Oligonucleotide Mapping via LC-UV-MS/MS to Enable Comprehensive Primary Structure Characterization of mRNA Drug Substance

Brian C. Gau[1], Andrew W. Dawdy[1], Leah Hanliu Wang[1], Bradley Bare[1], Carlos H. Castaneda[1], Olga V. Friese[1], Matthew S. Thompson[2], Thomas F. Lerch[1], David J. Cirelli[2], and Jason C. Rouse[2]

[1]Biotherapeutics Pharm. Sci., Pfizer Inc, St Louis, MO, USA and [2]Andover, MA, USA

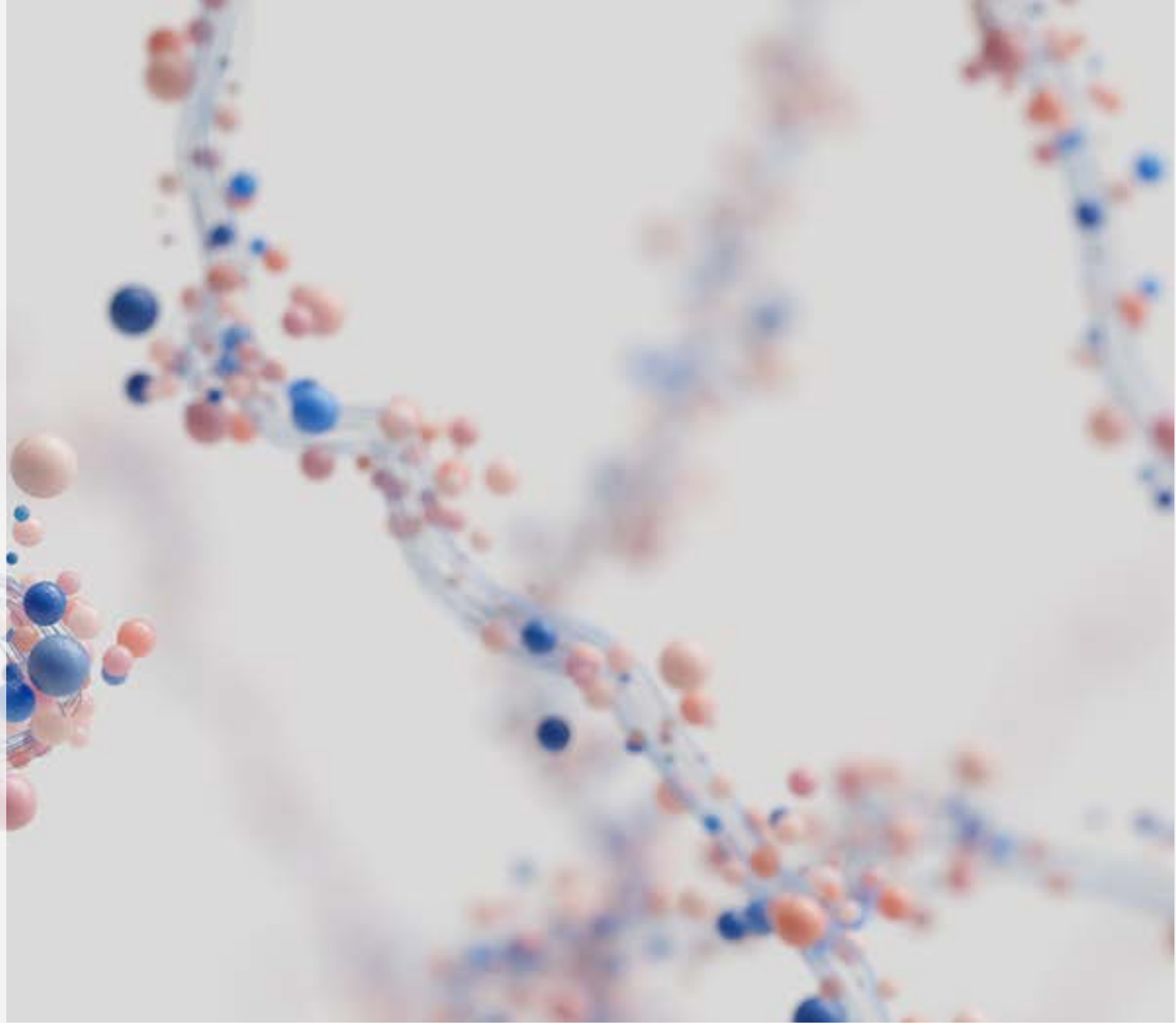
September 2023

Outline

- Overview of the Comirnaty mRNA Vaccine Against SARS-CoV-2
- Oligonucleotide Mapping Considerations
- Oligonucleotide Mapping of BNT162b2 mRNA Primary Structure by LC-UV-MS/MS
  - Link:

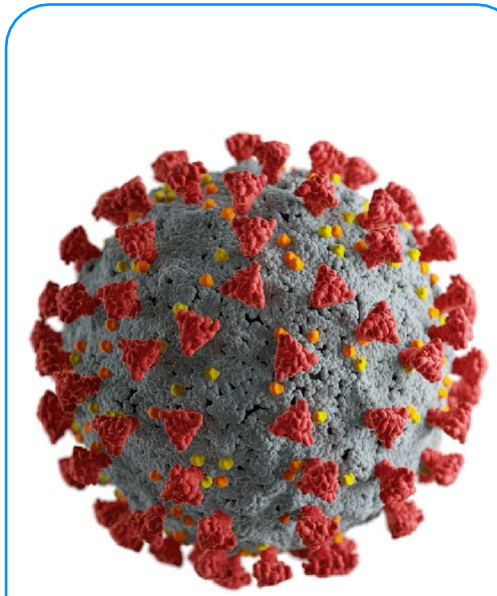
    Gau, B.C. et al. Oligonucleotide mapping via mass spectrometry to enable comprehensive primary structure characterization of an mRNA vaccine against SARS-CoV-2. *Scientific Reports* ***13,*** 9038 (2023) **https://rdcu.be/di05D**
- Utility of Oligonucleotide Mapping
- Ensuring Optimum Chromatographic Separation
- Ensuring Optimum MS/MS
- Data Analysis Workshop

Overview of the Comirnaty mRNA Vaccine Against SARS-CoV-2

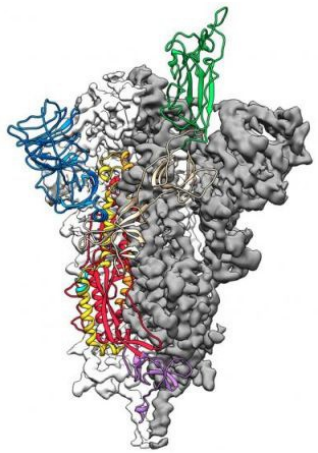


Basic Design of Pfizer/BioNTech mRNA Vaccine(s) against SARS-CoV-2

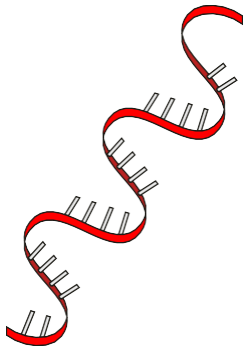
- Train patient's immune system to recognize the virus, specifically the spike protein on the surface
- Give the "code" or "recipe" of the spike protein to your cells
- The original mRNA construct in the Comirnaty Vaccine is "BNT162b2"



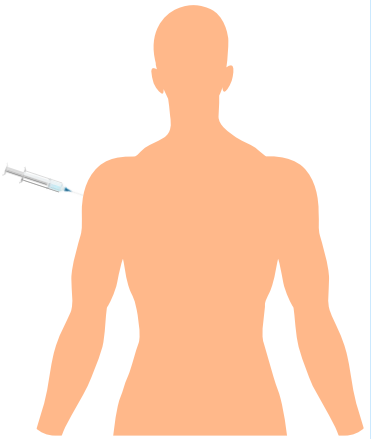
SARS-CoV-2
(3D Model)



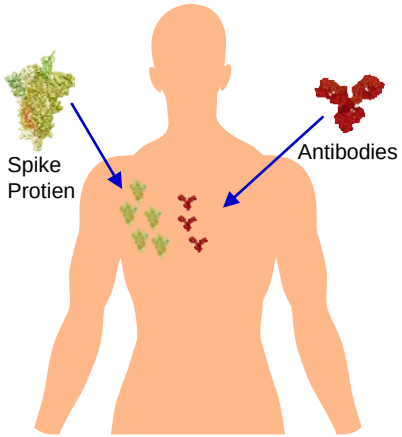
Spike Protein



mRNA for
spike protein of
coronavirus



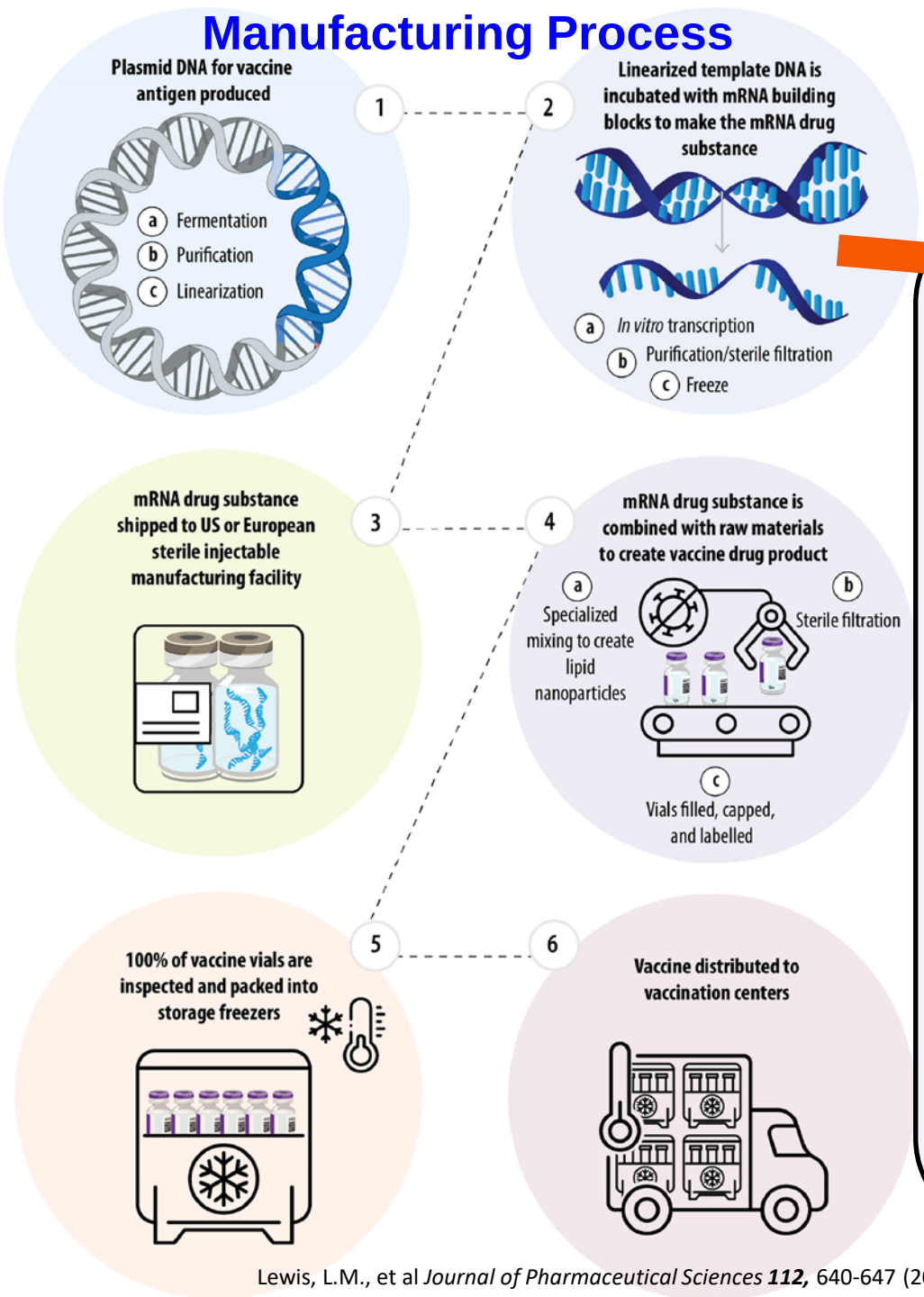
mRNA is
injected into
muscle



The vaccine triggers production of the spike protein and antibodies specific to it

Wrapp, D. et al *Science* ***367,*** 1260-1263 (2020)

Manufacturing Process



Lewis, L.M., et al *Journal of Pharmaceutical Sciences* ***112,*** 640-647 (2023)

Analytical Characterization of the Drug Substance (mRNA) is Critical for Development of a High-Quality Manufacturing Process & Product

Drug Substance (mRNA)

Platform QC Assays

- **Compendial methods**
- **Purity** by **Capillary Gel Electrophoresis**
- **Purity** by **Immunoblot**
- **Concentration** by **UV spectroscopy**
- **Identity, Impurities** by **PCR-based methods**

Heightened Characterization

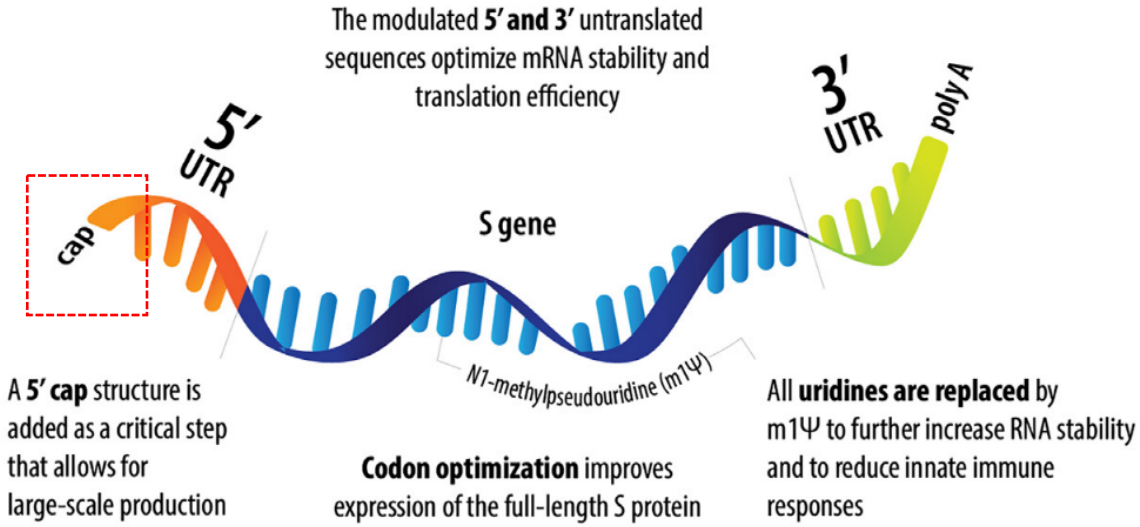
Primary Structure
- **Oligonucleotide mapping (LC-UV-MS/MS)**
- **Nucleoside Analysis (LC-UV-MS)**
- **NextGen Sequencing (NGS)**

Higher Order Structure
- **Circular Dichroism (CD)**

Protein Expression
- **FACS**
- **Western Blot**

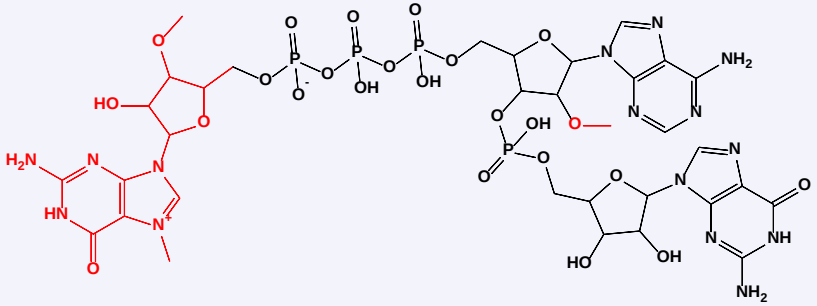
BNT162b2 mRNA is Capped at its 5′ End

- The 5′ end of endogenous mRNA is covalently modified with a 5′-5′ linked N^7-methyl guanosine (m^7G) cap
  - Protects 5′ end of the mRNA from exonucleolytic attack and promotes translation[1]
  - Multiple cap-specific enzymes involved
- In vitro transcription of the mRNA vaccine from linearized plasmid DNA mimics this by reaction control of four bases and a special 5’ cap

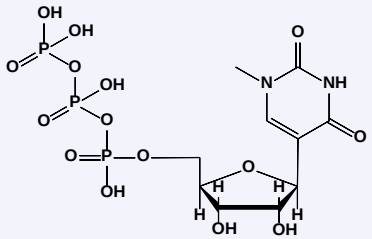


- In vitro nucleotide raw materials:

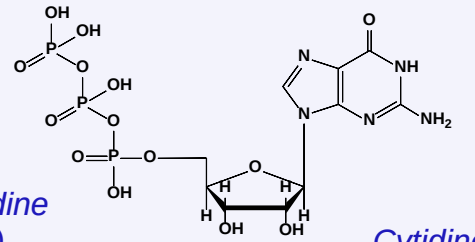
5'-cap analog ($m2^{7,3'\text{-}O}$Gppp(m12’-O)ApG)



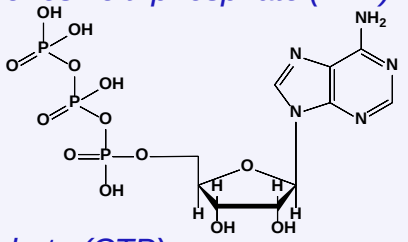
N1-methylpseudouridine triphosphate ($^{m1}\Psi$TP)



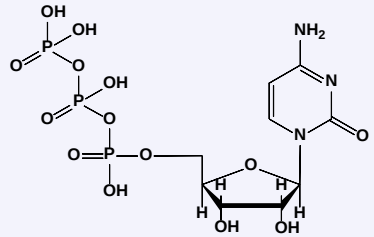
Guanosine triphosphate (GTP)



Adenosine triphosphate (ATP)



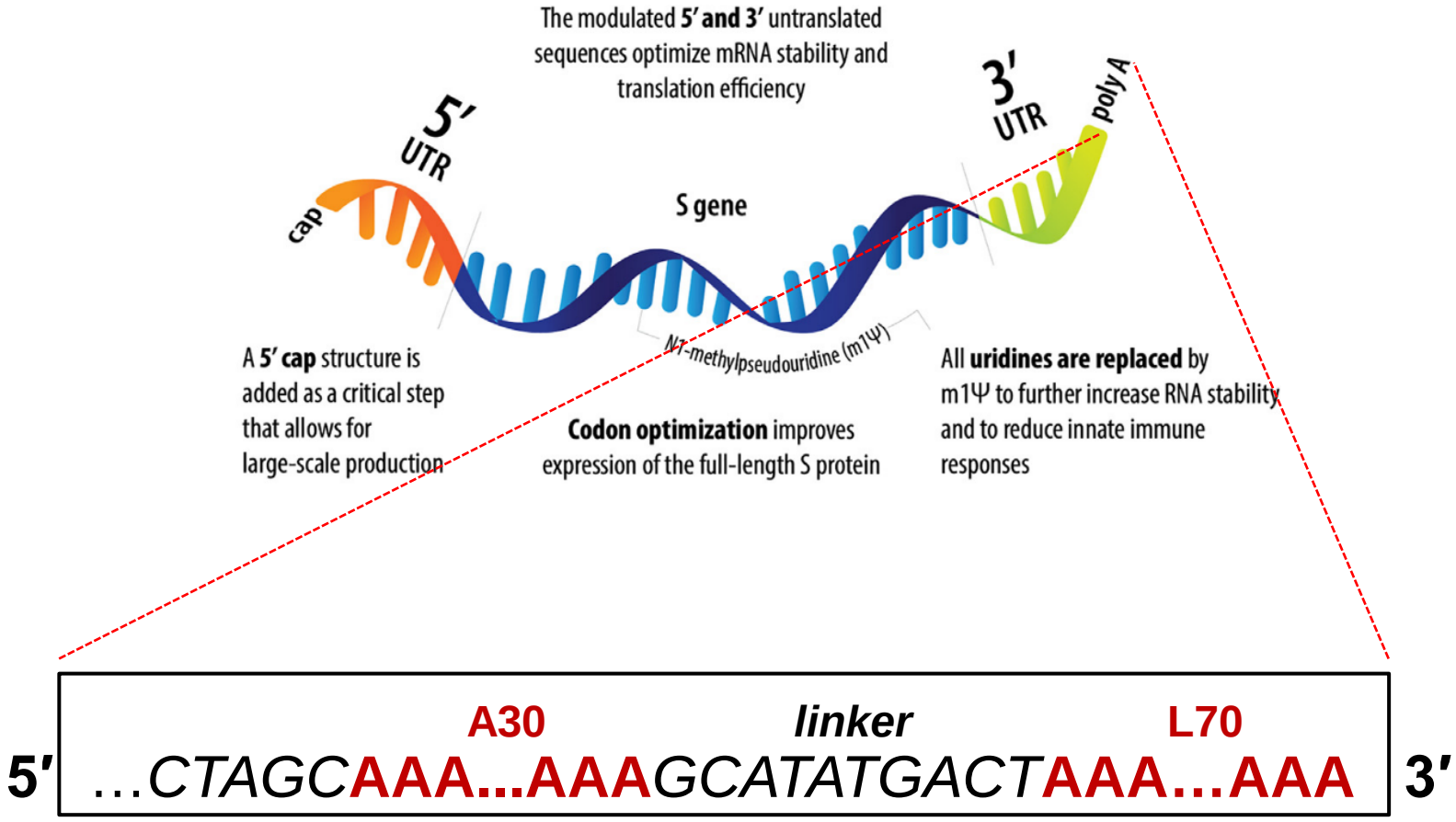
Cytidine triphosphate (CTP)



[1]Furuichi, Y., et al *Nature* ***266,*** 235-239 (1977)

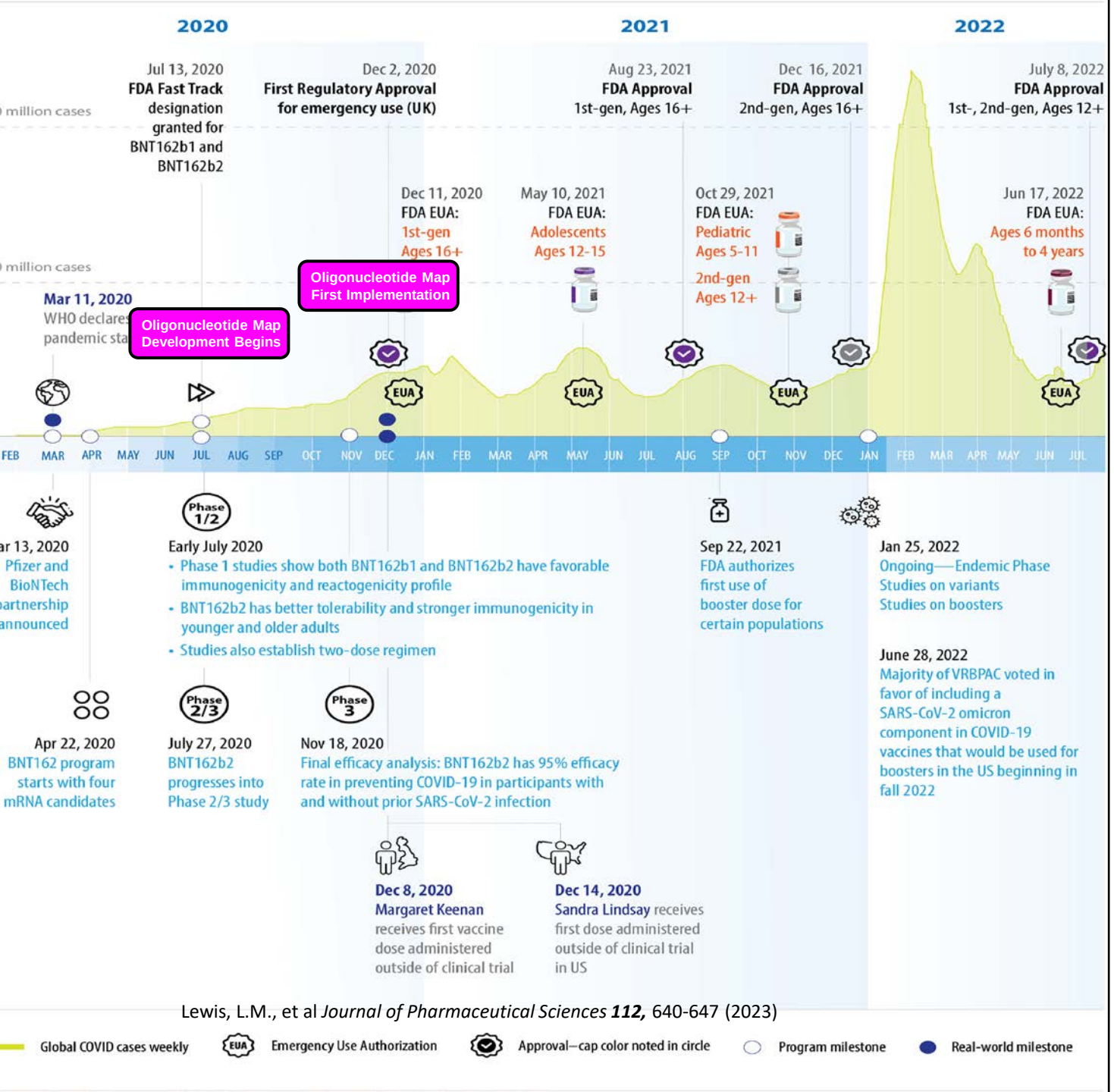
BNT162b2 mRNA has a Poly(A) Tail at Its 3′ End

- The poly(A) tail is important for nuclear export, RNA stability and translational efficiency[1]
- The DNA plasmid encodes for the poly(A) tail
- Polymerase transcriptional slippage gives rise to multiple poly(A) species[2]
  - Usually a series of species, each different from the previous smaller by the incorporation of a single adenosine nucleotide
  - Bias towards more A than coded for by DNA



