# Oligonucleotide mapping using BioPharma Finder software

Authors: Haichuan Liu<sup>1</sup>, Kevin Guo<sup>2</sup>, Jennifer Sutton<sup>1</sup>, Min Du<sup>2</sup> <sup>1</sup>Thermo Fisher Scientific, San Jose, CA <sup>2</sup>Thermo Fisher Scientific, Boston, MA

Keywords: Oligonucleotide, DNA, RNA, data dependent acquisition, tandem mass spectrometry, LC-MS, ion pairing chromatography, IR RP

## **Application benefits**

- High resolution accurate mass mass spectrometry (HRAM MS) coupled with data dependent tandem MS (ddMS<sup>2</sup>) allows accurate mass measurement and confident identification and mapping of oligonucleotides.
- Thermo Scientific<sup>™</sup> BioPharma Finder<sup>™</sup> 4.0 software offers a complete toolset for characterizing oligonucleotides using ddMS<sup>2</sup> data.
- Comparative analysis of ddMS<sup>2</sup> data leads to fast method optimization for analyzing oligonucleotides in different sizes.

## Goal

This work aims to develop and optimize a liquid chromatography (LC) ddMS<sup>2</sup> method for the untargeted analysis of oligonucleotides using Orbitrap-based HRAM MS. We will describe oligonucleotide identification and mapping using the new Oligonucleotide Analysis workflow



available in BioPharma Finder software. The key features that enable comparative analysis of multiple raw files and facilitate optimization of ddMS<sup>2</sup> methods will be highlighted.

## Introduction

Since the first antisense oligonucleotide (ASO) drug was approved by the U.S. Food and Drug Administration (FDA) in 1998, therapeutic oligonucleotides, such as ASO and siRNA, have gained substantial interest in both academia and pharmaceutical industry.<sup>1-3</sup> Currently, eleven oligonucleotide drugs have received approval from the FDA, including nine since 2013, with over 150 drugs in the clinical pipeline (Phase I to Pre-registration). Oligonucleotide therapeutics have been successfully developed for the treatment of rare diseases such as neurological and



hepatic diseases. The applications of oligonucleotide drugs to treat muscle and cardiovascular diseases are being pursued.<sup>1-3</sup> The potential of oligonucleotide therapies for alleviating the coronavirus disease-19 (COVID-19) pandemic was also recently discussed.<sup>4</sup>

The advancement of oligonucleotide therapeutics drives the need to develop robust analytical approaches for characterizing these novel drugs. While the LC-MS method using MS<sup>1</sup> only data has proven to be a viable tool for mass measurement and confirmation of therapeutic oligonucleotides,<sup>5,6</sup> it does not provide information about their sequences base-by-base or exact location of the modifications. By comparison, the ddMS<sup>2</sup> method allows confident identification and mapping of oligonucleotides, thereby enabling umambiguous sequence determination and localization of the modifications. Additionally, ddMS<sup>2</sup> does not require construction of a target mass list, as in the case of a parallel reaction monitoring (PRM) experiment, and hence is easy to execute. Currently, the application of ddMS<sup>2</sup> to oligonucleotide analysis is challenging due to the lack of a complete software package for automatic and confident MS/MS interpretation. To address this need, new tools are added in BioPharma Finder software that can process ddMS<sup>2</sup> data for fast identification and mapping of oligonucleotides. It offers the following key features for oligonucleotide analysis:

- Sequence Manager supports RNA and DNA sequences and provides ease-of-use and flexibility in sequence building with user-defined custom building blocks;
- Oligonucleotide Analysis workflow offers a powerful comparative analysis of multiple raw files, automatic MS<sup>2</sup> annotation, confident sequence identification, and site localization of modifications;
- 3. Intact Mass Analysis workflow supports modified oligonucleotides and provides automatic annotation using sequence or formula.

The first two features will be highlighted in this application note.

## Experimental

## Equipment

- Thermo Scientific<sup>™</sup> Q Exactive<sup>™</sup> Plus Hybrid Quadrupole-Orbitrap<sup>™</sup> Mass Spectrometer
- Thermo Scientific<sup>™</sup> Vanquish<sup>™</sup> Horizon UHPLC System

## Data analysis software

• BioPharma Finder 4.0 software

## Oligonucleotide samples

The lyophilized oligonucleotides were purchased from Integrated DNA Technologies (IDT) (Coralville, IA, U.S.A.). Table 1 lists the six DNA and RNA oligonucleotides analyzed in this study. A stock solution of 1 mg/mL was prepared for each oligonucleotide per quantity specified by IDT. This solution was diluted to 10 µg/mL before ddMS<sup>2</sup> analysis. A mixture of 15mer, 20merAC, and 25merMod at 10 µg/mL each were also analyzed by ddMS<sup>2</sup>.

Table 1. List of synthetic oligonucleotides from IDT. The sequencesfollow a 5'-1'-2' notation defined in BioPharma Finder software, asillustrated in Figure 1. The sequence difference of two isomeric 20mers(20merAC and 20merCA) is highlighted in bold. The mass values in thetable correspond to the monoisotopic peaks.

Name	Sequence	MW (Da)
15mer	Cd-pAd-pGd-pTd-pCd-pAd-pGd-pTd- pCd-pAd-pGd-pTd-pCd-pAd-pGd	4,574.808
20merAC	Cd-pAd-pGd-pTd-pCd-pAd-pGd-pTd- pCd-pAd-pGd-pTd- <b>pAd-pCd</b> -pGd-pTd- pCd-pAd-pGd-pTd	6,114.057
20merCA	Cd-pAd-pGd-pTd-pCd-pAd-pGd-pTd- pCd-pAd-pGd-pTd- <b>pCd-pAd</b> -pGd-pTd- pCd-pAd-pGd-pTd	6,114.057
25merMod	Cd-pAd-pGd-pTd-pMe-pAe-pGd-pld-pTd- pld-pCd-pAf-pGd-pCd-pld-pAd-pGd-pld- pTd-pCd-pld-pAd-pGd-pTd	7,864.345
45mer	Cd-pAd-pGd-pTd-pCd-pAd-pGd-pTd- pCd-pAd-pGd-pTd-pCd-pAd-pGd-pTd- pCd-pAd-pGd-pTd-pAd-pAd-pAd- pAd-pAd-pCd-pAd-pGd-pTd-pCd-pAd- pGd-pTd-pCd-pAd-pGd-pTd-pCd-pAd- pGd-pTd-pCd-pAd-pGd	13,864.369
RNA20mer	Cr-pAr-pGr-pUr-pCr-pAr-pGr-pUr-pCr- pAr-pGr-pUr-pCr-pAr-pGr-pUr-pCr-pAr- pGr-pUr	6,363.877



**Figure 1. Oligonucleotide sequence notation in BioPharma Finder software.** (a) An oligonucleotide sequence is annotated using a 5'-1'-2' notation. (b) List of common annotations of 5' backbone linker, 1' base, and 2' ribose building blocks as well as 5' and 3' terminus. For example, an unmodified CAG is annotated as Cd-pAd-pGd for DNA and Cr-pAr-pGr for RNA in BioPharma Finder software. The CAG with phosphorothioate at 5' backbone linkers is annotated as Cd-sAd-sGd for DNA.

## Solvents

- Water, UHPLC-MS Grade, Thermo Scientific (P/N W81)
- Methanol (MeOH), UHPLC-MS Grade, Thermo Scientific (P/N A456-1)
- 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP), >99.0%, TCI Chemicals (P/N H0424)
- N,N-Diisopropylethylamine (DIPEA), >99.0%, TCI Chemicals (P/N D1599)
- Solvent A: 2% (~190 mM) HFIP and 0.1% (~5.7 mM) DIPEA in water (pH 7.8)

For a 500 mL solution, add 489.5 mL  $\rm H_{2}O,$  10 mL HFIP, and 0.5 mL DIPEA.

