Applicability of Tandem Mass Spectrometry to the Automated Comparative Sequencing of Long-Chain Oligonucleotides

Herbert Oberacher and Walther Parson
Institute of Legal Medicine, Leopold Franzens University, Innsbruck, Austria

Peter J. Oefner
Stanford Genome Technology Center, Palo Alto, California, USA

Bettina M. Mayr and Christian G. Huber
Instrumental Analysis and Bioanalysis, Saarland University, Saarbruücken, Germany

An algorithm for the comparative sequencing (COMPAS) of oligonucleotides is shown to be suitable for the sequence verification of nucleic acids ranging in length from a few to 80 nucleotides. The algorithm is based on the matching of a fragment ion spectrum generated by collision-induced dissociation to m/z values predicted from a known reference sequence employing established fragmentation pathways. Prior to mass spectrometric investigation, the oligonucleotides were on-line purified by ion-pair reversed-phase high-performance liquid chromatography using monolithic separation columns. This study evaluated the potential and the limits of COMPAS regarding the length and the charge state of oligonucleotides, the selected collision energy, and the analyzed amount of sample using a quadrupole ion trap mass spectrometer. The results revealed that oligonucleotides could be very reliably re-sequenced up to 60-mers, although the algorithm was successfully used to even verify sequences up to 80-mers. The relative collision energy was typically in the range between 13 and 20%, which allowed in most cases a verification of the reference sequence in a window of at least three consecutive collision energies. To select a proper charge state for fragmentation, a compromise had to be found between high signal intensity and low charge state. Furthermore, by reducing the eluent flow rate during elution of the oligonucleotide, the sequence of a 50-mer was successfully verified from the analysis of 295 fmol of the raw product. COMPAS was proven to be reproducible and was applied to the genotyping of the polymorphic, Y-chromosomal locus M9 contained in a 62-base pair polymerase chain reaction product. (J Am Soc Mass Spectrom 2004, 15, 510–522) © 2004 American Society for Mass Spectrometry

The generation of sequence data represents an essential part of nucleic acids research and applications. The quality and integrity of synthetic oligonucleotides has to be controlled carefully before utilization in therapeutic and diagnostic protocols [1]. Equally important is the discovery and characterization of sequence variation in nucleic acids of biological origin. Valuable insights in inherited disposition to disease as well as in human origin and gene migration are facilitated by information about genetic diversity [2, 3]. In due consequence, the discovery of small dissimilarities in DNA sequences of different individuals, so-called single nucleotide polymorphisms (SNP), is rapidly gaining significance [4]. Since polymorphisms between two randomly chosen chromosomes occur only at a frequency of one in 800–62,000 base pairs [5], their discovery by conventional Sanger sequencing would dissipate a lot of time and experimental effort for the determination of already known sequences. Hence, a considerable number of different methodologies for screening for polymorphisms and for determining individual allelic states have been developed, which have been reviewed recently [6, 7]. Matrix-assisted laser desorption/ionization mass spectrometry MALDI-MS [8, 9] and electrospray ionization mass spectrometry (ESI-MS) [10–12] have been successfully applied to the investigation of sequence variations in nucleic acids.

In this context, we have introduced ion-pair reversed-phase high-performance liquid chromatography on-line interfaced to electrospray ionization mass spectrometry (ICEMS) as an analytical tool for the charac-

terization of DNA sequences chain reaction (PCR) [13–18]. ICEMS enables the detection of SNPs based on the determination of the intact molecular mass using low femtomol amounts of DNA fragments up to lengths of more than 100 base pairs. However, ICEMS suffers from the limitation that knowledge about the molecular mass is not sufficient to unequivocally determine the position of a variation within a given DNA sequence. In this regard, tandem mass spectrometry (MS^2), in which sequence information is gained from fragmentation by collisions in the gas phase, has high potential for evaluating DNA sequences for unknown variations relative to a given genomic sequence.

The groundwork for sequencing approaches by MS^2 has been laid by several studies investigating the fragmentation reactions of oligonucleotides following on the gas-phase collision-induced activation (reviewed in [1, 19–22]). Nevertheless, manual spectrum interpretation is challenging and time consuming, especially with fragment ion spectra of long oligonucleotides showing a plethora of signals. Hence, McCloskey and coworkers have developed the Simple Oligonucleotide Sequencer (SOS) algorithm for the automated determination of unknown oligonucleotide sequences by electrospray ionization tandem mass spectrometry (ESI-MS^2) [23, 24]. The algorithm works by extending from both the 5' (a-B ions) and 3' (w ions) ends ion series encoding the complete DNA sequence. Mass ladders are identified by sequentially adding each of the four possible nucleotide masses and searching the spectrum for the best match of expected ions. Because the algorithm relies on the presence of fragment ions for each position in the oligonucleotide, SOS is restricted to oligomers at approximately the 20-mer level and below [24].

We have recently introduced a comparative sequencing algorithm (COMPAS) for the computer-aided interpretation of fragment ion spectra generated by ESI-MS^2 [25]. By combining ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization tandem mass spectrometry (ICEMS^2) with the COMPAS algorithm, the sequences of synthetic oligonucleotides ranging in length up to 39-mers were successfully verified. Moreover, the alleles of two SNPs were unequivocally genotyped in a 73 bp fragment containing the human STS G-101769 site [16]. Extending these recent communications, we report here on the use of COMPAS for the sequence verification of nucleic acids with lengths up to 80 nucleotides (nt). Moreover, we demonstrate that experimental parameters like precursor ion charge state, sample concentration, and collision energy have a significant impact on sequencing results and thus have to be selected carefully. As a result of these optimization efforts, COMPAS was proven to be highly reproducible and was applied to the sequence verification of a 62 base pair (bp) product of polymerase chain reaction (PCR).

Experimental

Chemicals and Oligonucleotides

Acetonitrile (HPLC gradient-grade) was obtained from Merck (Darmstadt, Germany). A stock solution (0.50 M) of butyldimethylammonium bicarbonate was prepared by passing carbon dioxide gas (AGA, Vienna, Austria) through a 0.50 M aqueous solution of butyldimethylamine (Fluka, Buchs, Switzerland) at 5 °C until pH 8.4 was reached. For preparation of all solutions, HPLC-grade water (Merck) was used. The synthetic oligonucleotides listed in Table 1 were ordered from Microsynth (Balgach, Switzerland) or from GIBCOBRL (Rockville, MD) and used without further purification. The 62 bp PCR product of the Y-chromosomal locus M9 was amplified according to a recently published protocol [17].

Capillary HPLC Coupled to ESI-MS^2

The HPLC system consisted of a low-pressure gradient micro pump (model Rheos 2000, Flux Instruments, Basel, Switzerland) controlled by a personal computer, a vacuum degasser (Knauer, Berlin, Germany), a column thermostat made from 3.3 mm o.d. copper tubing which was heated by means of a circulating water bath (model K 20 KP, Lauda, Lauda-Königshofen, Germany), and a microinjector (model C4-1004, Valco Instruments Co. Inc., Houston, TX) with 500 nL internal sample loop. ESI-MS^2 was performed on a Finnigan LCQ quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with an electrospray ion source. The 60 × 0.2 mm i.d. monolithic capillary column was prepared according to the published protocol [13] and connected directly to the spray capillary (fused silica, 90 μm o.d., 20 μm i.d., Polymicro Technologies, Phoenix, AZ) by means of a microtight union (Upchurch Scientific, Oak Harbor, WA). A syringe pump equipped with a 250 μL glass syringe (Unimetrics, Shorewood, IL) was used for the post-column addition of a 3.0 μL/min flow of acetonitrile by means of a tee-piece. For analysis with pneumatically assisted ESI, an electrospray voltage of 5.0 kV and a nitrogen sheath gas flow of 100 arbitrary units were employed. The temperature of the heated capillary was set to 250 °C. Total ion chromatograms and mass spectra were recorded on a personal computer with the Xcalibur software version 1.1 (Thermo Finnigan).

