Accurate Base Composition of Double-Strand DNA by Mass Spectrometry

David J. Aaserud, Neil L. Kelleher, Daniel P. Little, and Fred W. McLafferty

Baker Chemistry Laboratory, Cornell University, Ithaca, New York, USA

An accurate molecular weight (M_r) assignment for a double-strand (ds) DNA determines or greatly restricts the possible number of each of its four bases, while the compositions for its two single-strand (ss) components can also be derived from their M_r values. For a ds 64-mer (39 kDa), the ss-M_r values (± 0.5 Da) of its high-resolution mass spectrum from an electrospray ionization/Fourier transform instrument yield only the correct ds- and ss-base compositions. Literature mass spectra of lower mass accuracy show that such data can also restrict their possible composition assignments, with further discrimination using the abundance vs. base composition of small fragment ions from the dissociation of the ss molecular ions.
 (J Am Soc Mass Spectrom 1996, 7, 1266–1269)

Accurate, fast characterization of nucleotides is increasingly important in many research fields, and mass spectrometry (MS) has recently shown high promise for this [1–10]. McCloskey et al. [2] have pointed out that an accurate value of molecular weight (M_r) for an unknown single-strand (ss) oligonucleotide greatly restricts its possible base compositions, the number of each of its four constituent bases. However, the required mass accuracy increases rapidly with increasing M_r, with ± 0.5 Da accuracy giving 20 possible DNA compositions for M_r = 7584 Da [7]. Routine measurements [11] with combined electrospray ionization [12] (ESI)/Fourier transform (FT) MS [13] have achieved mass errors of < 0.5 Da for up to 43 kDa M_r values [7, 9–11, 14–16] (e.g., 0.2 Da for a 15 kDa DNA [9] and 25 kDa tRNA [10], 0.3 Da for a 31 kDa 100-mer DNA) [10]. Recently ESI/MS has been applied to double-strand (ds) DNA from polymerase chain reactions [5, 6], with complementary ss 49-mers measured with 0.7 and 0.8 Da errors [5]. We show here that such accuracy will define or greatly restrict the base pair composition of dsDNA as large as 370 kDa (600 base pairs). Furthermore, such accuracy for a 39 kDa dsDNA [17] will define also the base compositions of each of its Watson–Crick complementary ssDNAs.

The negative ion ESI/FTMS spectrum of this dsDNA of two 64-mer ss monomers [17] (Figure 1) yields ${}_{ss}M_r$ values [9, 10] in error by < 0.1 Da, as is their sum, the ${}_{ds}M_r$ value, 39,423.6-*18* [20]. The accuracy of measuring the ds molecular ion is only ± 1 Da because of the greater uncertainty in fitting the isotopic abundances to the predicted distribution [21]. The monoisotopic M_r value (M_{r0}) [20] of a ssDNA ($A_aT_tC_cG_g$) is the sum of all its base unit masses (base + sugar + phosphate: **A** = 313.0576, **T** = 304.0460, **C** = 289.0463, and **G** = 329.0525) [7, 9] plus H (1.008 Da) added at the 5′-terminus, minus PO_2 (62.964 Da) missing at the 3′-terminus, or a**A** + t**T** + c**C** + g**G** − 61.956. For dsDNA (a = t, c = g), ${}_{ds}M_{r0}$ = a(**A** + **T**) + c(**C** + **G**) − 123.912; combining these values gives an expression for the number of base pairs, a + c, in dsDNA (eq 1).

$$a + c = ({}_{ds}M_{r0} - 0.9952c + 123.91)/617.1036 \quad (1)$$

Because the base pair mass values are so similar, with (**C** + **G**) = (**A** + **T**) + 0.9952 Da, ${}_{ds}M_{r0}$ depends mainly on the number of pairs and only marginally (0.9952c) on their identity. For the dsDNAs $(AT)_{600}$ and $(CG)_{600}$, M_{r0} = 370,138 and 370,735, respectively; any ${}_{ds}M_{r0}$ value between these also must represent a 600 base pair DNA, as $(CG)_{599}$ = 370,117 and $(AT)_{601}$ = 370,755.