 Solvent B: 0.075% (~7.1 mM) HFIP and 0.0375% (~2.1 mM) DIPEA in methanol

For a 500 mL solution, add 499.4 mL MeOH, 375  $\mu L$  HFIP, and 187.5  $\mu L$  DIPEA.

## Ion-pairing reversed phase liquid chromatography

Oligonucleotide separations were performed with a Thermo Scientific<sup>™</sup> DNAPac<sup>™</sup> RP column, (4 µm, 2.1 × 50 mm, P/N 088924) using a Vanquish Horizon UHPLC system. The autosampler was held at 5 °C while the column was maintained at 60 °C with the column oven Thermostatting Mode set to Still Air. The solvents were prepared in the original UHPLC solvent bottles to minimize salt contamination. The LC gradient used in this study is shown in Figure 2.



Figure 2. LC gradient for oligonucleotide separation at a flow rate of 0.4 mL/min

## Mass spectrometry

The Q Exactive Plus mass spectrometer was operated in negative ion mode with calibration performed using the negative ion CalMix within the Exactive Series Tune software. The source and MS method settings are provided in Table 2. The data was acquired in the standard pressure mode using Thermo Scientific<sup>™</sup> Chromeleon<sup>™</sup> Chromatography Data System (CDS). (Note: Data can also be acquired with Thermo Scientific<sup>™</sup> Xcalibur<sup>™</sup> software.) Data analysis was performed using BioPharma Finder software. **Table 2. Source and Full MS/ddMS**<sup>2</sup> **method settings.** <sup>I</sup>S-lens RF level was varied from 40 to 100 when optimized for 20merAC but was kept at 60 for the rest of experiments. <sup>II</sup>A resolution setting of 70,000 (at *m/z* 200) was used for the Full MS portion of Full MS-ddMS<sup>2</sup> experiments. However, a resolution setting of 140,000 was utilized for Full MS only experiments. <sup>III</sup>(N)CE or stepped (N)CE was extensively optimized for 20merAC and 25merMod, as listed in Table 3.

MS source setting	Value
Sheath gas flow rate	40
Aux gas flow rate	20
Sweep gas flow rate	2
Spray voltage (kV)	2.8
S-Lens RF Level	60 <sup>i</sup>
Aux gas heater temp. (°C)	320
Capillary temp. (°C)	300
Sweep water flow	100 mL/min
Properties of the method	Value
Globa	al settings
Use lock masses	Off
Chrom. peak width (FWHM)	6 s
	Time
Method duration	9.0 min
Properties of Full MS/ dd-MS <sup>2</sup> (Top5)	Value
G	eneral
Runtime	0 to 6 min
Polarity	Negative
Default charge state	2
Fu	ull MS
Resolution	70,000 <sup>ii</sup>
AGC target	1e6
Maximum IT	100 ms
Scan range	<i>m/z</i> 550 to 2000
	1
Full MS	→ dams²))
dd-MS	<sup>2</sup> / dd-SIM
Resolution	35,000
AGC target	1e5
Maximum IT	150 ms
Loop count	5
TopN	5
Isolation window	<i>m/z</i> 2.0
Fixed first mass	<i>m/z</i> 300.0
(N)CE / stepped (N)CE	nce: 13, 15, 17 <sup>iii</sup>
dd s	settings
Minimum AGC target	5,000
Intensity threshold	3.3e4
Charge exclusion	1
Peptide match	Preferred
Exclude isotopes	On
Dynamic exclusion	3.0 s

Table 3. List of stepped (N)CEs tested for 20merAC, 25merMod, RNA20mer, and 45mer

Oligonucleotides	Stepped (N)CE	Parameters
20merCA	NCE	10-12-14, 11-13-15, 12-13-14, 12-14- 16, 12-15-18, 13-14-15, 13-15-17, 14-15-16, 14-16-18, 15-17-19, 16-18- 20, 17-19-21, 18-20-22, 22-24-26, 24-26-28, 26-28-30 (16 in total)
	NCE	10-11-12, 10-12-14, 11-13-15, 12-14- 16, 12-16-20, 13-15-17, 14-16-18, 15- 17-19, 16-18-20, 17-19-21, 18-20-22 (11 in total)
25merMod	CE	16-17-18, 17-18-19, 18-20-22, 19-20- 21, 24-26-28, 26-28-30, 28-30-32, 31-32-33, 32-33-34, 33-34-35, 34- 35-36, 35-36-37, 37-38-39, 39-40-41 (14 in total)
RNA20mer	NCE	10-12-14, 12-14-16, 13-15-17, 14-16- 18, 15-17-19 (5 in total)
45mer	NCE	10-12-14, 11-13-15, 12-14-16, 13-15- 17, 14-16-18, 15-17-19, 16-18-20 (7 in total)

## Oligonucleotide analysis in BioPharma Finder software

A new oligonucleotide sequence can be created in Sequence Manager by importing a FASTA file or entering it manually. In this study, all the sequences were added manually in the Manual Input Sequence pane (Figure 3a) as plain format (e.g., CAG TCA GTC AGT ACG TCA GT for 20merAC), which was automatically converted into the Thermo triplet format (Table 1 and Figure 1) once entered. For 25merMod, the sequence of the unmodified 25mer was entered as described above, which was then followed by choosing the modified base and 2' ribose in the Edit Sequence pane (Figure 3b) using drop-down menus. The default variable modifications, including dephosphorylation at 5' terminal, phosphorylation at 3' terminal (for DNA oligonucleotides), and backbone linker and base-specific modifications (e.g., base loss), were used for searching ddMS<sup>2</sup> data (Figure 3c). BioPharma Finder software also allows users to edit the existing building blocks or add new variable modifications (Figure 3d).