Mass calibration and coarse tuning were performed in the positive ion mode by direct infusion of a solution of caffeine (Sigma, St. Louis, MO), methionyl-arginyl-phenylalanyl-alanine (Finnigan), and Ultramark 1621 (Finnigan). Fine tuning for ESI-MS of oligonucleotides in the negative ion mode was performed according to reference [26]. For MS^2 experiments, the isolation width and the relative collision energy were set to 4.0 mass units and 13–20%, respectively. 0–100% relative collision energy correspond to a high frequency alternating

Table 1. Sequences, molecular masses M_r, and analyzed amounts of the oligonucleotides (1–9) and the PCR amplicon (M9) analyzed in this study

No.	Sequence	Length [nt]	M_r [Da]	Amount [pmol]
1	CAGTGATGCG AGCGTTCTTA TCGATGAATT CGAGCTCG	38	11724.6	85.3
2	CGAGCTCGAA TTCATCGATA AGAACGCTCG CATCACTG	38	11631.6	86.0
3	CGGTGTCGGT CTCGTAGCAG AAGGCTCAAT TTCCGTAAT	39	12013.8	83.2
4	CGATTACGGA AATTGAGCCT TCTGCTACGA GACCGACACC G	41	12579.2	79.5
5	CTTCTGTGGG TAATTTGTTT GTTCCTGTCA ATCAGCTATC AGGAGGACTG	50	15419.0	64.9
6	AAACCACATT CTGAGCATAC CCCCAAAAAA TTTCATGCCG AAGCTGTGGT C	51	15580.2	40.8
7	CCGGTGAACG GCTCGTGCCA AGTTCTGCTC CAGTCTTGTC TTGCCTTTTG CCAGTAACTT	60	18341.9	62.4
8	GACGATCAGC TTCAAGTACA GTCCCGTATG CTGGTGATAG CTGAGTAAGC TAGTAATCCC AAGTTTCAGA	70	21612.1	53.4
9	CCCTTCCGGC TGGCTGGTTT ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGCG GTATCATTGC AGCACTGGGG	80	24788.0	16.5
M9	AACGGCCTAA GATGGTTGAA T(G/C)CTCTTTAT TTTTCTTTAA TTTAGACATG TTCAAACGTT CA	62	38174.9	0.1

voltage for resonance excitation from 0 to 5 V maximum-to-maximum. Helium, present in the ion trap at a pressure of 0.1 Pa, served as collision gas.

Computer-Aided Data Interpretation

All calculations were performed on a personal computer under Windows NT operating system (400 MHz Pentium, 128 MB RAM). Measured MS^2 spectra of oligonucleotides were exported from the Xcalibur Software (Thermo Finnigan) as ASCII files. Automated comparison of measured and predicted spectra was performed with a program written in ActivePerl 5.6.1 (Active State Corporation, Vancouver, BC, Canada). A copy of the program for academic use is available upon request from the corresponding author. All data were collected, evaluated, and prepared for final output in Microsoft Excel for Windows 2000. Correlation diagrams were prepared from the Excel data using Origin 6.0 (OriginLab Corporations, Northampton, MA).

Results and Discussion

Comparative Sequencing Strategy and Algorithm

The starting point of our sequencing strategy relies on the fact that for many applications there is no need for a complete de novo determination of an unknown sequence but merely to confirm identity or to reveal small differences compared to a known sequence. In this context, sequencing by MS^2 represents an attractive alternative to conventional Sanger sequencing, because sequence-specific data can be obtained rapidly upon

collision-induced fragmentation of nucleic acids and subsequent mass analysis of the fragments. Nevertheless, manual interpretation of the fragment spectra remains time consuming and prone to error, and hence, automated spectrum interpretation using suitable algorithms is strongly desirable.

The details of the COMPAS algorithm for the automated interpretation of oligonucleotide MS^2 spectra have been described in detail elsewhere [25]. Briefly, COMPAS involves the comparison of a measured MS^2 spectrum to a set of fragment ion m/z values predicted from a reference sequence. The closeness of matching between the measured spectrum and the predicted set of ions is characterized by a value FS for the fitness. The fitness value is calculated taking into account the difference (Δ) between measured and predicted m/z values, the relative intensity (I%) of the assigned fragment ions, the number of fragments assigned (K), and the number of nucleotide positions not covered by fragment ions in the experimental spectrum (M) according to the following relationship:

$$FS = a \cdot 1/K \cdot \sum \left(100 \cdot \frac{\Delta}{I\%}\right) - b \cdot K + c \cdot M$$

The coefficients a, b, and c are introduced to weigh the relative contributions of the three terms in the formula (a = 1, b = $1.3/K^2 \times \Sigma[100 \times \Delta/I\%]$, c = 0.1, for details, refer to [25]). The smaller the value for fitness, the closer the match between measured spectrum and predicted m/z values. In order to identify sequences that possibly better fit the experimental data, one nucleotide in the

first reference sequence is exchanged with an alternative nucleotide followed by recalculation of the fitness value using fragment ions predicted from that alternative sequence. A smaller fitness value indicates that the alternative sequence fits the experimental data better than the original reference sequence, whereas a larger fitness value identifies the original reference sequence as the better fit. Analogous nucleotide exchanges are then performed sequentially at all positions in the sequence, incorporating all possible nucleotides. The sequence fitting the experimental data best is identified by that sequence having the lowest absolute fitness value.

Comparative Sequencing of an 80-mer Oligonucleotide

The success of MS² sequencing largely depends on the purity of the sample presented to the mass spectrometer, because cation adduction and/or ion suppression by coeluting species significantly reduce the abundance of ions to be selected for fragmentation. Therefore, we have utilized ion-pair reversed-phase high-performance liquid chromatography (IP-RP-HPLC) as the sample preparation method which not only efficiently removes adducted cations and other low molecular mass components coming from solid-phase oligonucleotide synthesis or PCR, but also fractionates mixtures of nucleic acids prior to their mass spectrometric investigation. Figure 1a illustrates the reconstructed ion current chromatogram derived from an ICEMS² experiment investigating an 80-mer sequence (Oligonucleotide 9, Table 1). 16.5 pmol of the synthetic raw product were purified using a gradient of acetonitrile in 25 mM butyldimethylammonium bicarbonate. The MS² spectrum originated from fragmentation of the 20- charge state at 17% relative collision energy (Figure 1b) and was extracted as the average of 25 scans from the peak eluting from the monolithic, 200 μm i.d. poly-(styrene/divinylbenzene) capillary column at 6.5 min.

It is obvious that manual interpretation of the mass spectrum shown in Figure 1b is practically impossible because of the high number and relatively low intensity of observed fragment ions. For automated spectrum interpretation, the mass spectral data, namely m/z values and relative intensities of the fragments, were fed into the COMPAS algorithm and the fitness values using the reference sequence and alternative sequences were calculated. The results of COMPAS are conveniently presented in a correlation diagram in which the fitness values of the reference sequence as well as those of alternative sequences are plotted against the positions at which the nucleotides were exchanged. While the fitness value for the original reference sequence remains at a constant value throughout the diagram, fitness values for alternative sequences fall below or above that of the reference sequence, indicating better or worse fits of the data.

