

使用离子对液相色谱和质谱法进行合成 mRNA 寡核苷酸映射

Maissa M. Gaye, Jonathan Fox, Johannes P.C. Vissers, Ian Reah, Chris Knowles, Matthew A. Lauber

Waters Corporation

摘要

信使RNA已迅速成为人类医学的重要方式，目前应用于癌症治疗评估和FDA批准的COVID-19 mRNA疫苗。随着分析方法的进步，mRNA疫苗和其他类别的mRNA疗法也得到了快速发展。这类方法的重点之一在于，它是通过液质联用法(LC-MS)进行序列映射，来鉴定治疗性mRNA并确认纯度和修饰。与模板化RNA测序相比，使用质谱法进行RNA测序的优势在于，这种方法能够直接检测片段分子，可用于定位核苷杂质并鉴定重要的结构属性（5'端帽和poly-A尾）。因此，我们提出了一种全面的LC-UV-MS方法，通过自下而上的工作流程来表征mRNA，该工作流程可产生源自精确质量数匹配的mRNA组分注释色谱图。

优势

- 将离子对反相色谱与ACQUITY™ Premier BEH™ C₁₈ 300 Å色谱柱结合使用，可实现高色谱分离度和MS灵敏度
- 在计算机模拟mRNA酶解计算和waters_connect/UNIFI™科学数据库应用程序的支持下，实现基于精确质量数匹配的自动化mRNA酶解注释

简介

新冠疫情(SARS-CoV-2)促进了核酸药物的快速发展，特别是合成mRNA¹。如今，距离1961年Brenner等人发现这种药物已经过去了40年² [i]，mRNA已经发展成为一种具有巨大潜力的重要方式，分别在2021年8月和2022年1月用于癌症治疗的人体临床试验¹⁻³以及获得美国食品药品监督管理局全面批准的两种COVID-19 mRNA疫苗。随着分析方法的进步，mRNA疫苗和其他类别的mRNA疗法也得到了快速发展。这类方法的重点之一在于，它是通过液相色谱串联质谱(LC-MS)进行寡核苷酸映射并测序，来鉴定治疗性mRNA并确认纯度和修饰。Sanger和下一代测序(NGS)等核酸测序技术为药物开发人员提供了有价值的信息。不过，也可以通过将LC与基于串联质谱(LC-MS/MS)⁴或MS^E（高低碰撞能量交替扫描）⁵的碎裂结合使用来提高分析水平。与自下而上的蛋白质组学方法类似，基于LC-MS/MS或MS^E的测序方法具有直接检测RNA片段分子的优势，包括核苷杂质和重要结构属性的检测和定位，如脂质化核酸碱基⁶、封端残基和polyA尾部修饰^{4,7}。与存在大量数据处理解决方案的自下而上蛋白质组学工作流程不同，RNA映射的选择有限。我们提出了一种基于自下而上方法的寡核苷酸映射工作流程，用于在包括LC、UV检测和MS测量的单一平台内表征给定的合成mRNA。使用内部开发的、可免费获得的计算机模拟酶解物库计算器“mRNACalcondemand”与waters_connect™结合处理酶解组分，以产生带注释的色谱图。此处展示了使用RNase T1酶解的荧光素酶mRNA进行mRNA序列映射的分析方法。

实验

样品信息

本研究采用大约90 µg合成Cypridina荧光素酶mRNA（未加盖且未用polyA尾修饰）；由Bijoyita Roy（纽英伦生物技术公司，美国马萨诸塞州伊普斯维奇）赠送。使用3'-鸟苷特异性核糖核酸酶RNaseT1（Worthington Biochemical Corporation，美国新泽西州莱克伍德）进行酶解。请注意，使用10 µg萤火虫荧光素酶mRNA（CleanCap® FLuc mRNA，TriLink Biotechnologies，加州圣地亚哥）（非翻译序列为专有序列）重复此工作流程获得了相当的结果。酶解前，使用在无核酸酶缓冲液（10 mM Tris，0.1 mM EDTA，水溶液，pH 7.5，Integrated DNA Technologies, Inc，爱荷华州克拉尔维尔）中制备的20 µL尿素(8 M)，在80 °C下将荧光素酶mRNA处理5分钟使其变性。接下来，将重悬于无核酸酶缓冲液中的24 µg (~10kU) RNase T1（Worthington Biochemical Corporation，美国新泽西州莱克伍德）添加到室温下的变性mRNA中，然后将混合物在37 °C下孵育30分钟。在孵育期结束时加入无核酸酶缓冲液(40 µL)，使总样品体积达到80 µL。将最后的等分试样转移到300 µL聚丙烯自动进样器样品瓶中(P/N: 186002639 <<https://www.waters.com/nextgen/global/shop/vials-containers--collection-plates/186002639-polypropylene-12-x-32-mm-screw-neck-vial-with-cap-and-preslit-pt.html>>)。在使用BioAccord™ RDa™检测器在负离子模式下进行MS检测之前，无需任何进一步操作即可对所得酶解物进行离子对反相色谱分析(IP-RPLC)。

液相色谱条件

液相色谱系统:	ACQUITY UPLC™ Premier BSM系统 (作为BioAccord系统的一部分)
检测器:	ACQUITY UPLC TUV检测器
波长:	260 nm
色谱柱:	ACQUITY Premier BEH C ₁₈ 肽分析专用柱, 2.1 X 150 mm, 300 Å, 1.7 μm (P/N: 186009495)
柱温:	70 °C
样品温度:	4 °C
进样:	5 μL
流速:	0.4 mL/min
流动相A	含0.1% N,N-二异丙基乙胺(DIPEA) (作为IP试剂) 和1% 1,1,1,3,3,3-六氟异丙醇(HFIP)的去离子水
流动相B	含0.0375% DIPEA和0.075% HFIP的65:35乙腈/水

梯度表

时间(min)	mL/min	A (%)	B (%)	曲线
初始	0.4	97	3	*
60	0.4	70	30	6
60.5	0.4	5	95	6
61	0.4	97	3	6
70	0.4	97	3	6

运行时间 = 70 分钟

质谱条件

质谱系统:	BioAccord LC-MS系统
检测器:	ACQUITY RDa检测器
模式:	碎裂模式下的全扫描
极性:	负
锥孔电压:	40 V
碎裂锥孔电压:	80-200 V
质量范围:	高(400~5000 m/z)
扫描速率:	2 Hz
毛细管电压:	0.80 kV

脱溶剂气温度：

400 °C

结果与讨论

使用C₁₈固定相的离子对反相色谱法(IP-RPLC)已成为一种经过验证的寡核苷酸分析方法^{4,7}。流动相含有一种离子对试剂，通常是烷基胺，它吸附在C₁₈固定相上^{8,9,10}，从而引入了一种类似混合模式的保留机制⁸⁻¹⁰。本应用中所用的N,N-二异丙基乙胺(DIPEA)/1,1,1,3,3,3-六氟异丙醇(HFIP)流动相体系与光学UV检测和负离子模式质谱均兼容^{4,7,-10}。HFIP用于增强电喷雾电离作用⁸。