(a)								(c)					
▼ Sequ	ence Informatio	on			▼ Manual Ir	nput Sequence		▼ Assign Variable More	difications				
Target (	ligonucleotid	ie	Chain		Chain No.	- Information		Max # Modifications			1		
,	Name	25merMod	Chain	1 ~	Chain Na	me 25merMod		Oligonucleotides	1				
	Description		Managinatania Mana	7064 2440	Input plain fo	ormat (ATCGA) or thermo triple se for BASE	et format (Ad-pTd-pCd-pGd-pAd) no	Modifications	Modific	ations Selected for	r Search		
1	ample Type	Oligonucleoti	Monoisotopic Mass	7004,5440	CAG TCA GAT	T ACA GTC AAG ATC AAG T		Dephosphorylation		5' Terminal		Dephosphorylation	
	ampie type	ongonucieou	Average Mass	7868.13				Dephosphorothiola	tion	Mono. Mass 0			
	Category	Seque Y								Avg. Mass	dd		
Monois	otopic Mass	7,864.3448								Ret	move		
A	verage Mass	7,868.13								Load De	ault Mods		
	Formula	C Apply								Louid DC			
		15 0000						Phosphorylation Phosphorothiolatio	n	3' Terminal		Phosphorylation	
<ul> <li>Oligo</li> </ul>	Sequence Maj	P						Deoxy		Avg. Mass 0			
>1: 1 Cd	-pAd-pGd-pT	d-pMe-pAe-pGd	-pId-pTd-pId-	<u>^</u>						-	ldd		
11 pCc	l-pAf-pGd-pT	d-pCd-pId-pAd-	pGd-pId-pTd-							Rer	move		
21 pCd	-pId-pAd-pG	id-pTd								Load Del	fault Mods		
(b)								Adenine loss(A)		Oligo Variable Mod	dification	Adenine loss(A)	~
(D)								Uracil loss(U)		Mono. Mass 0		Uracil loss(U)	
▼ Edi	t Sequence							Guanine loss(G)		Avg. Mass 0		Guanine loss(G)	
	Select Chai	in 1 v				Apply	Cancel	Hypoxanthine loss	(1)	Residues	udid	Thymine loss(T)	
						CODE OF CODE O	Concer	Deamination(AC)			laa	Oxidation(ACGU)	
	=			151.				Oxidation(ACGU)		Ker	nove	G Depurination (G)	
	ninal			<ul> <li>(5' termina</li> </ul>	hal has prece	edence over linker.)		A Depurination(F)		Load De	fault Mods	C Depyrimidination (C)	~
Seq	uence shou tomized bas	uld contain sar se, 2' ribose o	me oligo building t r backbone linker r	block for a give means these b	en symbol. Duilding bloc	Highlighted in orange cks have same symbol i	but are	(d)					
Seq cust usin	uence shou tomized bas Ig different Triplet	uld contain sar se, 2' ribose o oligo building Backbor	me oligo building t r backbone linker r g block. ne linker	block for a give means these bi Base	en symbol. building bloc	Highlighted in orange tks have same symbol   2' ribose	but are	(d) • Building Block and	Variable Modification Editor				
5' Terr Seq cus usin	uence shou tomized bas g different Triplet	uld contain sar se, 2' ribose o oligo building Backboi	me oligo building b r backbone linker r g block. ne linker	block for a give means these bi Base	en symbol. building bloc	Highlighted in orange cks have same symbol I 2' ribose	but are	(d) • Building Block and Oligo building block Subunit	Variable Modification Editor	v	Oligo Variable Modi Subunit	fication	
5' Terr Seq cusi usir	uence shou tomized bas og different Triplet Cd	uld contain sar se, 2' ribose o oligo building Backboi	me oligo building b r backbone linker r g block. ne linker	block for a give means these bl Base C - Cytosine(	en symbol. building bloc (C4H5 •	Highlighted in orange cks have same symbol I 2' ribose d - Deoxy (DNA)(H	but are	(d) • Building Block and Oligo building block Subunit Name	Variable Modification Editor Backbone linker Phosohorothicate	v	Oligo Variable Modi Subunit Name	fication Base	v
5' Terr Seq cus' usir 1 2	uence shou tomized bas g different Triplet Cd pAd	uld contain sar se, 2' ribose o oligo building Backbor p - Pho	me oligo building b r backbone linker r g block. ne linker v sphate(H3P •	block for a give means these bi Base C - Cytosine( A - Adenine(C	en symbol. building bloc (C4H5 ▼ (C5H5 ▼	Highlighted in orange cks have same symbol i 2' ribose d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) • Building Block and Oligo building block Subunit Name Symbol	Variable Modification Editor Backbone linker Phosphorothioate S Y	v	Oligo Variable Modi Subunit Name Site	fication Base Select	
S' Terr Seq cus' usir 1 2 3	uence shou tomized bas gg different Triplet Cd pAd pGd	uld contain sar se, 2' ribose o oligo building Backbox p - Pho p - Pho	me oligo building t r backbone linker g block. 	Base C - Cytosine( G - Guanine( G - Guanine(	en symbol. building bloo (C4H5 ▼ (C5H5 ▼ (C5H5 ▼	Highlighted in orange cks have same symbol 2' ribose d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(C) v Bulding Block and Oligo bulding Block Suburit Name Symbol Formula	Variable Modification Editor Backbone linker Phosphorothioate s v HAPO35	×	Oligo Variable Modi Subunit Name Site Monoisotopic Mass	floation Base Select	u V
S' Terr Seq cus' usir 1 2 3 4	uence shou tomized bas g different Triplet Cd pAd pGd pTd	uld contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho p - Pho	me oligo building t r backbone linker g block. sphate(H3P • sphate(H3P •	Base C - Cytosine( A - Adenine( G - Guanine( T - Thymine()	(C4H5 (C5H5 (C5H5 (C5H5 (C5H6	Highlighted in orange tks have same symbol I 2' ribose d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) v Euding Block and Oligo building Block Subunit Name Symbol Formula Monistrative Kass	Westelle Modification Editor Backbone linker Phosphorothicate H3POS5 113.854	v	Oligo Variable Modi Subunit Name Site Monoisotopic Mass Average Mass	flation Base Select	v
5' Terr Seq cusi usir 1 2 3 4 5	uence shou tomized bas g different Triplet Cd pAd pGd pTd pMe	p - Pho p - Pho p - Pho p - Pho	me oligo building t r backbone linker r g block. ne linker sphate(H3P * sphate(H3P * sphate(H3P *	Base C - Cytosine( A - Adenine( G - Guanine( T - Thymine( M - 5-Methyl	(C4H5 (C5H5 (C5H5 (C5H5 (C5H6 (C5H6 (C5H6 (C5H6 )	Highlighted in orange tks have same symbol I 2' ribose d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., e - 2'-Methoxyethy,	but are	(d) Coligo building block Subunit Name Symbol Formula Monositopic Mass Average Mass	Variable Modification Editor Backbone linker Photophonothicate s + H40005 113354 11406	۷ 	Oligo Variable Mod Subunt Name Monoisotopic Mass Average Mass Formula	Reation Base Select	v v
S' Terr Seq cus' usir 1 2 3 4 5 6	uence shou tomized basing different Triplet Cd pAd pGd pTd pMe	p - Pho p - Pho p - Pho p - Pho	me oligo building t r backbone linker r g block. me linker sphate(H3P, ¥ ssphate(H3P, ¥ ssphate(H3P, ¥ ssphate(H3P, ¥	Base C - Cytosine( A - Adenine( G - Guanine( T - Thymine( M - 5-Methyl	(C4H5 • (C5H5 • (C5H5 • (C5H5 • (C5H6 • (C5H6 •	Highlighted in orange cks have same symbol I 2' ribose d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., e - 2'-Methosyther, a' T. Methosyther, b' T. Methosyther, b' T. Methosyther, c - 2' Methosythe	but are	(C) • Building Block and Oligo building Block Suburit Name Symbol Formula Monoistorpic Mass Average Mass	Vanakle Modification Editor Backbone linker Photophotothioate H3P035 113554 11455		, Oligo Variable Modi Subunt Name Site Monoisotopic Mass Average Mass Formula	Reston Base Select	- Apply
S' Terr Seq cusi usir 1 2 3 4 5 6	uence shou tomized bas g different Triplet Cd pAd pAd pAd pTd pMe pAe	uld contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho p - Pho p - Pho p - Pho	me oligo building l r backbone linker r g block. ne linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	Base C - Cytosine( A - Adenine( G - Guanine( T - Thymine( M - 5-Methyl A - Adenine(	(C4H5 (C5H5 (C5H5 (C5H6.	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy e - 2'-Methoxyethy	but are	(d) * Building Block and Oligo building Block Suburit Name Symbol Formula Monoistotpic Mass Average Mass Base, backbone linker	Winkle Modification Editor Bactoone linker Phosphorothisate #3P035 113354 114.06 2' ribose and variable motificati	- Apply bn list	Oligo Variable Modi Subunit Name Site Monoisotopic Mass Average Mass Formula	Reation Base Select	Apply