In Figure 1c, the fitness value for the reference sequence remains the lowest throughout the correlation diagram of fitness versus position of nucleotide exchange, while none of the alternative sequences indicated by the position and type of base exchange in the diagram resulted in a better fit. Hence, the correlation diagram for the 80-mer oligonucleotide (Figure 1c) proves that the reference sequence of oligonucleotide 9 (Table 1) and the sequence of the investigated oligonucleotide were identical and that none of the alternative sequences showed a better fit. This example verifies that the fully automatable algorithm can be used to confirm the identity of an oligonuclotide sequence with a reference sequence even for long-chain oligonucleotides yielding highly complex MS² spectra. In order to investigate the general applicability of the comparative sequencing algorithm, we subsequently evaluated the influence of several experimental parameters including precursor ion charge state, collision energy, concentration, and length of the oligonucleotide, respectively, on the results of COMPAS for the sequence verification of 38- to 80-mers.

Charge State Effects on the Sequence Verification of a 60-mer Oligonucleotide

A general result of electrospray ionization of polyionic biopolymers is a bell-shaped distribution of multiply charged ion species. The distribution and relative abundances of the charge states have been shown to depend primarily on the length and base composition of the oligonucleotide [27, 28] as well as on the composition of the electrosprayed solution [29]. It is known from literature that for single- and mixed-nucleobase oligonucleotides the relative abundances of the product ions derived from a given oligomer are highly dependent upon the charge state of the precursor ion [30–32]. It has been also noted that a lower overall charge of the precursor ion increases the observed number of different sequence specific fragments. Although these studies were performed with oligonucleotides having a length equal or smaller than 12 nt, we can assume that the precursor ion charge state has to be selected even more carefully for verifying the sequence of long-chain oligonucleotides because of its significant impact on the fragmentation pattern.

To study the influence of the precursor ion charge state on COMPAS, we utilized a 60-mer of mixed sequence (Oligonucleotide 7, Table 1) for a series of ICEMS² experiments. The 60-mer was separated from its byproducts using a gradient of 2 to 28% acetonitrile in 25 mM butyldimethylammonium bicarbonate in 10 min. In a first step, the raw mass spectrum (Figure 2a) was extracted as the average of 20 scans from the peak eluting at 5.3 min. Deconvolution of this spectrum yielded an intact molecular mass of 18,341, which correlated well with the theoretical mass of 18,341.9 and verified the base composition of the oligonucleotide. In the ensuing experiments different charge states, namely 12-, 15-, 17-, 19-, and 22-, were isolated and fragmented
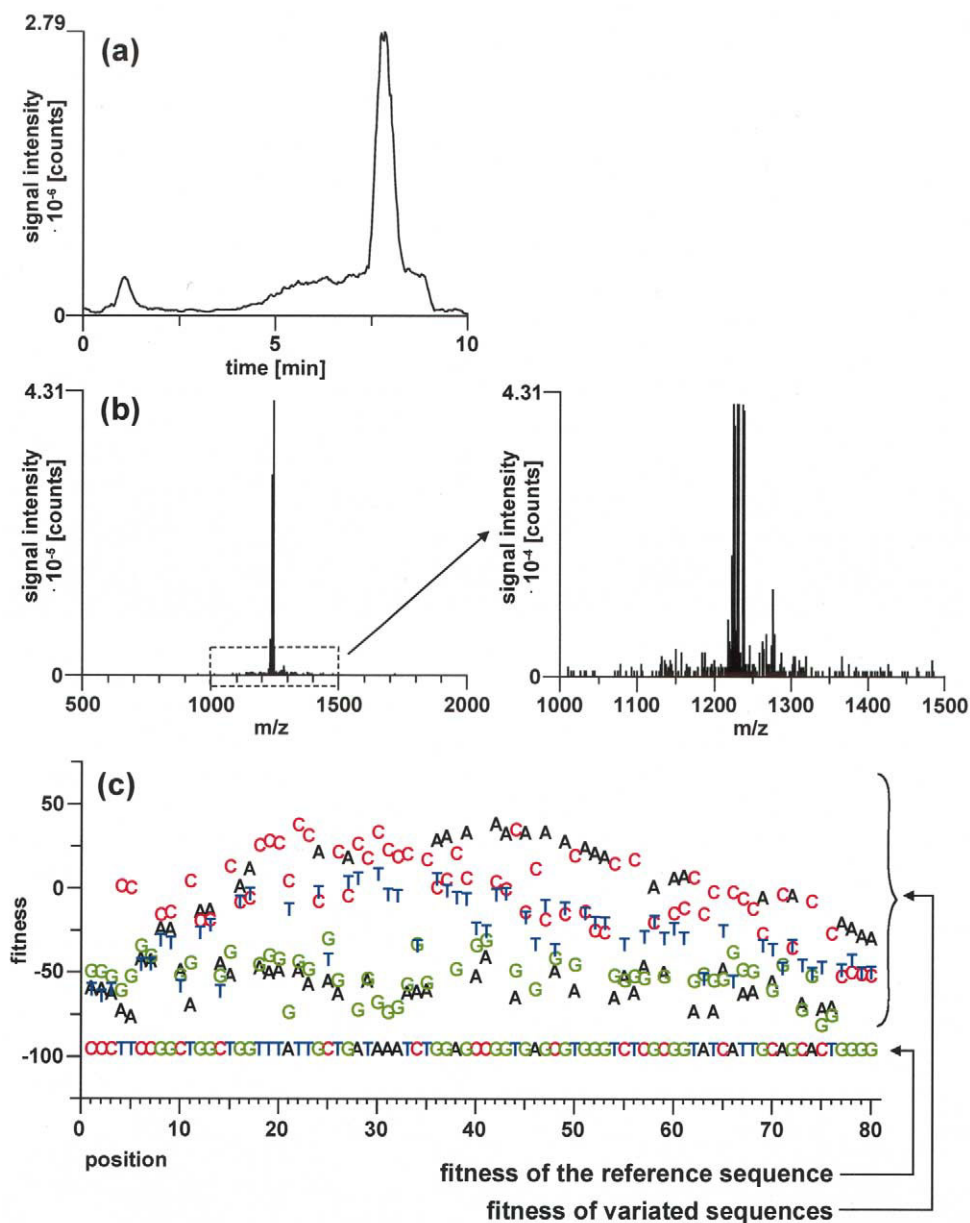
Figure 1. Sequence verification of a synthetic 80-mer. (**a**) Reconstructed total ion current chromatogram, (**b**) extracted MSMS spectrum, and (**c**) correlation diagram of fitness versus position for the 80-mer oligonucleotide. Column, PS-DVB monolith, 60 × 0.20 mm i.d.; mobile phase, (A) 25 mM butyldimethylammonium bicarbonate, pH 8.40, (B) 25 mM butyldimethylammonium bicarbonate, pH 8.40, 40% acetonitrile; linear gradient, 10–50% B in 10.0 min; flow rate, 2.0 μL/min reduced to 500 nl/min after 5.5 min; temperature, 70 °C; product ions of the 20- charged species at m/z 1238.1, 4.0 u isolation width, 16% relative collision energy; scan, 340–2000 u; electrospray voltage, 5.0 kV; sheath gas, 100 units; sheath liquid, 500 nL/min acetonitrile; sample, 16.5 pmol of the 80-mer.

using a relative collision energy of 15%. The corresponding mass spectral data served as input for the COMPAS algorithm, and the resultant correlation diagrams of fitness versus position of the 60-mer are depicted in Figure 2b–f. The sequence was successfully verified with the 15-, 17-, and 19- charge states.