将RNase T1酶解的荧光素酶mRNA进样至ACQUITY Premier BEH C₁₈ (2.1 x 150 mm, 300 Å, 1.7 μm)肽分析专用柱上，并使用配备ACQUITY UPLC TUV检测器的ACQUITY Premier二元LC开发梯度。本研究使用的ACQUITY Premier BEH C₁₈肽分析专用柱与ACQUITY Premier BEH C₁₈寡核苷酸分析专用柱相似，但孔径更宽；即将推出孔径为300 Å的寡核苷酸测试色谱柱。使用BioAccord台式LC-MS系统的ACQUITY RDa检测器，通过负离子模式质谱重复采集三份数据。此外，将ACQUITY RDa检测器设置为采集MS^E数据，这样每隔一次扫描就会产生一个高能量碎片离子谱图，随后可用于证实LC峰的鉴定结果。图1描绘了RNase T1对照样品（上方迹线）、mRNA对照样品（下方迹线）和酶解mRNA（中间迹线）的总离子流色谱图(TIC)。根据4-sigma峰容量为613的分离，很容易分离由RNase T1酶解荧光素酶mRNA产生的寡核苷酸片段。总体而言，色谱峰尖锐且对称，三次重复进样之间的保留时间(RT)差异约为0.01分钟。使用60分钟梯度时间法时，酶解物组分在2~23分钟之间洗脱；一些未完全酶解的mRNA在29分钟左右洗脱，而完整的RNaseT1在54分钟左右洗脱。如图1中的上方迹线所示，RNase T1对照样品在50分钟保留时间后才开始显示信号，这证实RNase T1不会在mRNA酶解物组分的保留窗口内引入干扰。（图1，中间迹线）。同样，图1中的下方迹线显示完整的荧光素酶mRNA在大约38分钟时洗脱，证实29分钟时在酶解样品中观察到的峰（图1，中间迹线）对应于未完全酶解的mRNA。我们注意到，对于包含5'端帽和poly-A尾结构的合成mRNA，在酶解后约29分钟和37分钟观察到峰。约38分钟至约37分钟之间的微小变化可能指示未酶解的poly-A结构（未来会有一篇应用纪要详细描述对此色谱行为的研究）。

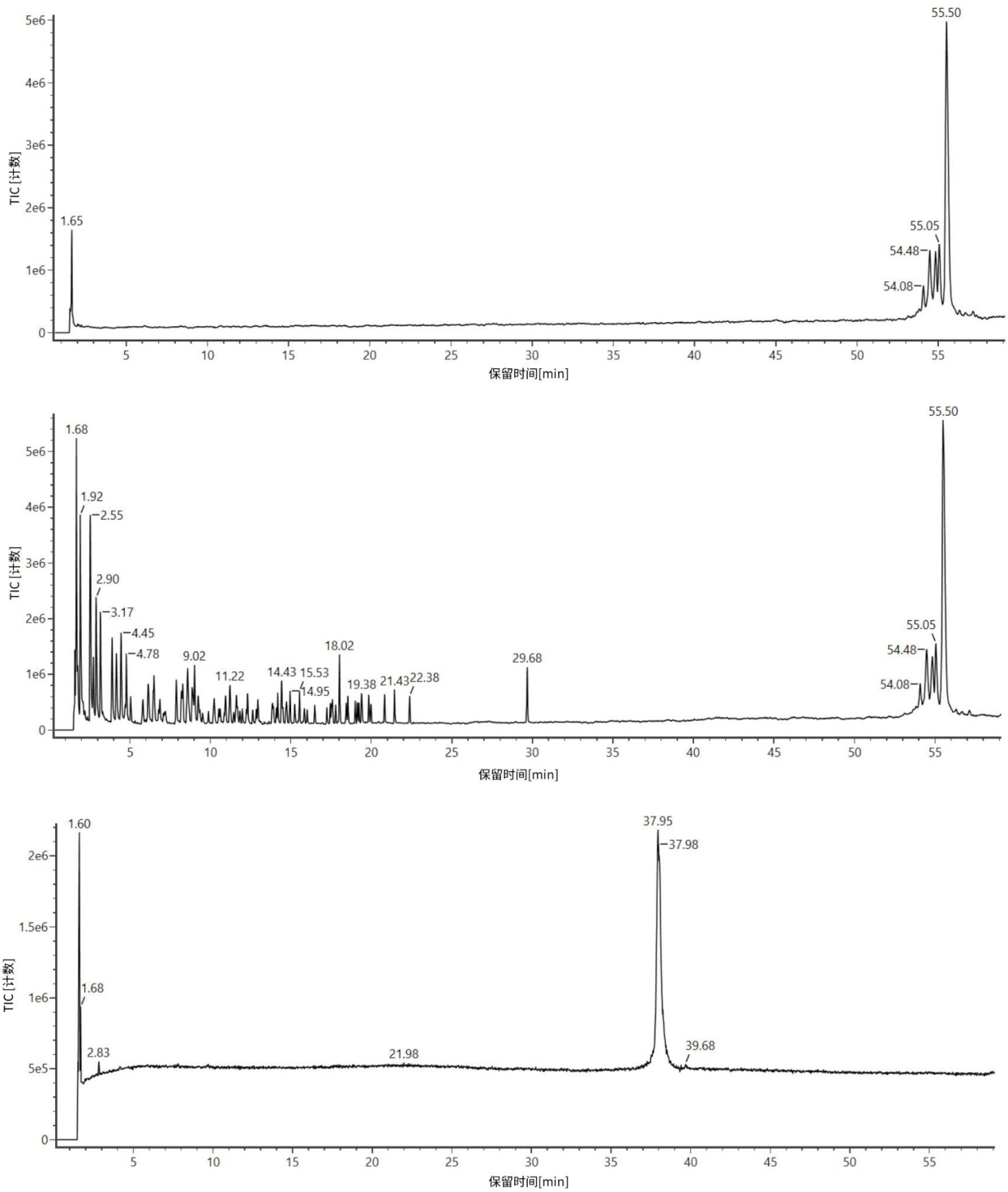


图1: 离子对反相色谱法(IP-RPLC)分析RNase T1对照样品 (上方迹线)、mRNA对照样品 (下方迹线) 和使用

*Rnase T1*酶解的荧光素酶mRNA酶解物（中间迹线）的TIC；使用ACQUITY UPLC I-Class系统（ACQUITY Premier BEH C₁₈肽分析专用柱, 2.1 x 150 mm, 300Å, 1.7 μm）和BioAccord ACQUITY RDa检测器在负离子模式下进行分析。

计算机模拟酶解mRNA计算器“mRNAlcondemand”的图形用户界面(GUI)如示意图1所示。可以在碱基序列旁边指定许多酶解参数，例如修饰、酶和漏切位点。基于此处输入的内容，计算器将默认一些MS特定设置，包括电荷态和m/z范围，以及基于单同位素或平均质量数进行计算的能力。生成的输出内容为纯文本csv文件形式，可用于UNIFI或waters_connect软件，或用于补充下游分析。

示意图1: 用于计算机模拟mRNA酶解质量数计算的mRNA计算器GUI。

接下来, 创建一个库, 将计算器的输出内容以电子表格的形式导入UNIFI科学库, 在库中将酶解物组分视为单个分析物。创建的库可用于HRMS筛查分析方法, 即使用用户指定的容差靶向查询采集后的酶解化合物。使用质量数容差进行库匹配的方法如下所示。库搜索的结果是自动生成带注释的色谱图(图2)。图2特别显示了17~20分钟保留时间窗口的放大图。