5' Terr Seq cusi usir 1 2 3 4 5 6 7	uence shou tomized bas gg different Triplet Cd pAd pGd pGd pMe pAe pGd	ld contain sar se, 2' ribose o oligo building p - Pho p - Pho p - Pho p - Pho p - Pho p - Pho p - Pho	me oligo building l r backbone linker r g block. ne linker ssphate(H3P ¥ ssphate(H3P ¥ ssphate(H3P ¥ ssphate(H3P ¥ ssphate(H3P ¥	Base C - Cytosine(C A - Adenine(C G - Guanine(C T - Thymine(C M - 5-Methyl A - Adenine(C G - Guanine(C G - Guanine(C	en symbol. uilding bloc (C4H5 • (C5H5 • (C5H6 • (C5H6 • (C5H5 • (C5H5 •	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy e - 2'-Methoxyethy d - Deoxy (DNA)(H	but are	(d) P Building Block and Oligo building Block Subunit Name Symbol Formda Monolicotopic Mass Average Mass Base, backbone linker	Walable Modification Editor Bactione linker Phosphorothiste K × HaPOSS 111354 114.05 2* those and variable modification	Apply an list	Oligo Variable Modi Subunt Name Monoisotopic Mass Average Mass Formula	Reation Base Select Contemport	- Apply Ddete
5' Terr Seq cus' usir 1 2 3 4 5 6 7 8	uence shou tomized bas og different Triplet Cd pAd pGd pAe pAe pAe pGd pId	uld contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho	me oligo building t r backbone linker r g block. me linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	Base C - Cytosine(( A - Adenine() G - Guanine() T - Thymine() M - 5-Methyl A - Adenine() G - Guanine() I - Inosine(C5	(C4H5 • (C5H5 • (C5H5 • (C5H6 • (C5H6 • (C5H6 • (C5H6 • (C5H5 • (C5H5 • (C5H5 • (C5H5 •	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy e - 2'-Methoxyethy d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) I totaling Block and Oligo building Block Suburit Name Symbol Formula Monolisotopic Mass Average Mass Base, backbone linker	Nariable Modification Editor Bacibone linker Phosphorothists	Apply Apply gendel er ste	Oligo Variable Modi Subunit Name Site Monosotopic Mass Average Mass Formula	fication Base Safect Safect Ent Kare Society Mass According According Mass According A	ر مراجع مراحع مراحم مراحع مراحع مراحع مراحم مراحع مراحم مارم مراحم مراحم مراحم مراحم مراحم مراحم مراحم مراحم مراحم مراحم مراحم مراحم ممارما ممارمم مما ممارم ممارم ممارمم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارم ممارمما ممامم ممامم ممماممام ممامم ممارمما ممامم ممامم ممامم ممامممام
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9	uence shou tomized bas gridifferent Cd pAd pGd pTd pMe pAe pGd pId pId	Id contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho	me oligo building t r backbone linker r g block. ne linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	Base C - Cytosine( G - Guanine( T - Thymine( M - 5-Methyl A - Adenine( G - Guanine( G - Guanine( I - Inosine( T - Thymine( T - Thymine( S - Thymine(	(C4H5 (C5H5 (C5H5 (C5H6 (C5H6 (C5H6 (C5H5 (C5H5 (C5H5 (C5H5 (C5H5 (C5H5 (C5H6 (C5H6 )	Highlighted in orange cks have same symbol I 2' ribose d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(c) v Rusleng Block and Oligo building block Suburit Name Symbol Formula Monoisotopic Mass Average Mass Base, backbone linken Suburit	Anable Modification Editor  Backbone linker  Photophototicate   Photophototicate  Photophototicate  Photophototicate  Photoe and variable modificatio  Photoe and variable modificate  Photoe	v Apply n list Symbol ar she As •	Oligo Varishe Mod Subunt Name Site Monoisotopic Mass Average Mass Formula	fication Base Select Se	ب م کونونی کونونی
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9 10	uence shou tomized bas gg different Cd pAd pGd pGd pMe pAe pGd pId pId pId pId	Id contain sar se, 2' ribose o oligo building P - Pho P - Pho	me oligo building b r backbone linker r g block. me linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	lock for a give means these bit Base C - Cytosine(( G - Guanine() T - Thymine() A - Adenine() G - Guanine() I - Inosine(C5 T - Thymine() I - Inosine()	en symbol. i uuilding bloc (C4H5 (C5H5 (C5H6 (C5H5 (C5H5 (C5H5 (C5H5 (C5H5 (C5H5 (C5H5 (C5H6	Highlighted in orange cks have same symbol I 2' ribose d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) v Building Block and Oligo building Block Suburit Name Symbol Formula Monoistorpic Mass Average Mass Base, backbone linker Solumit 5 al 10 Backbone linker 2 al bose	Vanake Modification Editor	e Appy n lat Sproba er ale At d	Oligo Variable Mod Subunt Name Site Monoisotopic Mass Average Mass Formula Formula Mass Horpozs H-QH	Reation  Base  Select  Address Subger Mass  Higher Subger Mass  Higher Subger Mass  -1113554  -15995	<ul> <li>Apply</li> <li>Deter</li> <li>roge Mast</li> <li>114.06</li> <li>-16.00</li> </ul>
5' Terr Seq cuss usir 1 2 3 4 5 6 7 8 9 10	uence shou tomized bas gdifferent Cd pAd pGd pTd pMe pAe pGd pId pId pTd pId	Id contain sar se, 2' ribose o oligo building p - Pho p - Pho	me oligo building b r backbone linker r g block. ne linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	lock for a give mass these bit Base C - Cytosine(( A - Adenine() G - Guanine() T - Thymine() M - 5-Methyl A - Adenine() G - Guanine() T - Thymine() T - Thymine() I - Inosine(C5 C - Castion)	en symbol. i uuilding bloc (C4H5 C5H5 C5H6 C5H6 C5H6 C5H5 C5H5 C5H5 C5H4 SH4N4 SH4N4 C5H4	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H e - 2'-Methoxyethy d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) * Ruideng Block and Oligo building Block Suburit Name Symbol Formula Monoistotpic Mass Average Mass Base, backbone linker * dat * 100 Baskbone linker * onbore	Winable Modification Editor Bactione linker Phosphorothioate H3P035 111354 11406 21 ribose and variable modificatio Phosphorothicate Devy (DNA) Hydrox (RA)	v Apply In list Special or site S d r	Oligo Variable Modi Subunt Name Site Monoisotopic Mass Average Mass Formula Formula Formula HIPO35 H-CH OH OH	Restion Base Select Select Managesideption Mass 113.594 -15.595 0.000	<ul> <li>Apply</li> <li>Delete</li> <li>rage Mass</li> <li>11406</li> <li>0.00</li> </ul>
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9 10 11	uence shou tomized basi g different Cd pAd pGd pGd pAe pGd pId pId pId pId pId pId pCd	Id contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho	me oligo building l r backbone linker r g block. ne linker ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P *	Base C - Cytosine(/ A - Adenine(/ G - Guanine(/ T - Thymine(/ M - 5-Methyl A - Adenine(/ G - Guanine(/ I - Inosine(CS T - Thymine(/ I - Inosine(CS C - Cytosine()	en symbol. i c(24H5 • (C5H5 • (C5H5 • (C5H6 • d cytos • (C5H5 • (C5H5 • (C5H5 • SH4N4 • (C5H6 • SH4N4 •	Highlighted in orange ks have same symbol d - Deoxy (DNA)(H d - Deoxy (DNA)(H	but are	(d) P Building Block and Oligo building Block Suburit Name Symbol Formula Monosiotopic Mass Average Mass Base, backbone linker P 10 Backbone lin	Weakle Modification Editor Bactione linker Phosphorothisate  x  haP0035 113054 114.05 21 tibose and variable modificatio er Decoy (DNA) Hydroxy (RNA) Methoxy Envine	Apply Apply Symbol or Me Bat S S d d r m m	Oligo Variable Modi Subunt Name Monoisotopic Mass Average Mass Formula Formula Mass H-QH OH-OH OCH3-OH S-CH	Reation Base Select Select Commonweak Select	• • • • • • • • • • • • • •