On the contrary, the sequence could not be proven correct by using the data obtained from the isolation and fragmentation of the 12- and 22- charge states. The 12- charge state has the lowest signal intensity of all charge states used for this study (Figure 2a). Therefore, the number of fragment ions and in particular the abundance of sequence specific fragment ions was too small for a correct sequence verification using COMPAS (Figure 2b). In contrast, the signal intensity of the 22- charge state was sufficient to yield enough detectable fragment ions. Nevertheless, due to the high number of negative charges or the low number of protons trans-
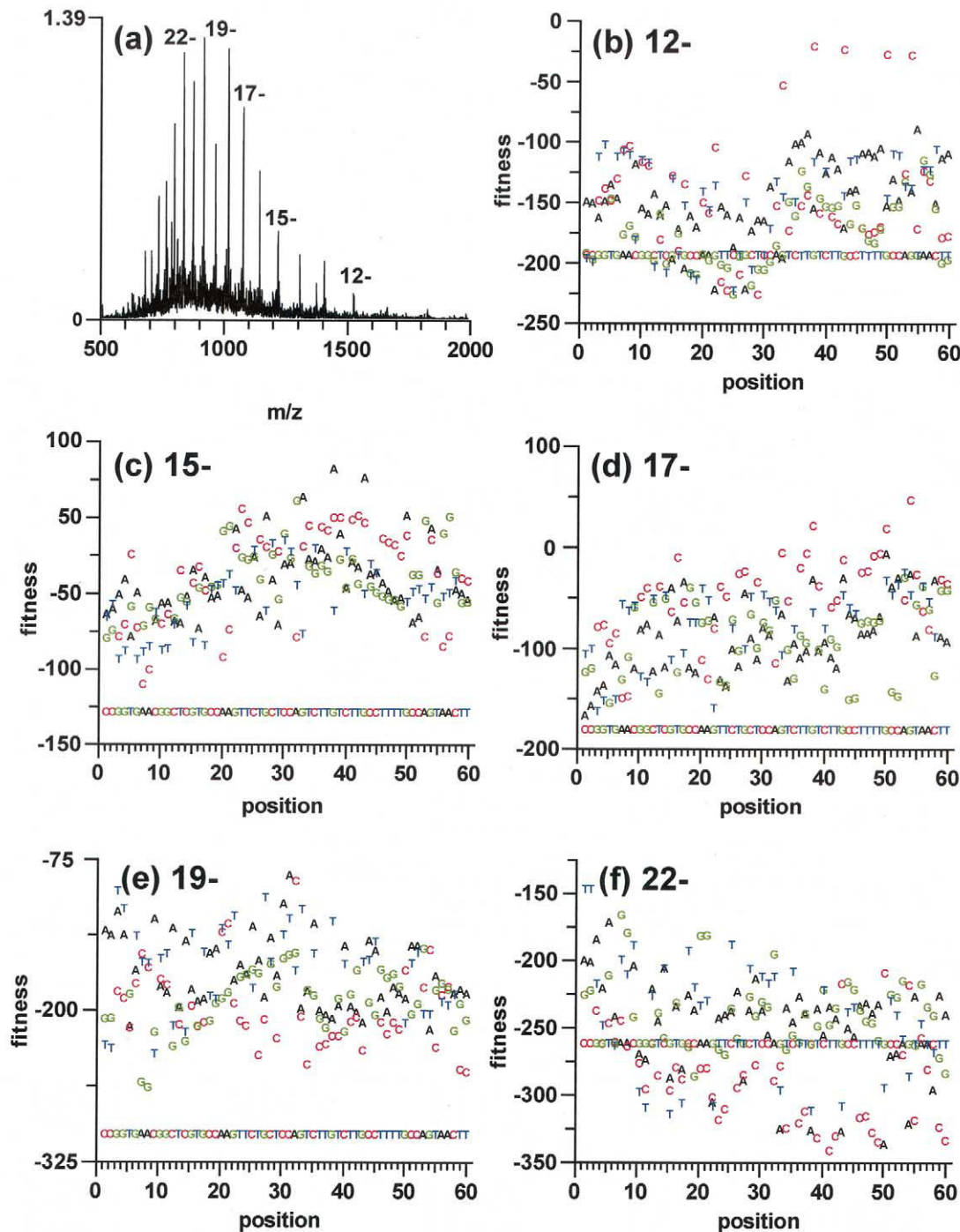
Figure 2. Charge state effects on the sequence verification of a 60-mer oligonucleotide; (**a**) mass spectrum of the 60-mer, (**b**)–(**f**) correlation diagrams obtained by fragmentation of different charge states of the 60-mer. Linear gradient, 10–70% B in 10.0 min; flow rate, 2.0 μL/min; product ions of (**b**) the 12- charged species at m/z 1526.8, (**c**) 15- at m/z 1221.2, (**d**) the 17- at m/z 1077.4, (**e**) the 19- at m/z 963.9, (**f**) the 22- at m/z 832.3, 4.0 u isolation width, 15% relative collision energy; scan, 500–2000 u; sample, 62 pmol of oligonucleotide 7; other conditions as in Figure 1.

ferable from the phosphate backbone to any one of the bases, which could initiate base loss and subsequent fragmentation of the oligonucleotide, only few and dominant fragmentation pathways were observed upon CID. This charge state effect was observed by

McLuckey and Vaidyanathan for small, highly charged oligonucleotides [31] and is most probably responsible for the observed failure to confirm the sequence with the 22- charge state of the 60-mer. In consequence, for successful application of COMPAS the lowest possible

Table 2. Results of 172 ICEMS² - COMPAS experiments with 9 different oligonucleotides.

Length [nt]	Charge state			Relative collision energy [%]							
	observed	most intense	selected	13	14	15	16	17	18	19	20
38	6- - 15-	11-	6-		x	x	✓				
			9-		x	✓	✓	✓	✓	✓	x
			11-			x	✓	✓	✓	✓	
38	6- - 14-	10-	7-		✓	✓	✓	✓	✓	✓	
			10-			x	✓	✓	✓	✓	
39	7- - 15-	13-	7-			x	✓	✓	✓		
			9-			✓	✓	✓	✓	✓	
			12-			x	✓	✓	✓	✓	✓
41	8- - 15-	13-	8-		x	x	x	✓	✓	✓	x
			13-			x	✓	x	✓	✓	x
50	9- - 23-	17-	11-		✓	✓	xx	xx	xx	✓	x
			13-	xx	x	x	x	✓xx	✓✓x	✓✓✓✓✓	x
			15-	xxx	x	x	xx	✓x	✓x	✓	x
51	9- - 17-	13-	9-		✓	✓	✓	✓	✓	✓	✓
			11-		✓	x	xx	✓	✓	✓	
			13-		✓	x	xx	✓	✓	x	
60	10- - 27-	20-	12-	xx	xx	xxx	xxx	✓x	✓	✓x	x
			15-	x	x	✓	✓	xx	✓x	✓	x
			17-	xxx	xx	✓x	xx	xx	x	x	
			19-		x	✓	x	x	x	x	
			22-								
70	11- - 31-	21-	15-	x	✓	xx	xx	✓	✓	x	x
			18-		x	xx	✓	xx	✓	✓	x
80	14- - 34-	24-	18-			x	xx	x	x	✓	x
			20-			x	✓	✓	x	x	x

Legend:
- ☐ not tested
- ■ x sequence not confirmed
- ■ ✓ sequence confirmed

Conditions, collision energies and selected charge states are given in the Table, other conditions as in Figure 1. Oligonucleotide sequences are given in Table 1.

charge state having a relative abundance of at least 30-50% should be selected for fragmentation.