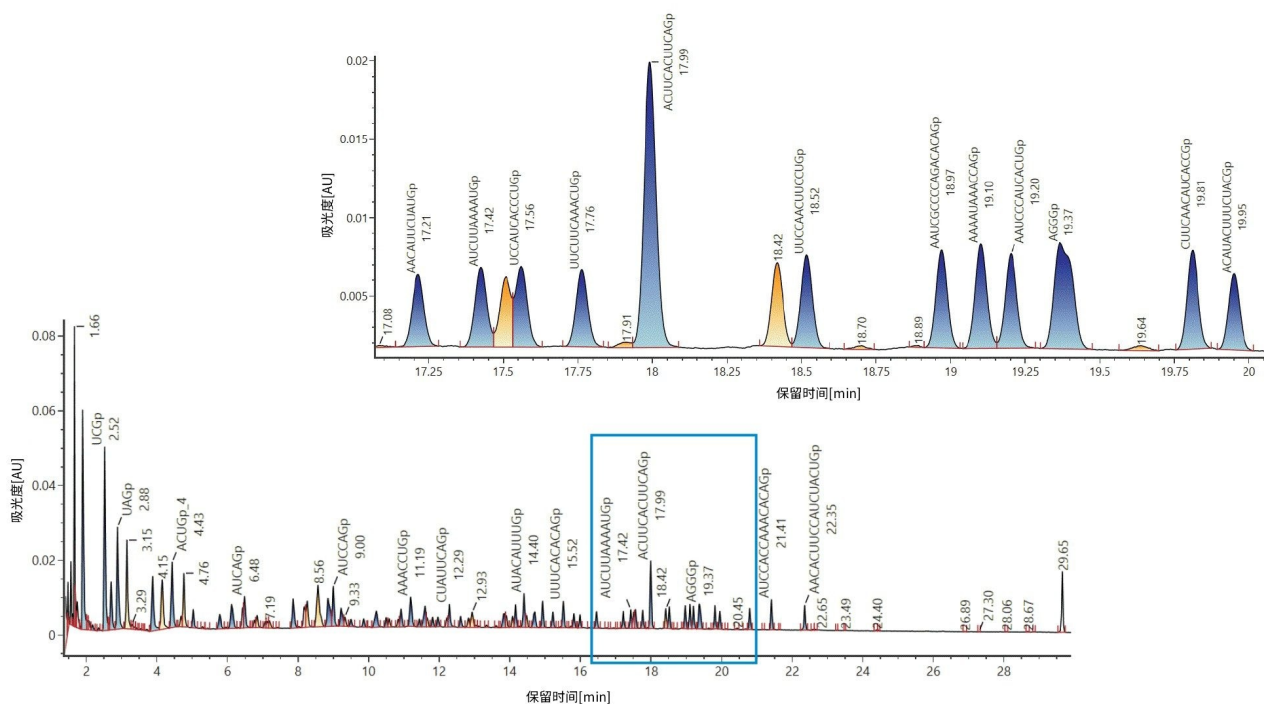


图2：基于精确质量数与目标组分库匹配后生成的荧光素酶mRNA酶解物的带注释TIC。使用Rnase T1酶解荧光素酶mRNA，并使用ACQUITY Premier BSM LC（ACQUITY Premier BEH C₁₈ 肽分析专用柱，2.1 x 150 mm, 300 Å, 1.7 μm）和BioAccord ACQUITY RDa检测器在负离子模式下进行分析。使用mRNA MASS计算器计算目标组分。

在以最多2个漏切位点为容差创建的计算机模拟库中进行数据筛查后，从每个技术重复样中产生了总共436、428和441个酶解物组分的潜在鉴定结果(ID)。在人工验证时考虑了几项标准。根据丰度和峰形（例如，低丰度色谱峰肩），在441个已鉴定组分中，有40个被拒绝。这些被拒绝的ID大多数(27/40)位于24~60分钟之间。总体而言，约60%的已鉴定和验证组分的质量数误差在10 ppm以内（261个酶解物组分）。对RNase T1对样品（图1，上方迹线）进行了相同的查询，如预期一样没有产生任何鉴定结果。除了峰形和丰度之外，我们还基于对同位素分布的高置信度解析进一步验证了结果，以推断电荷分配，并因此得到了139个酶解物组分（65%在5 ppm质量数误差内）。最后，其中某些分配结果似乎是由于重复检测展宽和包含肩峰的色谱峰触发的冗余分配，因此未纳入分析中。因此，又排除了16个鉴定结果。然而，在几个例子中，分析物质量数出现在多个保留时间，导致产生了两个独立的ID，对应于两个不同的、明确定义的色谱峰。在这里，我们还注意到存在同分异构体ID（化学成分相同但序列不同的物质）以及同量异位异构体ID（化学成分不同[精确质量数不同]，但名义质量数相似的物质）。最终，基于精确质量数匹配可以鉴定90种独特的组分，如表1所示，包括同分异构核苷酸序列和同量异位离子（表1中以灰色突出显示的单元格）。