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9 10 111 12	uence shou tomized basi g different Cd pAd pGd pGd pAe pGd pId pId pId pId pId pCd pAf	Id contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho	me oligo building t r backbone linker r g block. me linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	Base Base C - Cytosine( A - Adenine( G - Guanine( T - Thymine( M - 5-Methyl A - Adenine( G - Guanine( I - Inosine(C5 T - Thymine( I - Inosine(C5 C - Cytosine( A - Adenine( C - Cytosine( A - Adenine( C - Cytosine( C	en symbol. i wuilding bloc (C5H5 • (C5H5 • (C5H5 • (C5H5 • (C5H5 • (C5H6 • 5H4N4 • (C4H5 •	Highlighted in orange ks have same symbol d - Deoxy (DNA)(H d - Deoxy (DNA)(H f - Fluorine(F-OH, 1)	but are	(c) v Ruiding Block and Olipo building block Suburit Name Symbol Formula Monoisotopic Mass Average Mass Base, backbone linker 10 2 ribose 12 2 ribose 13 2 ribose 14 2 ribose 14 2 ribose 15 2 ribose	Variable Modification Editor	v Appij n lat Spekar in sk Se S s d r r m f f c	Oligo Variable Modi Subunt Name Site Monosotopic Mass Average Mass Formula Formula H3P035 H-CH OH-OH OH-OH OCH3-OH F-CH COHSO	Reation      Base      Select      Manusatage Mass      Manusatage Mass	• Apply Delete 1110,05 -16,00 0,00 14,03 1,09 5,5,08
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9 10 11 12 2' Terr	uence shou tomized basi g different Cd pAd pGd pGd pAe pGd pId pId pId pId pCd pAf	Id contain sar se, 2' ribose o oligo building Backbor P - Pho P - Pho	me oligo building b r backbone linker r g block. sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	lock for a give means these bit Base C - Cytosine(C G - Guanine(C T - Thymine(C T - Thymine(C G - Guanine(C G - Guanine(C T - Thymine(C T - Thymine(C T - Thymine(C C - Cytosine(C A - Adenine(C	en symbol. J wuilding bloc (C4H5 • (C5H5 • (C5H6 • (C5H6 • (C5H5 • (C5H6 • (	Highlighted in orange cks have same symbol I 2' ribose d - Deoxy (DNA)(H., d - Deoxy (DNA)(H., f - Fluorine(F-OH, 1, d - Deoxy (DNA)(H., f - Fluorine(F-OH, 1, d - Deoxy (DNA)(H., f - Fluorine(F-OH, 1, d - Deoxy (DNA)(H., f - Fluorine(F-OH, 1, f - Fluorine(F-OH, 1) f - Fluorine(F-OH	but are	(C) Clipo building block and Suburit Name Symbol Formula Monoistopic Mass Average Mass Base, backbone linker <u>Suburit</u> <u>10</u> Backbone linker <u>10</u> Backbone linker <u>10</u> Jackbone linker <u>10</u>	Vanake Modification Editor	v Aqvýr n lat Systel de ak Adv s s d t r r f t e e s	Oligo Variable Mod Subunt Name Site Monosotopic Mass Average Mass Formula Hepos Hepos Hepos Hepos Hepos OcH-OH OcH-OH OcH-OH OcH-OH OcH-OH	Reation  Base  Extent  Select  Moreorport (Select)  Moreorport (Select)  Moreorport (Select)  Moreorport (Select)  Moreorport (Select)  Moreorport (Select)  Select  Select Select  Select Selec	ب ب کواچل تعوی Mass - 1600 - 1600 - 1600 - 1600 - 1403 - 1500 - 1403 - 1500 - 1403 - 1500 -
5' Terr Seq cus usir 1 2 3 4 5 6 7 8 9 10 11 12 2 3' Terr	uence shou tomized basy og different Cd pAd pGd pGd pAe pGd pId pId pId pId pCd pAf	Id contain sar se, 2' ribose o oligo building p - Pho p - Pho	me oligo building b r backbone linker r g block. ne linker sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P * sphate(H3P *	lock for a give means these bit Base C - Cytosine(( A - Adenine() G - Guanine() T - Thymine() A - Adenine() G - Guanine() I - Inosine(C5 T - Thymine() I - Inosine(C5 C - Cytosine() A - Adenine() V	en symbol. J wuilding bloc (C4H5 • (C5H5 • (C5H6 • (C5H6 • (C5H6 • (C5H5 • 5H4N4 • (C5H6 • 5H4N4 • (C4H5 • (C4H5 •	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H) d - Deoxy (DNA)(H d - Deoxy (DNA)(H d - Deoxy (DNA)(H) d - De	but are	(d) v Ruideng Nock and Glipo building Mock Subunt Name Symbol Formula Monoistotpic Mass Average Mass Base, backbone linker V and P 100 Backbone ninker 12 2' riboze 13 2' riboze 14 2' riboze 15 2' riboze 15 2' riboze 15 2' riboze 15 2' riboze 16 2' riboze 17 3' riboze 17 3' rimnial 18 3' riboze 18 3'	Venake Modification Editor	Apply Apply n lat s t t t t t t t t t t t t t t t t t t	Oligo Variable Modi Subunt Name Site Monoisotopic Mass Average Mass Formula Pormula MBPO35 H-POH OH-OH OH-OH OH-OH CHHO HISPO4 HISPO4 HISPO55	Reston Base Select Select Con Advances degree Mars 113554 113554 1595 0000 14.016 1.996 58.042 97577 113554 -79566	<ul> <li>Apply</li> <li>Delete</li> <li>rage Mass</li> <li>114.06</li> <li>1-600</li> <li>0.00</li> <li>1.403</li> <li>1.99</li> <li>54.08</li> <li>98.00</li> <li>114.08</li> <li>98.00</li> <li>114.08</li> <li>7.99.8</li> </ul>
5' Terr Sec cus usii 1 2 3 4 5 6 7 8 9 10 11 12 3' Terr	uence shou tomized basi g different Cd pAd pGd pGd pAe pGd pId pId pId pId pId pId pCd pAf	Id contain sar se, 2' ribose o oligo building p - Pho p - Pho	me oligo building l r backbone linker r g block. ne linker ssphate(H3P ssp	olock for a give means these by Base C - Cytosine(i A - Adenine(i G - Guanine(i T - Thymine(i A - Adenine(i G - Guanine(i C - Guanine(i C - Guanine(i C - Cytosine(i A - Adenine(i V	en symbol. J puilding bloc (C4H5 • (C5H5 • (C5H5 • (C5H5 • (C5H5 • 5H4N4 • 5H4N4 • (C5H5 •	Highlighted in orange ks have same symbol I d - Deoxy (DNA)(H d - Deoxy (DNA)(H) d - Deoxy (DNA)(H d - Deoxy (DNA)(H) d - D	but are	(d) r Building Block and Olipo building Block Suburit Name Symbol Formula Monoistopic Mass Average Mass Base, backbone linker <u>Suburit</u> P 100 Backbone linker <u>Suburit</u> P 100 Backbone linker <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Suburit</u> <u>Subu</u>	Musike Modification Editor Bactoone linker Phosphorothioate H42PO25 111354 114205 21 ribose and variable modificatio Phosphorothioate Denyo (DNA) Hydroxy (RA) Methoay Hydroxy (RA) Methoay Picurine 21	Apply Apply on lat South on the Southernorm F F S S	Oligo Variable Modi Subunt Name Site Monoisotopic Mass Average Mass Formula H3P035 H-OH OH OH OCH3-OH F-OH CH460 H3P04 H3P035 -4P025	Reation Base Select Sel	<ul> <li>Apply</li> <li>Delete</li> <li>114.06</li> <li>-16.00</li> <li>14.03</li> <li>199</li> <li>58.08</li> <li>98.00</li> <li>114.06</li> <li>-7998</li> <li>-96.05</li> </ul>
5' Terr Sec cus usir 1 2 3 4 5 6 7 8 9 10 11 12 3' Terr	uence shou tomized basi g different Cd pAd pGd pGd pGd pGd pId pId pId pId pCd pAf	Id contain sar se, 2' ribose o oligo building Backbor p - Pho p - Pho	me oligo building l r backbone linker r g block. ne linker ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P * ssphate(H3P *	Block for a give means these bit Base C - Cytosine(i A - Adenine(i G - Guanine(i T - Thymine(i A - Adenine(i G - Guanine(i I - Inosine(C5 C - Cytosine(i A - Adenine(i X - Adenine(i)	en symbol. J wilding bloc (C5H5 • (C5H5 • (C5H5 • (C5H5 • (C5H5 • (C5H5 • (C5H6 • (C4H5 • (C5H5 •	Highlighted in orange ks have same symbol d - Deoxy (DNA)(H d - Deoxy (DNA)(H f - Fluorine(F-OH, 1) r - Hydroxy (RNA m - Methoxy(OC f - Fluorine(F-OH) e z - 2' Methowat	but are	(c) v Ruiding Block and Olipo building block Suburit Name Symbol Formula Monoiscropic Mass Average Mass Average Mass Base, backbone linker 11 2 r hobe 12 2 r hobe 13 2 r hobe 14 2 r hobe 15 2 reminal (h 15 2 reminal (h 16 3 2 reminal (h 17 3 2 reminal (h 18 3 2 reminal (h 19 3 2 reminal (h) 19 3 2 reminal (h 19 3 2 reminal (h) 19 4 2 reminal (h) 19 4 2 reminal (h) 19 4 2 reminal (h) 19 4 2 reminal (h) 19 5 2 reminal (h)	Variable Modification Editor	v Apply n list Spectar to ste d f f f f f f f f f f f f f f f f f f	Oligo Variable Modi Subunt Name Site Monosotopic Mass Average Mass Formula Harmale Harmale H-CH OH-OH OH-OH OH-OH OH-OH OH-OH OH-OH HarPod HarPod HarPod HarPod HarPod HarPod HarPod	Reation	• • • • • • • • • • • • • •