Furthermore, a ${}_{ds}M_{r0}$ value of ± 1 Da accuracy defines the pairs composition to one c value, e.g., the M_{r0} values of $(AT)_{599}(CG)_1$ = 370,139, $(AT)_{598}(CG)_2$ = 370,140, etc. Using the Figure 1 measured value of ${}_{ds}M_{r0}$ = 39,405.59 (from 39,423.65 − *18*-18 × 1.0034) [20] and eq 1, a + c = 64.0565; of these 64 base pairs, 35 (0.0565 × 617.1/0.9952 = 35.04) must be CG pairs. A measurement error, $\Delta_{ds}M_{r0}$, of +(−)0.9952 Da can only lead to a +(−)1 error in the calculated c value, but (if a + c = < 600) will give no base pair error, Δ(a + c) = 0. Thus $\Delta c = -\Delta a = \Delta_{ds}M_r/0.9952$ (note that $\Delta M_{r0} = \Delta M_r$).

Address correspondence to Dr. Fred W. McLafferty, Baker Chemical Laboratory, Cornell University, Department of Chemistry, Ithaca, NY 14853-1301.

1044-0305/96/$15.00
PII S1044-0305(96)00194-8

Revised September 3, 1996
Accepted September 3, 1996

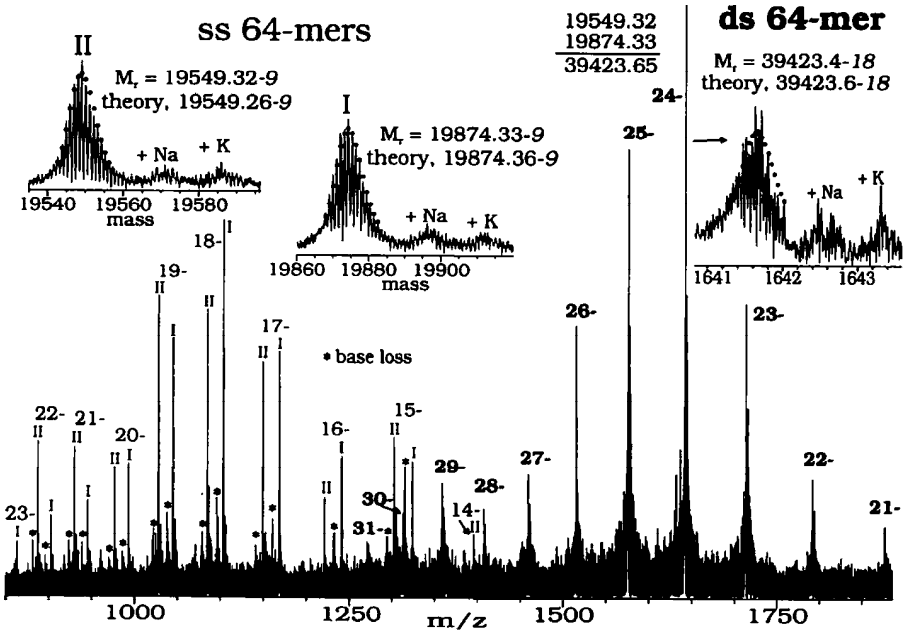

Figure 1. ESI/FTMS spectrum (10 scans) of ds 64-mer, ds charge states in bold. *: neutral adenine loss. Insets: isotopic peaks (ss charge states combined by deconvolution), with solid circles as the best fit of the predicted isotopic abundances [11, 21].