组分名	分子式	实测RT (min)	中性质谱数 (Da)	实测 m/z	电荷	质量误差 (ppm)	观察到的额外离子
UGUGp_1	C38H48N14O31P4	1.64	1320.15604	659.0689	-2	-4	-2H, -H
CCUUGCGCCGp	C392H120N32O73P10	1.65	3150.40375	1049.1302	-3	0.4	-3H
UCUUAUCGp	C394H121N33O69P9	1.7	2870.35772	1434.1633	-2	-6.3	-2H
CCGCUCAAAGp	C395H121N34O68P10	1.83	3261.45559	1086.1526	-3	3.9	-3H
GAAGp	C404H50N20O27P4	2.1	1366.21047	682.0924	-2	-9.2	-2H
CCGp_6	C28H38N10O23P3	2.54	973.14057	485.567	-2	6.6	-2H, -H
GAGACCCUAGp	C115H45N49O82P12	2.54	3895.55339	1297.5089	-3	-3.2	-3H
CUCGp	C47H61N18O37P5	2.55	1624.21331	1623.2076	-1	1	-H
CUGp	C28H37N10O23P3	2.55	974.12459	473.1177	-1	0.4	-H, -2H
GAGACUACUAGp	C115H43N47O84P12	2.55	3897.52142	1298.1735	-3	3.5	-3H
UAUUAUAGACGp	C115H42N46O85P12	2.55	3898.50544	1948.2401	-2	-3.1	-2H
ACGp	C29H38N13O21P3	2.73	997.1518	996.1407	-1	-3.8	-H, -2H
CAGp	C29H38N13O21P3	2.73	997.1518	996.1407	-1	-3.8	-H, -2H
AUGp	C29H37N12O22P3	2.92	998.13582	997.1285	-1	0	-H, -2H
AAg 4	C30H38N15O20P3	3.18	1021.16304	1020.153	-1	-2.7	-H, -2H
ACACCCUAGAAg	C115H45N49O81P12	3.44	3879.55848	1292.17	-3	-8.8	-3H, -6H, -9H
UGAUUCUGp	C75H94N26O60P8	3.9	2566.30045	1282.1387	-2	-3.9	-2H
CCCGp	C37H50N14O29P4	3.91	1278.18186	638.0888	-2	6.9	-2H, -3H, -H
CCUGp	C37H49N13O30P4	3.91	1279.16587	638.5822	-2	9	-2H, -3H, -H
CUCUGUGp	C75H95N27O60P8	3.91	2581.31135	1289.6535	-2	3.4	-2H
GGUCCUGp	C75H96N28O59P8	3.91	2580.32733	1289.1652	-2	6.2	-2H
AAAGCGUGp	C77H97N33O56P8	4.17	2627.36578	1312.675	-2	-1.1	-2H
UAAGp_1	C39H49N17O29P4	4.79	1327.18834	1326.1802	-1	-0.7	-H, -2H, -3H
AAAGp	C404H50N20O28P4	5.06	1350.21556	1349.2049	-1	-2.5	-H, -2H, -3H
CUCUGp	C404H60N18O38P5	5.81	1658.19118	791.5862	-2	9	-2H, -3H, -H
UCCCGp	C464H61N18O37P5	5.83	1584.20716	791.1011	-2	5.1	-2H, -3H, -H
CACCGp	C47H62N19O35P5	6.18	1607.23438	802.6146	-2	4.8	-2H, -3H, -H
AUCUGp	C47H60N17O37P5	6.51	1609.20241	803.6021	-2	9.2	-2H, -3H, -H
AUCCGp	C47H61N18O36P5	6.53	1608.21839	803.1076	-2	6.1	-2H, -3H, -H
AAAAACUAGUCCGp	C144H71N96O103P15	6.52	4860.68775	1619.2101	-3	-4.7	-3H
CAAGp	C49H62N23O33P5	6.79	1655.25684	826.625	-2	3.7	-2H, -3H, -H
CCUUCUGp	C55H72N18O45P6	7.89	1890.32346	944.1145	-2	5.1	-2H, -3H, -4H, -H
CUCGCGp	C55H73N19O44P6	7.91	1889.24845	943.6212	-2	3.7	-2H, -3H, -4H, -H
CCCAAGp	C56H74N22O42P6	8.29	1912.27566	955.1342	-2	3	-2H, -3H, -4H, -H
AUCCAAGp	C57H73N23O42P6	8.58	1937.27091	967.6315	-2	2.6	-2H, -3H, -4H, -H
CAUUAAGp	C57H72N22O43P6	8.58	1938.25493	1937.2514	-1	1.9	-H
AUCUUAAGp	C56H71N19O45P6	8.63	1915.22771	956.614	-2	6.9	-2H, -3H, -4H, -H
AUAUAAGp	C58H73N25O41P6	8.88	1961.28215	979.6375	-2	3	-2H, -3H, -4H, -H
UAUAAGp	C58H73N25O41P6	8.88	1961.28215	979.6375	-2	3	-2H, -3H, -4H, -H
AUCCAAGp	C57H73N23O42P6	8.89	1937.27091	967.6353	-2	6.5	-2H, -3H, -H
AACCUAGp	C57H73N23O42P6	9.03	1937.27091	967.634	-2	5.2	-2H, -3H, -4H, -H
ACACAAGp	C58H74N26O40P6	9.26	1960.29813	979.1457	-2	3.2	-2H, -3H, -4H, -H
ACAAAAGp	C59H74N28O39P6	9.52	1984.30936	991.1508	-2	2.7	-2H, -3H, -4H, -H
UCUUCUGp	C64H82N19O54P7	9.89	2197.24778	1097.6182	-2	3.5	-2H, -3H, -4H
AUCUUCUGp	C65H83N20O53P7	10.23	2220.269	1099.1359	-2	7.1	-2H, -3H, -4H
UCUUCUGp	C65H83N20O52P7	10.23	2220.269	1099.1359	-2	7.1	-2H, -3H, -4H
UCUUCUGp	C65H83N20O52P7	10.61	2220.269	1099.1302	-2	2	-2H, -3H, -4H
AUCCAAGp	C66H84N25O50P7	10.53	2243.29621	1120.643	-2	1.2	-2H, -3H, -4H
CAUUAAGp	C66H84N25O50P7	10.53	2243.29621	1120.643	-2	1.2	-2H, -3H, -4H
CAUUAAGp	C66H84N25O50P7	10.96	2243.29621	1120.6441	-2	2.2	-2H, -3H, -4H