[1]Wu, X. & Brewer, G. *Gene* ***500,*** 10-21 (2012)
[2]Beverly, M., et al *Analytical and Bioanalytical Chemistry* ***410***, 1667-1677 (2018)

Oligonucleotide Mapping of mRNA Primary Structure by LC-UV-MS/MS has Supported Regulatory Filings And Launches in 180+ Markets Globally



Lewis, L.M., et al *Journal of Pharmaceutical Sciences* ***112,*** 640-647 (2023)

Oligonucleotide Mapping Provides

- **Direct Primary Structure Understanding**
  - 5′ terminus cap heterogeneity
  - 3′ terminus poly(A) tail heterogeneity
  - Full-length mRNA
- **Orthogonal Identity**
  - BNT162b2 (Original)
  - Variant constructs (Delta, Omicron)
- **Batch Comparability Assessment**
  - Process changes
  - Scale-up
  - Scale-out

Supporting Regulatory Leaflets for Numerous EUAs/MAAs/BLAs

- **3.2.S.3.1 (Elucidation of Structure)**
- **3.2.S.2.6 (Comparability)**

Fully Annotated Oligonucleotide Map Generated by a Robust Workflow

Rapid One-Pot, One-Enzyme Sample Prep

Ribonuclease T_1
Cleave 3′ to G

+

IP-RP-HPLC-UV

MS-MS/MS

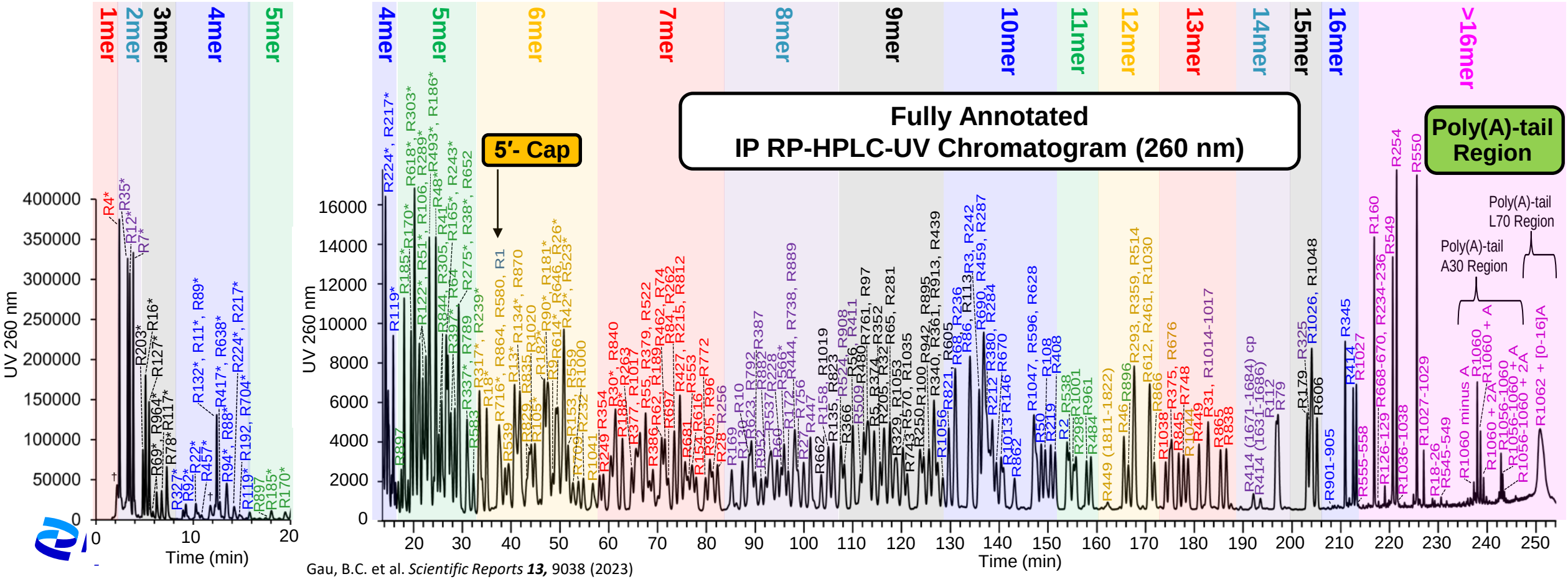
+

Semi-Automated Data Analysis

Commercial Software

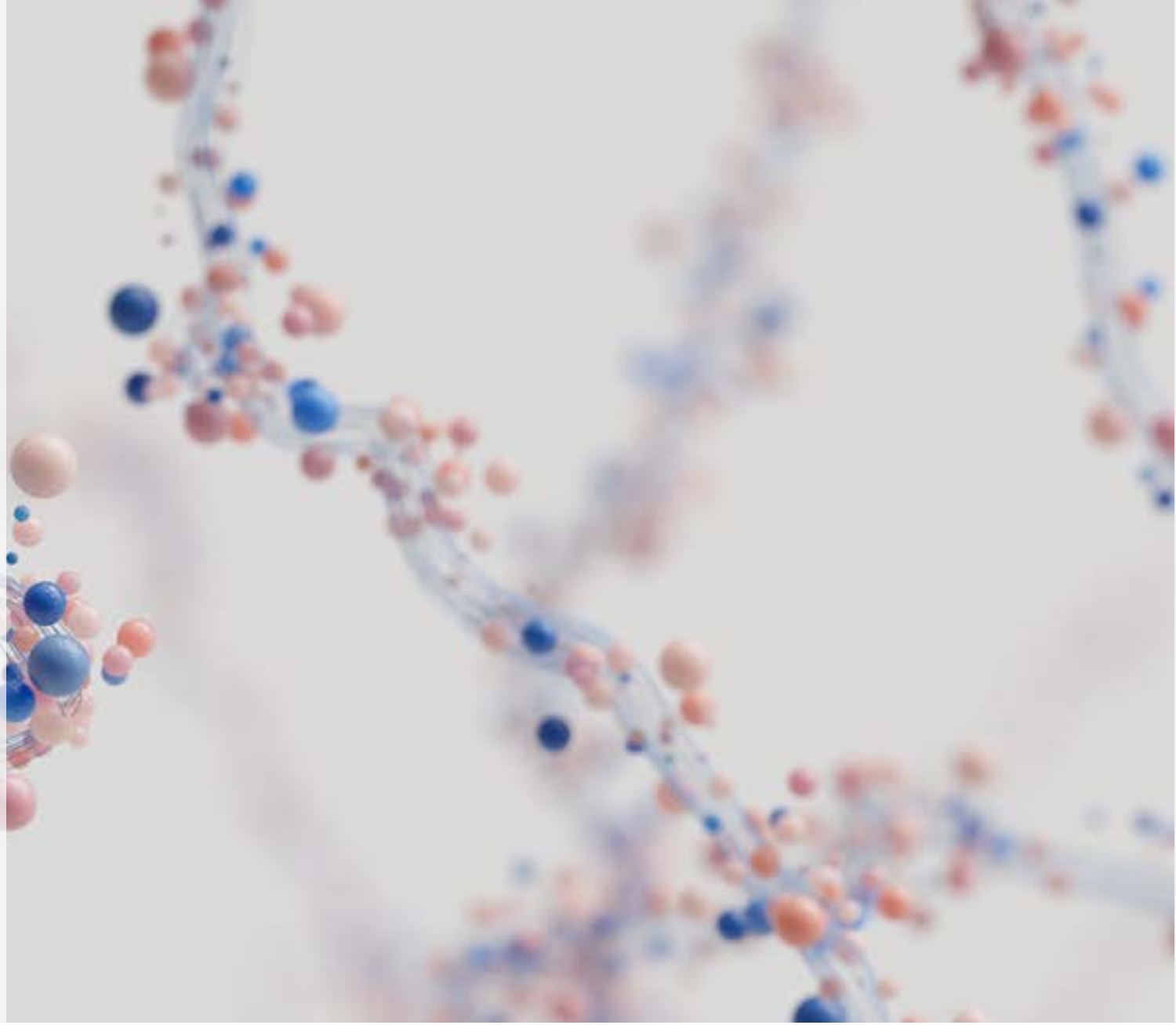
+

In-House VBA Tools



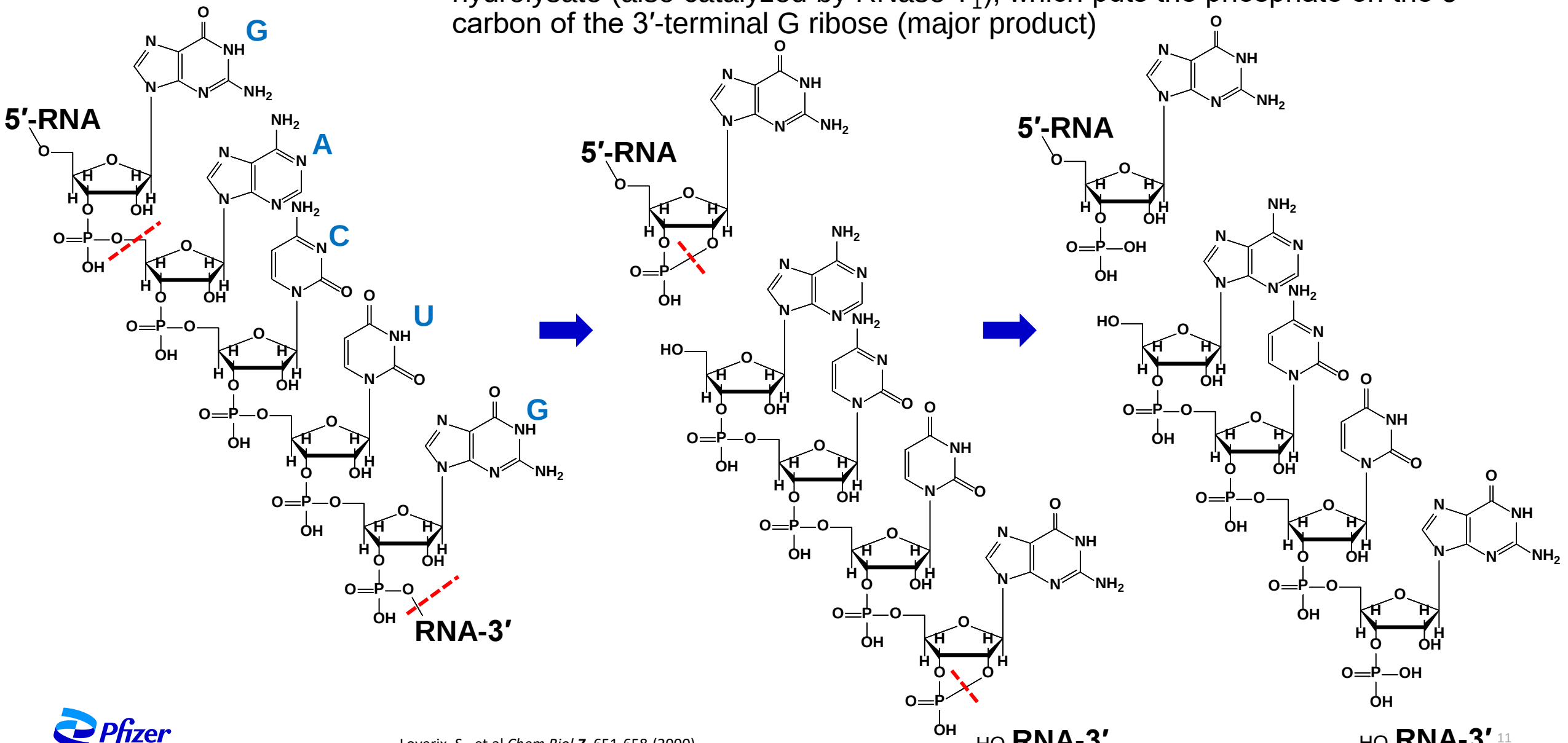
Gau, B.C. et al. *Scientific Reports* ***13,*** 9038 (2023)

Oligonucleotide Mapping Considerations



RNase T_1 is the Trypsin Analog

- RNase T_1 cuts after every G
- Two reaction products: the 3′ cyclic phosphate (minor product), and its hydrolysate (also catalyzed by RNase T_1), which puts the phosphate on the 3′ carbon of the 3′-terminal G ribose (major product)

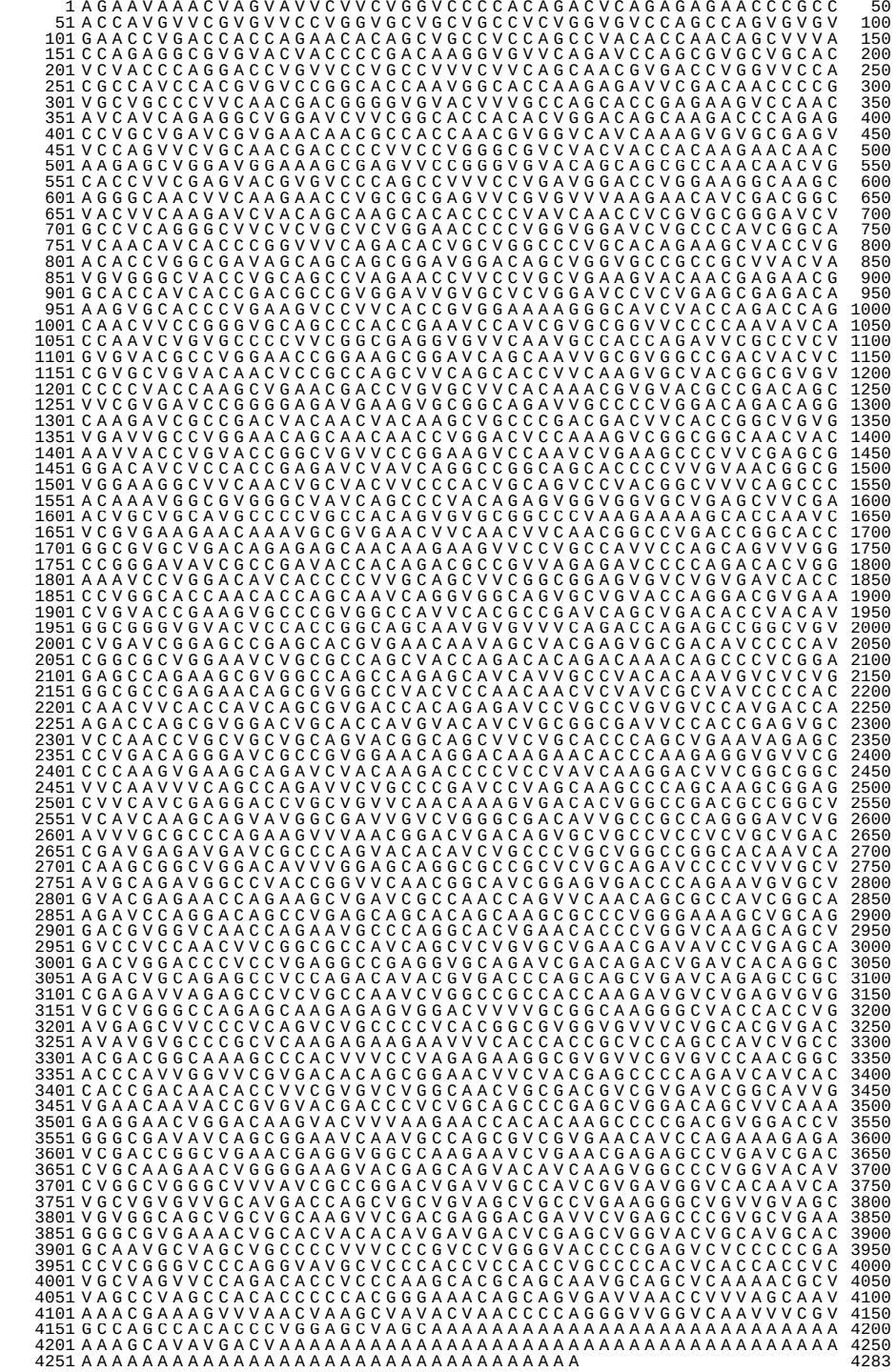


Loverix, S., et al *Chem Biol* **7,** 651-658 (2000)

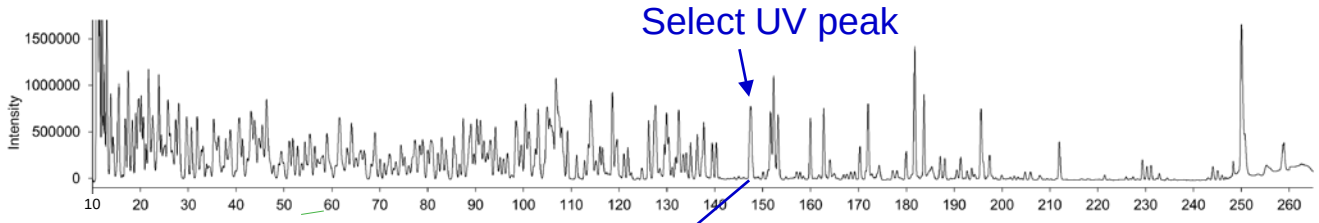
BNT162b2 mRNA has 4283 Residues, From 4 Nucleotide Building Blocks

- Oligonucleotide mapping is highly analogous to peptide mapping
  - RNase T_1 is the Trypsin analog: it cuts after every G
  - The RNase T_1 digest product oligonucleotides are analyzed by LC-UV-MS/MS, just as Trypsin digest product peptides are
- *But,* recognize the inherent chemical differences between protein and RNA
  - 4 nucleotides, not 20 amino acids
  - Much longer (4000 nt) than typical protein (200 AA)
- Whereas it is highly unlikely a peptide will have any sequence isomers in a protein's trypsin digest, it is assured that many oligonucleotides will have sequence isomers in the ensemble of a construct's RNase T_1 digestion products

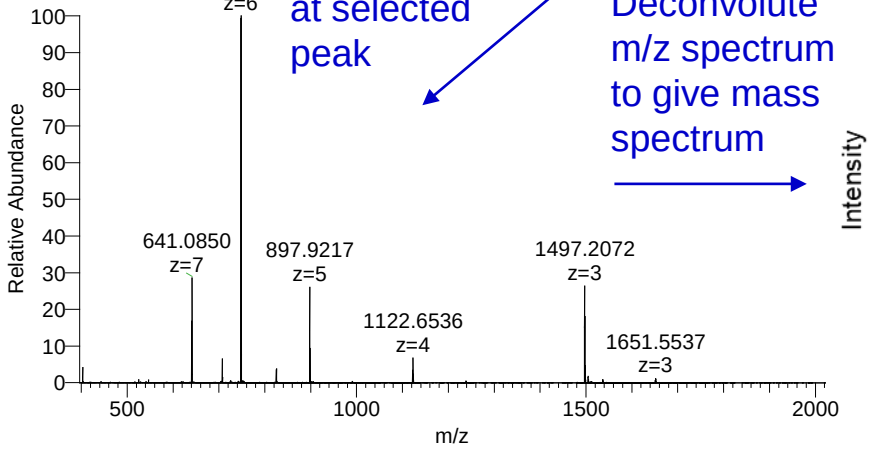
"V" = N1-methylpseudouridine



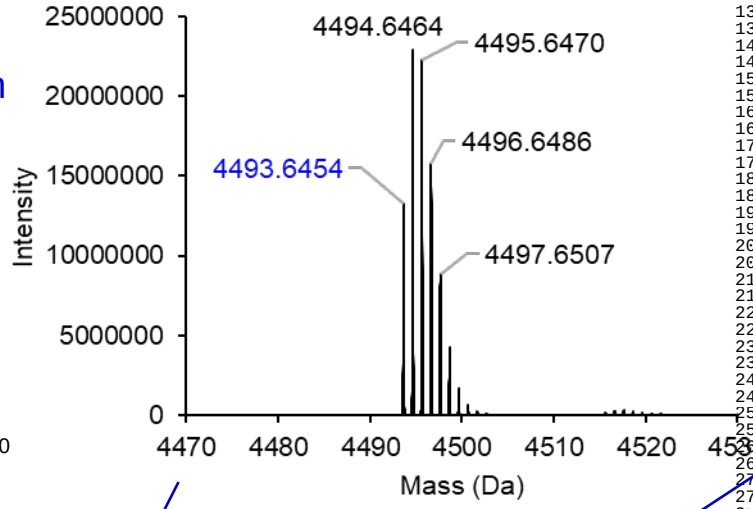
LC-UV-MS Peak ID Narrows the Possibilities, But Is Not Definitive



m/z spectrum at selected peak



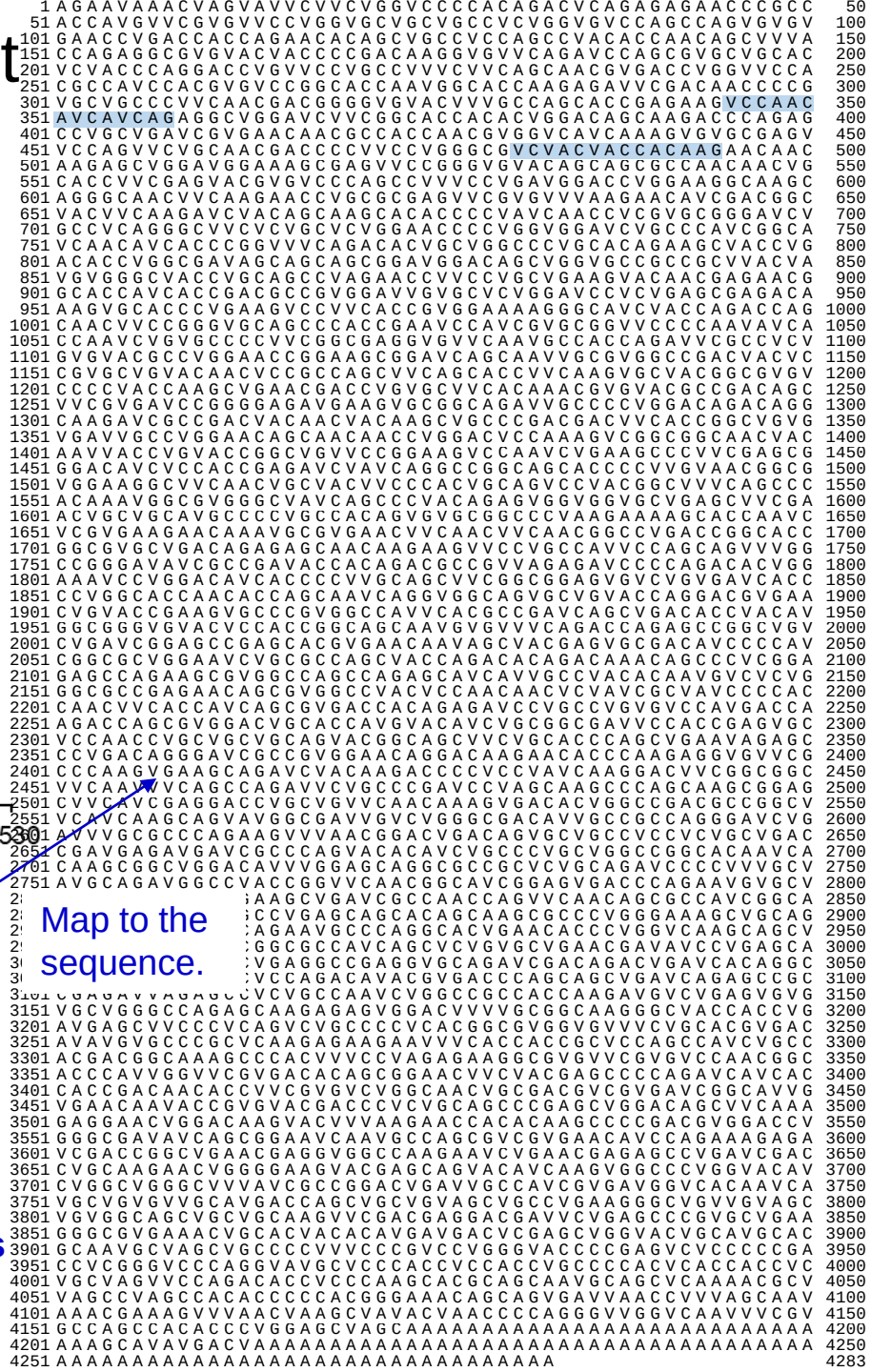
Deconvolute m/z spectrum to give mass spectrum



Match high resolution monoisotopic mass w/ theoretical RNase T_1 digest of construct

Observed Mass	Sequence	Start	End	Length	Theoretical Mass	Error (ppm)
4493.6454	VCCAACAVCAVCAG	345	358	14	4493.6499	-1.0
4493.6454	VCVACVACCACAAG	481	494	14	4493.6499	-1.0

Map to the sequence.



Is the UV peak VCCAACAVCAVCAG, VCVACVACCACAAG, or a mixture of peak isomers?