Influence of Collision Energy and Charge State on COMPAS of 38- to 80-mer Oligonucleotides

To get a more general impression of the applicability of COMPAS for the sequence verification of long-chain oligonucleotides, nine oligonucleotides differing in size and base sequence were analyzed in 172 different ICEMS² experiments in which both selected charge state and collision energy were systematically varied. The proper setting of collision energy was essential in order to avoid misinterpretation of internal or secondary fragment ions formed at higher collision energies as regular a-B or w ions. Hence, the collision energies were set high enough so that the sequence was covered as completely as possible by ions of the a-B and w series but low enough to keep secondary and/or alternative fragmentation pathways to a minimum. This compromise was found to be in the region of the collision energy just sufficient for complete fragmentation of the precursor ion [25]. The relative collision energies were typically varied in the range between 13 and 20%.

Additionally, the charge states of the selected oligonucleotide precursor ions were varied to study its influence on the success of COMPAS in combination with the selected collision energy. As can be deduced from Table 2, the minimum and maximum charge states of the observed species were in the range of 6- to 34- for the investigated oligonucleotides. Various charge states were selected for fragmentation by ICEMS² based on the following criteria: charge states having signal intensities close to the maximum signal intensity observed in the spectrum, lower charge states having a relative signal intensity of approximately 20–40%, and in some cases, the lowest charge state observed in the mass spectrum.

Table 2 summarizes the results of ICEMS²–COMPAS with oligonucleotides ranging in lengths from 38 to 80 nt (Oligonucleotides 1–9, Table 1). Positive and negative sequence verifications are identified in the table by check marks and crosses, respectively. For repetitive experiments under identical conditions, the results of the individual experiments are indicated separately. It can be seen that COMPAS yielded positive sequence verifications for all oligonucleotides up to 80-mers utilizing at least three different combinations of charge state and collision energy. In most cases, a verification of the reference sequence was possible in a window of two or more consecutive collision energies (Table 2). At 15% relative collision energy, all oligonucleotides except the 70- and the 80-mer were successfully verified, while the latter showed positive results at slightly higher collision energies, indicating that higher collision energy is required for the fragmentation of long-chain oligonucleotides. A relative collision energy of 20% enabled positive sequence verifications only in three experiments and presumably represents the upper limit of applicable collision energies. In general, the

window of suitable collision energies was significantly broader with smaller oligonucleotides. As already discussed above, very low charge states were usually not intense enough for confident sequence verification. The most suitable charge states selected for ICEMS²– COMPAS were generally located somewhere between the lowest observable and the most abundant charge state.

The results presented above indicate that COMPAS is basically applicable for the re-sequencing of oligonucleotides up to 80-mers using quadrupole ion trap mass analyzers. However, the range of suitable experimental conditions for successful sequence verification becomes narrower with oligomers longer than 60-mers. Hence, a couple of experiments with variation in precursor ion selection and collision energy are recommended for the investigation of longer oligomers. The upper size limit of about 80 nucleotides for COMPAS is in part due to the limited mass accuracy of the utilized quadrupole ion trap mass analyzer. High mass accuracy is mandatory for the correct assignment of calculated and measured fragment ion masses, especially because generally only one charge state is utilized for the calculation of the molecular mass of a fragment ion. Accordingly, a mass accuracy typical for an ion trap mass analyzer of ± 0.3–0.5 mass units translates into an absolute mass difference of ± 6–10 mass units for the 20- charge state of an 80-mer fragment ion ($M_r \sim 25,000$) showing a signal at m/z ~1250 in the mass spectrum. Such a mass accuracy makes the differentiation of A and T having a difference of nine mass units impossible.

Whether or not the upper size limit for COMPAS can be increased with mass analyzers offering higher mass accuracy, such as time-of-flight or Fourier transform ion cyclotron resonance mass analyzers [33], remains to be investigated. Nonetheless, earlier studies have shown that fragmentation conditions in quadrupole ion traps favor low-energy fragmentation reactions [34–36], yielding mainly a-B and w fragments. Although the fragmentation in linear multipole collision cells under low-energy conditions has not been studied in detail so far, the published data indicate that several alternative fragmentation pathways as well as more unspecific fragmentations are observed, which make the interpretation of the MS² spectra of oligonucleotides longer than 20–30 nucleotides difficult, if not impossible [36]. Hence, it may be the case that for the COMPAS algorithm less selective fragmentation resulting in an increased number of unspecific fragments outweighs the advantage of improved mass accuracy offered by alternative mass analyzers.

Reproducibility of COMPAS Results

A prime prerequisite for any analytical method besides accuracy is reproducibility. In the context of COMPAS this means that the sequence verification can be repeated several times using the same set of experimental parameters. To evaluate the reproducibility of COMPAS, the sequence of a 50-mer (Oligonucleotide 5,
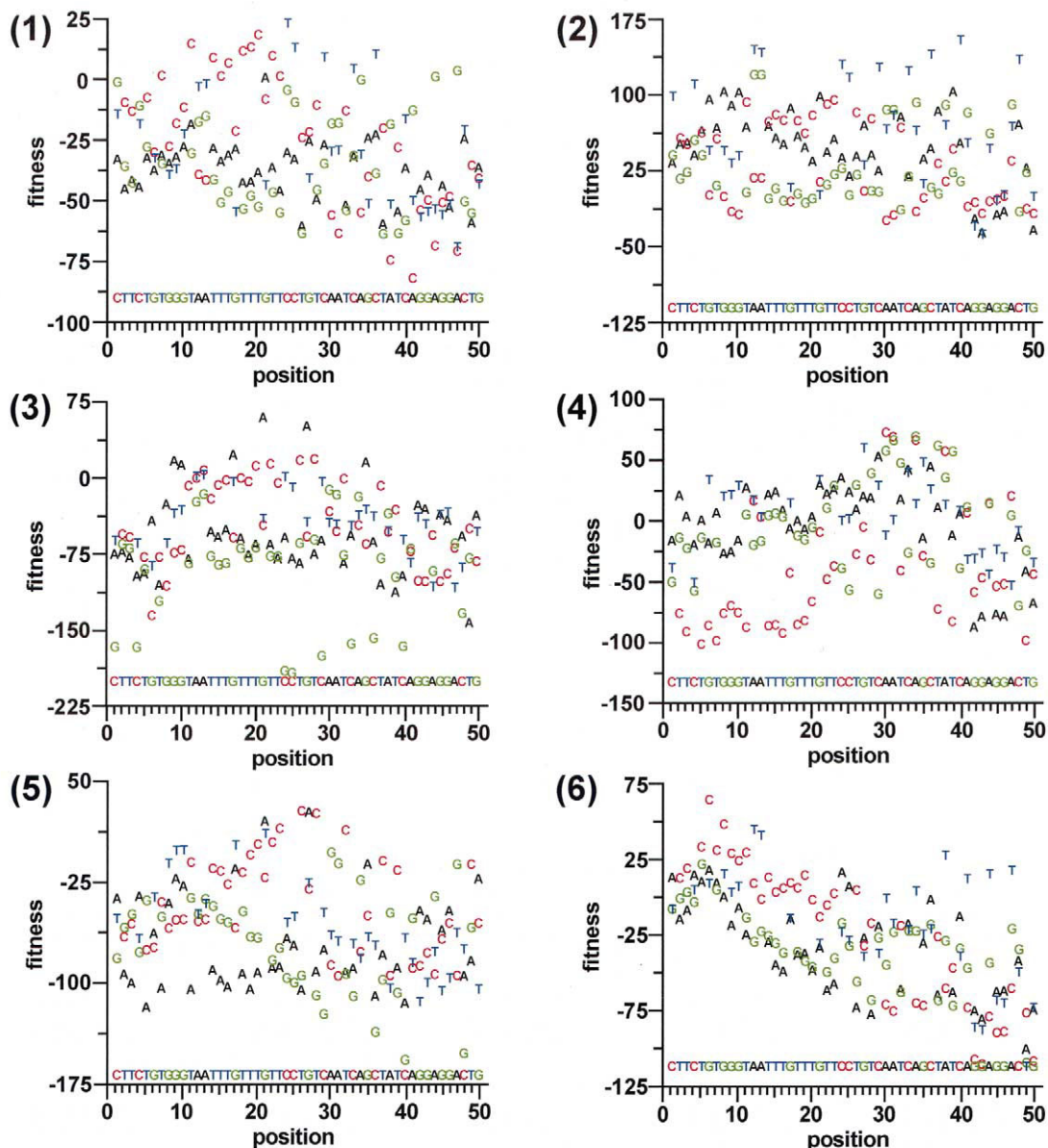
Figure 3. Six consecutive sequence verifications of a synthetic 50-mer by ICEMS^2–COMPAS. (1)–(6) correlation diagrams for the 50-mer oligonucleotide. Product ions of the 13- charged species at m/z 1185.1, 4.0 u isolation width, 19% relative collision energy; scan, 500–2000 u; sample, 65 pmol of the 50-mer; other conditions as in Figure 1.