AUUAUAAGp	C66H83N24O51P7	10.95	2244.28023	1121.141	-2	6.6	-2H, -3H, -4H, -H
AAACCUAGp	C67H85N28O48P7	11.22	2266.32343	1132.1603	-2	4.5	-2H, -3H, -4H, -H
ACUAUAAGp	C67H85N28O48P7	11.22	2266.32343	1132.1603	-2	4.5	-2H, -3H, -4H, -H
CCACAAGp	C67H86N29O47P7	11.25	2265.33942	1131.6655	-2	2	-2H, -3H, -4H, -H
UCUUAUAAGp	C73H94N32O61P8	11.69	2502.28307	1250.1365	-2	1.1	-2H, -3H, -4H, -5H
AAAAACGp	C69H86N33O45P7	11.84	2313.36188	1155.6762	-2	1.5	-2H, -3H, -4H, -H
UCCUUAUAAGp	C74H95N32O59P8	11.99	2525.31028	1261.6501	-2	1.2	-2H, -3H, -4H, -5H
CCACCUAGp	C75H97N29O56P8	12.32	2547.35349	1272.6742	-2	3.1	-2H, -3H, -4H, -H
AAAUUAUAAGp	C76H96N30O56P8	12.63	2572.34873	1285.1683	-2	0.3	-2H, -3H, -4H, -5H
ACAUUAUAAGp	C76H96N30O56P8	12.63	2572.34873	1285.1683	-2	0.3	-2H, -3H, -4H, -5H
ACAUUAUAAGp	C76H96N30O56P8	12.85	2572.34873	1285.1676	-2	-0.2	-2H, -3H, -4H
ACAUUAUAAGp	C76H96N30O56P8	12.96	2572.34873	1285.1705	-2	2.1	-2H, -3H, -4H
CUCGCGp	C56H73N21O44P6	12.64	1929.25459	642.0863	-3	9.8	-3H
GGCGp	C39H50N18O29P4	13.37	1358.19415	1357.1889	-1	1.5	-H
CCUACACGp	C84H109N32O63P9	13.84	2852.39477	1425.1991	-2	5	-2H, -3H, -4H, -5H
ACAUUCUCCGp	C84H108N31O64P9	13.87	2853.37879	1425.6931	-2	7.1	-2H, -3H, -4H, -5H
UCCACUCCGp	C84H108N31O64P9	13.87	2853.37879	1425.6931	-2	7.1	-2H, -3H, -4H, -5H
UCCACUCCGp	C84H108N31O64P9	14.1	2853.37879	1425.6993	-2	4.5	-2H, -3H, -4H, -5H
AUUAUUUAAGp	C83H104N25O69P9	13.93	2833.30362	1415.653	-2	5.4	-2H, -3H, -4H, -5H
ACAUUAUAAGp	C85H107N32O64P9	14.19	2878.37404	1438.1666	-2	-9.7	-2H
AUUAUAUAAGp	C84H105N28O67P9	14.19	2856.33083	1427.1719	-2	9.1	-2H, -3H, -4H
UGAUUAUUUAAGp	C131H182N44O105P14	14.19	4464.50161	1487.1526	-3	-6.6	-3H
AUAUAUAUAAGp	C85H106N31O65P9	14.43	2879.35805	1438.6666	-2	9.8	-2H, -3H, -5H
ACACCUAGp	C85H109N34O62P9	14.44	2876.406	1437.2048	-2	5.8	-2H, -3H, -4H, -5H
AAACAUAUAAGp	C86H109N36O61P9	14.72	2900.41724	1449.2093	-2	5	-2H, -3H, -4H, -5H
AAACAUAUAAGp	C87H109N38O60P9	14.96	2924.42847	1461.2135	-2	3.9	-2H, -3H, -4H, -5H
CAACAUAUAAGp	C88H110N41O58P9	14.96	2947.45569	1472.7265	-2	3.5	-2H, -3H, -4H, -5H
AAAAACUUAUAAGp	C88H109N40O59P9	15.25	2948.4397	1473.2225	-2	6.2	-2H, -3H, -4H, -5H
CUGACCUUAUAUAAGp	C140H178N51O107P15	15.25	4749.61192	1582.19	-3	-5.8	-3H
UUUAUAUAUAAGp	C94H119N35O71P10	15.55	3183.41532	1590.7086	-2	4.7	-2H, -3H, -4H, -5H
UUAUAUAUAUAAGp	C95H118N36O71P10	15.84	3208.41057	1603.2007	-2	1.2	-2H, -3H, -4H, -5H
AUAUAUAUAUAUAAGp	C144H178N50O103P15	16.02	4845.65685	1614.2191	-3	3.1	-3H, -6H
CUACAUAUAUAAGp	C96H120N40O68P10	16.02	3230.45377	1614.2191	-2	-0.8	-2H, -3H, -4H, -5H
GGCUAGp	C48H61N20O37P5	16.02	1664.21946	1663.208	-1	-2.5	-H
UCCACUUAUAUAAGp	C102H130N35O80P11	16.49	3465.42939	1731.7095	-2	0.8	-2H, -3H, -4H, -5H, -6H
AAUAUAUAUAUAAGp	C104H130N35O79P11	17.24	3513.45186	1755.7214	-2	1.1	-2H, -3H, -4H, -5H, -6H
AUAUAUAUAUAUAAGp	C105H130N41O77P11	17.45	3537.46309	1767.7284	-2	1.9	-2H, -3H, -4H, -5H
ACUAUAUAUAUAUAAGp	C114H139N45O83P12	18.45	3841.52036	1919.7549	-2	0.6	-2H, -3H, -4H, -5H, -6H
CUGUGGp	C57H72N22O45P6	18.45	1970.24476	1969.2473	-1	5	-H
QUUGp	C39H48N17O30P4	18.55	1359.17817	1358.1705	-1	-0.3	-H
AGGGp	C40H50N20O28P4	19.39	1382.20539	690.0948	-2	-2	-2H, -H
GAGGp_1	C40H50N20O28P4	19.39	1382.20539	690.0948	-2	-2	-2H, -H
UCAUAUAUAUAUAUAUAAGp	C188H234N68O144P20	19.39	6366.78303	2121.2642	-3	3.8	-3H
UAUAUAUAUAUAUAUAUAAGp	C124H156N18O88P13	19.43	4169.58887	1388.8648	-3	4.8	-3H, -2H, -4H, -5H, -6H
CAGGp	C39H50N18O28P4	22.38	1342.19924	1341.1865	-1	-4.1	-H, -2H

