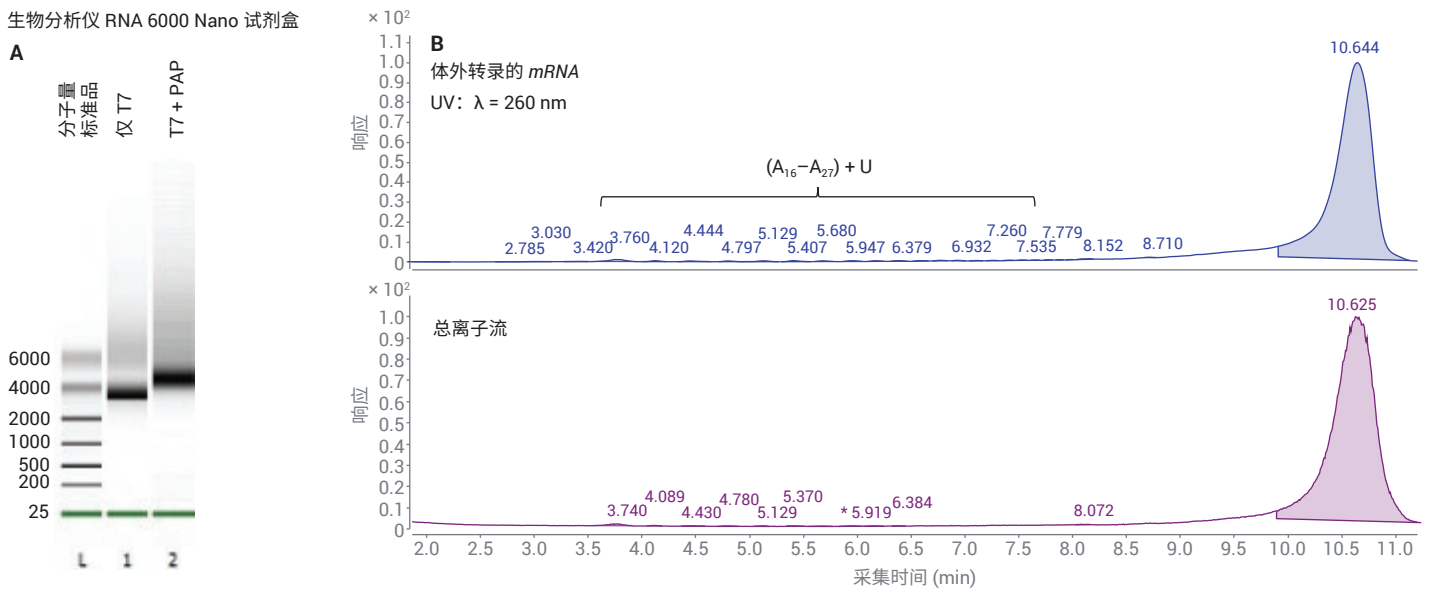


图 8. (A) 用 PAP 加尾之前（第 1 道）和加尾之后（第 2 道）的体外转录 mRNA 的生物分析仪分析。(B) 在四种核苷酸磷酸盐前体均存在的条件下，添加到体外转录 mRNA 的 poly-A 序列在 260 nm 处的 UV 吸光度（上图）和总离子流色谱图（下图）在 PLRP-S 色谱柱上的分离