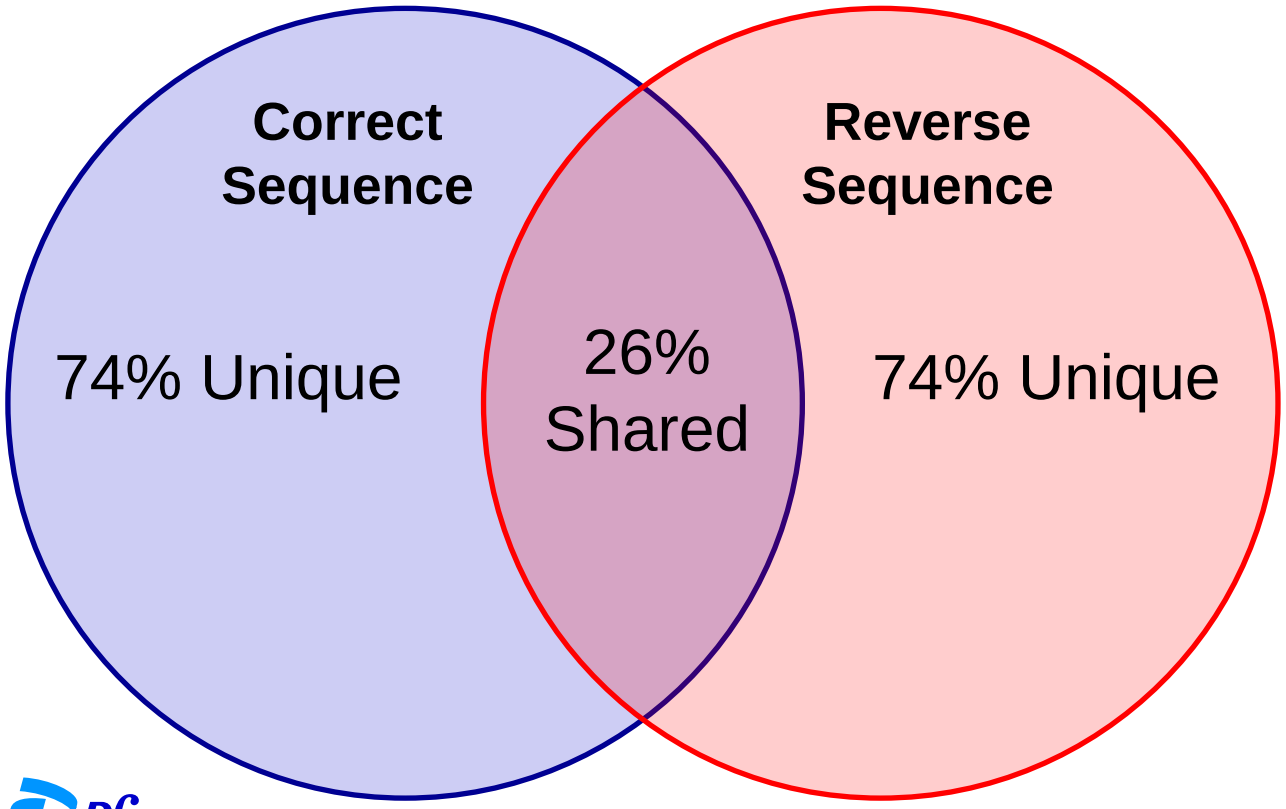
MS/MS fragmentation sequencing is key for determining the identity of sequence isoforms

Pfizer

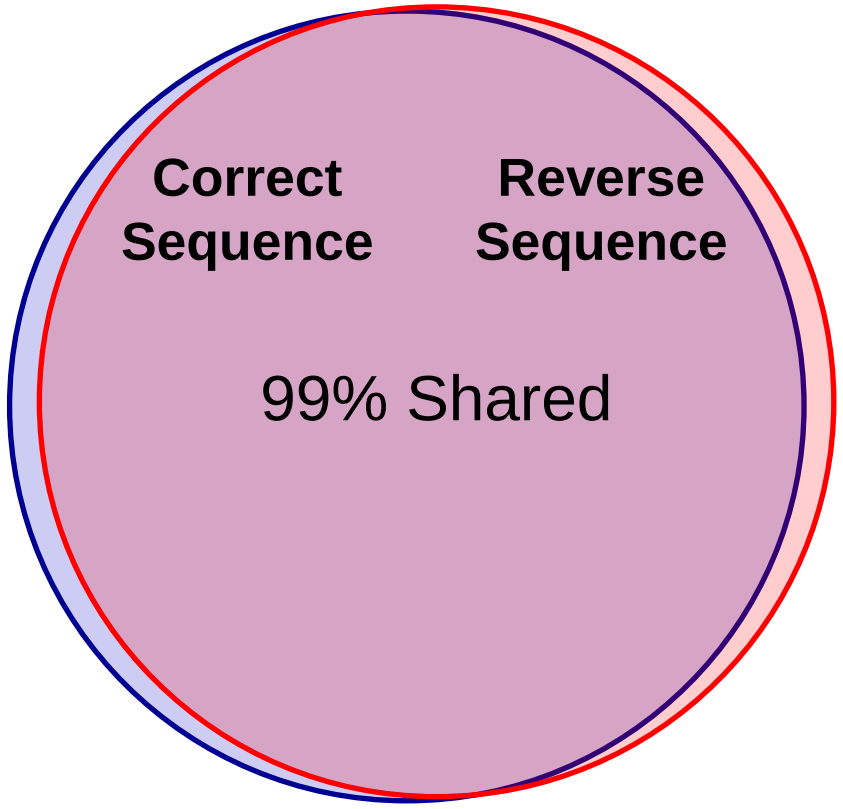
Sequence Isomers Present a Significant Analytical Challenge

- To illustrate: compare the set of 302 theoretical RNaseT$_1$ digestion products from BNT162b2 to the 302 theoretical digestion products from a construct having the reverse sequence (not the complement):

RNaseT$_1$-generated oligonucleotide sequence overlap

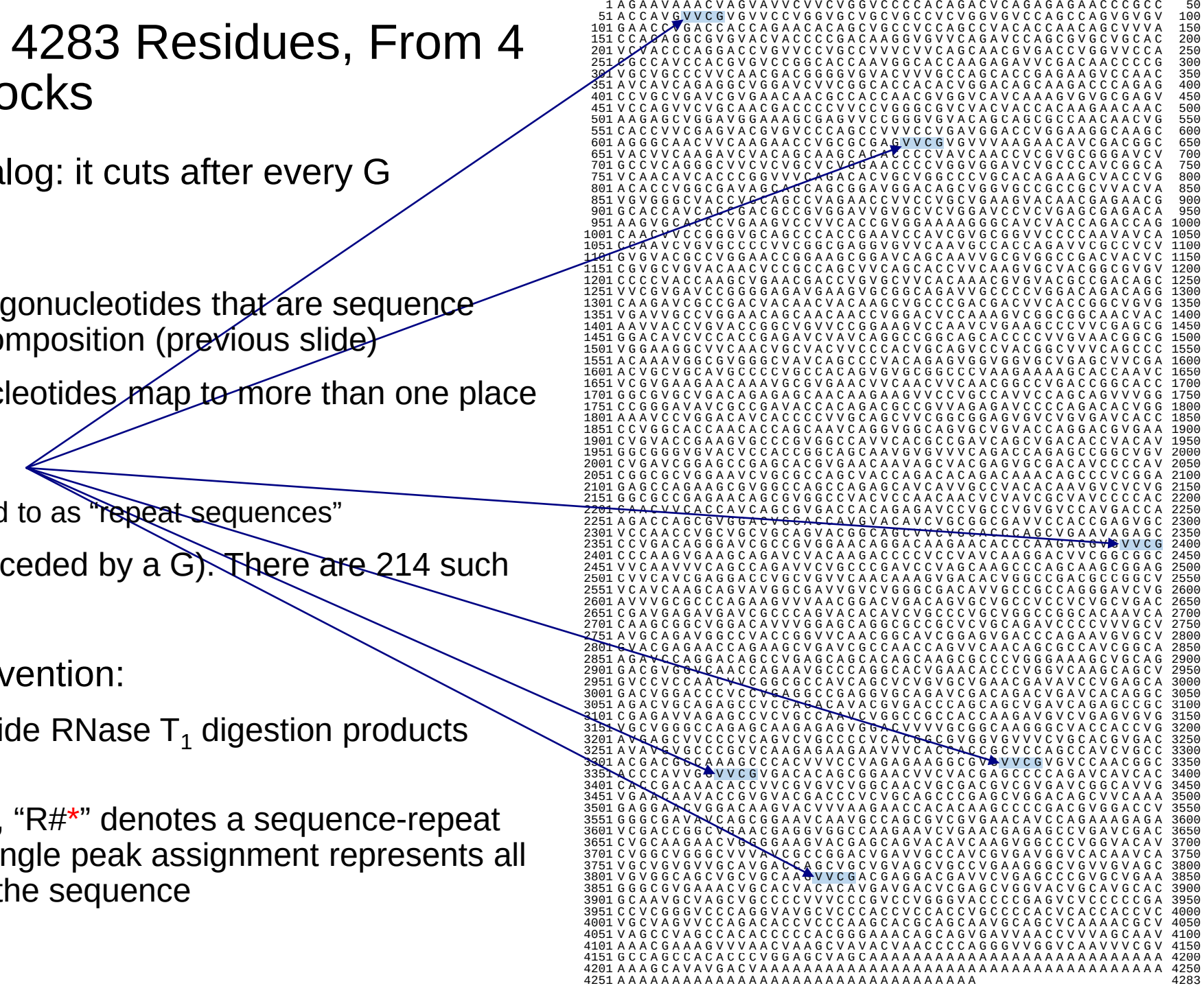


RNaseT$_1$-generated oligonucleotide mass overlap

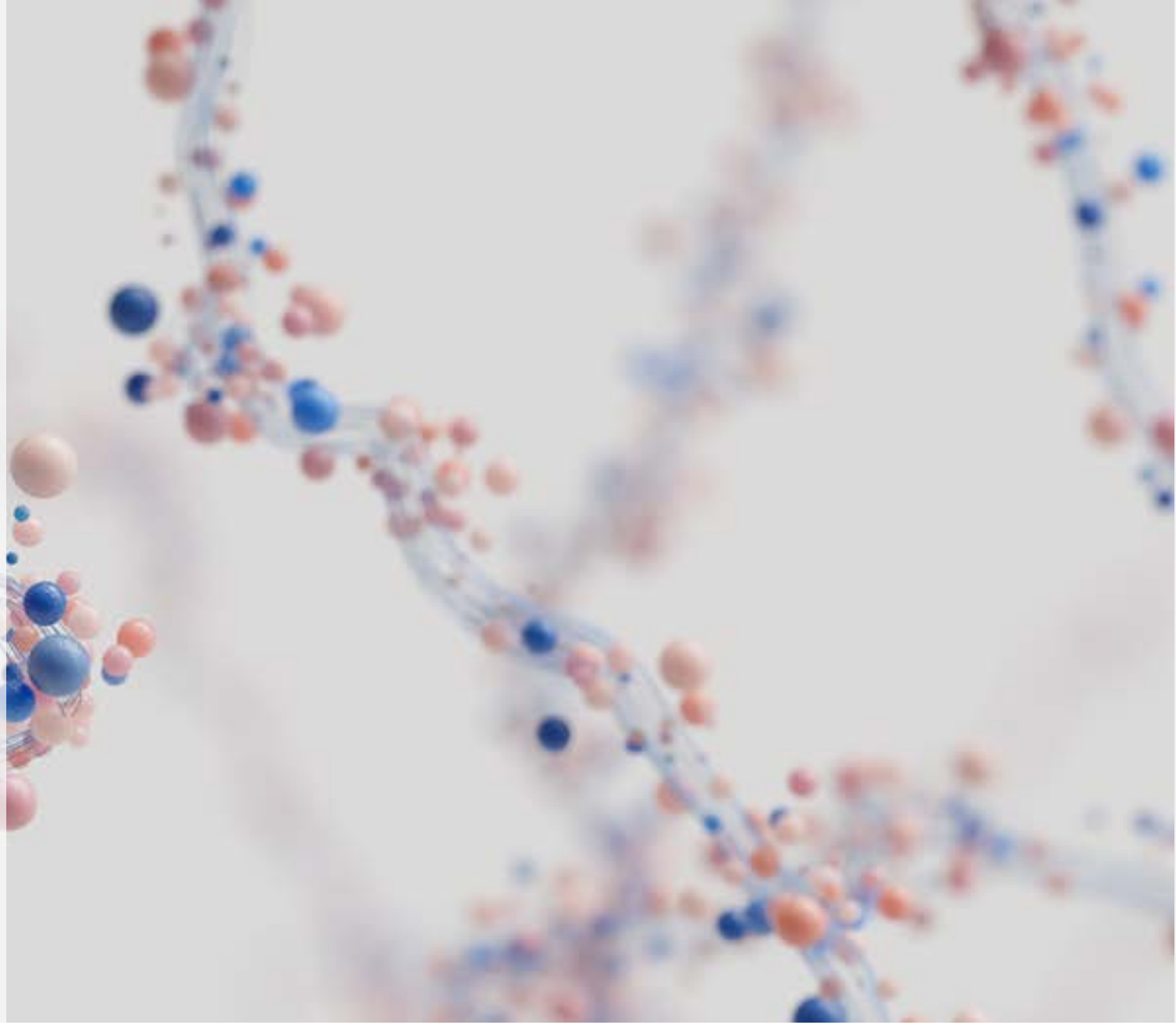


BNT162b2 mRNA has 4283 Residues, From 4 Nucleotide Building Blocks

- RNase T_1 is the Trypsin analog: it cuts after every G
- In an RNase T_1 digest,
  - There will be many digest oligonucleotides that are sequence isomers sharing the same composition (previous slide)
  - Many shorter digest oligonucleotides map to more than one place (locus) in the sequence
    - For example, **VVCG** has 6 loci
    - These are sometimes referred to as “repeat sequences”
  - The shortest is G (that is preceded by a G). There are 214 such loci
- Annotation and naming convention:
  - “R#” represents oligonucleotide RNase T_1 digestion products indexed from the 5′ to 3′ end
  - In chromatogram annotation, “R#*” denotes a sequence-repeat oligonucleotide, where the single peak assignment represents all identical oligonucleotides in the sequence



Oligonucleotide Mapping of BNT162b2 mRNA Primary Structure by LC-UV-MS/MS



Oligonucleotide Mapping Sample Handling & Acquisition

- Sample Handling
  - 50 µg + 2500 U RNase T_1 + buffer/EDTA to a final vol of 35 µL, 50 mM Tris pH 7.5, 20 mM EDTA, in glass total-recovery autosampler vial
  - Digest progressed 1 h at 37 °C, then stored at -80 °C until analysis
- IP-RP-UHPLC-UV
  - System: Agilent 1290 BioInert
  - Column: Waters ACQUITY PREMIER Oligonucleotide C18 Column, 130Å, 1.7 µm, 2.1 x 150 mm
  - Mobile phase A: 0.1% TEA (triethylamine), 1% HFIP (hexafluoroisopropanol), Water
  - Mobile phase B: 0.1% TEA, 1% HFIP, 50% Methanol
  - 5 h method gradient: 1% → 17% B, 195 min, then 17% → 38% B, 70 min, 0.2 mL/min, 60 °C
  - PDA detector; monitor 260 nm, 4 nm bandwidth with reference at 360 nm, 20 nm bandwidth
- HRMS/MS
  - System: Orbitrap Thermo Eclipse
  - Source: negative mode, 2700 V, 40 Sheath Gas, 10 Aux Gas, 320 °C Ion Transfer Tube, 300 °C Vaporizer
  - HRMS, main segment: 120000 RP (at 400 m/z), 50 ms max inj time, 100% AGC target, 1 microscan, 450 – 2000 m/z, 0-240 min
  - HRMS, poly(A) segment: 120000 RP, 300 ms max inj time, 250% AGC target, 5 microscans, 700 – 2000 m/z, >240 min
  - MS/MS, main segment: HCD fragmentation, 17/21/25 stepped collision energy (%), 30000 RP Orbitrap fragment scan, 2 min cycle time, DDA precursor selection, dynamic exclusion 6 sec, 300 ms max inj time, 250% AGC target, 1 microscan

Comprehensive, Semi-Automated, High-Fidelity Data Analysis Workflow

Semi-Automated Data Analysis Workflow

1. Automated Search

- Mass table by retention time
- Identifications (72% Coverage)

2. Automated LC-UV Annotation

- Match Peak IDs to Chromatogram
- Reformatted Mass Table

3. Supplement LC-UV Annotation

- Data mining & MS/MS Analysis Tools

4. Supplement Missing Coverage

- Data mining & MS/MS Analysis Tools

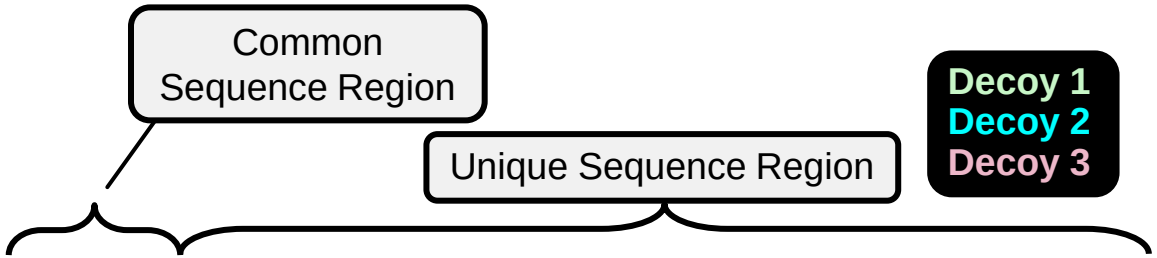
5. Add 5' & 3' Termini Characterization

Final Reportables

- Fully-Annotated Chromatographic Map
- Sequence Coverage Calculation & Map
- Curated Mass Table
- 5′ & 3′ terminus characterization

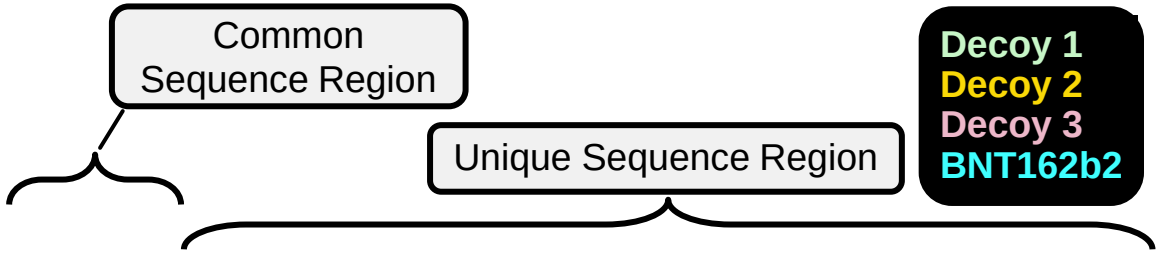
Verification by Decoy Searching

Decoy search excluding BNT162b2 mRNA construct



No Preferential Match

Decoy search including BNT162b2 mRNA construct



Preferential Match to BNT162b2

Fully Annotated Oligonucleotide Map Generated by a Robust Workflow

Rapid One-Pot, One-Enzyme Sample Prep

Ribonuclease T_1
Cleave 3′ to G

+

IP-RP-HPLC-UV

MS-MS/MS

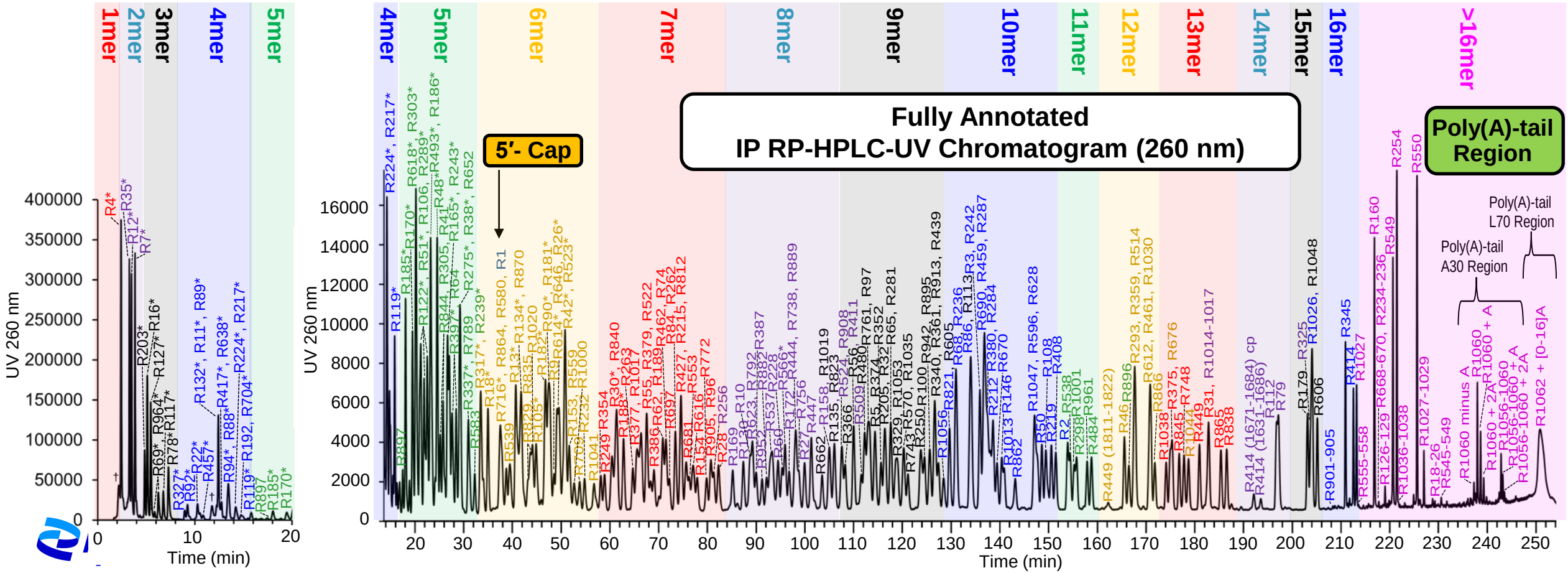
+

Semi-Automated Data Analysis

Commercial Software

+

In-House VBA Tools



100% BNT162b2 Sequence Coverage Observed

Approximately half of consensus RNase T_1 cleavages map to one locus and half contain multiple loci

232 oligonucleotides (48.8% Coverage) — **One Locus**

70 oligonucleotides (51.2% Coverage) — **Multiple Loci**

46 oligonucleotides / **14 oligonucleotides** — **Missed-Cleavages & Non-Consensus Cleavages**

1 AGAAVAAACVAGVAVVCVVCVGGVCCCCACAGACVCAGAGAACCCGCCACCAVGVVCGVGVVCCVGGVGCVGCVGCCVCVGGVGVCCAGCCAGVGVGAACCVGACCACCAGAACAC 120

121 AGCVGCCVCCAGCCVACACCAACAGCVVVACCAGAGGCGVGVACVACCCCGACAAGGVGVVCAGAVCCAGCGVGCVGCACVCVACCCAGGACCVGVVCCVGCCVVVCVVCAGCAACGVGA 240

241 CCVGGVVCCACGCCAVCCACGVGVCCGGCACCAAVGGCACCAAGAGAVVCGACAACCCCGVGCVGCCCVVCAACGACGGGGVGVACVVVGCCAGCACCGAGAAGVCCAACAVCAVCAGAG 360

361 GCVGGAVCVVCGGCACCACACVGGACAGCAAGACCCAGAGCCVGCVGAVCGVGAACAACGCCACCAACGVGGVCAVCAAAGVGVGCGAGVVCCAGVVCVGCAACGACCCCVVCCVGGGCG 480

481 VCVACVACCACAAGAACAACAAGAGCVGGAVGGAAAGCGAGVVCCGGGVGVACAGCAGCGCCAACAACVGCACCVVCGAGVACGVGVCCCAGCCVVVCCVGAVGGACCVGGAAGGCAAGC 600

601 AGGGCAACVVCAAGAACCVGCGCGAGVVCGVGVVVAAGAACAVCGACGGCVACVVCAAGAVCVACAGCAAGCACACCCCVAVCAACCVCGVGCGGGAVCVGCCVCAGGGCVVCVCVGCVC 720

721 VGGAACCCCVGGVGGAVCVGCCCAVCGGCAVCAACAVCACCCGGVVVCAGACACVGCVGGCCCVGCACAGAAGCVACCVGACACCVGGCGAVAGCAGCAGCGGAVGGACAGCVGGVGCCG 840

841 CCGCVVACVAVGVGGGCVACCVGCAGCCVAGAACCVVCCVGCVGAAGVACAACGAGAACGGCACCAVCACCGACGCCGVGGAVVGVGCVCVGGAVCCVCVGAGCGAGACAAAGVGCACCC 960

961 VGAAGVCCVVCACCGVGGAAAAGGGCAVCVACCAGACCAGCAACVVCCGGGVGCAGCCCACCGAAVCCAVCGVGCGGVVCCCCAAVAVCACCAAVCVGVGCCCCVVCGGCGAGGVGVVCA 1080

1081 AVGCCACCAGAVVCGCCVCVGVGVACGCCVGGAACCGGAAGCGGAVCAGCAAVVGCGVGGCCGACVACVCCGVGCVGVACAACVCCGCCAGCVVCAGCACCVVCAAGVGCVACGGCGVGV 1200

1201 CCCCVACCAAGCVGAACGACCVGVGCVVCACAAACGVGVACGCCGACAGCVVCGVGAVCCGGGGAGAVGAAGVGCGGCAGAVVGCCCCVGGACAGACAGGCAAGAVCGCCGACVACAACV 1320

1321 ACAAGCVGCCCGACGACVVCACCGGCVGVGVGAVVGCCVGGAACAGCAACAACCVGGACVCCAAAGVCGGCGGCAACVACAAVVACCVGVACCGGCVGVVCCGGAAGVCCAAVCVGAAGC 1440

1441 CCVVCGAGCGGGACAVCVCCACCGAGAVCVAVCAGGCCGGCAGCACCCCVVGVAACGGCGVGGAAGGCVVCAACVGCVACVVCCCACVGCAGVCCVACGGCVVVCAGCCCACAAAVGGCG 1560

1561 VGGGCVAVCAGCCCVACAGAGVGGVGGVGCVGAGCVVCGAACVGCVGCAVGCCCCVGCCACAGVGVGCGGCCCVAAGAAAAGCACCAAVCVCGVGAAGAACAAAVGCGVGAACVVCAACV 1680

1681 VCAACGGCCVGACCGGCACCGGCGVGCVGACAGAGAGCAACAAGAAGVVCCVGCCAVVCCAGCAGVVVGGCCGGGAVAVCGCCGAVACCACAGACGCCGVVAGAGAVCCCCAGACACVGG 1800

1801 AAAVCCVGGACAVCACCCCVVGCAGCVVCGGCGGAGVGVCVGVGAVCACCCCVGGCACCAACACCAGCAAVCAGGVGGCAGVGCVGVACCAGGACGVGAACVGVACCGAAGVGCCCGVGG 1920