表1: 基于精确质量数匹配在10 ppm质量数误差内初步鉴定和验证的荧光素酶mRNA酶解物组分。对应于同分异构

序列或同量异位离子的单元格以灰色突出显示。

图3显示了产生酶解物组分UCCACUCUAUGp鉴定结果的示例数据。如左上色谱图所示，该组分在16.49分钟洗脱，基于 m/z 576.5732 ($[M-6H]^{6-}$)、692.0876 ($[M-5H]^{5-}$)、865.3603 ($[M-4H]^{4-}$)、1154.1438 ($[M-3H]^{3-}$)和1731.7095 ($[M-2H]^{2-}$)处5个携带2~6个负电荷的离子（图3，左下图）通过计算机模拟酶解物库鉴定得出。图中展示了 $[M-5H]^{5-}$ 和 $[M-2H]^{2-}$ 离子的同位素分布以显示支持电荷态分配的数据（图3，右）。

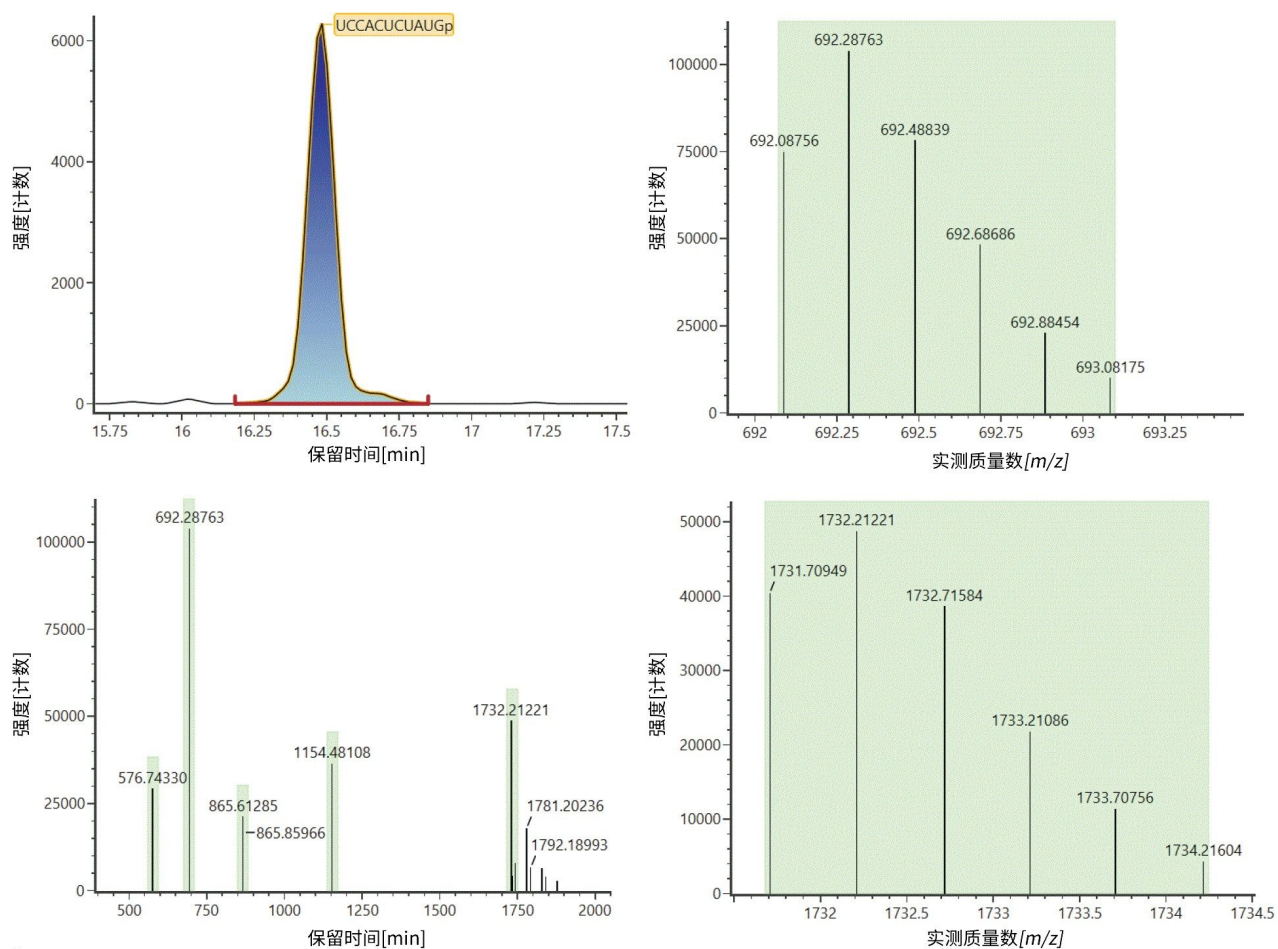


图3: 在16.49分钟洗脱的酶解物组分UCCACUCUAUGp的鉴定结果 (左上图)。基于 m/z 576.5732 ($[M-6H]^{6-}$)、692.0876 ($[M-5H]^{5-}$)、865.3603 ($[M-4H]^{4-}$)、1154.1438 ($[M-3H]^{3-}$)和1731.7095 ($[M-2H]^{2-}$)处5个携带2~6个负电荷的离子 (左下图) 通过生成的计算机模拟酶解物鉴定得出该化合物。实验中观察到的 $[M-5H]^{5-}$ 和 $[M-2H]^{2-}$ 离子的同位素分布分别见右上图和右下图。

如表2所示, 使用waters_connect CONFIRM Sequence™应用程序解析 MS^E 谱图, 消除因存在由同分异构体或同量异位离子而造成不确定性。第623-631位(ACAUCCUCGp)和第551-559位(UCACCAUCGp)的同分异构体序列均由RNase T1计算机模拟酶解物预测得来, 并且两者都分配至相同的保留时间13.87分钟。无法通过完整质量数分析进行正确分配。使用waters_connect™ CONFIRM Sequence应用程序, 基于同一进样的 MS^E 数据来阐明此分配的正确序列, 预测其中每个序列的高能量碎片离子, 并通过定制算法将这些离子与集成原始数据的同位素簇相匹配。利用点图表示已确认的碎片离子, 从而快速评估序列覆盖率 (图4, B): UCACCAUCGp的序列覆盖率达到

100%，因此可以很容易地验证它是正确的分配（表2，第63行）。

手动验证的酶解物组分（表1）获得的中性母离子质量数从973.1406 Da（CCGp，RT 2.54分钟）到6366.7830 Da的20 mer核苷酸（UCAUUGAGUUCUCAAACUGp，RT 19.39分钟）不等。最早洗脱的组分是UGUGp，经鉴定该组分的观测中性质量数为1320.1560，保留时间为1.64分钟。在手动验证的ID组（表1）中洗脱的最后一个荧光素酶mRNA酶解物组分是CAGGp（1342.1992Da），观察到它在22.38分钟洗脱。鉴于与CAGGp密切相关的序列UAAGp在约4分钟时洗脱，因此该洗脱时间明显晚于预期且相差较大。为了解决假阳性问题，我们使用waters_connect CONFIRM Sequence应用程序来进一步表征组分ID。基于MS数据和MS^E数据进行验证。从表2中可以看出，包括CAGGp在内的34/90个组分没有生成足够数量的MS^E片段来进一步验证序列分配，尽管精确质量数匹配支持该分配。另一个值得关注的例子是在同一数据集中观察到的指定序列AAAAACAUGUUGCCGp（4860.6678 Da，15-mer，9个嘌呤）和AUACAUUUGACAAAGp（4845.6568 Da，15-mer，9个嘌呤），二者具有相同的嘌呤含量但洗脱时间相隔10分钟。使用MS^E数据，我们能够将AAAAACAUGUUGCCGp作为假阳性结果排除，并确认AUACAUUUGACAAAGp的鉴定结果。此例证明了战略性地使用精确质量数匹配和碎片离子质谱图来帮助明确鉴定mRNA序列酶解产生的组分的重要性。此外，我们通过观察结果认识到，mRNA酶解物组分的保留可能不如预期那样可预测，并且迫切需要进一步研究和建模来确认寡核苷酸与色谱固定相之间的相互作用。

最后，我们通过比较匹配的、手动验证的酶解物组分与mRNA序列来手动估计序列覆盖率。初步估计401个初始匹配的序列覆盖率得到的覆盖率值大约为76%。根据mRNA序列检查了经过严格验证的匹配后，得到的覆盖率值大约为22%。观察到的许多酶解物组分映射到荧光素酶mRNA序列的多个位置（表2），由于在修饰或完全修饰的核酸序列中只有四个单独的残基，因此预计会出现冗余。