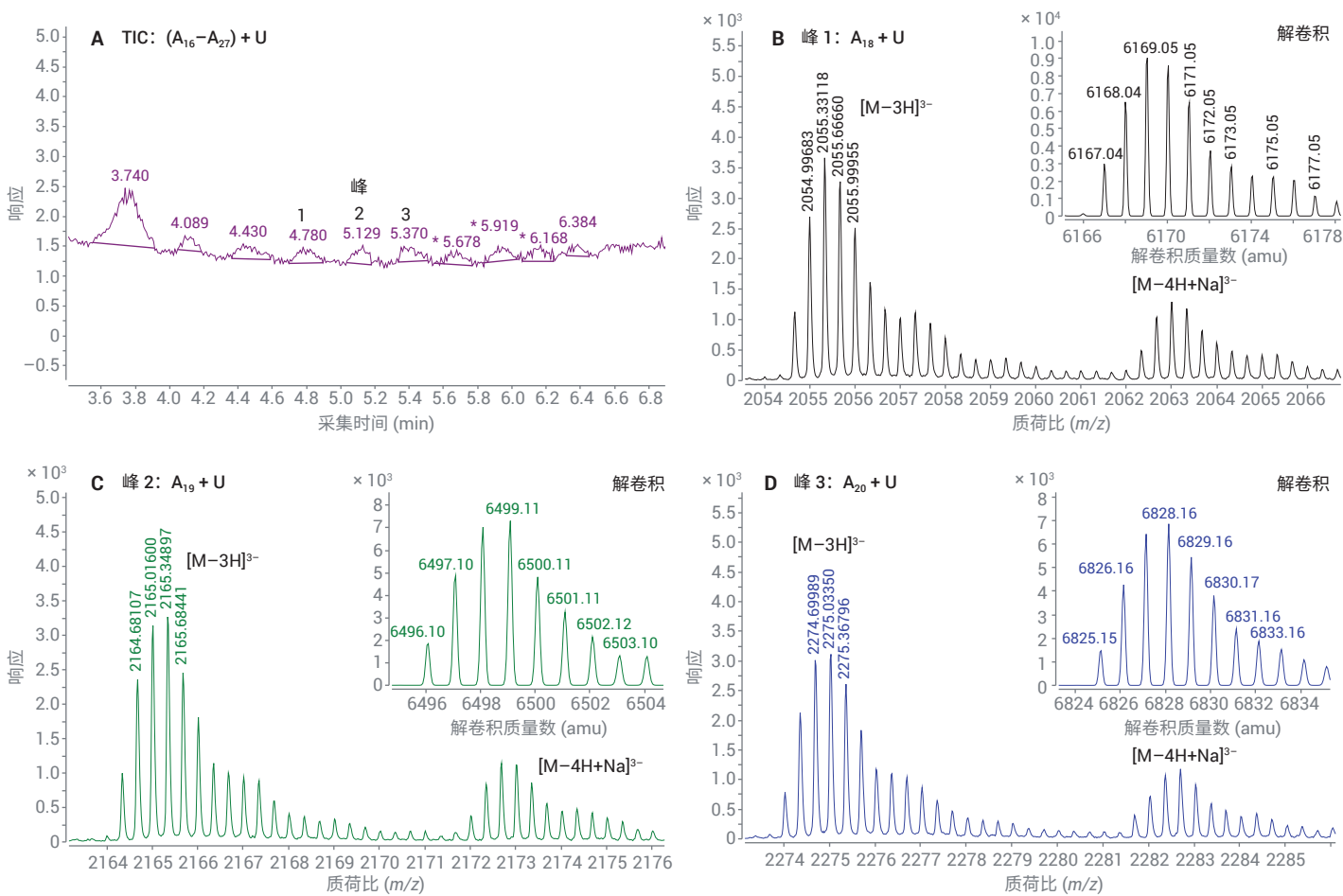


图 9. 包含错误结合的 U 核苷酸的 A<sub>16</sub>-A<sub>27</sub> 寡核苷酸。(A) 显示三个选定峰的总离子流色谱图，提取质谱图如 (B-D) 所示。A<sub>18</sub> + U (M<sub>obs</sub> = 6167.04 Da, M<sub>theo</sub> = 6167.01 Da)、A<sub>19</sub> + U (M<sub>obs</sub> = 6496.10 Da, M<sub>theo</sub> = 6496.07 Da) 和 A<sub>20</sub> + U (M<sub>obs</sub> = 6825.15 Da, M<sub>theo</sub> = 6825.12 Da) 的解卷积质谱图以内插图的形式显示。M<sub>obs</sub>: 实测单一同位素质量数; M<sub>theo</sub>: 理论单一同位素质量数