1921 CCAVVCACGCCGAVCAGCVGACACCVACAVGGCGGGVGVACVCCACCGGCAGCAAVGVGVVVCAGACCAGAGCCGGCVGVCVGAVCGGAGCCGAGCACGVGAACAAVAGCVACGAGVGCG 2040

2041 ACAVCCCCAVCGGCGCVGGAAVCVGCGCCAGCVACCAGACACAGACAAACAGCCCVCGGAGAGCCAGAAGCGVGGCCAGCCAGAGCAVCAVVGCCVACACAAVGVCVCVGGGCGCCGAGA 2160

2161 ACAGCGVGGCCVACVCCAACAACVCVAVCGCVAVCCCCACCAACVVCACCAVCAGCGVGACCACAGAGAVCCVGCCVGVGVCCAVGACCAAGACCAGCGVGGACVGCACCAVGVACAVCV 2280

2281 GCGGCGAVVCCACCGAGVGCVCCAACCVGCVGCVGCAGVACGGCAGCVVCVGCACCCAGCVGAAVAGAGCCCVGACAGGGAVCGCCGVGGAACAGGACAAGAACACCCAAGAGGVGVVCG 2400

2401 CCCAAGVGAAGCAGAVCVACAAGACCCCVCCVAVCAAGGACVVCGGCGGCVVCAAVVVCAGCCAGAVVCVGCCCGAVCCVAGCAAGCCCAGCAAGCGGAGCVVCAVCGAGGACCVGCVGV 2520

2521 VCAACAAAGVGACACVGGCCGACGCCGGCVVCAVCAAGCAGVAVGGCGAVVGVCVGGGCGACAVVGCCGCCAGGGAVCVGAVVVGCGCCCAGAAGVVVAACGGACVGACAGVGCVGCCVC 2640

2641 CVCVGCVGACCGAVGAGAVGAVCGCCCAGVACACAVCVGCCCVGCVGGCCGGCACAAVCACAAGCGGCVGGACAVVVGGAGCAGGCGCCGCVCVGCAGAVCCCCVVVGCVAVGCAGAVGG 2760

2761 CCVACCGGVVCAACGGCAVCGGAGVGACCCAGAAVGVGCVGVACGAGAACCAGAAGCVGAVCGCCAACCAGVVCAACAGCGCCAVCGGCAAGAVCCAGGACAGCCVGAGCAGCACAGCAA 2880

2881 GCGCCCVGGGAAAGCVGCAGGACGVGGVCAACCAGAAVGCCCAGGCACVGAACACCCVGGVCAAGCAGCVGVCCVCCAACVVCGGCGCCAVCAGCVCVGVGCVGAACGAVAVCCVGAGCA 3000

3001 GACVGGACCCVCCVGAGGCCGAGGVGCAGAVCGACAGACVGAVCACAGGCAGACVGCAGAGCCVCCAGACAVACGVGACCCAGCAGCVGAVCAGAGCCGCCGAGAVVAGAGCCVCVGCCA 3120

3121 AVCVGGCCGCCACCAAGAVGVCVGAGVGVGVGCVGGGCCAGAGCAAGAGAGVGGACVVVVGCGGCAAGGGCVACCACCVGAVGAGCVVCCCVCAGVCVGCCCCVCACGGCGVGGVGVVVC 3240

3241 VGCACGVGACAVAVGVGCCCGCVCAAGAGAAGAAVVVCACCACCGCVCCAGCCAVCVGCCACGACGGCAAAGCCCACVVVCCVAGAGAAGGCGVGVVCGVGVCCAACGGCACCCAVVGGV 3360

3361 VCGVGACACAGCGGAACVVCVACGAGCCCCAGAVCAVCACCACCGACAACACCVVCGVGVCVGGCAACVGCGACGVCGVGAVCGGCAVVGVGAACAAVACCGVGVACGACCCVCVGCAGC 3480

3481 CCGAGCVGGACAGCVVCAAAGAGGAACVGGACAAGVACVVVAAGAACCACACAAGCCCCGACGVGGACCVGGGCGAVAVCAGCGGAAVCAAVGCCAGCGVCGVGAACAVCCAGAAAGAGA 3600

3601 VCGACCGGCVGAACGAGGVGGCCAAGAAVCVGAACGAGAGCCVGAVCGACCVGCAAGAACVGGGGAAGVACGAGCAGVACAVCAAGVGGCCCVGGVACAVCVGGCVGGGCVVVAVCGCCG 3720

3721 GACVGAVVGCCAVCGVGAVGGVCACAAVCAVGCVGVGVVGCAVGACCAGCVGCVGVAGCVGCCVGAAGGGCVGVVGVAGCVGVGGCAGCVGCVGCAAGVVCGACGAGGACGAVVCVGAGC 3840

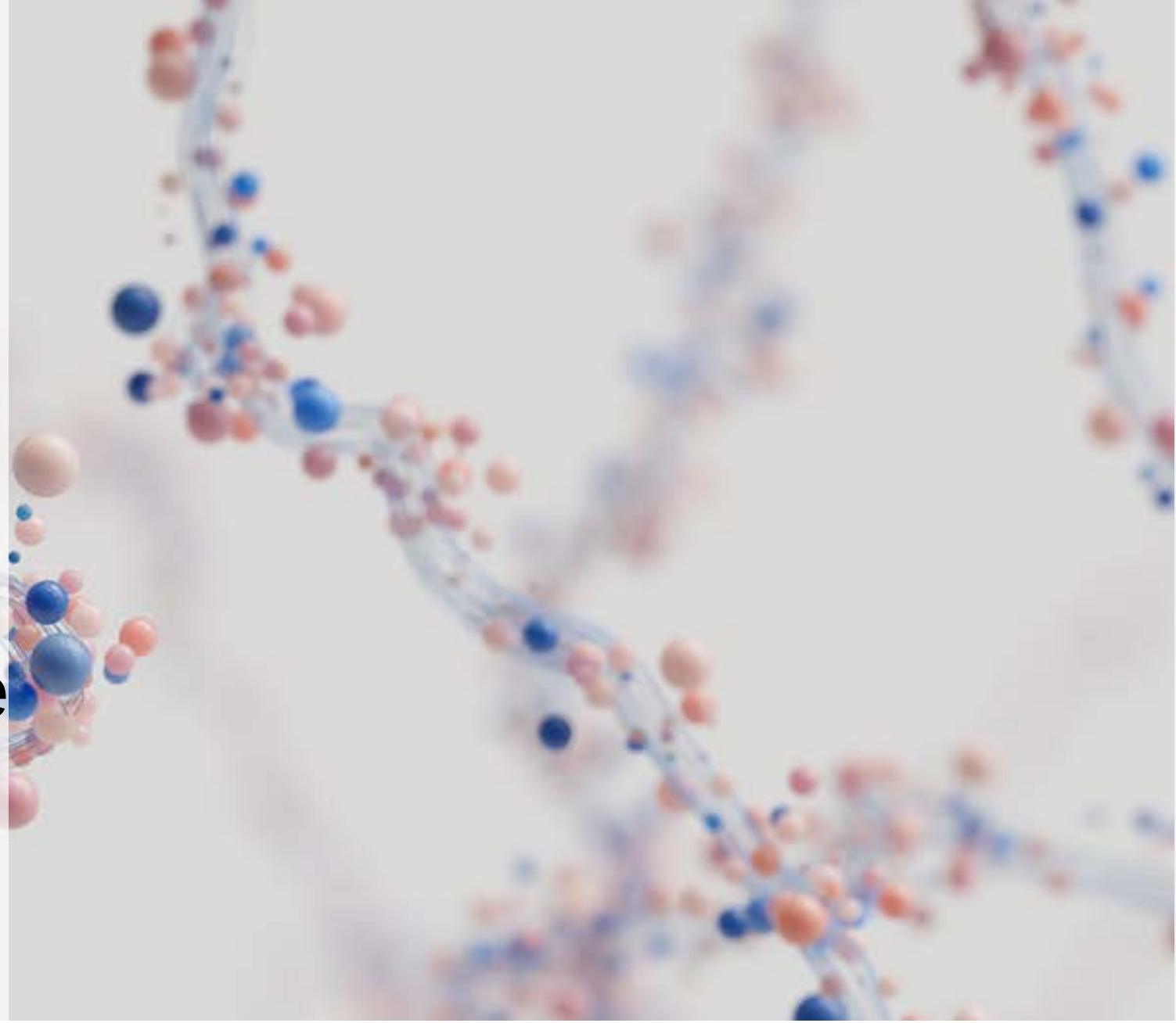
3841 CCGVGCVGAAGGGCGVGAAACVGCACVACACAVGAVGACVCGAGCVGGVACVGCAVGCACGCAAVGCVAGCVGCCCCVVVCCCGVCCVGGGVACCCCGAGVCVCCCCCGACCVCGGGVCC 3960

3961 CAGGVAVGCVCCCACCVCCACCVGCCCCACVCACCACCVCVGCVAGVVCCAGACACCVCCCAAGCACGCAGCAAVGCAGCVCAAAACGCVVAGCCVAGCCACACCCCCACGGGAAACAGC 4080

4081 AGVGAVVAACCVVVAGCAAVAAACGAAAGVVVAACVAAGCVAVACVAACCCCAGGGVVGGVCAAVVVCGVGCCAGCCACACCCVGGAGCVAGCAAAAAAAAAAAAAAAAAAAAAAAAAAA 4200

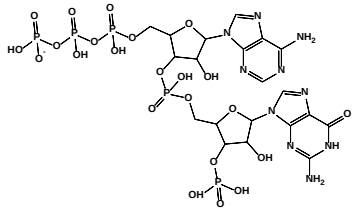
4201 AAAGCAVAVGACVAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4283

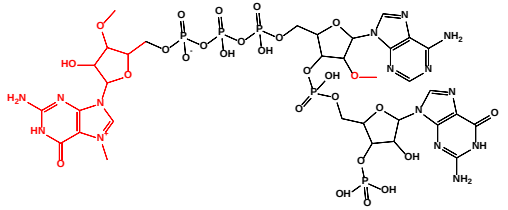
Utility of Oligonucleotide Mapping



Oligonucleotide Mapping Enables Simultaneous Characterization of the 5′ Terminus Without Affinity Purification

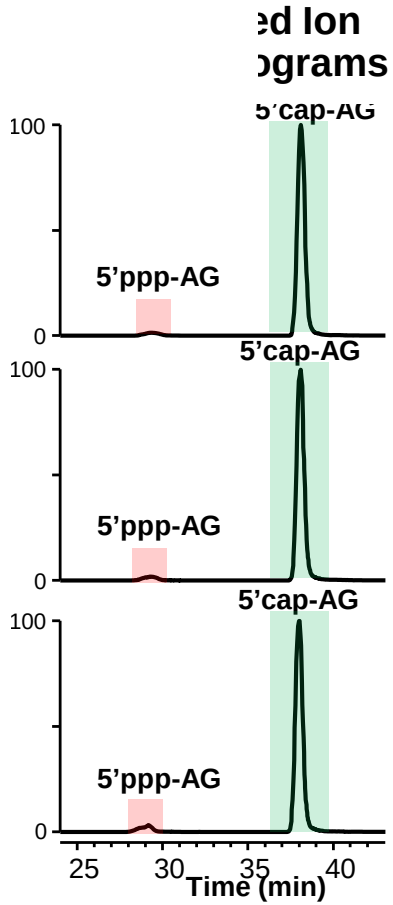
- Translationally-competent mRNA is capped at its 5′ end
  - Degree of capping is a CQA
- Three constructs shown
  - Comirnaty “Original” is BNT162b2 mRNA. It encodes the first spike protein with 2 two stabilizing proline mutations
  - “Delta” mRNA encodes the first Delta Covid-19 spike protein variant
  - “Omicron” mrNA encodes the first Omicron Covid-19 spike protein variant
- Majority of 5′ terminus is capped

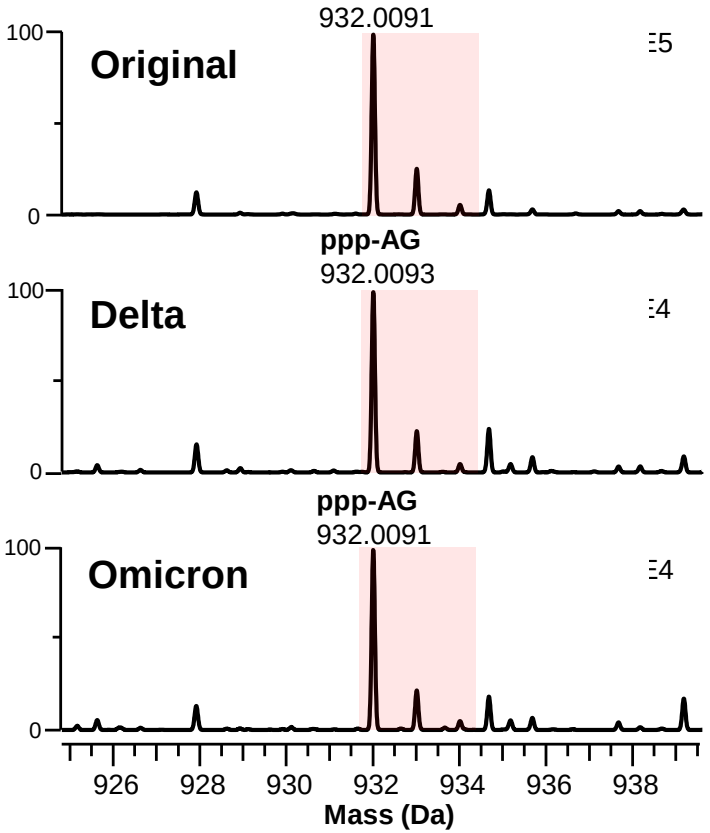


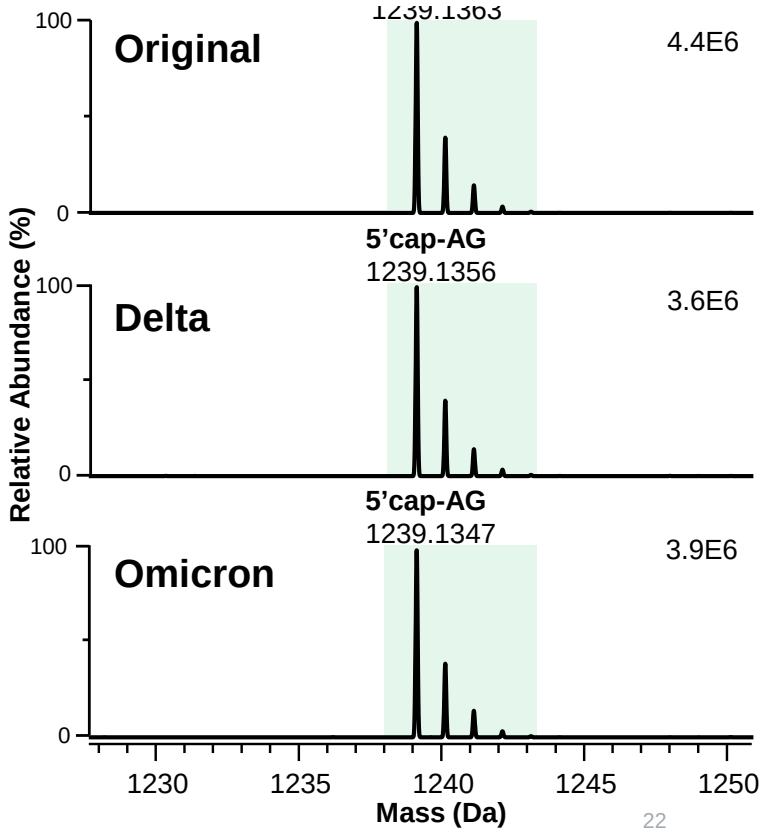


C **Capped 5´ terminus**

Deconvolved Mass Spectra

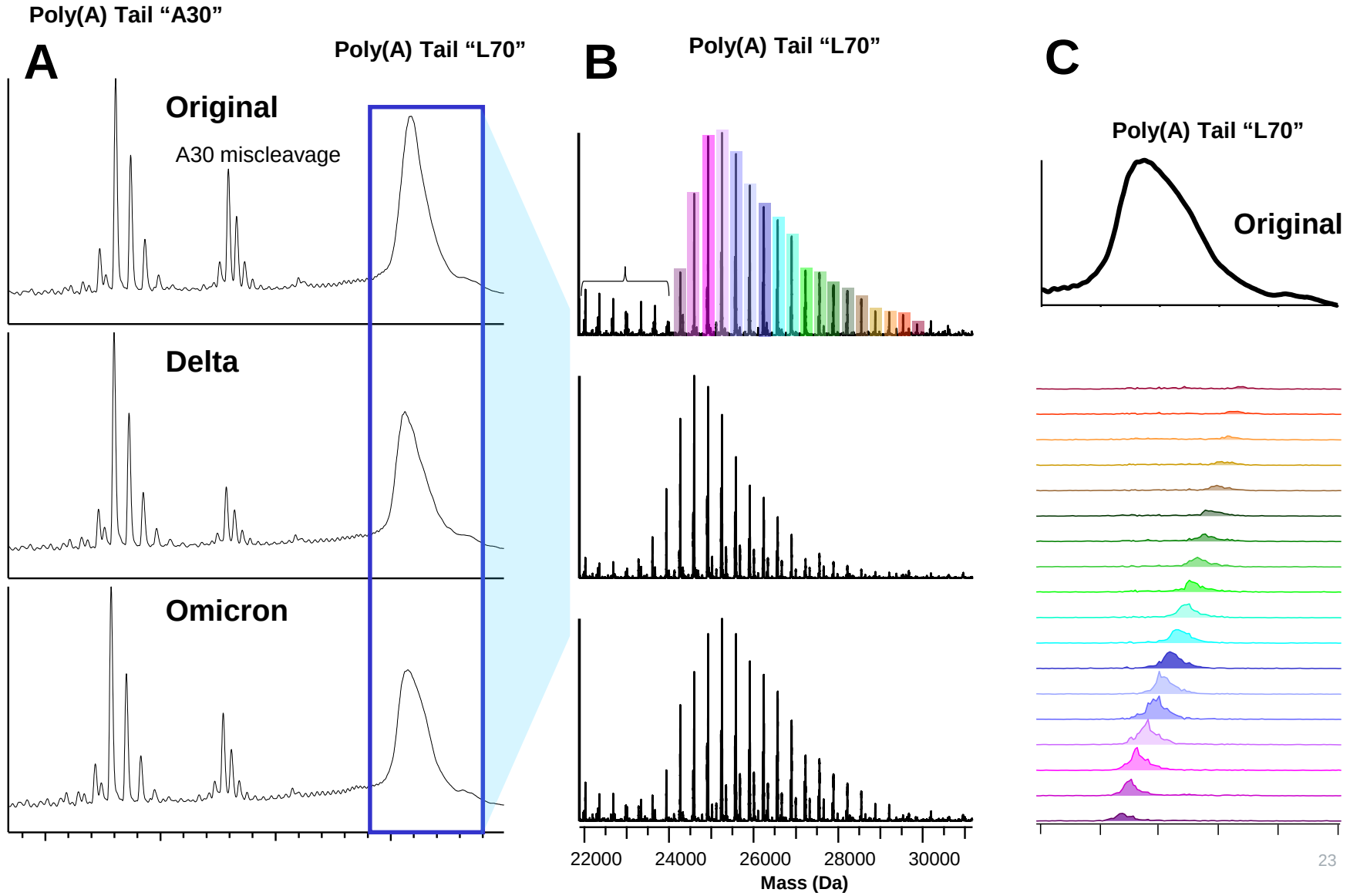




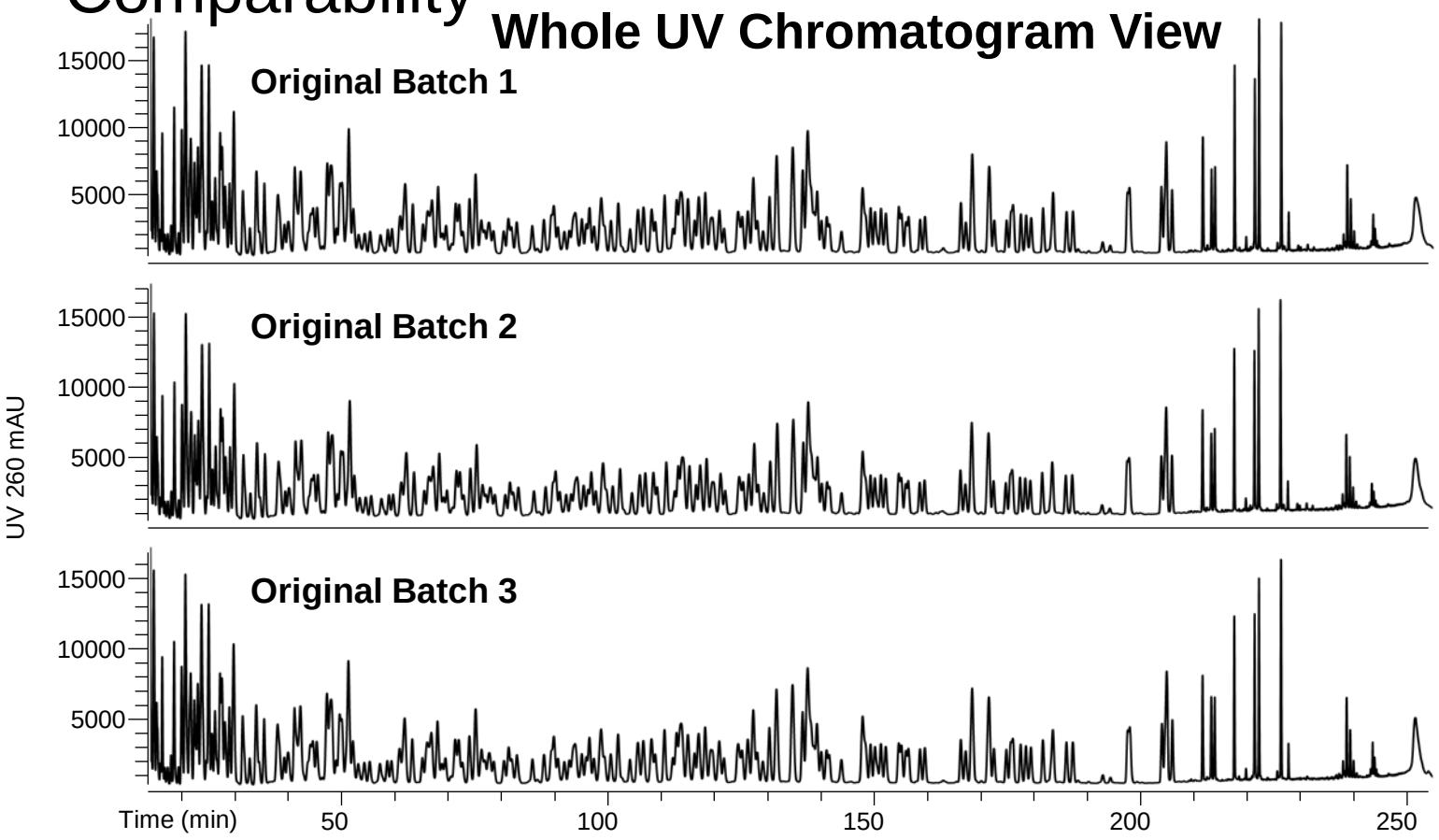


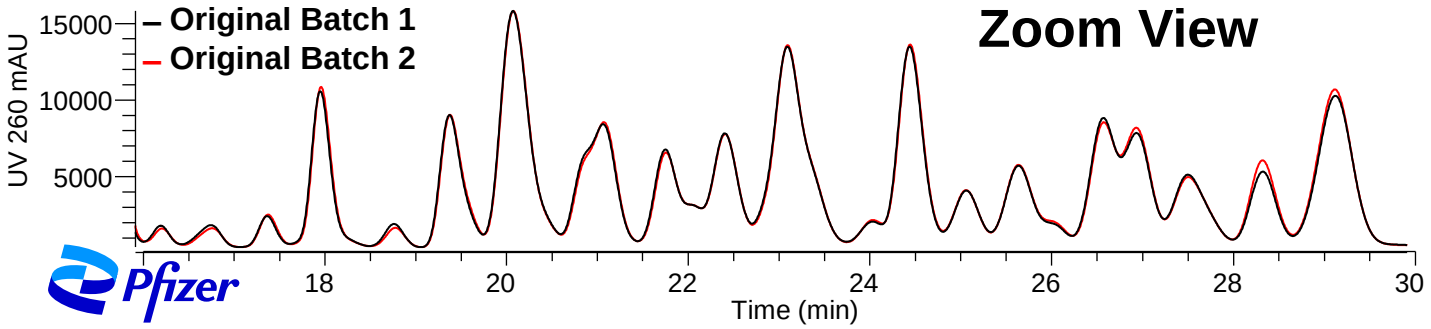
Oligonucleotide Mapping of mRNA Enables Simultaneous Characterization of the 3′ Terminus Without Affinity Purification

- Translationally-competent mRNA needs a 3′ Poly(A) tail
  - The Comirnaty and variant constructs' 3′ Poly(A) tail is designed to be 100 A's split by a short oligonucleotide linker to "A30" (30 A's) and "L70" (70 A's) segments
- Poly(A) tail heterogeneity from transcriptional slippage profiled by LC-UV and LC-MS
- IP-RPUHPLC-UV cannot resolve longer L70 poly(A); HRMS is needed