Table 1) was checked by six consecutive analyses. In each run, 65 pmol of the synthetic raw product were purified by IP-RP-HPLC using a gradient of 4 to 20% acetonitrile in 25 mM butyldimethylammonium bicarbonate in 10 min. The MS^2 spectra originated from fragmentation of the 13- charge state at 19% relative collision energy and were extracted as the average of approximately 20 scans from the peak eluting at 6.2 min. Input of the mass spectral data and application of the COMPAS algorithm gave the correlation diagrams depicted in Figure 3. Although the patterns of fitness values for alternative sequences varied in the correla-

tion diagrams for the 50-mer oligonucleotide due to slight differences in the observed fragment ion spectra, all six calculations proved correctly and unequivocally the sequence of Oligonucleotide 5. This observation indicates that COMPAS is reproducible and reliable for the sequence verification even of long-chain oligonucleotides.

Determination of the Mass Detection Limit of ICEMS^2–COMPAS

In order to be applicable to the analysis of real biological DNA samples amplified by the polymerase chain

Table 3. Dependence of the COMPAS result of a 50-mer on the amount of oligonucleotide used to generate the fragment ion spectrum

Amount analyzed [fmol]	Reduced flow applied	Number of detected ions/ assignments	Sequence confirmed	Reduced flow applied	Number of detected ions/ assignments	Sequence confirmed
65000	no	384/412	yes			
6500	no	309/307	yes			
3250	no	263/243	yes	yes	285/276	yes
1300	no	224/239	no	yes	204/251	yes
650	no	154/152	no	yes	211/214	yes
433	no	99/155	no	yes	161/160	yes
371				yes	123/149	yes
325				yes	106/114	yes
295				yes	117/131	yes

reaction [37], COMPAS has to be capable of sequence verification utilizing fmol amounts of sample material. For the determination of the effect of sample amount on COMPAS results, a number of ICEMS² experiments were performed by analyzing 500 nL of a solution that contained a synthetic 50-mer oligonucleotide (Sequence 5, Table 1) at decreasing concentrations. The obtained fragment ion spectra were evaluated with respect to positive or negative confirmation of the reference sequence using COMPAS. The results of this study are summarized in Table 3. It is apparent that two different limits for the applicability of COMPAS can be distinguished. Without changing any experimental parameter, the sequence verification of the 50-mer was possible with amounts down to 3.25 pmol.

Interestingly, by applying reduced-flow conditions, in which the eluent flow rate was reduced from 2 μL/min to 500 nL/min during elution of the oligonucleotide, COMPAS was successful in checking the oligonucleotide sequence down to 295 fmol of the raw 50-mer. The reduction in the flow-rate enabled an about 2 to 3-fold increase in the time window available for MS² investigation, resulting in a considerable improvement in the quality of the recorded MS² spectra. The amount of 295 fmol is equivalent to the mass detection limit for oligonucleotide analysis by ICEMS². At this point it is also important to note that the synthetic oligonucleotide used for this dilution series was a raw product from solid-phase synthesis, containing up to 40–50% of byproducts. Hence, the detection limits of MS² and of COMPAS for single compounds are in fact significantly lower than 295 fmol and consequently in a range where PCR products can be successfully analyzed.

Comparison of Normal Flow- and Reduced Flow Analyses

Figure 4 illustrates the tandem mass spectra and the corresponding correlation diagrams of 65 pmol and 295 fmol amounts of a synthetic 50-mer oligonucleotide, respectively. It is obvious that upon reducing the sample amount the signal intensity as well as the number of detected fragment ions decreased continuously (Figure 4 and Table 3). Table 3 also lists the number of assignments which may be higher than the number of detected ions, because several fragments having different charge states may match with the same m/z value. The number of sequence specific fragment ions was high enough in both cases to unequivocally verify the correct sequence of the 50-mer.

However, sample amounts equal or less than 1.3 pmol could be successfully re-sequenced only under reduced flow but not under normal flow conditions. This is a consequence of the higher number of recorded MS² spectra, resulting in better spectrum quality, a higher number of fragment ions, higher number of assignments, and higher mass accuracy because of better spectrum statistics. Finally, 117 fragment ions were detected upon analysis of 295 fmol, which were still specific enough to correctly verify the sequence of the 50-mer. At this point, it is difficult to reveal the mass spectral details that enabled a positive sequencing result by using 131 assignments obtained from 295 fmol of sample in reduced flow mode, while 155 assignments with 433 fmol in normal flow mode could not verify the sequence. We believe that the most probable explanation for the different result rests within a better mass accuracy and better statistics of signal intensity which may be achieved at reduced flow because of longer measurement times.