索引	使用精确质量数匹配 (10 ppm质量误差) 验证的荧光素酶 mRNA 酶解物组分	实测RT	mRNA序列上的位置	mRNA序列上的其他位置	使用MSe数据进行验证 (CONFIRM Sequence)			
					是/否	覆盖率(%)		
						峰1	峰2	峰3
1	UGUGp_1	1.64	217-229	1553-1556	是	50		
2	CCUUCGCCCGp	1.65	886-905	NA	否	0		
3	UCUACUCCGp	1.7	71-79	NA	否	0		
4	CCGGUCAAGp	1.93	1520-1529	NA	否	10		
5	GAAp	2.1	378-381	445-448, 635-638, 712-715	是	25		
6	CCGp_6	2.54	59-61	246-248, 284-286, 838-840, 855-857, 1073-1075, 1076-1078, 1223-1225, 1520-1522	否	0		
7	GAGACACCCUAAGp	2.54	1667-1678	NA	否	0		
8	CUCGp	2.55	25-29	NA	否	0		
9	CUGp	2.55	177-179	422-424, 427-429, 452-454, 512-514, 560-562, 852-854, 906-908, 940-942, 965-967, 1430-1432, 1493-1495	是	50		
10	GAGACAUCAUUGp	2.55	455-466	NA	否	16.67		
11	UAUUGAUAGCAg	2.55	165-176	NA	否	0		
12	ACGp	2.73	779-781	809-811, 920-922, 958-960	否	0		
13	CAGp	2.73	174-176	273-275, 287-289	是	50		
14	AUGp	2.92	209-211	291-293, 407-409, 571-573, 709-711, 806-808, 1100-1102, 1193-1195, 1271-1273, 1367-1369, 1415-1417, 1427-1429, 1467-1469, 1581-1583, 1688-1690	是	100		
15	AAGp_4	3.18	43-45	146-148, 155-157, 379-381, 446-448, 636-638, 713-715, 1105-1107, 1139-1141, 1220-1222, 1304-1306, 1433-1435, 1475-1477, 1490-1492, 1679-1681	是	100		
16	ACACCCUAGAAGp	3.44	1670-1681	NA	否	16.67		
17	UGAUUCUGp	3.9	1530-1537	NA	是	50		
18	CCCGp	3.91	902-905	NA	否	0		
19	CCUGp	3.91	1439-1442	1439-1442, 1564-1567, 1599-1602	否	25		
20	CUUCUGUGp	3.91	1307-1314	NA	否	12.5		
21	GUUCCUGp	3.91	1048-1055	NA	否	12.5		
22	AAGCCGUGp	4.17	1220-1227	NA	否	12.5		
23	UAAGp_1	4.79	644-647	1371-1374	是	50		
24	AAAGp	5.06	312-315	1147-1150, 1645-1648	是	100		
25	CUCUGp	5.81	1307-1311	NA	是	60		
26	UCCGp	5.83	1232-1236	NA	否	0		
27	CACCGp	6.18	1470-1474	NA	否	0		
28	AUCUGp	6.51	639-643	1445-1449	是	100		
29	AUCCGp	6.53	30-34	NA	是	20		
30	AAAAACAUGUUGCCGp	6.52	234-248	NA	否	0		
31	CAAAg	6.79	822-826	NA	是	40		
32	CCUCUGp	7.89	687-692	NA	是	50		
33	CUCCCGp	7.91	968-973	NA	是	33.33		
34	CCCCAGp	8.29	653-658	NA	否	16.67		
35	AUCCAGp	8.58	734-739	NA	是 (共洗脱)	33.33	33.33	
36	CAUUAAGp	8.58	65-70	NA	是	100		
37	AUUCUGp	8.63	1204-1209	NA	是	66.67		
38	AAUACAGp	8.88	1155-1160	NA	是	66.67		
39	UAACAAGp	8.88	841-846	NA	是	100		
40	AUCCAGp	8.89	734-739	NA	是	33.33	33.33	
41	AACCGp	9.03	110-115	332-337, 400-405, 984-989	是	100		
42	ACACAAGp	9.26	659-664	NA	否	16.67		
43	ACAAGp	9.52	1025-1030	NA	是	33.33		
44	UCUUCUGp	9.89	828-834	NA	是	71.43		
45	AUCUCUGp	10.23	679-685	NA	否	14.29		
46	UCUACUGp	10.23	71-77	NA	是	57.14	28.57	
47	AUCCAAGp	10.53	318-324	NA	是	14.29		
48	CAUACUGp	10.53	815-821	NA	是	57.14		
49	AUUAUCAGp	10.95	1408-1412	NA	是	71.43		
50	AAACCUAGp	11.22	1483-1489	NA	是	71.43		
51	ACUACAGp	11.22	947-953	1285-1291	是	71.43		
52	CCACAAGp	11.25	1557-1563	NA	是	14.29		
53	UCUCUUCGp	11.69	1512-1519	NA	是	37.5		
54	AAAACAAGp	11.84	1197-1203	NA	是	71.43		
55	UCCUACGp	11.99	102-109	NA	是	37.5		
56	CCACCAUGp	12.32	35-42	NA	是	25		
57	AAAUCCUGp	12.63	798-805	NA	否	12.5		
58	ACAUAUCGp	12.63	1277-1284	NA	是	14.29	62.5	40
59	CUGCCGp	12.64	852-857	NA	否	0		
60	GGCGp	13.37	367-370	NA	否	0		
61	CCUCACACGp	13.84	929-937	NA	是	22.22		
62	ACAUCCUGp	13.87	623-631	NA	是	33.33		
63	UCACCAUCGp	13.87	551-559	NA	是	100	33.3	
64	AUUCUUCUGp	13.93	1418-1426	NA	是	55.56		
65	ACAUCUUCGp	14.19	458-466	NA	是	22.22		
66	AUUCUUAUAGp	14.19	1063-1071	NA	是	44.44		
67	UGAUUAUCUUCUUGp	14.19	1413-1426	NA	否	7.14		
68	AUACAUUUGp	14.43	1016-1024	NA	是	55.56		
69	ACACCCUAGp	14.44	1670-1678	NA	否	11.11		
70	AAACAUCGp	14.72	740-748	NA	是	22.22		
71	AAACAUCGp	14.98	1162-1170	NA	是	55.56		
72	CAACAACAGp	14.98	665-673	NA	是	77.78		
73	AAAAACAUGp	15.25	234-242	NA	是	77.78		
74	CUGACCCUUAUCUAGp	15.25	512-526	NA	否	0		
75	UUUCACACAGp	15.55	1109-1118	NA	是	80		
76	UACAUUAAGp	15.84	1348-1355	NA	是	80		
77	AUACAUUUGACAAAGp	16.02	1018-1030	NA	是	20		
78	CUACAUAAGp	16.02	1653-1662	NA	是	50		
79	GGCUGp	16.02	420-424	NA	否	0		
80	UCCACUCUAUGp	16.49	786-796	NA	是	72.73		
81	AACAUUCUAUGp	17.24	300-310	NA	是	63.64		
82	AUCUUAUAAGp	17.45	698-708	NA	是	63.64		
83	ACAUAUCUACAGp	18.45	197-208	NA	是	50		
84	CUUGUGp	18.45	177-182	NA	否	0		
85	GUGGp	18.55	506-509	914-917, 1312-1315	否	0		
86	AGGp	19.39	1603-1606	NA	否	0		
87	GAGp_1	19.39	632-635	1214-1217	否	0		
88	UCAUUGAGUUCUACAACUGp	19.39	593-612	NA	否	0		
89	UAACUACAACAGp	19.43	1392-1404	NA	是	69.23		
90	CAGp	22.38	287-290	NA	否	0		