In contrast, the complementary bases (here designated as I and II) have large mass differences: A − T = 9.0116 Da and G − C = 40.0062 Da. Switching a G_IC_{II} pair to $G_{II}C_I(\Delta g_I = -1 = -\Delta c_I)$ causes complementary changes in the ss M_r values of $\Delta_I M_r = -40$ (larger ssDNA) and $\Delta_{II} M_r = +40$. Thus, eq 2,

$$40.01c_I - 9.012a_I = 304.05a + 329.05c - {}_IM_{r0} - 61.96 \quad (2)$$

derived using these restrictions and ${}_IM_{r0} = a_I\mathbf{A} + t_I\mathbf{T} + c_I\mathbf{C} + g_I\mathbf{G} - 61.96$, defines the possible ssI compositions; these can be calculated by exhaustive trials. Alternatively, after finding one consistent composition, the others possible within a defined M_r error ($\Delta_I M_r$, $\Delta_{II} M_r$) can be derived by modifying eq 2 to use differential values to define Δm_I, the mass difference between the original and the new composition of ssI (eq 3; note that $\Delta c =$

$$\Delta m_I = 25.01\Delta a + \Delta_I M_r - 9.012\Delta a_I + 40.01\Delta c_I \quad (3)$$

$-\Delta a$). The compositions possible from eq 3 for $\Delta_{I,II} M_r$ and $\Delta m_I \leq \sim 1$ Da are listed in Table 1.

Alternative Compositions for Δa = 0

If all Δm_r values are correct, the minimum Δm_I value is found for $\Delta a_I = +40(-40)$ and $\Delta c_I = +9(-9)$: $\Delta m_I = -(+)360.48 + +(-)360.09 = -0.39$ Da (+0.39 Da); these compositions need not be considered if a = < 40. Also, $\Delta a_I = +9(-9)$ and $\Delta c_I = +2(-2)$ yield $\Delta m_I = -(+)81.11 + +(-)80.02 = -1.09(+1.09)$ Da (Table 1). For $\Delta_{ds} M_r = 0$, but with compensating errors such as $\Delta_I M_r = +1(-1)$ and $\Delta_{II} M_r = -1(+1)$, unchanged composition values ($\Delta a_I = 0$, etc.) in eq 3 yield corresponding inequalities, $\Delta m_I = +1(-1)$. In the same way, these $\Delta_{I,II} M_r$ errors for $\Delta a_I = +9(-9)$ and $\Delta c_I = +2(-2)$ yield $\Delta m_I = -0.1(+0.1)$. Thus, summing all the values of rows 1 and 2 (Table 1) yields those of row 3. Many of these obvious combinations are omitted from Table 1, as are $\Delta a_I = +31(-31)$, $\Delta c_I = +7(-7)$ and other Δm_I compositions possible for $\Delta a_I = > 21$.

Table 1. Possible ssI composition changes vs. measurement errors for $\Delta a_I < 22$[a]

$\Delta_I M_r$	$\Delta_{II} M_r$	Δ_{a_I}	Δt_I	Δc_I	Δg_I	Δm_I[b]
0	0	+9(−9)	−9(+9)	+2(−2)	−2(+2)	−1.1(+1.1)
+1(−1)	−1(+1)	0	0	0	0	+1.0(−1.0)
+1(−1)	−1(+1)	+9(−9)	−9(+9)	+2(−2)	−2(+2)	−0.1(+0.1)
0	−1(+1)	−6(+6)	+7(−7)	−2(+2)	+1(−1)	−0.9(+0.9)
+1(−1)	0	−7(+7)	+6(−6)	−1(+1)	+2(−2)	−0.9(+0.9)
−0.5(+0.5)	−0.5(+0.5)	−15(−16)	+16(+15)	−4(−3)	+3(+4)	−0.4(−0.4)
−1(+1)	−1(+1)[c]	+1(−1)	+1(−1)	−1(+1)	−1(+1)	0.0(0.0)
−1(+1)	−1(+1)[c]	+10(−10)	−8(+8)	+1(−1)	−3(+3)	−1.1(+1.1)

[a] Other possible compositions can be found by summing the values in two rows. [b] $\Delta m_I = \pm < 1.1$ Da; Δm_{II} value the same but of opposite sign. [c] ${}_{ds}M_r$ error 2 Da.