**Figure 3. BioPharma Finder software provides flexibility in creating and modifying oligonucleotide sequences.** The sequence can be added as a plain format in (a) Manual Input Sequence and modified in (b) Edit Sequence using drop-down menus, with variable modifications defined in (c) Assign Variable Modifications. The Building Block and Variable Modification Editor (d) allows the user to modify the building blocks or add new variable modifications.

The identification and mapping of oligonucleotides were performed in the Oligonucleotide Analysis workflow within BioPharma Finder software. The processing method for searching ddMS<sup>2</sup> data was created (as a new method) from the Basic Default Method by making two minor changes to the identification parameters: 1) increase maximum oligonucleotide mass from 11,000 to 15,000 under Oligonucleotide Identification, and 2) uncheck "Enable Mass Search for Unspecified Modifications" under Advanced Search. Other identification parameters include: Use MS/MS = Use All MS/MS, Mass Accuracy = 5 ppm for all oligonucleotides except 45mer and 10 ppm for 45mer, and minimum confidence = 0.80. For component detection. the task of "Final All lons in the Run" was chosen. All other component detection parameters such as MS noise level and S/N threshold were automatically determined by BioPharma Finder software.

During method optimization, multiple raw files were processed in the multi-consensus format, which allows a comparative analysis of all data in one result file. The results were reviewed in the Process and Review and Mapping pages, where the comparative analysis was facilitated by a rich amount of information about the oligonucleotides identified, including extracted ion chromatogram (XIC), sequence annotations, trend ratio or MS area, fragment coverage map, full scan vs. deconvolution spectra, and experimental vs. predicted (or another experimental) MS<sup>2</sup> spectra. The comparisons of trend MS area, fragment coverage sequence, and MS<sup>2</sup> spectra helped determine the best method settings for a specific oligonucleotide.

## **Results and discussion**

## HRAM data of oligonucleotides

The oligonucleotides were separated based on ion-paring chemistry using a Vanquish Horizon UHPLC system coupled with a DNAPac RP column. The experimental condition employed in this study offers highly reproducible retention and separation of oligonucleotides in different sizes. Figure 4 displays typical mass spectra of 20merAC and 45mer. The inset in Figure 4b showcases excellent baseline separation for the isotopes of the 15- of 45mer at a resolution setting of 140,000 (at *m/z* 200), leading to the accurate mass measurement of the 45mer (0.5 ppm for the monoisotopic peak of the 15-).



Figure 4. Mass spectra of (a) 20merAC and (b) 45mer with the inset displaying the isotopic distribution of the 15-. Both spectra were acquired at a resolution setting of 140,000 (at m/z 200).

## **Optimization of S-lens RF level**

BioPharma Finder software provides an array of plots and graphs for comparative analysis of multiple raw files acquired with different method settings. This feature can be utilized to determine the optimal S-lens RF level for analyzing oligonucleotides. In this study, eight ddMS<sup>2</sup> raw files of 20merAC were acquired by varying S-lens RF levels from 40 to 100 and were then searched in a multi-consensus format in BioPharma Finder software. Figure 5 displays Trend MS Area plots of three different charge states (4-, 8-, and 10-) of 20merAC at eight different S-lens RF levels (Figure 5a-5c) and their selected ion chromatograms (SICs) at two extreme RF settings (RF = 40 and 100) (Figure 5d-5f). The plots show that while the S-lens RF level had little effect on the most abundant charge state (8-) of 20merAC (Figure 5b), it greatly affected the intensities of the 4- and 10- (Figure 5a and 5c), albeit in an opposite manner. An S-lens RF level of 60-65 provided a good signal for all the charge states of 20merAC.

## Optimization of HCD collision energy (CE) for oligonucleotides

A key method parameter for obtaining high-quality ddMS<sup>2</sup> data of oligonucleotides is the HCD CE. The strategy described above for optimization of S-lens RF level was employed to determine the optimal HCD CE (absolute) or normalized CE (NCE) for 20merAC, 25merMod, and 45mer (Table 3). To assess the quality of MS<sup>2</sup> data, the fragment coverage map and annotated MS<sup>2</sup> spectra were compared across multiple raw files acquired using different stepped (N)CE settings (Table 3). Figure 6 shows that a stepped NCE of 10-12-14 was insufficient for fragmenting the 8- of 20merAC (Figures 6a and 6b). By comparison, a complete fragment coverage was obtained for 20merAC using stepped NCE15-17-19 (Figure 6c and 6d).



Figure 5. Trend MS Area plots of the (a) 4-, (b) 8-, and (c) 10- of 20merAC at S-lens RF levels of 40–100 and SICs of three charge states (d-f) at RF settings of 40 and 100. The SICs confirmed the same peaks used for comparison

### (a)

· /																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cd-	-pAd-	-pGd-	-pTd-	-pCd-	pAd-	-pGd-	pTd-	-pCd-	pAd-	-pGd-	-pTd-	pAd-	-pCd-	pGd	-pTd-	pCd	-pAd-	pGd	-pTd
20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Cd-nAd-nGd-nGd-nCd-nAd-nGd-nGd-nCd-nAd-nGd-nCd-nGd-nCd-nCd-nAd-nGd-nCd

	a18 BI6 1/897	0							
	a10-D[0-](007	-)							
a14	-B[5-](822.3)	w6[3-](627.8)							
a13-B[	5-](760.1)	w7[4-](542.8)							
a10-B[3-](951.8)		w10[5-](623.5)							
a9-B[3-](855.1)		w11[5-](686.1)							
d8[4-](621.1)		z12[4-](905.	7)						
a6-B[2-](810.1)		w14[6-](725.3)							
			w5[3-](518.1)						
			w3[2_1						

Color Code for Ion Intensity >8.3e+05 >6.5e+05 >5.1e+05 >4.0e+05 >3.2e+05





Color Code for Ion Intensity >8.0e+05 >4.3e+05 >2.3e+05 >1.2e+05 >6.7e+04

Figure 6. Fragment coverage map and annotated MS<sup>2</sup> spectra of 20merAC (8-) acquired at stepped NCEs of 10-12-14 (a and b) and 15-17-19 (c and d). The two annotated MS<sup>2</sup> spectra (b and d) were compared in a stacked format in BioPharma Finder software.

The average structural resolution (ASR) is a score calculated from the total number of base residues and the number of oligonucleotide fragments detected. An ASR of 1.0 indicates the detection of fragments that cover all bond cleavages between oligonucleotide residues. The further the ASR deviates from 1.0, the less complete the fragment coverage map for an oligonucleotide is. The ASR can be used as a guide to determine the optimal range of stepped NCEs. Table 4 lists ASRs measured for the charge states 4- to 10- of 20merAC at eleven stepped NCEs. This table reveals that a broad range of stepped NCEs can provide complete or nearly complete fragment coverages for different charge states of 20merAC (ASR = 1.0-1.2, in the green background), with the best results obtained from stepped NCEs of 15-17-19, 16-18-20, and 17-19-21.