Oligonucleotide Mapping Enables Assessment of mRNA Batch Comparability

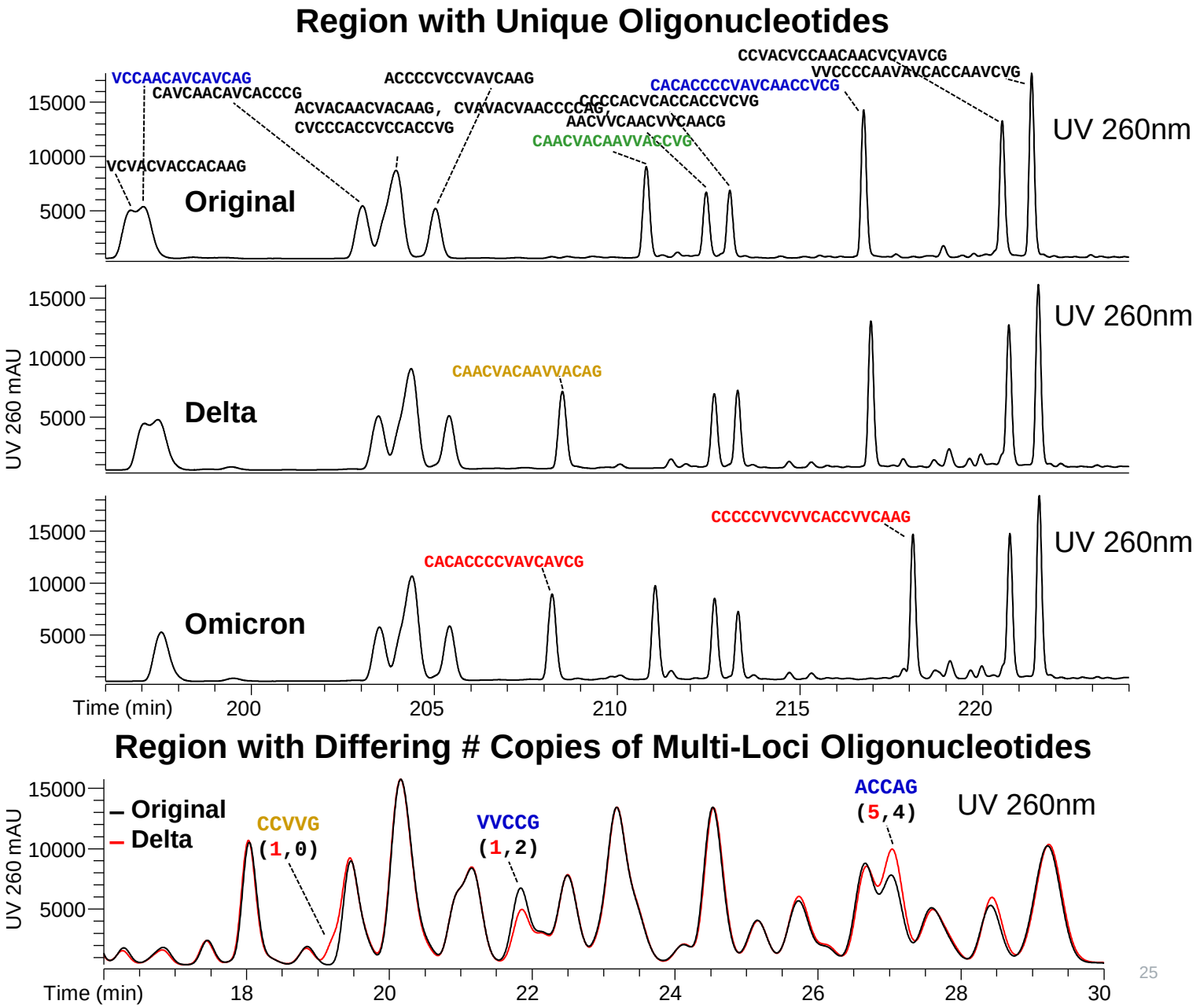




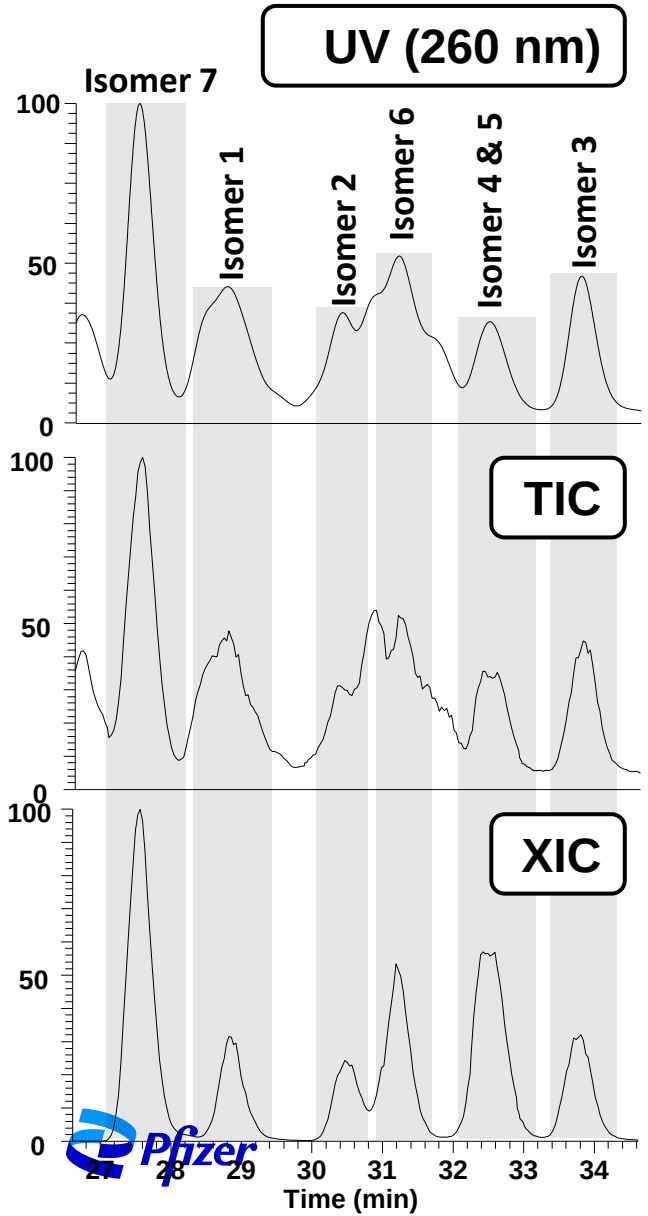
- Base-peak or total-ion chromatograms are not appropriate for overall comparability asssment
  - Background ions (esp. HFIP complexes)
  - Ionization efficiency sensitivities
- The LC-UV chromatogram provides a reliable fingerprint of mRNA digest
- Oligonucleotide Mapping Demonstrating Comparability of Multiple BNT162b2 mRNA Drug Substance Batches
- Side-by-side analyses are highly robust
- Chromatographic peaks overlay well

Oligonucleotide Mapping Enables Comparison of mRNA for Variant Constructs

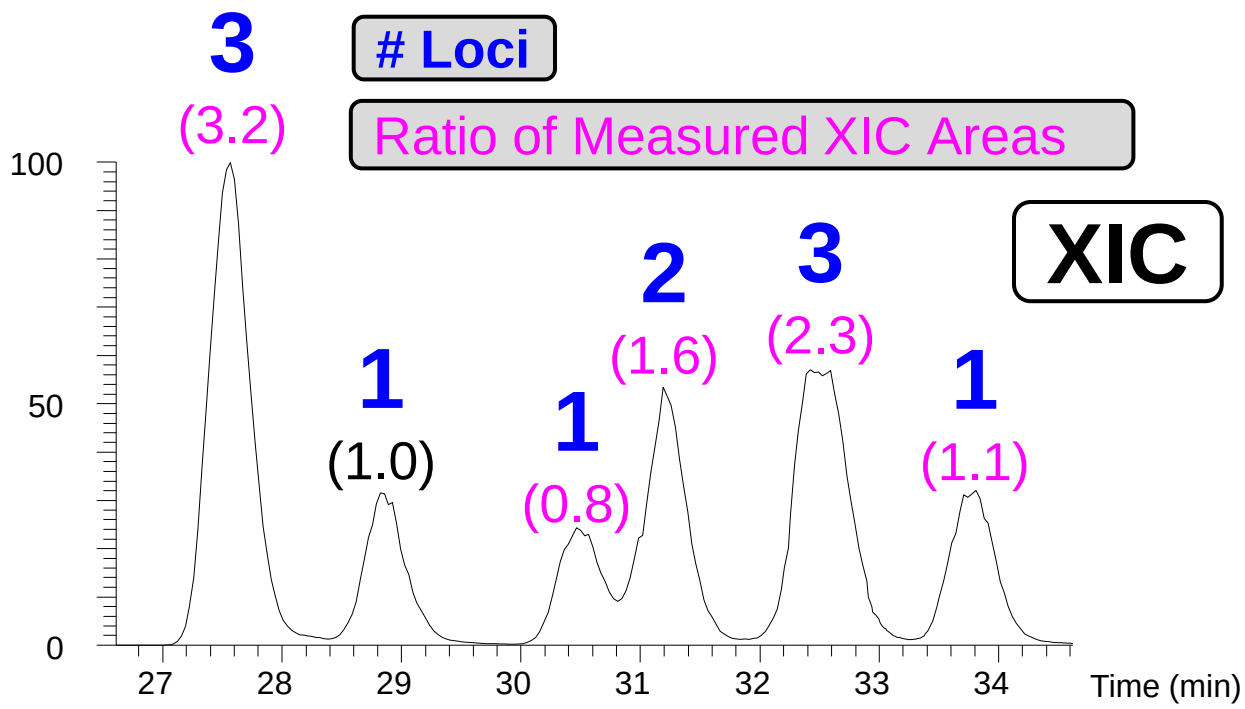
- The LC-UV chromatogram serves as an identity fingerprint
- ClustalW sequence analysis:
  - BNT162b2 Delta is 99.6% and BNT162b2 Original
  - BNT162b2 Omicron is 98.6% similar to BNT162b2 Original
- LC-UV is (often) conspicuously discerning in the unique-sequence chromatogram region
- Differences in copies of multi-loci oligonucleotides are also apparent
- This could serve as an alternative identity assay to ddPCR



Measured XIC Areas of Non-Unique Sequence Isomers Correlate with their Number of Loci in the Full Length mRNA Sequence



Theoretical				Observed	
Oligonucleotide	**Sequence**	**# of Loci**	**Monoisotopic Mass (Da)**	**Retention Time (min)**	Ratio of XIC Areas Normalized to Isomer 1
Isomer 1	CVAAG	1	1646.2453	28.9	1
Isomer 2	VCAAG	1		30.5	0.8
Isomer 3	AVCAG	1		33.8	1.1
Isomer 4	AVACG	1		32.5	2.3
Isomer 5	AACVG	2			
Isomer 6	VACAG	2		31.2	1.6
Isomer 7	CAAVG	3		27.6	3.2

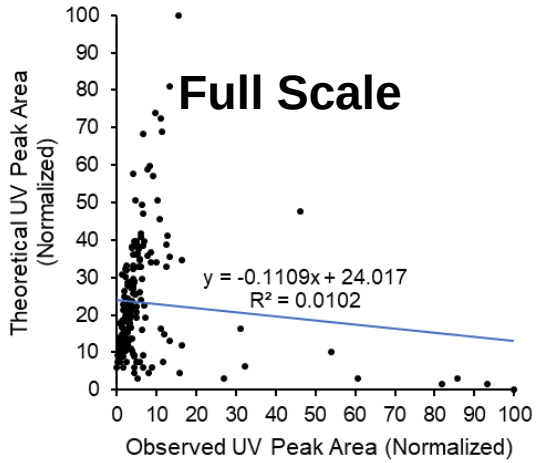


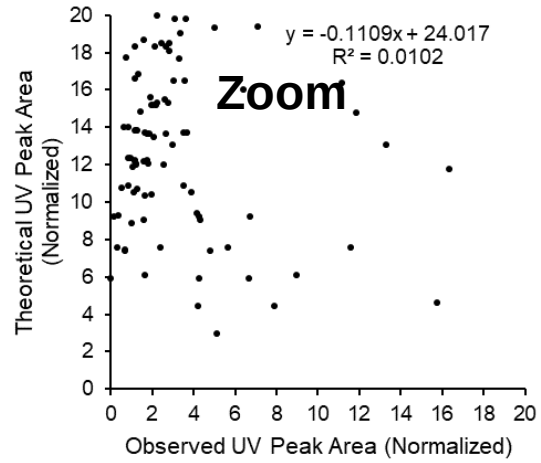
Measured UV Areas Across Oligonucleotide Map Correlate with Theoretical UV Areas With Proper Accounting

- Empirical peak areas were determined by
  1. ICIS peak detection optimized for detection→Table of UV Peak RTs
  2. Each end point was re-calculated as ½ distance between ICIS end points of neighboring peak and current peak
  3. Peak area = sum of intensity between end points after background subtraction

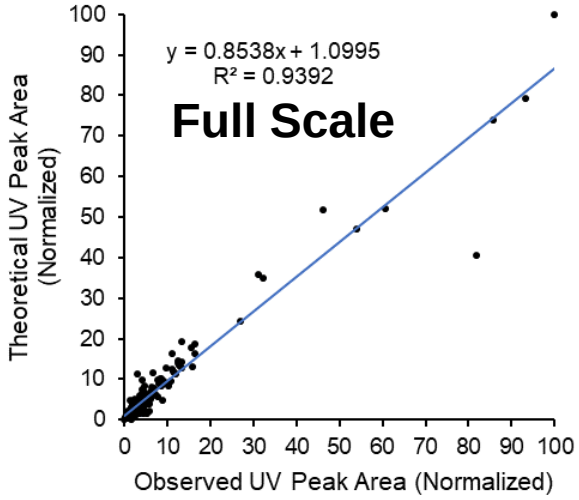
- Theoretical peak areas were calculated by
  1. Using the Table of UV Peak RTs.
  2. Assigning a UV peak's ID to the nearest MS-ID'd oligo; more than one oligo can map to a UV peak
  3. Dermining each oligo's theoretical extinction coefficient from its composition and based on NMR-derived extinction coefficients[1] for pdG, pdA, pdC, and N1-methylpseudouridine monophosphate[2]
  4. Summing these values for all oligos mapped to peak, if it is a mixture peak, and factoring the number of loci (bottom graphs)

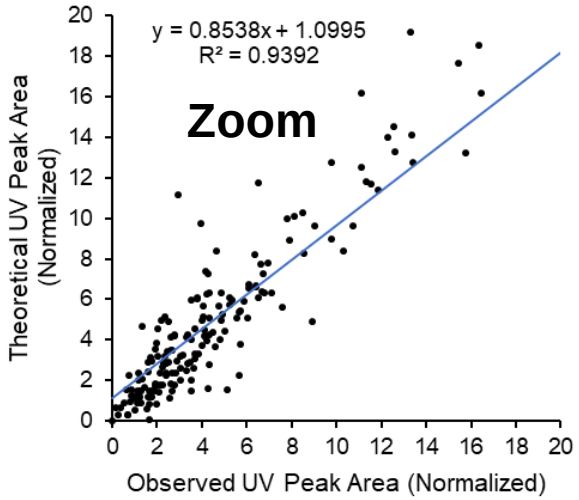
Not Accounting for Multiple Loci of Non-Unique Oligonucleotides





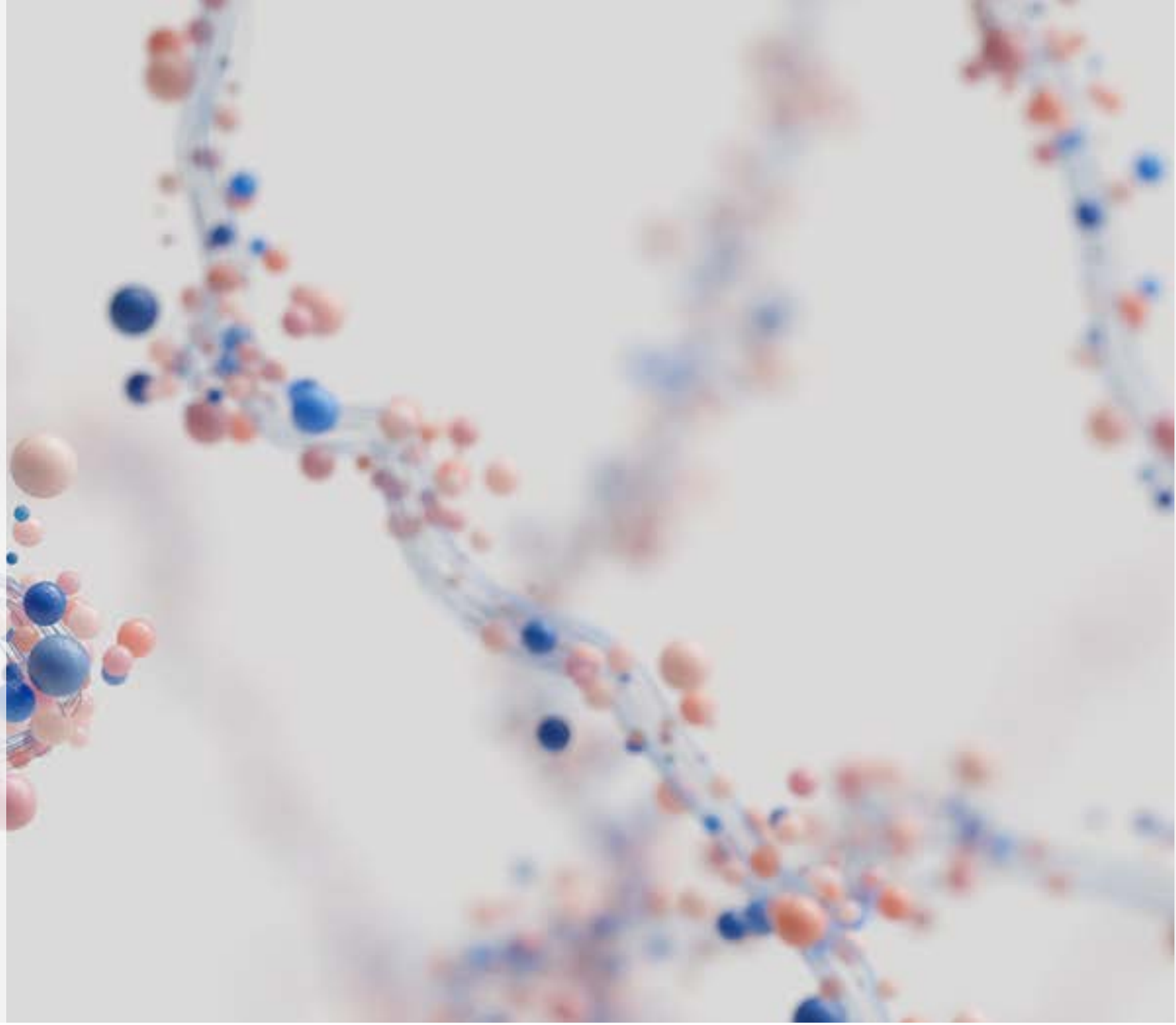
Accounting for Multiple Loci of Non-Unique Oligonucleotides





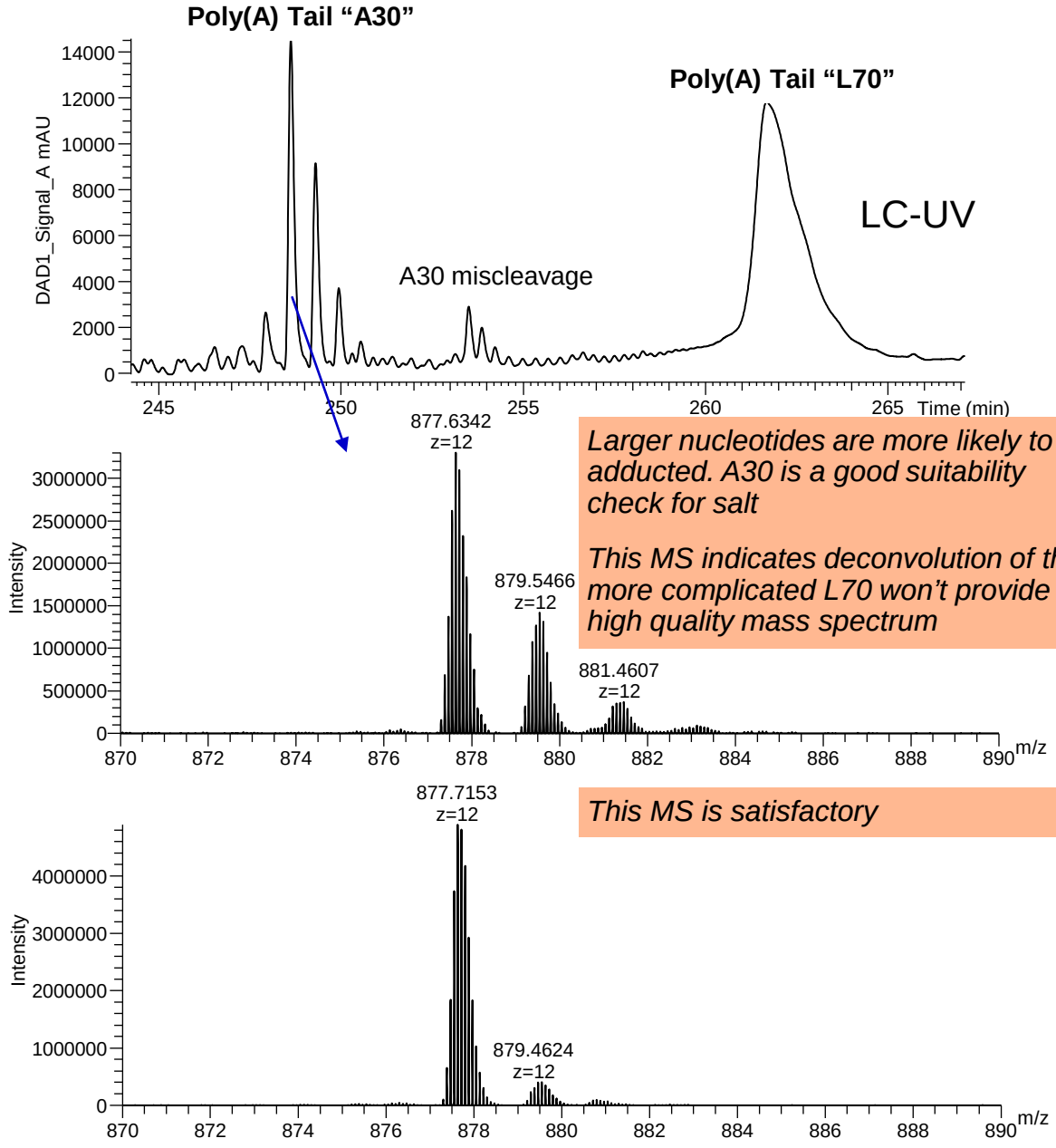
[1]Cavaluzzi, M.J. & Borer, P.N. *Nucleic Acids Res* ***32***, e13 (2004)
[2]Emperically determined at Pfizer

Ensuring Optimum Chromatographic Separation and MS Ionization



Solvents and Additives Must be of the Highest Quality

- Ion-paired reversed phase separation of 1-70 nt oligonucleotides best done with TEA/HFIP/Methanol/Water
  - 0.1% TEA 1% HFIP in both mobile phases
  - Shallow gradient to tease apart mixture peaks
  - Only LC-MS-grade solvents, TEA and HFIP are acceptable
    - MQ-Water with LC-PAK cartridge acceptable; run the system for several min before solvent prep
- The UHPLC should be passified to lessen secondary metal-phosphate interaction
  - 0.85% phosphoric acid, then lots of water/methanol, UHPLC **offline** from MS
  - "BioInert" classified UHPLC are ideal
- Early application notes suggest 400 mM (4.2%) HFIP, *but*
  - **LC/MS grade HFIP is hard to source**; 1% HFIP works
  - pH of solvents will change over time, warrenting a short shelf life—except—HFIP can be in short supply…don't discard!
  - Side-by-side analysis can be done with older solvents; it is only historical comparability that can be jeapordized using old solvents



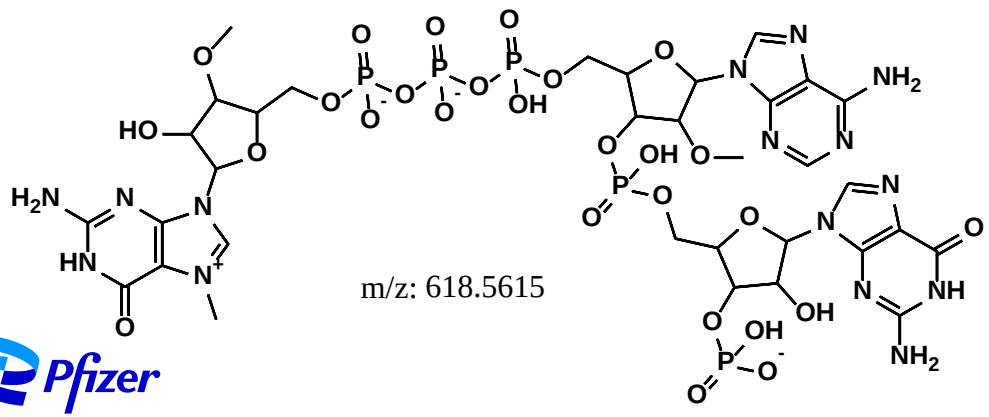
Larger nucleotides are more likely to be adducted. A30 is a good suitability check for salt

This MS indicates deconvolution of the more complicated L70 won't provide a high quality mass spectrum