$\Delta a = \pm 1$

The error combination $\Delta_I M_r = 0$, $\Delta_{II} M_r = -1(+1)$ yields the alternatives $\Delta a_I = -6(+6)$ and $\Delta c_I = -2(+2)$, $\Delta m_I = -0.9(+0.9)$. However, errors of $\Delta_I M_r = -0.5(+0.5)$, $\Delta_{II} M_r = -0.5(+0.5)$ give $\Delta a_I = -15(-16)$, $\Delta c_I = -4(-3)$, with $\Delta m_I = -0.4(-0.4)$. This composition change is also possible for $\Delta_I M_r = -1(+1)$, $\Delta_{II} M_r = 0(0)$, $\Delta m_I = -0.8(+0.2)$, and $\Delta_I M_r = 0(0)$, $\Delta_{II} M_r = -1(+1)$, $\Delta M_I = +0.2(-0.8)$ that can be derived from summing appropriate rows of Table 1.

$\Delta a = \pm 2$

If both $\Delta_I M_r$ and $\Delta_{II} M_r = -1(+1)$ (not possible if $\Delta_{ds} M_r = < 2$ Da), as would result from a $-1(+1)$ Da calibration error, changes of $\Delta a_I = +1(-1)$ and $\Delta c_I = -1(+1)$ in eq 3 yield $\Delta m_I = 0.0$ Da. Summing rows as above, an additional composition change of $\Delta a_I = +9(-9)$, $\Delta c_I = +2(-2)$ yields the compositions $\Delta a_I = -8(+8)$, $\Delta c_I = -3(+3)$ with $\Delta m_I = +1.1(-1.1)$, and those of Table 1, row 8.

39 kDa dsDNA

If the Figure 1 data were from an unknown, the only possible ssI composition (Table 2) for ≤ 0.5 Da errors in the M_r values and Δm_I is $A_{17}T_{12}C_{14}G_{21}$ (correct; ssII $A_{12}T_{17}C_{21}G_{14}$). Increasing $\Delta_{ds} M_r$ to 1 Da makes $A_2T_{28}C_{10}G_{24}$ and $A_1T_{27}C_{11}G_{25}$ possible, while this larger error for $\Delta_{I,II} M_r$ also makes the compositions of $\Delta a = \pm \sim 9$ possible with $\Delta m < 0.2$ Da. If all errors, including Δm_I, are ~ 1 Da, the compositions of $\Delta a_I = \pm 6$–10 are possible, such as $A_8T_{21}C_{12}G_{23}$ and $A_{23}T_5C_{16}G_{20}$. Ancillary data could distinguish such substantially different compositions. For example, in FT MS/MS spectra the mass accuracies for internal fragment ions containing one and two bases are sufficient to assign base compositions unequivocally [9]. Averaging these fragment ion compositions for the MS/MS spectrum of ssI (ssII) gave $A_{12}T_{14}C_{15}G_{23}$ ($A_8T_{21}C_{22}G_{13}$); because base A is much more (see Figure 1) easily lost in fragmentation, and T much less [1, 7, 9], these compositions support the correct assignments $A_{17}T_{12}C_{14}G_{21}$ ($A_{12}T_{17}C_{21}G_{14}$). Compositions with single base variations, such as $A_{18}T_{13}C_{13}G_{20}$, are only possible for same sign 1 Da errors in ${}_I M_r$ and ${}_{II} M_r$ that give a 2 Da error in ${}_{ds} M_r$.

30 kDa dsDNA [5]

ESI/MS of this PCR-prepared sample (45 base pairs plus 5′-A_2T_2 on each ss) gave no ds molecular ions, but the ss ions (isotopically unresolved) yielded M_r values of 15275.2 and 15029.0 (theory 15276.0 and 15029.7). If the ${}_{ss} M_r$ and Δm errors had been instead ≤ 0.5 Da, the only possible composition from eq 2 (accounting for 5′-phosphorylation) is the correct one, $A_{23}T_9C_7G_{10}$ using 15275.7 and 15029.5. For ≤ 1 Da assumed errors, $A_{24}T_{10}C_6G_9$ (actual errors $\Delta_I M_r = -1.3$, $\Delta_{II} M_r = -1.2$), $A_8T_{25}C_3G_{13}$, $A_{15}T_{19}C_4G_{11}$, and $A_{32}T_0C_9G_8$, plus those very similar (Table 1, row 7) to the last three, are possible.