Similarly, a range of stepped (N)CEs were determined for high charge states (8- to 12-) of 25merMod based on the ASRs obtained (Table 5). Figure 7 displays fragment coverage maps and annotated MS<sup>2</sup> spectra of the most abundant charge state (10-) of 25merMod at two-stepped (N)CEs that produced complete sequence coverage (ASR = 1.0). By contrast, none of the stepped (N)CEs tested in this study gave satisfactory fragment coverage for the 4- of 25merMod (Table 5), indicating the importance of using the right charge state(s) and (N)CE(s) for achieving complete fragment coverage. However, the MS<sup>2</sup> of the 4- generated abundant large fragments (e.g.  $b_{24}$  and  $y_{24}$ ) that may provide complementary information to that obtained from higher charge states.

Table 4. ASRs measured for the charge states 4- to 10- of 20merAC at eleven stepped NCEs (out of sixteen tested). Green: ASR = 1.0-1.2, Yellow: ASR = 1.3-1.5, Red: ASR > 1.5+.

NCEs	101214	111315	121416	131517	141618	151719	161820	171921	182022	222426	262830
4-	1.3	1.3	1.1	1.2	1.3	1.3	1.3	1.3	1.4	1.7	1.7
5-	1.3	1.3	1.3	1.3	1.2	1.1	1.1	1.1	1.3	1.3	1.3
6-	5.0	2.2	1.5	1.3	1.3	1.1	1.2	1.1	1.1	1.1	1.3
7-	5.0	2.0	1.3	1.3	1.3	1.1	1.0	1.1	1.0	1.1	1.3
8-	2.0	1.4	1.3	1.2	1.2	1.0	1.0	1.1	1.1	1.0	1.5
9-	1.8	1.2	1.1	1.1	1.0	1.1	1.0	1.0	1.2	1.3	1.7
10-	1.1	1.1	1.1	1.1	1.1	1.0	1.1	1.1	1.1	1.4	1.5

Table 5. ASRs measured for selected charge states (4-, 8-, 10-, and 12-) of 25merMod at twelve stepped (N)CEs (out of twenty-five tested). Green: ASR = 1.0-1.2, Yellow: ASR = 1.3-1.5, Red: ASR > 1.5+.

			Steppe	ed CEs			Stepped NCEs							
	161718	192021	242628	283032	343536	383940	121416	141618	151719	161820	171921	182022		
4-	25.0	25.0	6.3	3.1	2.5	1.9	1.9	1.6	1.8	1.9	1.9	2.1		
8-	8.3	1.9	1.1	1.0	1.0	1.5	1.7	1.3	1.1	1.1	1.1	1.1		
10-	1.1	1.0	1.0	1.3	1.6	2.1	1.3	1.1	1.0	1.0	1.0	1.0		
12-	1.0	1.0	1.6	1.8	2.1	2.3	1.0	1.0	1.0	1.0	1.1	1.0		

## (a)

1 2 3 4 5 6 Cd-pAd-pGd-pTd-pMe-pAe	7 8 9 10 11 - pGd- pId- pTd- pId- pCd-	12 13 14 pAf-pGd-pTd-	15 16 13 pCd-pId-pA	18 19 20 d-pGd-pId-pTd	21 2 - pCd- pI	2 23 24 25 d-pAd-pGd-pTc
	a	24-B[8-](923.2)				xl
	a23-1	B[8-](884.4)				w2
	a22-B[8	-](844.8)				w3[2-]
	a20[7-](898.9	)				w5[3-](521.4)
	a19-B[7-](836.1)				w6[	2-](934.6)
	a18[6-](946.1)				w7[3-]	(727.8)
	a17-B[6-](868.5)				w8[4-](62	7.8)
	a16-B[6-](816.1)			wS	[4-](706.1)	
	a15[5-](944.2)			w10[4	-](784.6)	
	a14[5-](886.0)			w11[5-](6	85.3)	
a13	-B[4-](994.2)			w12[5-](746.	1)	1
b12[5	·](763.1)					w4[2-]
al1-B[4-](	838.9)			w14[6-](731.6)		
a10-B[4-](760	.6)		w	15[6-](779.8)		
a9[3-](954.9)			w16[	7-](713.1)		
a8-B[3-](808.2)			w17[7-](	(756.8)		
a7-B[2-](1047.7)			y 18[7-](79	0.3)		
c6[3-](665.5)			y 19[7-](837.1)	( <u></u>		
c5[2-](805.2)		W	20[8-](790.9)			
a4(1154.2)		w21[	[8-](838.1)			
a3-B		w22[9-](	(778.7)			
a2-B		w23[8-](917	73)			
		w24[9-](849.8)	1			

Color Code for Ion Intensity

## (c)

 $\frac{1}{Cd} - \frac{2}{pAd} + \frac{3}{pAd} + \frac{p}{Pd} - \frac{2}{pAd} + \frac{p}{pAd} + \frac{p}{pBd} + \frac{p}{pBd} + \frac{19}{pHd} + \frac{19}{pHd} + \frac{12}{pHd} + \frac{13}{pHd} + \frac{13}{pHd} + \frac{14}{pHd} +$ 

1 2 3 4 5 6 Cd-pAd-pGd-pTd-pMe-pAe-p	7 8 9 10 11 1 Gd-pId-pTd-pId-pCd-pA	12 13 14 15 Af-pGd-pTd-pCd-p	16 17 18 19 Id-pAd-pGd-pId-	20 21 22 23 24 25 pTd-pCd-pId-pAd-pGd-pTd						
	a24-	·B[9-](820.6)		wl						
	a23-B[8	8-](884.1)		w2						
	a22-B[8-](8	344.6)		w3[2-]						
	c21[8-](832.6	5)		w4[2-]						
	a20[7-](899.0)			w5[3-](521.4)						
	a19-B[7-](836.1)			w6[3-](622.8)						
	a18-B[7-](789.0)			w7[3-](727.4)						
	a17-B[6-](868.5)			w8[4-](627.6)						
	a16-B[6-](816.1)			w9[4-](705.9)						
	a15[5-](944.2)			y 10[4-](764.6)						
a1	4[5-](886.2)		w11	[5-](685.3)						
a13-B[	4-](993.9)		w12[5-]	](746.1)						
b12[5-](7	63.1)	y 13[5-](795.9)								
a11-B[4-](838	3.9)	w14[6-](731.6)								
a10-B[4-](760.4)		w15[6-](779.8)								
a9[3-](954.9)			w16[7-](713.2)							
a8-B[3-](807.8)			w17[7-](756.5)							
a7-B[2-](1047.7)		wl	8[7-](801.4)							
c6[3-](665.5)		x19[7-	-](846.0)							
a5-B[2-]		z20[7-](89	90.1)							
a4(1154.2)		w21[9-](744.6	6)							
a3-B		w22[9-](778.4)								
a2-B		w23[8-](917.1)								
d1		w24[9-](849.8)								



Color Code for Ion Intensity >1.4e+05 >6.4e+04 >2.9e+04 >1.3e+04 >5.9e+03

Figure 7. Fragment coverage map and annotated MS<sup>2</sup> spectra of 25merMod (10-) at stepped NCE15-17-19 (a and b) and CE19-20-21 (c and d).



The optimal range of stepped NCEs determined for 20merAC was further applied to RNA20mer and 45mer. Complete or nearly complete sequence coverage was obtained for these two oligonucleotides (Figure 8).

## (a) RNA20mer, 9-

 $\frac{1}{Cr} + \frac{1}{PGr} + \frac{1}{$ 



## Identification of oligonucleotides by ddMS<sup>2</sup> in a mixture

From the result described above, it is possible to use a single stepped NCE setting to achieve complete coverages for oligonucleotides in similar sizes. In our further test, a stepped NCE of 15-17-19 was found to provide complete fragment coverages (ASR = 1.0) for three oligonucleotides (15merGT, 20merAC, and 25merMod) in a mixture (Figure 9). This result further demonstrates the power of employing BioPharma Finder software to facilitate the optimization of (N)CEs for oligonucleotide mapping.