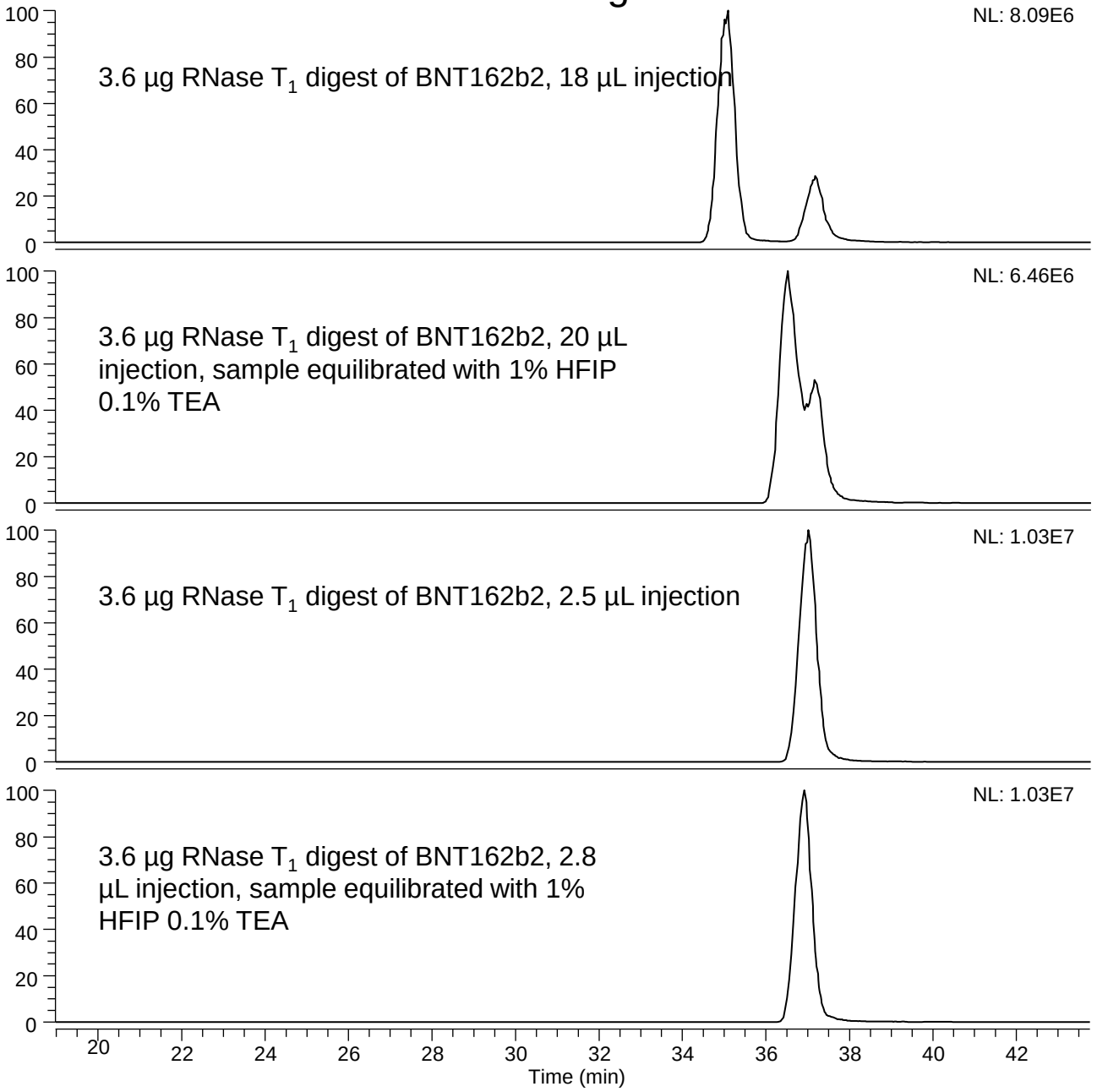
This MS is satisfactory

Peak Splitting

- IP-RPLC elution is directly proportional to the # of nucleotides
  - Ion pairing of triethylammonium to the negatively charged phophodiester backbone
- Samples may not fully equilibrate with the mobile phase in the time of passage from the autosampler to the head of the column
  - This gives rise to peak splitting
  - E.g., the capped R1 peptide
- Solution: spike the sample with TEA and HFIP to give their mobile phase levels



m/z: 618.5615

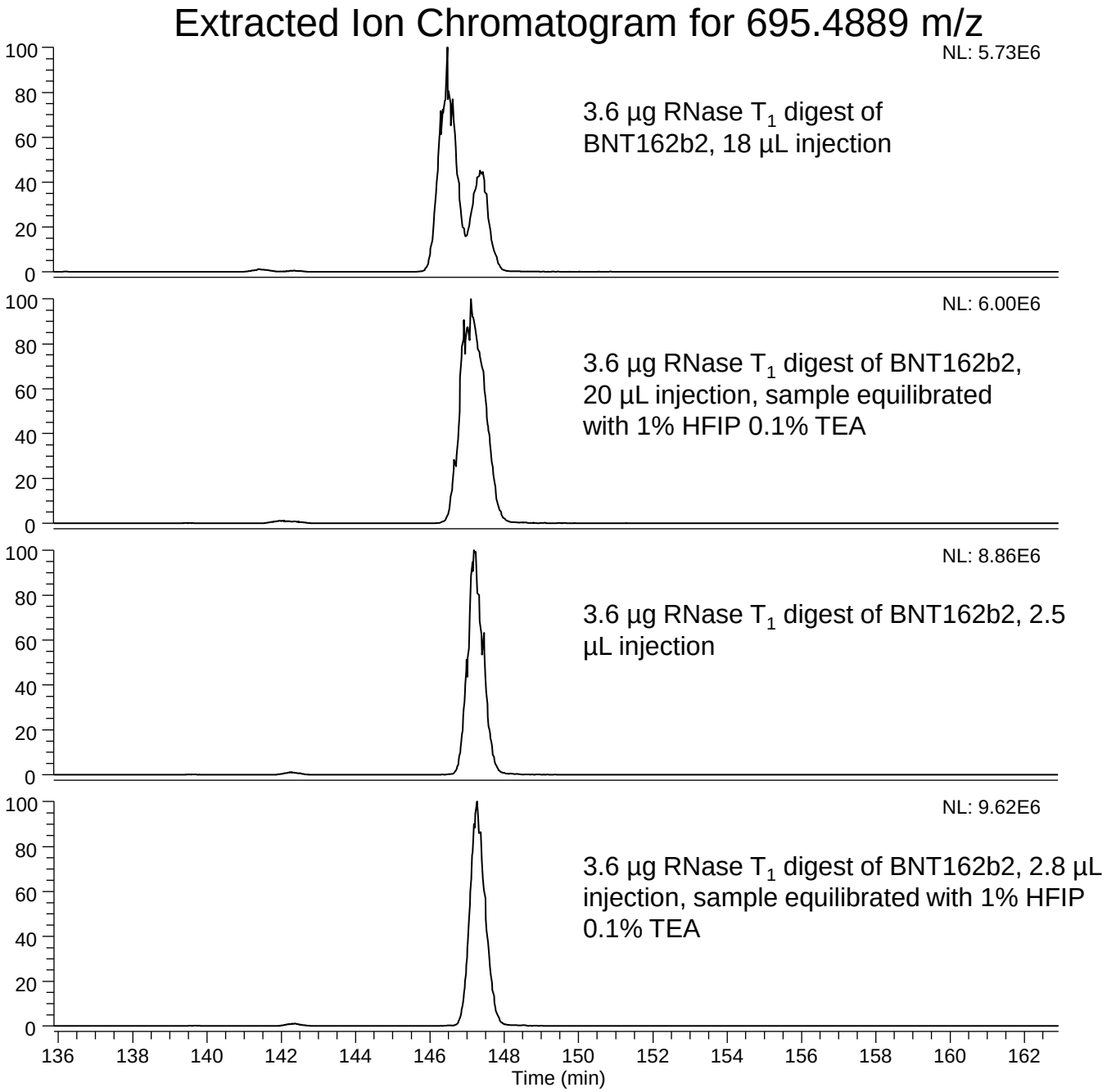


Peak Splitting

- 2nd example
- Smaller injection volume also helps, which predicates working with a more concentrated digest (described on Slide 17)

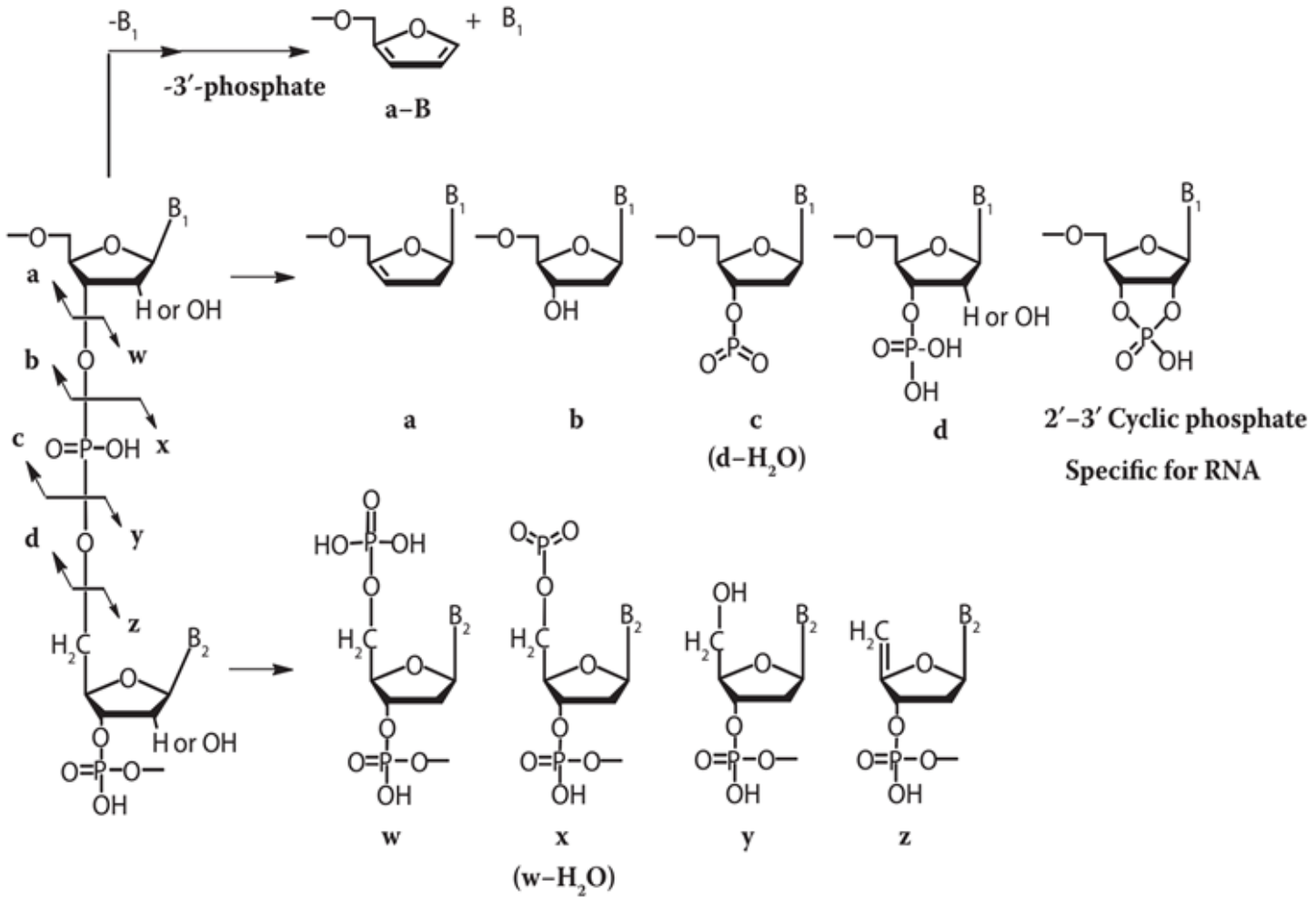
ACCCCTTCCTG

Monoisotopic Mass
3482.4811
Average Mass
3484.1243
Precursor Charge State
- 5
Precursor Monoisotopic m/z
695.4889



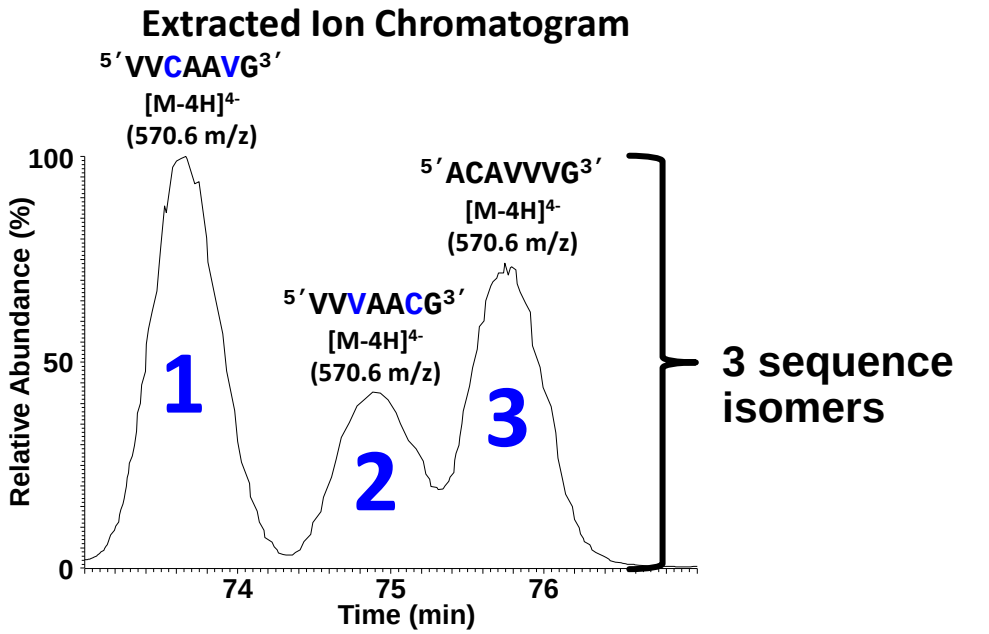
Ensuring Optimum MS/MS

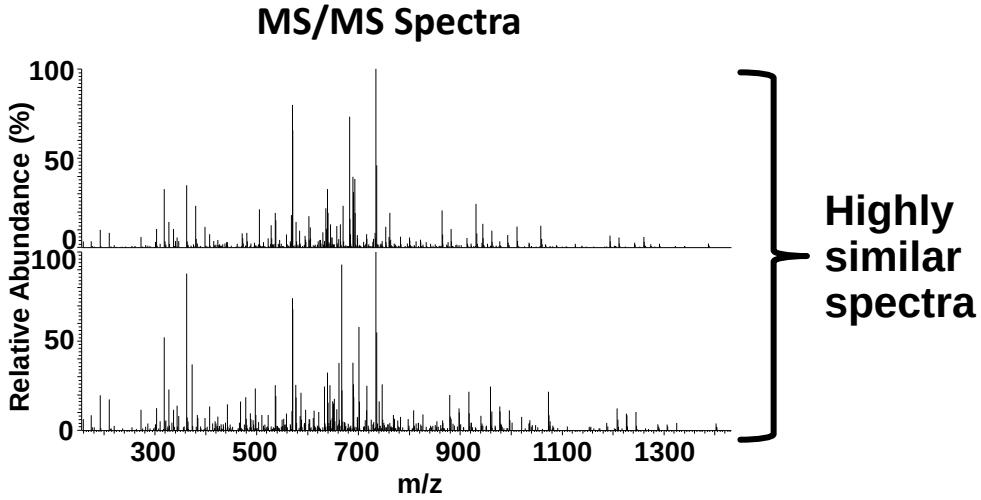
Higher-Energy Collisional Dissociation (HCD) Gives All Phosphorodiester Fragmentation Products

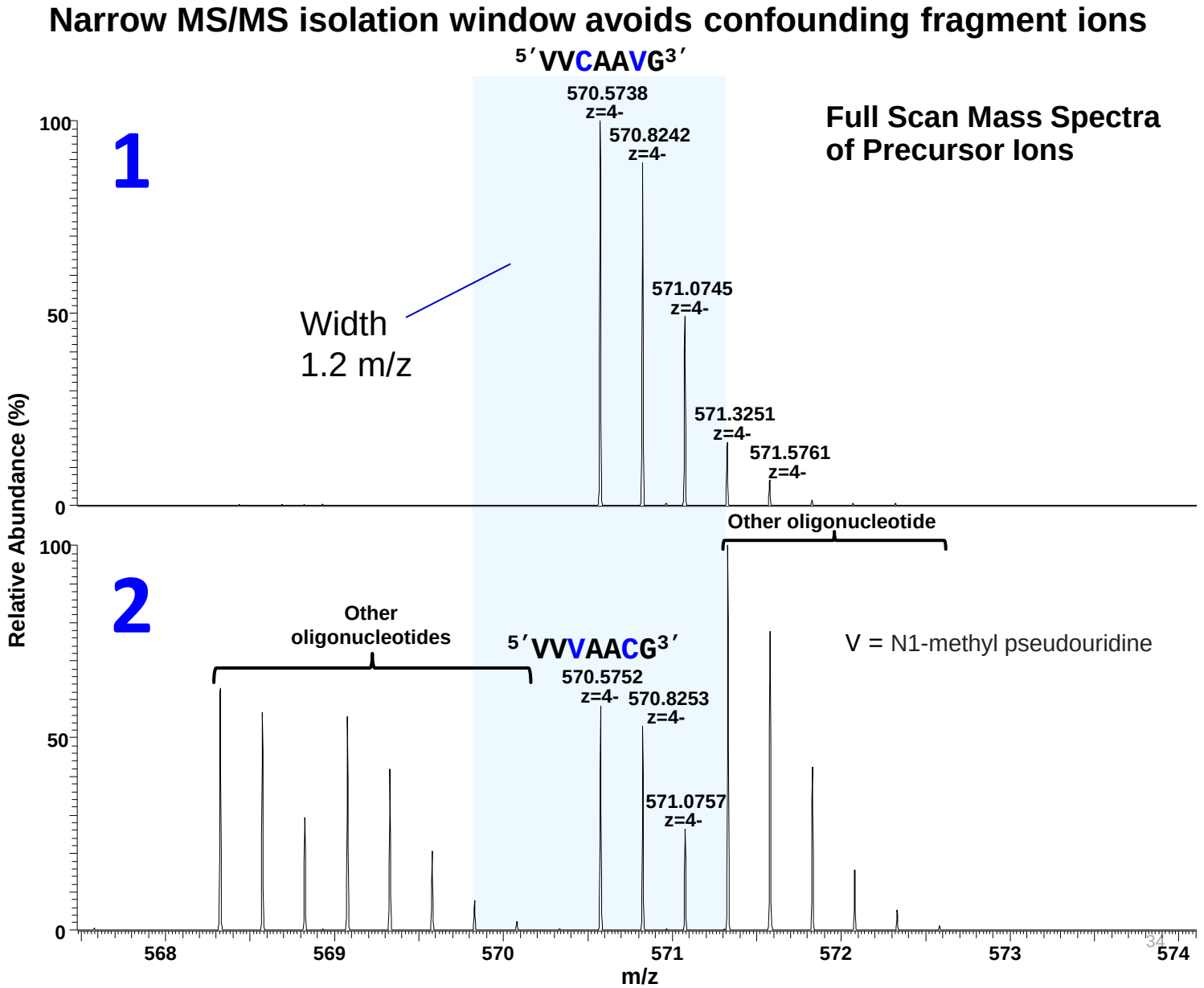


McLuckey, S.A., et al *Journal of the American Society for Mass Spectrometry* ***3,*** 60-70 (1992)
Figure: Timar, Z. Handbook of Analysis of Oligonucleotides and Related Products 10.1201/b10714-6. (eds. J.V. Bonilla & G.S. Srivatsa) 167-218 (CRC Press, 2011)

Applying Optimized HCD to Differentiate 2 Sequence Isomers Differing by a Single Exchange in Base Positions

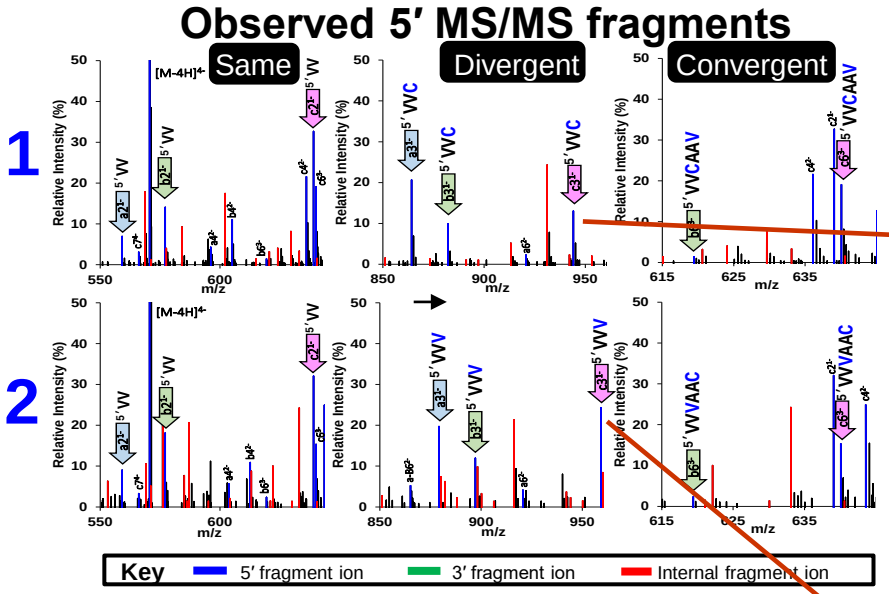






Optimal Fragmentation Enables Differentiation of Highly Similar Sequence Isomers

V = N1-methyl pseudouridine



+15 Da Shift

1 5′VVCAAVG3′ [M-4H]4-

Observed 5′ fragments					
a	b	c	d	#	
		$(319.0)^{1-}$	$(337.0)^{1-}$	1	V
$(559.1)^{1-}$	$(577.1)^{1-}$	$(639.1)^{1-}$	$(657.1)^{1-}$	2	V
$(864.2)^{1-}$	$(882.2)^{1-}$, $(440.6)^{2-}$	$(944.1)^{1-}$, $(471.6)^{2-}$	$(962.1)^{1-}$, $(480.6)^{2-}$	3	C
$(1193.2)^{1-}$, $(596.1)^{2-}$	$(1211.2)^{1-}$, $(605.1)^{2-}$	$(1273.2)^{1-}$, $(636.1)^{2-}$	$(1291.2)^{1-}$, $(645.1)^{2-}$	4	A
$(760.6)^{2-}$	$(769.6)^{2-}$	$(800.6)^{2-}$		5	A
$(920.6)^{2-}$	$(619.4)^{3-}$	$(640.1)^{3-}$	$(646.1)^{3-}$	6	V
$(728.4)^{3-}$	$(734.4)^{3-}$	$(566.1)^{4-}$		7	G

2 5′VVVAACG3′ [M-4H]4-

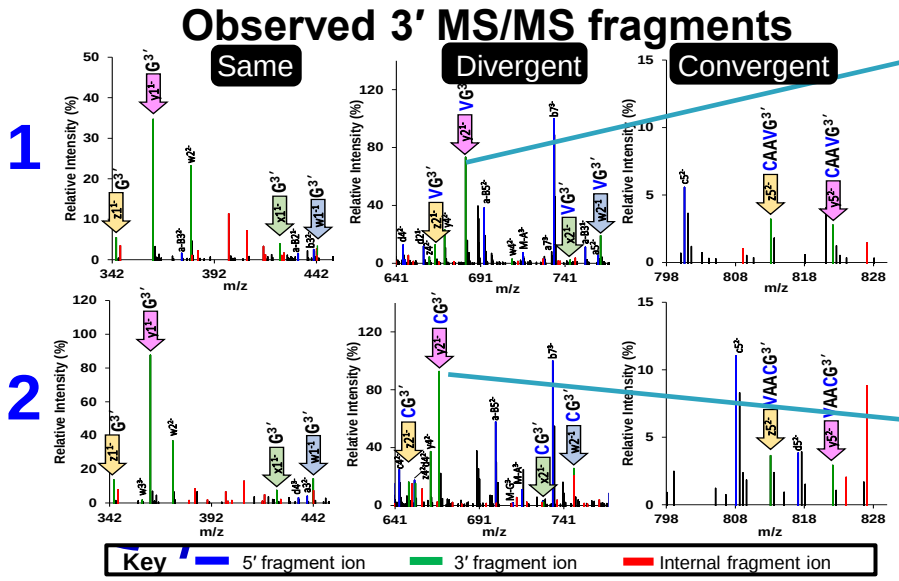
Observed 5′ fragments					
a	b	c	d	#	
		$(319.0)^{1-}$	$(337.0)^{1-}$	1	V
$(559.1)^{1-}$	$(577.1)^{1-}$	$(639.1)^{1-}$		2	V
$(879.1)^{1-}$, $(439.1)^{2-}$	$(897.2)^{1-}$	$(959.1)^{1-}$, $(479.1)^{2-}$	$(977.1)^{1-}$, $(488.1)^{2-}$	3	V
$(1208.2)^{1-}$, $(603.6)^{2-}$	$(1226.2)^{1-}$, $(612.6)^{2-}$	$(1288.2)^{1-}$, $(643.6)^{2-}$	$(1306.2)^{1-}$, $(652.6)^{2-}$	4	A
$(768.1)^{2-}$	$(777.1)^{2-}$	$(808.1)^{2-}$	$(817.1)^{2-}$	5	A
$(920.6)^{2-}$	$(619.4)^{3-}$	$(960.6)^{2-}$, $(640.1)^{3-}$		6	C
$(728.4)^{3-}$	$(734.4)^{3-}$	$(566.1)^{4-}$		7	G

Optimal Fragmentation Enables Differentiation of Highly Similar Sequence Isomers

V = N1-methyl pseudouridine

1 $^{5'}$VVCAAVG$^{3'}$ $[M-4H]^{4-}$

	Observed 3′ fragments				
	#	w	x	y	z
V	7				
V	6				
C	5			$(822.1)^{2-}$, $(547.7)^{3-}$	$(813.1)^{1-}$
A	4	$(709.6)^{2-}$, $(472.7)^{3-}$		$(669.6)^{2-}$	$(660.6)^{2-}$
A	3	$(545.1)^{2-}$	$(536.0)^{2-}$	$(1011.1)^{1-}$, $(505.1)^{2-}$	$(993.1)^{1-}$, $(496.1)^{2-}$
V	2	$(762.1)^{1-}$, $(380.5)^{2-}$	$(744.0)^{1-}$	(682.1), (340.5)	$(664.1)^{1-}$
G	1	$(442.0)^{1-}$	$(424.0)^{1-}$	$(362.1)^{1-}$	$(344.0)^{1-}$