SNP Genotyping Using COMPAS

The polymorphic locus present in a 62-bp fragment of the Y-chromosomal locus M9 was utilized to demonstrate the feasibility of direct genotyping of SNPs by ICEMS² and COMPAS. We have shown recently that the two possible genotypes of M9 containing either C or G at position 22 from the 5′ end of the forward strand can be determined unequivocally by intact molecular mass measurement using ICEMS [17]. Nevertheless, for the determination of the exact position of the base exchange sequencing information is mandatory. Therefore, the 19- charge state of the HPLC-purified forward strand was isolated and fragmented at 20% relative
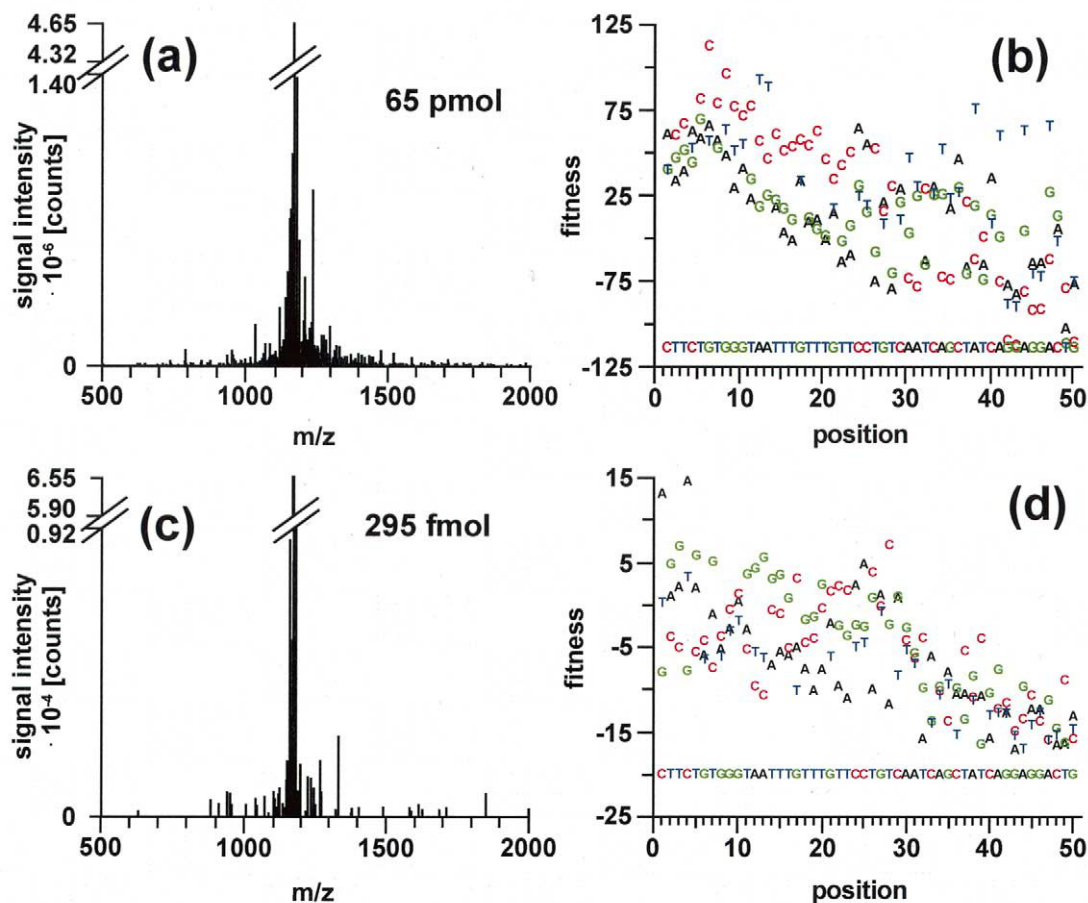
Figure 4. Tandem mass spectra (**a**, **c**) and correlation diagrams (**b**, **d**) for the 50-mer oligonucleotide using different sample amounts. Flow rate, (**a**) 2.0 μL/min, (**c**) 2.0 μL/min reduced to 500 nl/min after 6.1 min; product ions of the 13- charged species at *m/z* 1185.1, 4.0 u isolation width, 19% relative collision energy; scan, 500–2000 u; sheath liquid, (**a**) 2.0 μL/min, (**c**) 500 nL/min acetonitrile; sample, (**a**) 65 pmol, (**c**) 295 fmol of the 50-mer; other conditions as in Figure 1.

collision energy. The *m/z* values and relative signal intensities in the fragment ion spectrum served as input for the comparative sequencing algorithm. Figure 5 depicts the fitness diagrams using the forward strand sequences of the two possible genotypes as reference sequences. For clarity of presentation, only the nucleotides between the two primers from position 22–33 are included in the diagrams.

From Figure 5a it becomes clear that the sequence of the C allele did not match the experimental data. G instead of C at position 22 was most likely the correct base, because the fitness value of a sequence containing G at position 22 was the lowest observed in the correlation diagram. In fact, incorporation of G into position 22 of the sequence and recalculation of the correlation diagram showed a perfect match between the fragment ion spectrum and the reference sequence (Figure 5b). Hence, both fitness diagrams positively confirmed the allelic state of the investigated amplicon as the G allele at position 22 in the forward strand, which was additionally confirmed by restriction fragment polymorphism analysis [17].

Conclusions

This study explored both the potential as well as the limits of ICEMS[2] in combination with a computer-aided comparative sequencing algorithm called COMPAS for the sequence verification of investigation of sequence variation in synthetic oligonucleotides and PCR products. Given the fact that more than two SNPs are very unlikely to occur within a single, short PCR amplicon (<100 bp), the method can be used both to determine the allelic state of known polymorphisms and to identify previously unknown mutations. The method is applicable to the fully automable re-sequencing of nucleic acids ranging in size from a few to 80 nucleotides.

Experimental parameters were shown to have a significant impact on the success of the comparative sequencing algorithm. For successful sequencing it is necessary to find a compromise between high signal intensity and low charge state of the parent ion. Moreover, the collision energy has to be high enough to generate fragments covering the sequence as completely as possible by a-B and w ions but low enough to
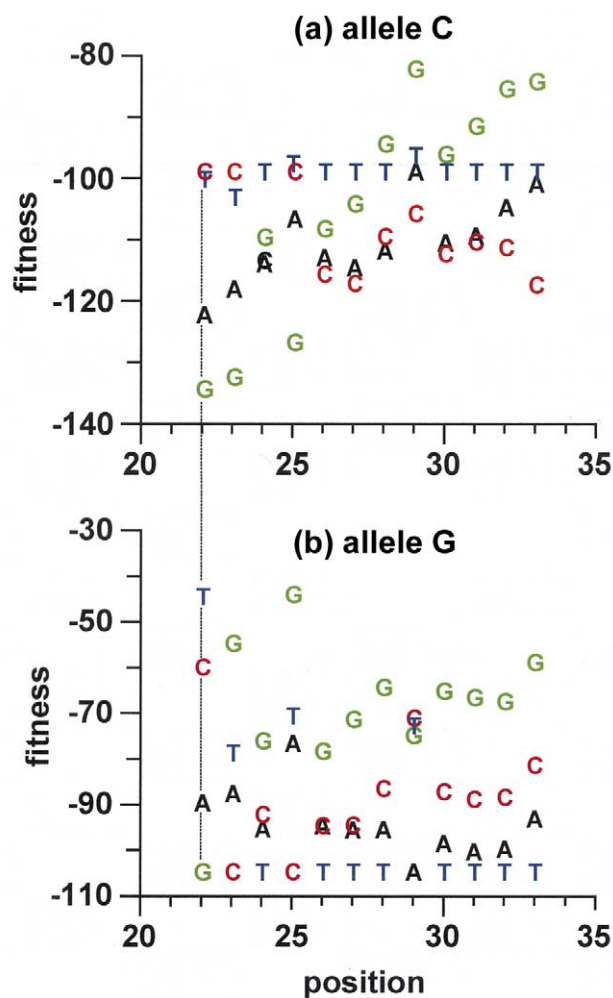
Figure 5. Genotyping of the PCR-amplified, Y-chromosomal locus M9 by ICEMS²–COMPAS. Linear gradient, 5–70% B in 10.0 min; flow rate, 2.0 μL/min reduced to 500 nl/min after 4.6 min; product ions of the 19- charged species at *m/z* 1001.9, 4.0 u isolation width, 20% relative collision energy; scan, 500–2000 u; sheath liquid, 500 nL/min acetonitrile; sample, 500 nl of the crude PCR solution; other conditions as in Figure 1.

suppress secondary and/or alternative fragmentation pathways. By using a reduced flow procedure, sequence verification is possible for a few hundred fmol of oligonucleotides by ICEMS²–COMPAS. Therefore, sequences amplified by PCR can be readily analyzed. Because there is no need for manual inspection of mass spectra, the method can also be utilized by non-experts in the interpretation of mass spectra. Due to its sensitivity, reliability, and short analysis times on a scale of a few minutes, ICEMS²–COMPAS has the potential for routine, medium-throughput genotyping.