## (b) 45mer, 16-



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 14 35 36 37 38 39 40 41 42 43 44 45 Cd-pAd-pGd-pTd-pCd-pAd-pGd-pTd-pCd-pAd-pGd-pTd-pCd-pAd-pAd-pAd-pAd-pAd-pAd-pAd-pAd-pCd-pAd-pGd-pTd-pCd-pAd-pGd

Color Code for Ion Intensity

Figure 8. Fragment coverage maps of (a) 9- of RNA20mer and (b) 16- of 45mer acquired at stepped NCEs of 15-17-19, and 13-15-17, respectively



#### (b) Cd + pAd + pGd + pTd + pCd + pAd + pGd + pTd + pAd + pAd + pGd + pTd + pAd +-pGd pGd pTd hAd

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 C4 = A4 = C4 = T4 = C4 = A4 = C4 = T4 = C4 = A4 = C4 = T4

Cu-pAu-pOu-	pra-pca-	pag-bog-	pra-pca-	pha-pc	na-bia-	pca-pAa-pa			
		a14-B[	5-](817.7)			W			
		a13-B[5-](1	159.7)			w2			
		a12-B[4-](874.	1)						
	a11	-B[4-](791.6)				y4(1172.2)			
	a10 B[	1 ](713.6)			w5[	3 ](526.4)			
	a9-B[3-](8	(55.5)			w6[3-](	(630.8)			
	c8[4-](616.6	)	y7[3-](700.5) z8[3-](796.1) w9[4-](703.6)						
d	7[4-](545.1)								
a6-B[2	-](810.1)								
a5-B[2	4		,	w10[5-](62	5.3)				
d4[2-]			z11	[4-](829.6)					
a3-B			w12[5-]	](744.1)					
a2-B	5								
		v	v14[6-](726.8)	)					

Color Code for Ion Intensity >5.4e+06 >2.2e+06 >8.6e+05 >3.5e+05 >1.4e+05

## (c)

Cd\_pAd\_pGd\_pTd\_pCd\_pAd\_pGd\_pTd\_pCd\_pAd\_pGd\_pTd\_pAd\_pCd -pTd -nGd

а-ряа-роа-рта-рес	-pAd-pot	-pru-	peu	. prid.	-pou-	pru	pAu	-pcu	-pou	-pro-	peu	priu-	pou	pre
			a19	-B[7-](	805.0)									wl
		-	a18-B[	6-](887	.3)								W	2
		a17	-B[6-](	839.1)									w3[2-]	
	al	5-B[5-]	(880.3)								wő	2-](777	.6)	
	a14-B[5-](822.5)										w6[3-]	(627.8)		
	a13-B[5-](760.1)										[3-](724	1.4)		
	a12-B[4-](874.4)										(801.8)			
al	1-B[4-](792.1	)							w9	[4-](69	7.4)			
a10-E	[4-](713.6)						w10[4-](779.6)							
a9-B[3-]	(855.1)						w11[5-](686.1)							
d8[4-](621	1)						z12[4-](905.7)							
a7-B[2-](966.7)							w13[6-](670.3)							
a6-B[2-](810.1)						,	w14[6-](725.4)							
a5-B[2-]							-					y4[	2-]	
a4(1154.2)	w16[7-](707.5)													_
a3-B	w17(7-)(751.1)													
a2-B					z18[6-]	914.8	)							-
d1				w19	[7-](84)	2.7)								-
							-							_

>1.4e+06 >6.8e+05 >3.2e+05 >1.5e+05 >7.2e+04

4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Cd-pAd-pGd-pTd-pM	Me-pAe-pGd-p	Id-pTd-pId-pCd-	pAf-pGd-pTd	l-pCd-pId-pA	d-pGd-pId-pTd-	pCd-pId-p	Ad-pGd-pTd	
a23-B[8-](884.0)							w2	
a22-B[8-](844.8)							w3[2-]	
a20[7-](899.0)						w5[3-](521.4)		
a19-B[7-](\$36.0)						w6[3-](622.8)		
a18[6-](946.1)						w7[3-](727.4)		
a17-B[6-](868.5)						w8[4-](627.8)		
a16-B[6-](816.1)					w9[4-](706.1)			
					w4[2-]			
a14[5-](886.2)				w11[5-](685.3)				
		w12[5-](745.9)						
b12[5-](763.3)							x1	
a11-B[4-](838.9)			w14[6-](731.8)					
a10-B[3-](1014.2)			w15[6-](779.8)					
a9[3-]		w16[7-](713.1)						
a8-B[3-](8		w17[7-](756.8)						
a7-B[2-](1047	7.7)							
c6[3-](665.5)			x19[7-](846.1)					
c5[2-](805.1)								
a4(1154.2)	a4(1154.2) w21[9-](744.9)							
a3-B	<b>a3-B</b> w22[9-](778.4)							
a2-B			w23[9-](8	14.9)				
			w24[9-](849.9	9)				

Color Code for Ion Intensity >1.6e+05 >8.6e+04 >4.5e+04 >2.4e+04 >1.3e+0

Figure 9. (a) Shaded chromatographic representation showing identification of three oligonucleotides in a mix; (b-d) Fragment coverage maps of 15mer 7- (b), 20merAC 8- (c), and 25merMod 10- (d), at a stepped NCE of 15-17-19

(d)

1 2 3

## thermo scientific

Fragmentation of isomeric 20merAC and 20merCA

BioPharma Finder software also offers tools to facilitate the characterization of isomeric oligonucleotides. Figure 10 displays the zoomed MS<sup>2</sup> spectra of the 6- of isomeric 20merAC and 20merCA as viewed in the stacked format in BioPharma Finder software. A mass difference of 8 Da observed for the  $w_7^{3-}$  peak aided unambiguous identification of these two isomers. The mass differences seen for the  $w_7$  fragments (2- to 4-) from two isomers were also displayed in their respective fragment coverage maps (data not shown).



Figure 10. Zoomed MS<sup>2</sup> spectra of the 6- of (a) 20merAC and (b) 20merCA at NCE15-17-19 showing the diagnostic  $w_7^{3-}$  peak. These two spectra were viewed in the stacked format in BioPharma Finder software.

## Conclusion

This application note describes a robust HRAM based ddMS<sup>2</sup> approach for accurate mass measurement and high confidence identification and mapping of oligonucleotides. The HRAM capability offered by the Orbitrap provides baseline resolving power and high mass accuracy for analyzing oligonucleotides. The ability of BioPharma Finder software to automatically interpret the ddMS<sup>2</sup> data of oligonucleotides opens up a new avenue for in-depth analysis of this novel modality. The powerful tools available in BioPharma Finder 4.0 software for comparative data analysis enables easy and fast method optimization, for which this application note can be used as a guide.

## References

- 1. Wang, F., *et al.* RNA therapeutics on the rise. *Nature Reviews Drug Discovery*, **2020**, *19*, 441.
- 2. Bajan, S., *et al.* RNA-based therapeutics: from antisense oligonucleotide to miRNAs. *Cells*, **2020**, *9*, 137.
- Yin, W., et al. Targeting RNA: a transformative therapeutics strategy. Clinical and Translational Science, 2019, 12, 98.
- Rossi, JJ, et al. Oligonucleotides and the COVID-19 pandemic: a perspective. Nucleic Acid Therapeutics, 2020, 30, 129.
- Pourshahian, S., *et al.* Therapeutic oligonucleotides, impurities, degradants, and their characterization by mass spectrometry. *Mass Spectrometry Reviews*, 2019. *doi: 10.1002/mas.21615.*
- Schurch, S., et al. Characterization of nucleic acids by tandem mass spectrometry the second decade (2004-2013): from DNA to RNA and modified sequences. *Mass Spectrometry Reviews*, 2016, *35*, 483.

### Acknowledgments

We would like to acknowledge all the colleagues who have contributed to the work and reviewed this application note.

## Find out more at thermofisher.com/biopharma

For Research Use Only. © 2020 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific Inc. products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details. AN73789-EN 0920S