-15 Da Shift

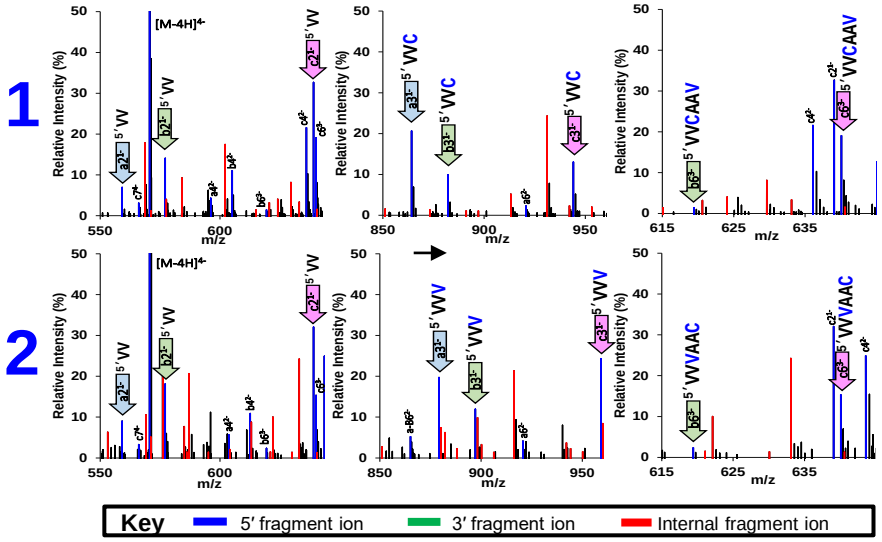
2 $^{5'}$VVVAACG$^{3'}$ $[M-4H]^{4-}$

	Observed 3′ fragments				
	#	w	x	y	z
V	7				
V	6				
V	5			$(822.1)^{2-}$	$(813.1)^{1-}$
A	4	$(467.7)^{3-}$		$(1325.2)^{1-}$, $(662.1)^{2-}$	$(1307.2)^{1-}$, $(653.1)^{2-}$
A	3	$(358.0)^{3-}$		$(996.1)^{1-}$, $(497.6)^{2-}$	$(978.1)^{1-}$, $(488.6)^{2-}$
C	2	$(747.1)^{1-}$, $(373.0)^{2-}$	$(729.0)^{1-}$	$(667.1)^{1-}$, $(333.0)^{2-}$	$(649.1)^{1-}$
G	1	$(442.0)^{1-}$	$(424.0)^{1-}$	$(362.1)^{1-}$	$(344.0)^{1-}$

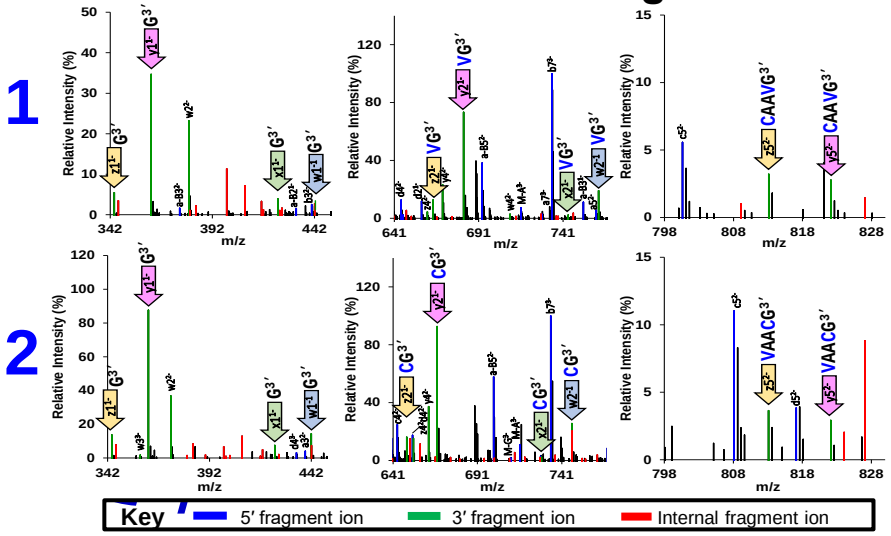
“Ladder Ions” Are Useful Sequencing Ions; Internal Fragment Ions Are Not

V = N1-methyl pseudouridine

Observed 5′ MS/MS fragments



Observed 3′ MS/MS fragments



1 5′VVCAAVG3′ [M-4H]4-

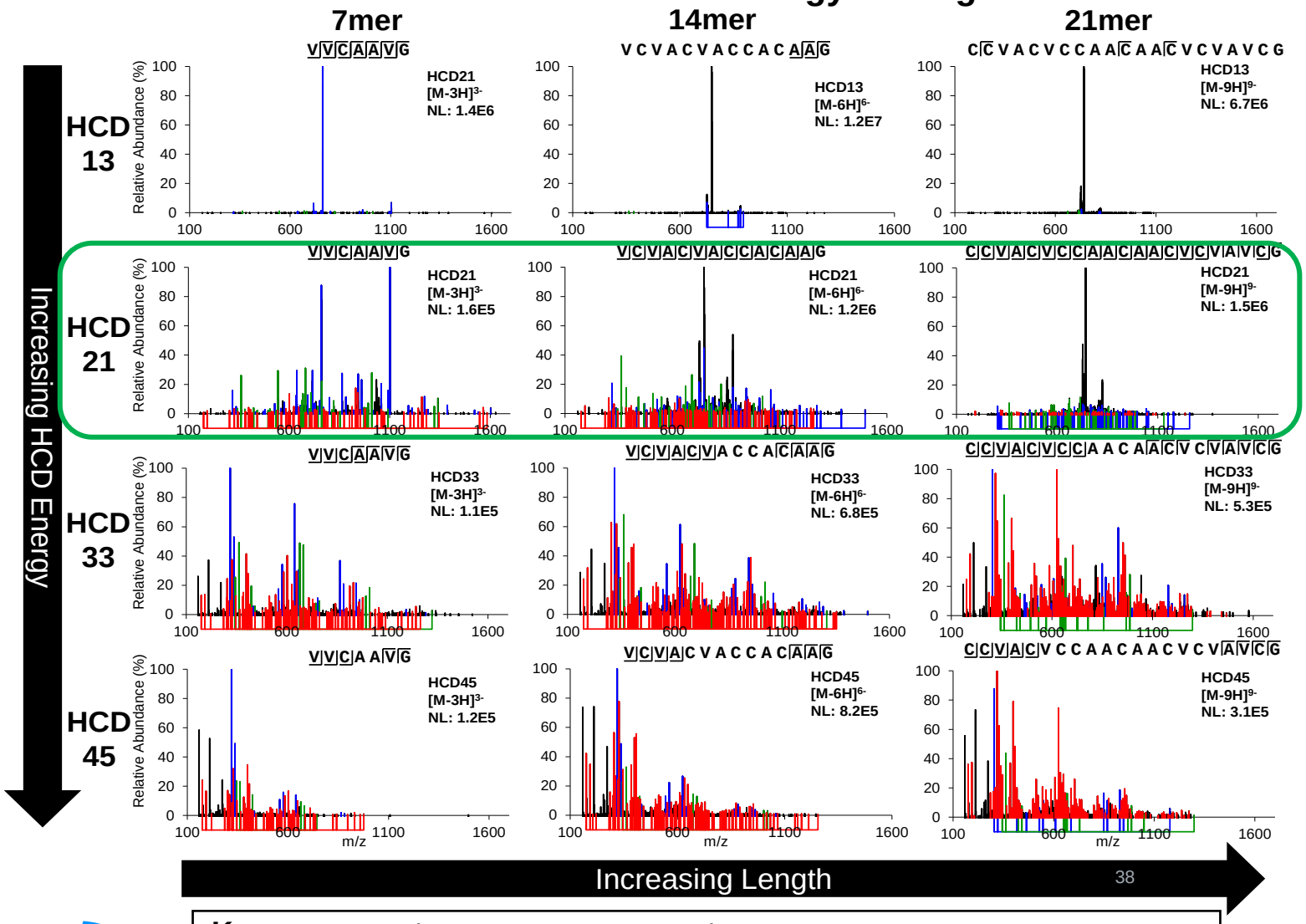
Observed 5′ fragments						Observed 3′ fragments				
a	b	c	d	#		#	w	x	y	z
		$(319.0)^{1-}$	$(337.0)^{1-}$	1	V	7				
$(559.1)^{1-}$	$(577.1)^{1-}$	$(639.1)^{1-}$	$(657.1)^{1-}$	2	V	6				
$(864.2)^{1-}$	$(882.2)^{1-}$, $(440.6)^{2-}$	$(944.1)^{1-}$, $(471.6)^{2-}$	$(962.1)^{1-}$, $(480.6)^{2-}$	3	C	5			$(822.1)^{2-}$, $(547.7)^{3-}$	$(813.1)^{1-}$
$(1193.2)^{1-}$, $(596.1)^{2-}$	$(1211.2)^{1-}$, $(605.1)^{2-}$	$(1273.2)^{1-}$, $(636.1)^{2-}$	$(1291.2)^{1-}$, $(645.1)^{2-}$	4	A	4	$(709.6)^{2-}$, $(472.7)^{3-}$		$(669.6)^{2-}$	$(660.6)^{2-}$
$(760.6)^{2-}$	$(769.6)^{2-}$	$(800.6)^{2-}$		5	A	3	$(545.1)^{2-}$	$(536.0)^{2-}$	$(1011.1)^{1-}$, $(505.1)^{2-}$	$(993.1)^{1-}$, $(496.1)^{2-}$
$(920.6)^{2-}$	$(619.4)^{3-}$	$(640.1)^{3-}$	$(646.1)^{3-}$	6	V	2	$(762.1)^{1-}$, $(380.5)^{2-}$	$(744.0)^{1-}$	(682.1), (340.5)	$(664.1)^{1-}$
$(728.4)^{3-}$	$(734.4)^{3-}$	$(566.1)^{4-}$		7	G	1	$(442.0)^{1-}$	$(424.0)^{1-}$	$(362.1)^{1-}$	$(344.0)^{1-}$

2 5′VVVAACG3′ [M-4H]4-

Observed 5′ fragments						Observed 3′ fragments				
a	b	c	d	#		#	w	x	y	z
		$(319.0)^{1-}$	$(337.0)^{1-}$	1	V	7				
$(559.1)^{1-}$	$(577.1)^{1-}$	$(639.1)^{1-}$		2	V	6				
$(879.1)^{1-}$, $(439.1)^{2-}$	$(897.2)^{1-}$	$(959.1)^{1-}$, $(479.1)^{2-}$	$(977.1)^{1-}$, $(488.1)^{2-}$	3	V	5			$(822.1)^{2-}$	$(813.1)^{1-}$
$(1208.2)^{1-}$, $(603.6)^{2-}$	$(1226.2)^{1-}$, $(612.6)^{2-}$	$(1288.2)^{1-}$, $(643.6)^{2-}$	$(1306.2)^{1-}$, $(652.6)^{2-}$	4	A	4	$(467.7)^{3-}$		$(1325.2)^{1-}$, $(662.1)^{2-}$	$(1307.2)^{1-}$, $(653.1)^{2-}$
$(768.1)^{2-}$	$(777.1)^{2-}$	$(808.1)^{2-}$	$(817.1)^{2-}$	5	A	3	$(358.0)^{3-}$		$(996.1)^{1-}$, $(497.6)^{2-}$	$(978.1)^{1-}$, $(488.6)^{2-}$
$(920.6)^{2-}$	$(619.4)^{3-}$	$(960.6)^{2-}$, $(640.1)^{3-}$		6	C	2	$(747.1)^{1-}$, $(373.0)^{2-}$	$(729.0)^{1-}$	$(667.1)^{1-}$, $(333.0)^{2-}$	$(649.1)^{1-}$
$(728.4)^{3-}$	$(734.4)^{3-}$	$(566.1)^{4-}$		7	G	1	$(442.0)^{1-}$	$(424.0)^{1-}$	$(362.1)^{1-}$	$(344.0)^{1-}$

HCD Collision Energy Optimized at Stepped CE 17, 21, 25

Oligonucleotide Fragment Ion Coverage as a Function of HCD Energy & Length



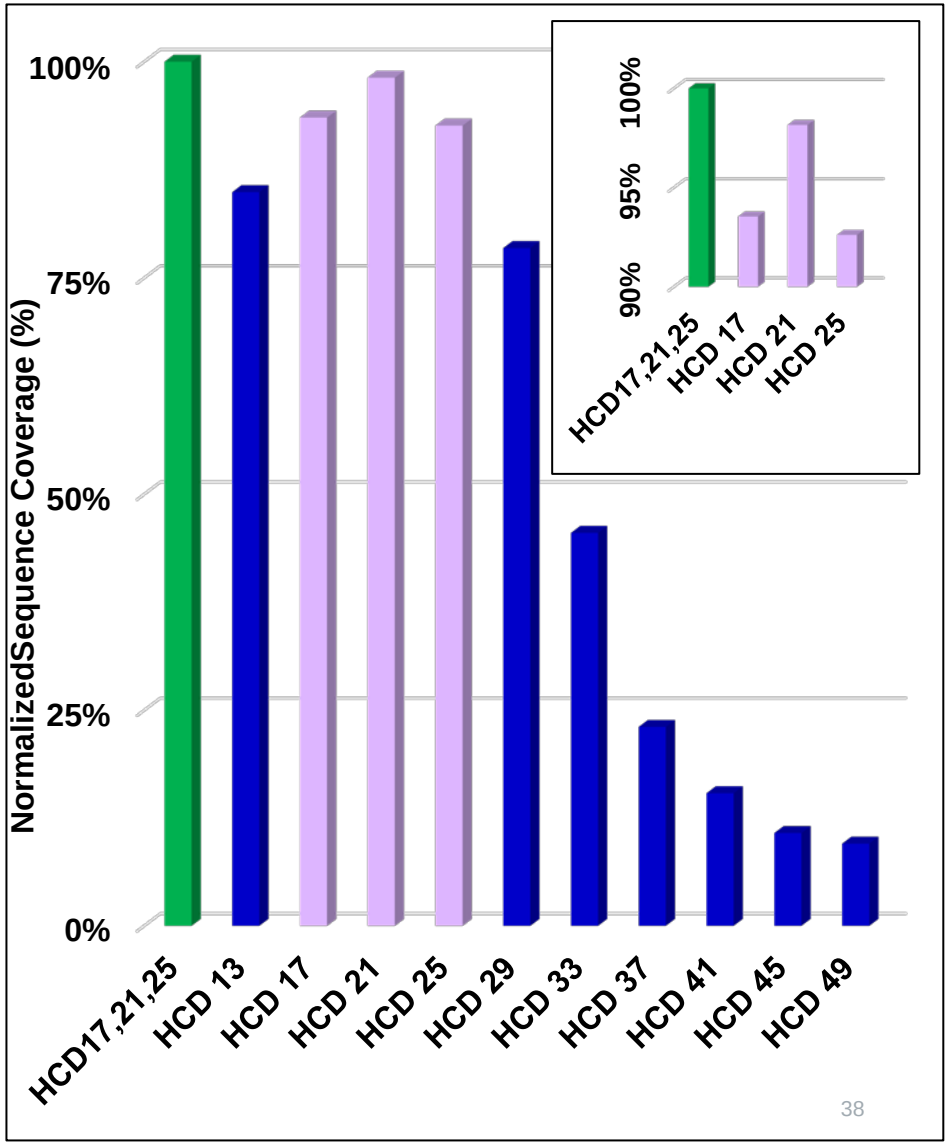
38

Key — 5′ fragment ion — 3′ fragment ion — Internal fragment ion

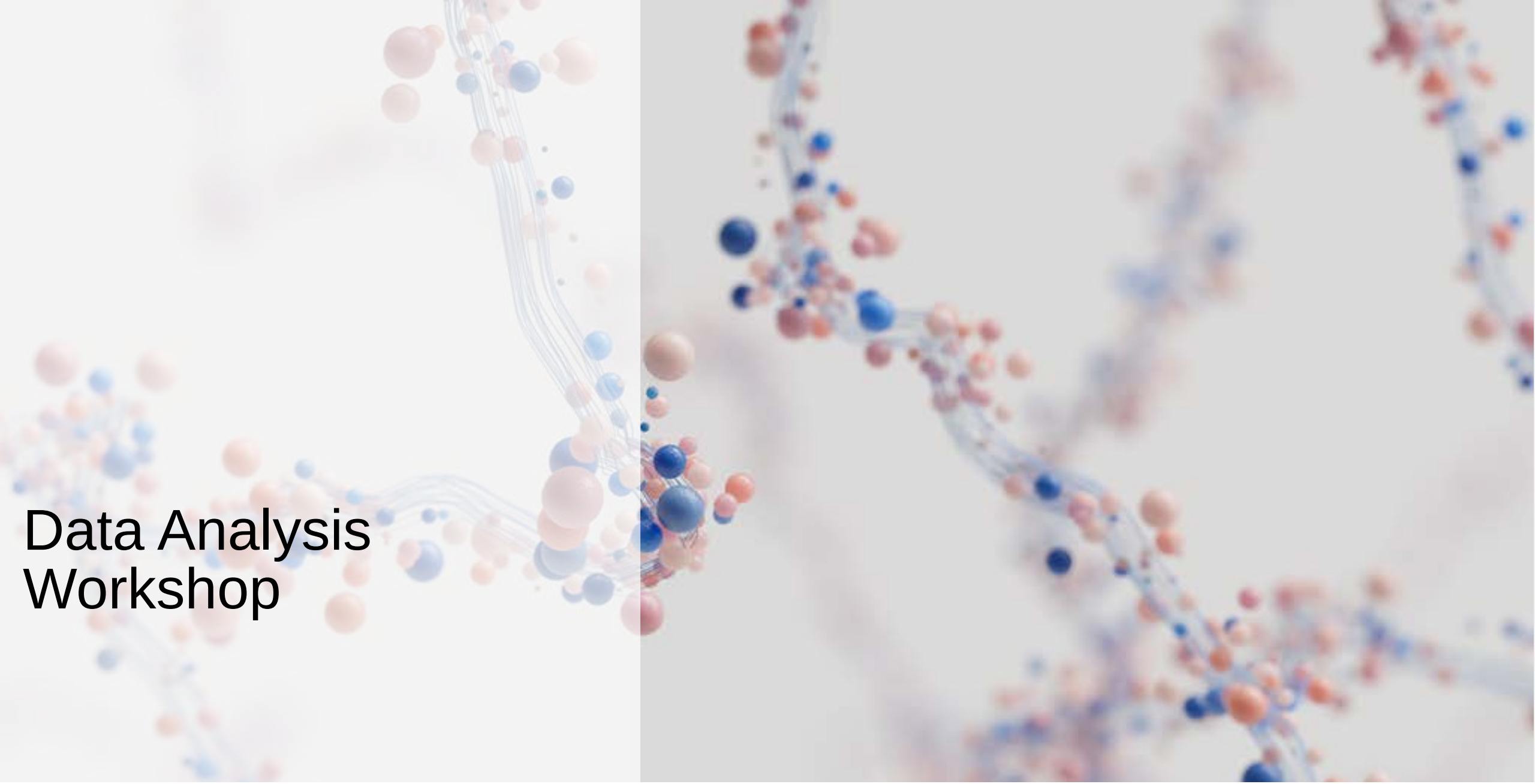
Charge densities are fixed at 0.4 charge / base

V = N1-methyl pseudouridine

Normalized BNT162b2 Sequence Coverage Across as a Function of HCD Energy



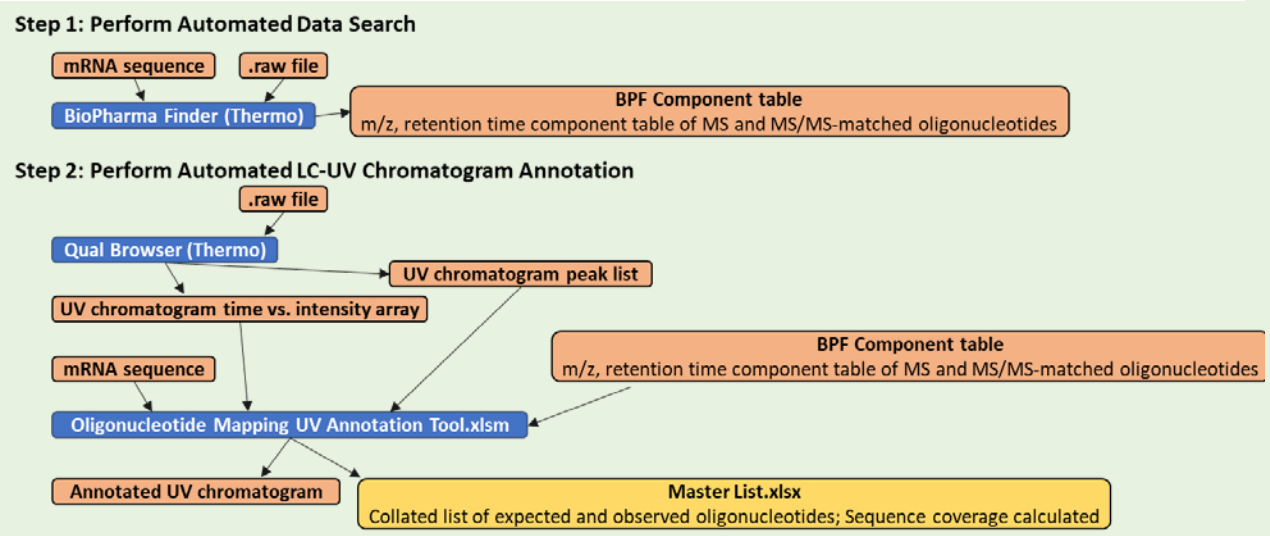
38

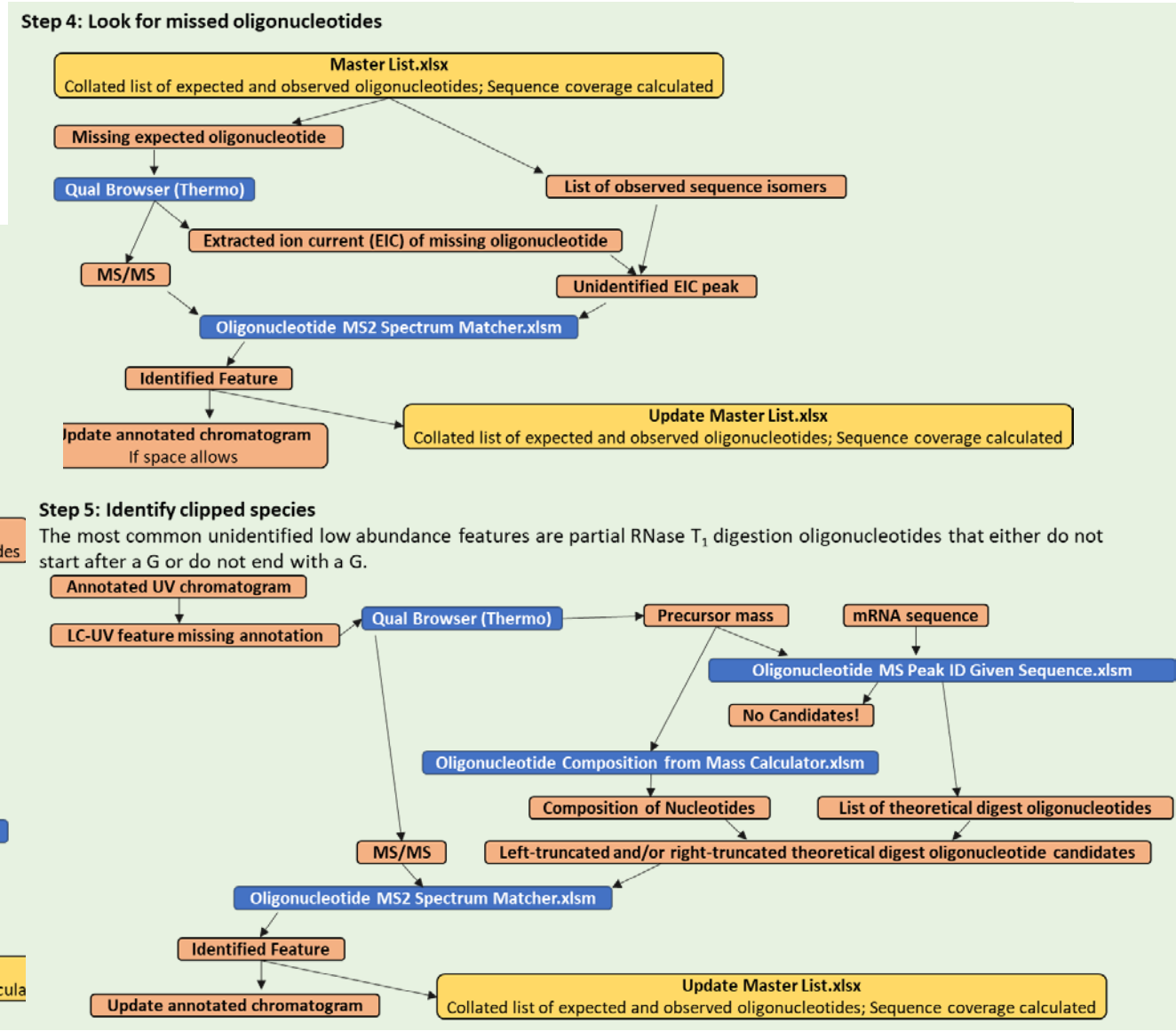


Data Analysis Workshop

Data Analysis Requires MS/MS Hypothesis Checking

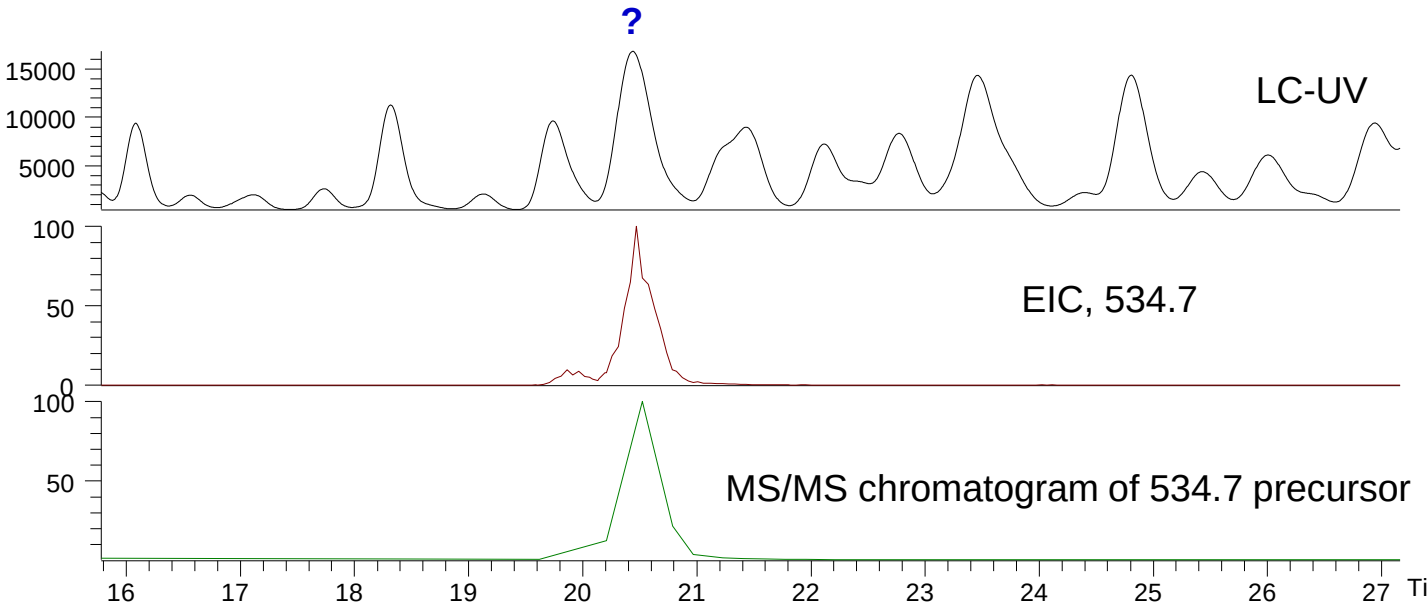
- Data analysis is semi-automated
  - 72-90% of sequence ID'd by commercial software (Steps 1 & 2)
  - Goal: 100% sequence coverage and ID all major & minor UV peaks (Steps 3-5)
- Un-ID'd UV features are often mixture peaks
- Missed oligonucleotides often reside in mixture peaks