Table 2. Possible ssI compositions from measured M_r values[a]

$\Delta_I M_r$, $\Delta_{II} M_r$	a_I	t_I	c_I	g_I	Δm_I	Δm_{II}
19874.33-*9*, 19549.32-*9*; actual errors −0.03, +0.06						
0.0, 0.0	*17*	*12*	*14*	*21*[b]	−0.03	+0.06
+1.0, −1.0	26	3	16	19	−0.12	+0.16
−1.0, +1.0	8	21	12	23	+0.07	−0.03
+1.0, 0.0	10	18	13	23	−0.96	+1.00
0.0, +1.0	23	5	16	20	+0.91	−0.86
−0.5, −0.5	2	28	10	24	−0.37	+0.40
−1.0, −1.0[c]	18	13	13	20	−0.03	+0.06
15275.2, 15039.0;[d] actual errors −0.8, −0.7						
0.0, 0.0	8	25	3	13	−0.5	+0.2
+0.5, +0.5	*23*	*9*	*7*	*10*	−0.2	−0.1
−0.5, −0.5	24	10	6	9	−0.2	−0.1
+1.0, 0.0	32	0	9	8	−0.8	+0.5
−1.0, 0.0	15	19	4	11	+0.4	−0.7
−1.0, −1.0[c]	9	26	2	12	−0.5	+0.2
6410, 5827; [de] actual errors −0.2, +0.2						
0.0, 0.0	*7*	*0*	*0*	*13*	0	0
0.0, 0.0	6	1	0	13	+9	−9
+0.5, +0.5	4	2	0	14	+2	−2
+1.0, +1.0[c]	2	3	0	15	−4	+4

[a] Other possibilities of very similar composition can be derived from Table 1, such as the effect of reversing the +, − signs of the first two columns. [b] Correct compositions in italics. [c] ${}_{ds} M_r$ error 2 Da. [d] Isotopically averaged M_r values (M_{ra}); $M_{r0} = 0.99952 M_{ra}$ for nucleotides [22]. [e] Possible compositions for $\Delta m = < 15$.

12 kDa dsDNA

In a pioneering example [3], ESI/MS of a 20 base pair dsDNA of M_r = 12237 (measured 12297) gave $_{ss}M_r$ values of 6410 ± 2 and 5827 ± 2 (theory 6410.2, 5826.8). Here, the smaller size and absence of two bases greatly limits the possible compositions; the correct one (ssI, A_7G_{13}) is favored (Table 2).

Conclusions

For the ≤ 0.5 Da errors of M_r achievable for dsDNA as large as 39 kDa (Figure 1), the possible mass values restrict the ssI and ssII compositions to only the correct assignments (Table 1). For 1 Da errors in $_{I,II}M_r$, but $\Delta_{ds}M_r = 2$, one alternative composition differs by only one in each base value, and Δa and Δt are ≥ 6 for all others; the elimination of such variant possibilities by MS/MS is under further investigation. ESI/FTMS can provide accurate M_r values even of proteins in complex mixtures [14, 15]; such characterization of unseparated restriction enzyme products from larger DNA should be of special importance with the increasingly widespread use of PCR for trace DNA amplification.

Acknowledgments

The authors thank Z. Guan, G. A. Valaskovic, and E. K. Fridriksson for valuable advice and assistance, T. W. Thannhauser for DNA synthesis and purification, and the National Institutes of Health (Grant No. GM16609; NLK, Cell & Molecular Biology training Grant 08-T2GM07273) for financial support.