## 结论

本研究表明: (1) 对长 (121-136 nt) 异质 poly-A 序列的完整质谱图进行解卷积可以准确测量其完整质量数, (2) 在标准体外转录条件下, PAP 对 ATP 并不具有完全的选择性, 导致 C 和 U 核苷酸被添加到 poly-A 加尾序列。

尽管这些序列变体对于体外研究而言可能无关紧要, 但从法规角度而言, 它们具有非常重要的意义。值得注意的是, 其他体外转录酶 (例如 T7 聚合酶) 也可能通过滑脱或转录阻滞等机制产生序列变体<sup>[9]</sup>, 从而突出了对检测这些杂质的高灵敏度和选择性方法的需求。

为了获得这样的灵敏度和选择性, 先前的一项研究使用放射性标记核苷酸证明了 PAP 的脱靶活性<sup>[10]</sup>。这种技术可能造成危害, 不适用于生产环境。LC/MS 无需使用此类试剂即可实现对单核苷酸的选择性。此外, LC/MS 可以检测和定量序列变体, 无需进行 RNA 测序特有的冗长的逆转录、连接和扩增步骤, 而这些步骤会引入偏差和干扰。

## 参考文献

1. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N. Engl. J. Med.* **2020**, *383*, 1920–1931
2. Mulligan, M. J. *et al.* Phase I/II Study of COVID-19 RNA Vaccine BNT162b1 in Adults. *Nature* **2020**, *586*, 589–593
3. Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study
4. Promising Interim Results from Clinical Trial of NIH-Moderna COVID-19 Vaccine.
5. Wadhwa, A. *et al.* Opportunities and Challenges in the Delivery of mRNA-Based Vaccines. *Pharmaceutics* **2020**, *12*, 102
6. Kahvejian, A. *et al.* Mammalian Poly (A)-Binding Protein Is a Eukaryotic Translation Initiation Factor, Which Acts via Multiple Mechanisms. *Genes Dev.* **2005**, *19*, 104-113
7. Beverly, M.; Hagen, C.; Slack, O. Poly A Tail Length Analysis of *In Vitro* Transcribed mRNA by LC/MS. *Anal. Bioanal. Chem.* **2018**, *410*, 1667-1677
8. Birdsall, R. E. *et al.* Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis. *Rapid Commun. Mass Spectrom.* **2016**, *30*, 1667-1679
9. Tateishi-Karimata, H.; Isono, N.; Sugimoto, N. New Insights into Transcription Fidelity: Thermal Stability of Non-Canonical Structures in Template DNA Regulates Transcriptional Arrest, Pause, and Slippage. *PLOS ONE* **2014**, *9*
10. Yehudai-Resheff, S.; Schuster, G. Characterization of the E. Coli Poly(A) Polymerase: Nucleotide Specificity, RNA-Binding Affinities and RNA Structure Dependence. *Nucleic Acids Res.* **2000**, *28*, 1139-1144

查找当地的安捷伦客户中心：

[www.agilent.com/chem/contactus-cn](http://www.agilent.com/chem/contactus-cn)

免费专线：

800-820-3278, 400-820-3278 (手机用户)

联系我们：

[LSCA-China\\_800@agilent.com](mailto:LSCA-China_800@agilent.com)

在线询价：

[www.agilent.com/chem/erfq-cn](http://www.agilent.com/chem/erfq-cn)

[www.agilent.com](http://www.agilent.com)

仅限研究使用。不可用于诊断目的。

RA44231.5456365741

本文中的信息、说明和指标如有变更，恕不另行通知。

© 安捷伦科技（中国）有限公司，2021  
2021年3月8日，中国出版  
5994-3005ZHCN