表2：基于精确质量数匹配的荧光素酶 mRNA 酶解物组分鉴定和验证结果，以及使用 waters_connect CONFIRM

Sequence应用程序和收集的MS^E谱图获得的进一步验证。

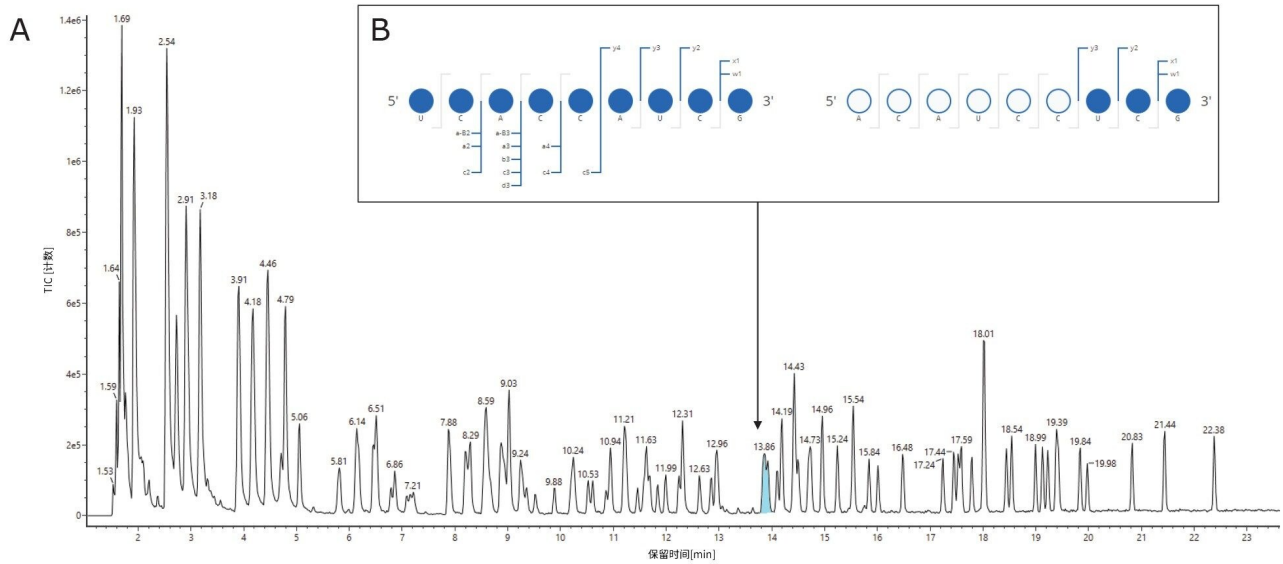


图4: (A)第623–631位(*ACAUCUCGp*)和第551–559位(*UCACCAUCGp*)的酶解片段组分基于RNase T1酶解物预测得出,并分配至TIC中的相同RT峰。无法通过完整质量数信息确定正确分配。(B)同一进样的MS^E数据可用于阐明此分配的正确序列。使用waters_connect CONFIRM Sequence应用程序,根据McLucky注释¹⁷预测每个序列的高能量碎片离子,并通过定制算法与集成原始数据的同位素簇相匹配。该软件在点图上显示了已确认的碎片离子,可用于快速评估序列覆盖率。

结论

在本研究中,我们建立了一个强大的分析工作流程,用于使用IP-RPLC和MS对合成mRNA进行寡核苷酸映射。

- 使用RNase T1可重现地酶解合成mRNA,包括低至10 μg的材料,并且无需额外的样品净化即可进样至ACQUITY Premier BEH C₁₈ (2.1 x 150 mm, 300 Å, 1.7 μm)肽分析专用柱上
- 在ACQUITY Premier LC上使用离子对反相色谱法实现了高色谱分离度,因此酶解物组分可以很容易地与未完全酶解的mRNA和残留酶分离,从而使用BioAccord ACQUITY RDa检测器进行有效检测

- 在计算机模拟mRNA酶解计算和waters_connect/UNIFI科学数据库应用程序的支持下，基于精确质量数匹配生成了带注释的mRNA酶解物色谱图
- 我们使用waters_connect CONFIRM Sequence应用程序根据MS^E谱图进一步验证了酶解组分的分配序列。此外，使用可视化点图快速检查了潜在分配的碎片离子覆盖率

本研究旨在建立促进mRNA分子自下而上表征所需的色谱、检测和数据解析方法。RNase T1酶解仅用作第一个示例，也是建立数据采集和分析工作流程的概念验证。也就是说，我们还有足够的机会通过以下方法更全面地探测给定的mRNA结构：(1)使用多种不同的核酸酶来生成正交和附加的序列映射信息和(2)采用多重方法进行数据采集。这些方法都旨在实现全面的序列覆盖，并且将在未来的工作中得到进一步研究。

参考资料

1. Xu, S.; Yang, K.; Li, R.; Zhang, L., mRNA Vaccine Era—Mechanisms, Drug Platform and Clinical Prospection. *International Journal of Molecular Sciences* 2020, 21 (18), 6582.
2. Brenner, S.; Jacob, F.; Meselson, M., An Unstable Intermediate Carrying Information From Genes to Ribosomes for Protein Synthesis. *Nature* 1961, 190 (4776), 576–581.
3. Weide, B.; Pascolo, S.; Scheel, B. Derhovannessian, E.; Pflugfelder, A.; Eigentler, T. K. Pawelec, G.; Hoerr, I.; Rammensee, H. G.; Garbe, C., Direct Injection of Protamine-Protected Mrna: Results of a Phase 1/2 Vaccination Trial in Metastatic Melanoma Patients. *J Immunother* 2009, 32 (5), 498–507.
4. Jiang, T.; Yu, N.; Kim, J.; Murgu, J.-R.; Kissai, M.; Ravichandran, K.; Miracco, E. J.; Presnyak, V.; Hua, S., Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs via Parallel Ribonuclease Digestions and LC-MS/MS. *Anal.Chem.* 2019, 91 (13), 8500–8506.
5. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K., UPLC/MS^E; A New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Communications in Mass Spectrometry* 2006, 20 (13), 1989–1994.
6. Packer, M.; Gyawali, D.; Yerabolu, R.; Schariter, J.; White, P., A Novel Mechanism for the Loss of mRNA Activity in Lipid Nanoparticle Delivery Systems. *Nat. Commun.* 2021, 12 (1), 6777.

7. Goyon, A.; Scott, B.; Kurita, K.; Maschinot, C.; Meyer, K.; Yehl, P.; Zhang, K., On-line Sequencing of CRISPR Guide RNAs and Their Impurities via the Use of Immobilized Ribonuclease Cartridges Attached to a 2D/3D-LC-MS System.*Anal.Chem.*2021.
8. Guo, L.; Worth, A. J.; Mesaros, C.; Snyder, N. W.; Glickson, J. D.; Blair, I. A., Diisopropylethylamine/Hexafluoroisopropanol-Mediated Ion-Pairing Ultra-High-Performance Liquid Chromatography/Mass Spectrometry for Phosphate and Carboxylate Metabolite Analysis: Utility for Studying Cellular Metabolism.*Rapid Commun Mass Spectrom* 2016, 30 (16), 1835–1845.
9. Birdsall, R. E.; Gilar, M.; Shion, H.; Yu, Y. Q.; Chen, W., Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis.*Rapid Commun Mass Spectrom* 2016, 30 (14), 1667-1679.
10. Fountain, K.; Gilar, M.; Budman, Y.; Gebler, J., Purification of Dye-Labeled Oligonucleotides by Ion-Pair Reversed-Phase High-Performance Liquid Chromatography.*Journal of chromatography.B, Analytical technologies in the biomedical and life sciences* 2003, 783, 61–72.
11. McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L., Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides.*J. Am.Soc.Mass Spectrom.*1992, 3 (1), 60-70.

致谢

我们在此感谢以下沃特世同事为本研究做出的宝贵贡献：Ana-Maria Rotaru、Emanuela Petreanu、Claudia Florea、Dave Jackson和Simon Jones。诚挚感谢以下的New England Biolabs合作者为我们提供荧光素酶 mRNA并与我们进行许多深入讨论：Bijoyita Roy、Siuhong Chan、Ivan R. Corrêa Jr.、Erbay Yigit和G. Brett Robb。

特色产品

ACQUITY UPLC I-Class PLUS系统 <<https://www.waters.com/134613317>>

生物制药领域专用的BioAccord LC-MS系统 <<https://www.waters.com/waters/nav.htm?cid=135005818>>

ACQUITY UPLC可变波长紫外检测器 <<https://www.waters.com/514228>>

[ACQUITY RDa检测器 <https://www.waters.com/waters/nav.htm?cid=135077027>](https://www.waters.com/waters/nav.htm?cid=135077027)

[UNIFI科学信息系统 <https://www.waters.com/134801648>](https://www.waters.com/134801648)

[waters_connect <https://www.waters.com/waters/nav.htm?cid=135040165>](https://www.waters.com/waters/nav.htm?cid=135040165)

720007669ZH, 2022年6月

© 2023 Waters Corporation. All Rights Reserved.

[使用条款](#) [隐私](#) [商标](#) [网站地图](#) [招聘](#) [Cookie](#) [Cookie设置](#)

[沪ICP备06003546号-2](#) [京公网安备 31011502007476号](#)