References

1. McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085–12095.
2. Pomerantz, S. C.; Kowalak, J. A.; McCloskey, J. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 204–209.
3. Light-Wahl, K. J.; Springer, D. L.; Winger, B. E.; Edmonds, C. G.; Camp, D. G., II; Thrall, B. D.; Smith, R. D. *J. Am. Chem. Soc.* **1993**, *115*, 803–804.
4. Ganem, B.; Li, Y.-T.; Henion, J. D. *Tetrahedron Lett.* **1993**, *34*, 1445–1448. Potier, N.; Van Dorsselaer, A.; Cordier, Y.; Roch, O.; Bischoff, R. *Nucleic Acids Res.* **1994**, *22*, 3895–3903. Gale, D. C.; Goodlett, D. R.; Light-Wahl, K. J.; Smith, R. D. *J. Am. Chem. Soc.* **1994**, *116*, 6027–6028. Doktycz, M. J.; Habibi-Goudarzi, S.; McLuckey, S. A. *Anal. Chem.* **1994**, *66*, 3416–3422. Ding, J.; Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 159–164. Doktycz, M. J.; Hurst, G. B.; Habibi-Goudarzi, S.; McLuckey, S. A.; Tang, K.; Chen, C. H.; Uzeil, M.; Jacobson, K. B.; Woychik, R. P.; Buchanan, M. V. *Anal. Biochem.* **1995**, *230*, 205–214.
5. Naito, Y.; Ishikawa, K.; Koga, Y.; Tsuneyoshi, T.; Terunuma, H.; Arakawa, R. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1484–1486.
6. Wunschel, D. S.; Fox, K. F.; Fox, A.; Bruce, J. E.; Muddiman, D. C.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 29–35.
7. Little, D. P.; Chorush, R. A.; Speir, J. P.; Senko, M. W.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1994**, *116*, 4893–4897.
8. Chen, R.; Cheng, X.; Mitchell, D. W.; Hofstadler, S. A.; Wu, Q.; Rockwood, A. L.; Sherman, M. G.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 1159–1163.
9. Little, D. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **1995**, *117*, 6783–6784. Little, D. P.; Aaserud, D. J.; Valaskovic, G. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **1996**, *118*, 9352–9359.
10. Little, D. P.; Thannhauser, T. W.; McLafferty, F. W. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2318–2322.
11. Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M. III; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 557–565. McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379–386.
12. Clegg, G. A.; Dole, M. *Biopolymers* **1971**, *10*, 821–826. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.
13. Marshall, A. G.; Comisarow, M. B. *Chem. Phys. Lett.* **1974**, *25*, 282–283. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991**, *63*, 215A–229A.
14. Wood, T. D.; Chen, L. H.; Kelleher, N. L.; Little, D. P.; Kenyon, G. L.; McLafferty, F. W. *Biochemistry* **1995**, *34*, 16251–16254.
15. Kelleher, N. L.; Costello, C. A.; Begley, T. P.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 981–984.
16. Speir, J. P.; Senko, M. W.; Little, D. P.; Loo, J. A.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 39–42.
17. The 64-mer CTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAG (ssI) and its complement synthesized by the Cornell Peptide/DNA Synthesis Facility were annealed in aqueous 5 mM piperidine solution at 95 °C for 5 min, slowly cooled to 25 °C, and diluted to 5 μM in 3:1 CH_3CN:H_2O (2.5 mM in piperidine) [18]. This solution was introduced at 1 μL/min into a previously described 6 T ESI/FTMS system [11], with noncovalent ion adducts minimized by 20 ms of infrared photodissociation (25 W Synrad CO_2 laser, 35% duty cycle) [19].
18. Greig, M.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 97–102.
19. Little, D. P.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 209–210.
20. Here, M_r is the mass of the most abundant isotopic peak, which differs by < 20 ppm from the M_r value based on natural isotopic abundances. Variations in the latter affect only the abundance, not the mass, of an isotopic peak [11]. The number of ^{13}C atoms, n, in the most abundant isotopic peak is given in italics following the mass value; $M_{r0} = M_r - n(1.0034)$.
21. Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
22. Zubarev, R. A.; Bondarenko, P. V. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 276–277.