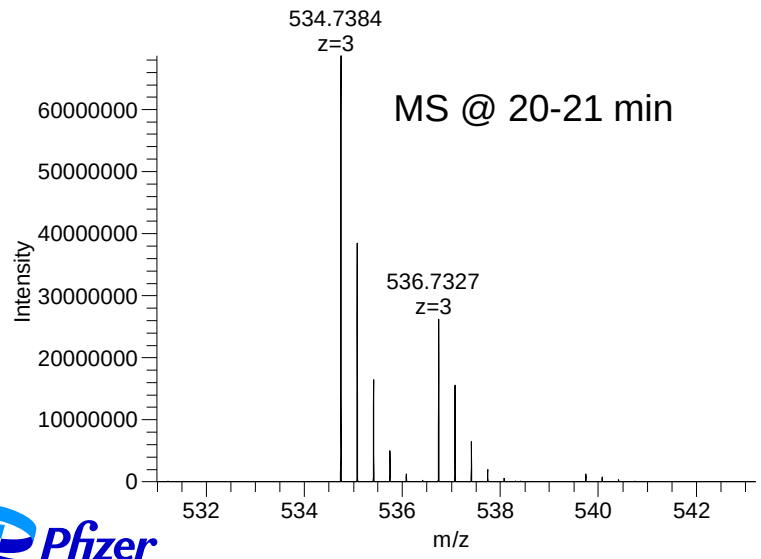
Gau, B.C. et al. *Scientific Reports* ***13,*** 9038 (2023)

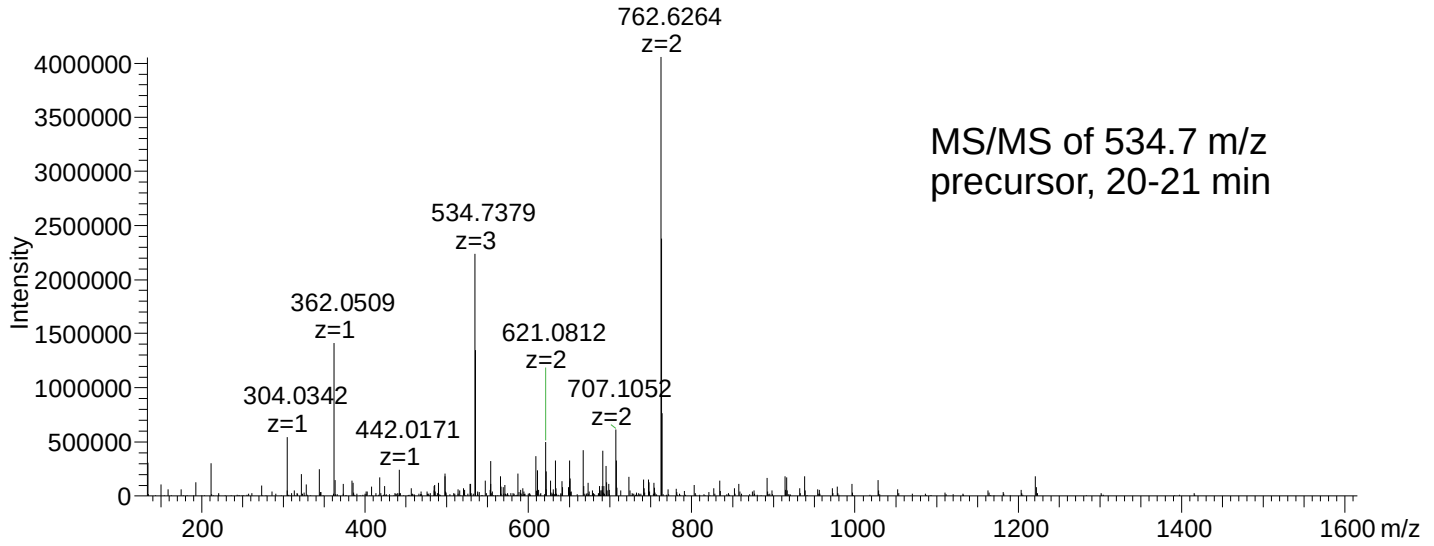
Mixture Peak



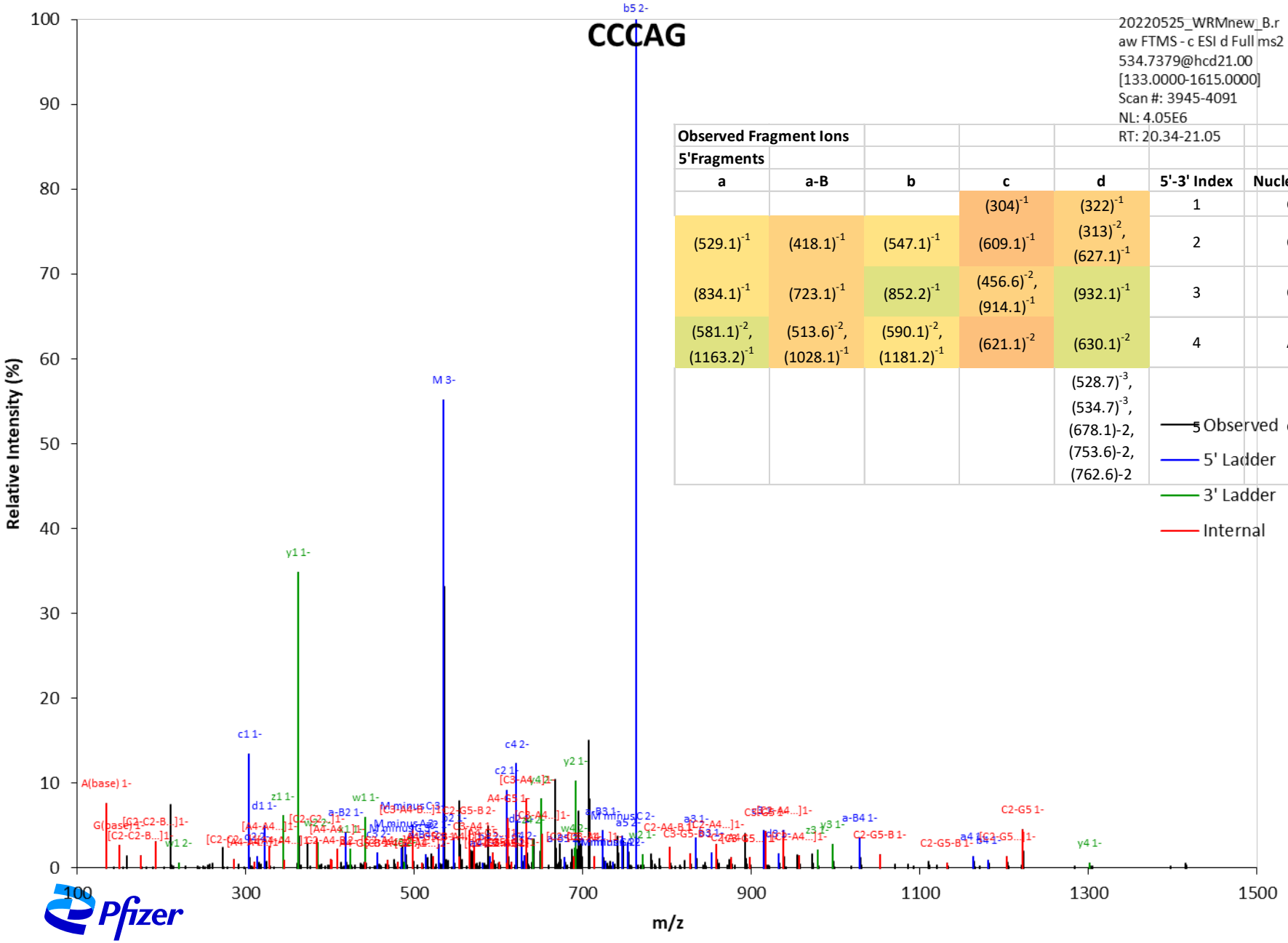
- 20.5 min peak is un-identified. There are two major species:
  - 534.7 m/z & 536.7 m/z
- This example IDs the 534.7 m/z species
  - 3 likely possibilities:

Observed Mass	Sequence	Theoretical Mass	Error (ppm)	Name	Frequency
1607.2370	CACCG	1607.2344	1.6	R76*	2
1607.2370	CCCAG	1607.2344	1.6	R618*	4
1607.2370	CCACG	1607.2344	1.6	R841	1



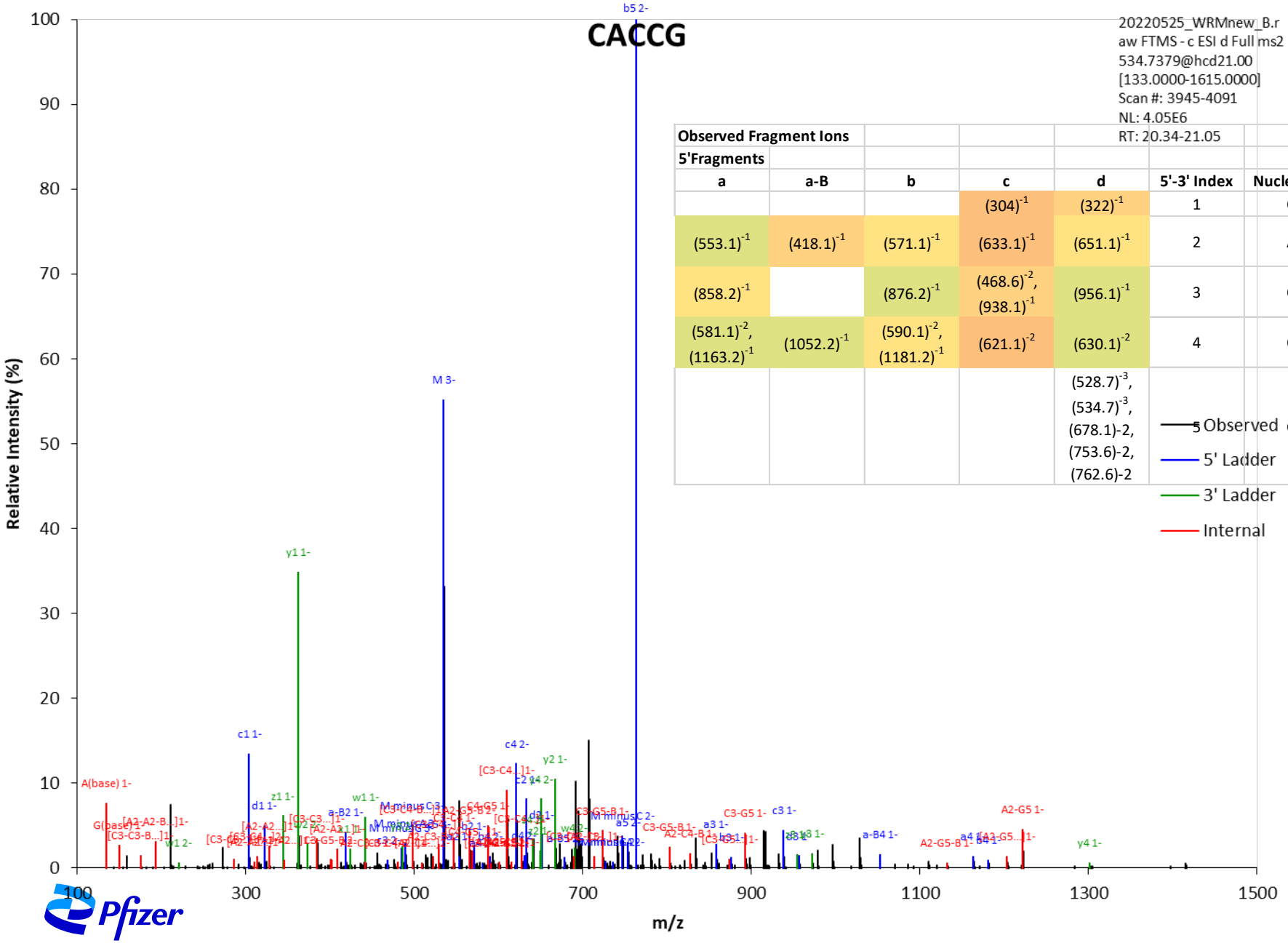


MS/MS 1st Match



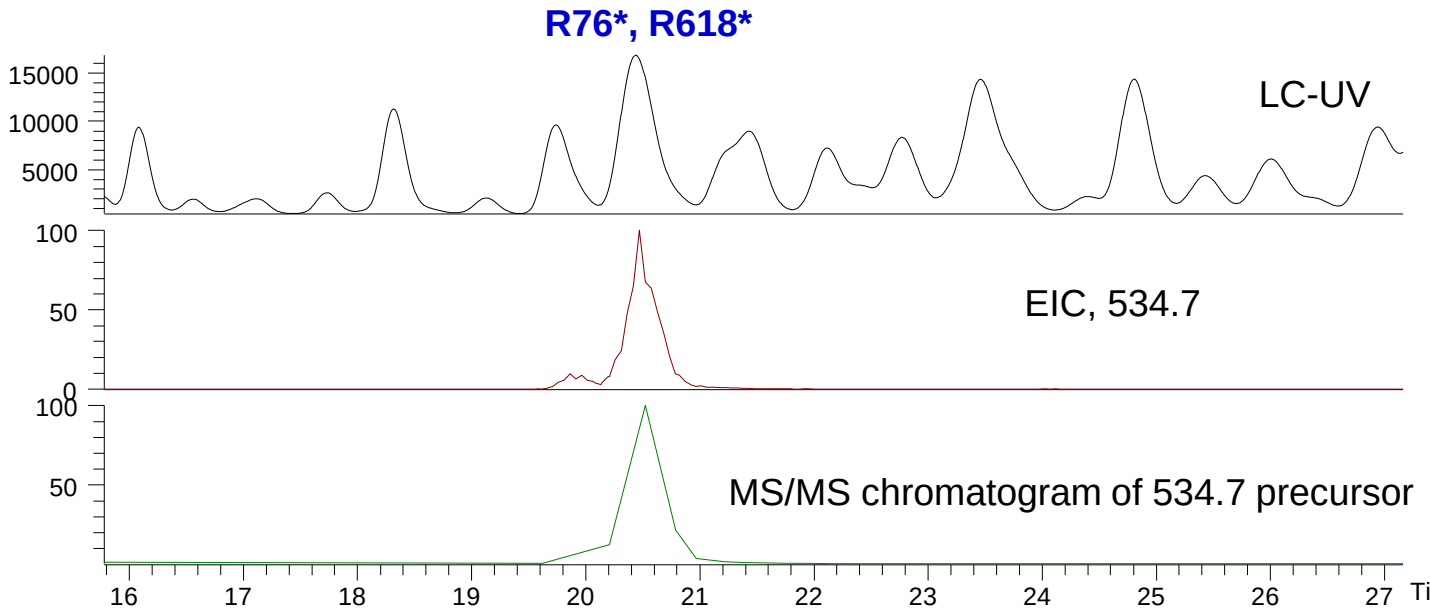
Observed Fragment Ions											
5'Fragments								3'Fragments			
a	a-B	b	c	d	5'-3' Index	Nucleotide	3'-5' Index	w	x	y	z
			$(304)^{-1}$	$(322)^{-1}$	1	C	5				
$(529.1)^{-1}$	$(418.1)^{-1}$	$(547.1)^{-1}$	$(609.1)^{-1}$	$(313)^{-2}$, $(627.1)^{-1}$	2	C	4	$(690.1)^{-2}$		$(650.1)^{-2}$, $(1301.2)^{-1}$	$(641.1)^{-2}$
$(834.1)^{-1}$	$(723.1)^{-1}$	$(852.2)^{-1}$	$(456.6)^{-2}$, $(914.1)^{-1}$	$(932.1)^{-1}$	3	C	3			$(996.1)^{-1}$	$(488.6)^{-2}$, $(978.1)^{-1}$
$(581.1)^{-2}$, $(1163.2)^{-1}$	$(513.6)^{-2}$, $(1028.1)^{-1}$	$(590.1)^{-2}$, $(1181.2)^{-1}$	$(621.1)^{-2}$	$(630.1)^{-2}$	4	A	2	$(385)^{-2}$, $(771.1)^{-1}$		$(691.1)^{-1}$	
				$(528.7)^{-3}$, $(534.7)^{-3}$, (678.1)-2, (753.6)-2, (762.6)-2	5	G	1	$(220.5)^{-2}$, $(442)^{-1}$	$(424)^{-1}$	$(362.1)^{-1}$	$(344)^{-1}$

MS/MS 2nd Match



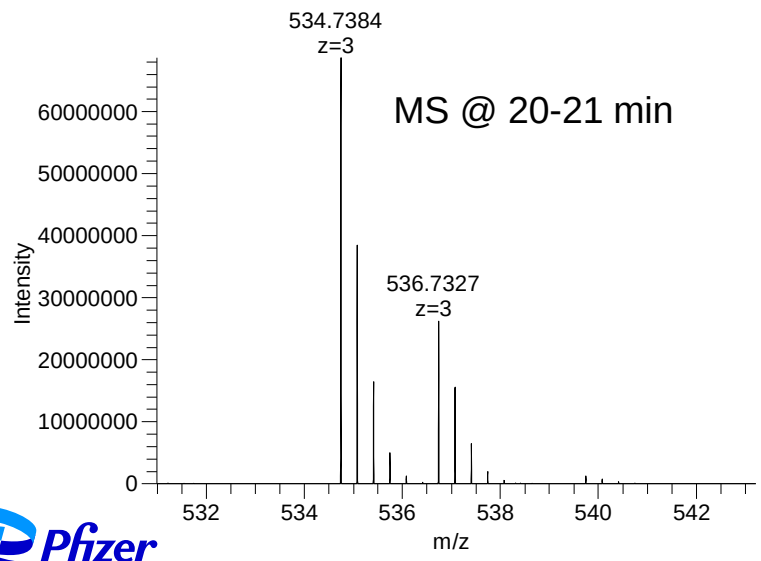
Observed Fragment Ions											
5'Fragments								3'Fragments			
a	a-B	b	c	d	5'-3' Index	Nucleotide	3'-5' Index	w	x	y	z
			$(304)^{-1}$	$(322)^{-1}$	1	C	5				
$(553.1)^{-1}$	$(418.1)^{-1}$	$(571.1)^{-1}$	$(633.1)^{-1}$	$(651.1)^{-1}$	2	A	4	$(690.1)^{-2}$		$(650.1)^{-2}$, $(1301.2)^{-1}$	$(641.1)^{-2}$
$(858.2)^{-1}$		$(876.2)^{-1}$	$(468.6)^{-2}$, $(938.1)^{-1}$	$(956.1)^{-1}$	3	C	3			$(485.6)^{-2}$, $(972.1)^{-1}$	$(954.1)^{-1}$
$(581.1)^{-2}$, $(1163.2)^{-1}$	$(1052.2)^{-1}$	$(590.1)^{-2}$, $(1181.2)^{-1}$	$(621.1)^{-2}$	$(630.1)^{-2}$	4	C	2	$(373)^{-2}$		$(667.1)^{-1}$	$(649.1)^{-1}$
				$(528.7)^{-3}$, $(534.7)^{-3}$, (678.1)-2, (753.6)-2, (762.6)-2	5	G	1	$(220.5)^{-2}$, $(442)^{-1}$	$(424)^{-1}$	$(362.1)^{-1}$	$(344)^{-1}$

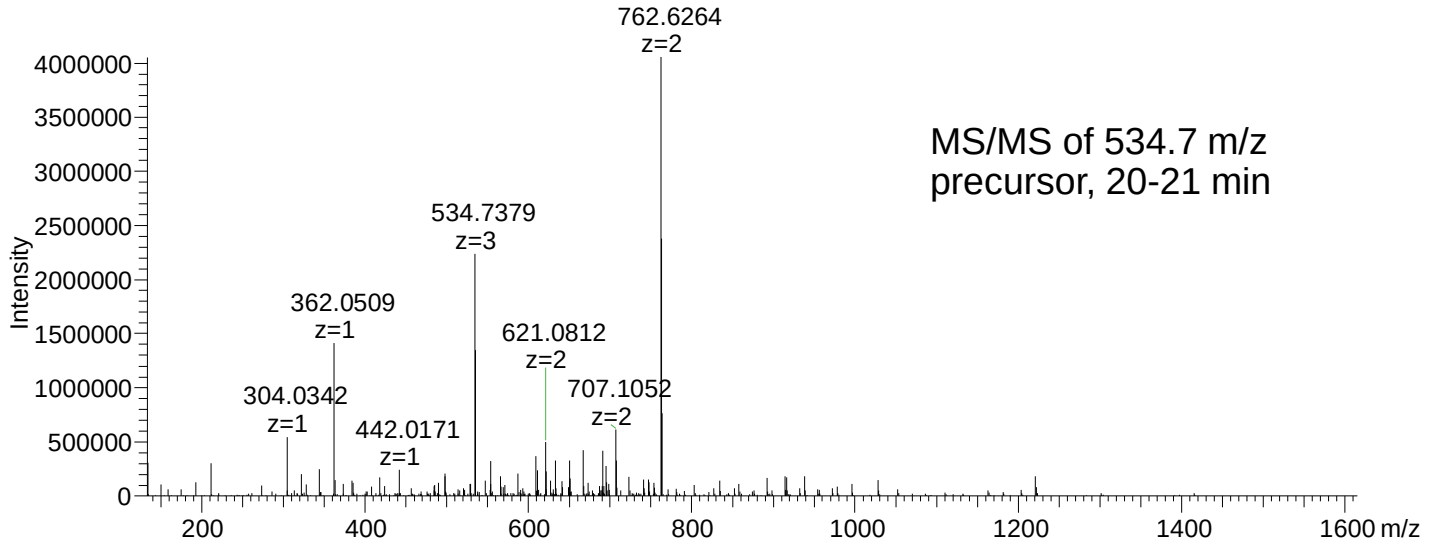
Mixture Peak



- The 534.7 precursor is a mixture of two sequence isomers
  - Its ID was not made by automated software at high confidence because it is a mixture

Observed Mass	Sequence	Theoretical Mass	Error (ppm)	Name	Frequency
1607.2370	CACCG	1607.2344	1.6	R76*	2
1607.2370	CCCAG	1607.2344	1.6	R618*	4
1607.2370	CCACG	1607.2344	1.6	R841	1





Conclusion

- Oligonucleotide mapping via LC-UV-MS/MS directly interrogates the primary structure of RNA, enabling enhanced structural understanding for mRNA vaccines, genetic therapies, and other RNA molecules
- Oligonucleotide mapping assisted the development and commercialization of the Comirnaty® vaccine against SARS-CoV-2
  - Elucidation of Structure (3.2.S.3.1)
  - Comparability (3.2.S.2.6)
  - Data supported regulatory filings to health authorities in 180+ markets
- Semi-automated workflow generates a reproducible and completely annotated oligonucleotide map
  - Annotated chromatographic map; 15-fold more species than a mAb peptide map
  - Sequence coverage map (up to 100% sequence coverage - e.g. BNT162b2)
  - Microheterogeneity assessment of 5′ terminus capping and 3′ terminus poly(A) tail length
- MS/MS fragmentation was optimized and fidelity of identifications verified by decoy sequence searching
- A step-by-step protocol and VBA-enabled data analysis tools are publicly available

Special Thanks

Andrew Dawdy (Pfizer)

Lead Oligonucleotide Mapping Co-Developer

BioNTech

ThermoFisher Scientific

Protein Metrics