Acknowledgments

References

1. Gross, J.; Hillenkamp, F. Encyclopedia of Analytical Chemistry. Meyers, R. A., Ed.; John Wiley and Sons Ltd.: Chichester, 2000, pp 5022–5051.
2. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J., et al. The Sequence of the Human Genome. *Science* **2001,** *291*, 1304–1351.
3. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C., et al. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001,** *409*, 860–921.
4. Brookes, A. J. The Essence of SNPs. *Gene* **1999,** *234*, 177–186.
5. Collins, F. S.; Patrinos, A.; Jordan, E.; Chakravarti, A.; Gesteland, R.; Walters, L. New Goals for the U.S. Human Genome Project: 1998–2003. *Science* **1998,** *282*, 682–689.
6. Kristensen, V. N.; Kelefiotis, D.; Kristensen, T.; Borresen-Dale, A. L. High-Throughput Methods for Detection of Genetic Variation. *Biotechniques* **2001,** *30*, 318–322.
7. Gut, I. G. Automation in Genotyping of Single Nucleotide Polymorphisms. *Hum. Mutat.* **2001,** *17*, 475–492.
8. Li, J.; Butler, J. M.; Tan, Y.; Lin, H.; Royer, S.; Ohler, L.; Shaler, T. A.; Hunter, J. M.; Pollart, D. J.; Monforte, J. A.; Becker, C. H. Single Nucleotide Polymorphism Determination Using Primer Extension and Time-of-Flight Mass Spectrometry. *Electrophoresis* **1999,** *20*, 1258–1265.
9. Tost, J.; Gut, I. G. Genotyping Single Nucleotide Polymorphisms by Mass Spectrometry. *Mass Spectrom. Rev.* **2002,** *21*, 388–418.
10. Laken, S. J.; Jackson, P. E.; Kinzler, K. W.; Vogelstein, B.; Strickland, P. T.; Groopman, J. D.; Friesen, M. D. Genotyping by Mass Spectrometric Analysis of Short DNA Fragments. *Nat. Biotechnol.* **1998,** *16*, 1352–1356.
11. Krahmer, M. T.; Walters, J. J.; Fox, K. F.; Fox, A. ; Creek, K. E.; Pirisi, L.; Wunschel, D. S.; Smith, R. D.; Tabb, D. L.; Yates, J. R. MS for Identification of Single Nucleotide Polymorphisms and MS/MS for Discrimination of Isomeric PCR Products. *Anal. Chem.* **2000,** *72*, 4033–4040.
12. Muhammad, W. T.; Fox, K. F.; Fox, A.; Cotham, W.; Walla, M. Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry and Quadrupole Mass Spectrometry for Genotyping Single Nucleotdie Substitutions in Intact Polymerase Chain Reaction Products in K-ras and p53. *Rapid Commun. Mass Spectrom.* **2002,** *16*, 2278–2285.
13. Premstaller, A.; Oberacher, H.; Huber, C. G. High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry of Single- and Double-Stranded Nucleic Acids Using Monolithic Capillary Columns. *Anal. Chem.* **2000,** *72*, 4386–4393.
14. Oberacher, H.; Parson, W.; Mühlmann, R.; Huber, C. G. Analysis of Polymerase Chain Reaction Products by On-Line Liquid Chromatography-Mass Spectrometry for Genotyping of Polymorphic Short Tandem Repeat Loci. *Anal. Chem.* **2001,** *73*, 5109–5115.
15. Oberacher, H.; Oefner, P. J.; Parson, W.; Huber, C. G. On-Line Liquid Chromatography-Mass Spectrometry: A Useful Tool for the Detection of DNA Sequence Variation. *Angew. Chem. Int. Ed.* **2001,** *40*, 3828–3830.
16. Oberacher, H.; Oefner, P. J.; Hölzl, G.; Premstaller, A.; Huber, C. G. Re-sequencing of Multiple Single Nucleotide Polymorphisms by Liquid Chromatography-Electrospray Ionization Mass Spectrometry. *Nucleic Acids Res.* **2002,** *30*, e67.
17. Berger, B.; Hölzl, G.; Oberacher, H.; Niederstätter, H.; Huber, C. G.; Parson, W. Single Nucleotide Polymorphism Genotyping by On-Line Liquid Chromatography-Mass Spectrometry in Forensic Science with the Y-Chromosomal Locus M9. *J. Chromatogr. B* **2002,** *782*, 89–97.
18. Oberacher, H.; Huber, C. G.; Oefner, P. J. Mutation Screening by Ion-Pair Reversed-Phase High-Performance Liquid Chro-

matography-Electrospray Ionization Mass Spectrometry (ICEMS). *Hum. Mutat.* **2003,** *21,* 86–95.

19. Murray, K. K. DNA Sequencing by Mass Spectrometry. *J. Mass Spectrom.* **1996,** *31,* 1203–1215.

20. Nordhoff, E.; Kirpekar, F.; Roepstorff, P. Mass Spectrometry of Nucleic Acids. *Mass Spectrom. Rev.* **1996,** *15,* 76–138.

21. Wan, K. X.; Gross, J.; Hillenkamp, F.; Gross, M. L. Fragmentation Mechanisms of Oligodeoxynucleotides Studied by H/D Exchange and Electrospray Ionization Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 193–205.

22. Wan, K. X.; Gross, M. L. Fragmentation Mechanisms of Oligodeoxynucleotides: Effects of Replacing Phosphates with Methylphosphonates and Thymines with Other Bases in T-Rich Sequences. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 580–589.

23. Ni, J.; Pomerantz, S. C.; Rozenski, J.; Zhang, Y.; McCloskey, J. A. Interpretation of Oligonucleotide Mass Spectra for Determination of Sequence Using Electrospray Ionization and Tandem Mass Spectrometry. *Anal. Chem.* **1996,** *68,* 1989–1999.

24. Rozenski, J.; McCloskey, J. A. SOS: A Simple Interactive Program for ab Initio Oligonucleotide Sequencing by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002,** *13,* 200–203.

25. Oberacher, H.; Wellenzohn, B.; Huber, C. G. Comparative Sequencing of Nucleic Acids by Liquid Chromatography-Tandem Mass Spectrometry. *Anal. Chem.* **2002,** *74,* 211–218.

26. Oberacher, H.; Walcher, W.; Huber, C. G. Effect of Instrument Tuning on Detectability of Biopolymers in Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **2003,** *38,* 108–116.

27. Bleicher, K.; Bayer, E. Various Factors Influencing the Signal Intensity of Oligonucleotides in Electrospray Mass Spectrometry. *Biol. Mass Spectrom.* **1994,** *23,* 320–322.

28. Huber, C. G.; Buchmeiser, M. R. On-line Cation-Exchange for Suppression of Adduct Formation in Negative-Ion Electrospray Mass Spectrometry of Nucleic Acids. *Anal. Chem.* **1998,** *70,* 5288–5295.

29. Cheng, X.; Gale, D.; Udseth, H. R.; Smith, R. D. Charge State Reduction of Oligonucleotide Negative Ions from Electrospray Ionization. *Anal. Chem.* **1995,** *67,* 586–593.

30. McLuckey, S. A.; Habibi-Goudarzi, S. Decompositions of Multiply Charged Oligonucleotide Anions. *J. Am. Chem. Soc.* **1993,** *115,* 12085–12095.

31. McLuckey, S. A.; Vaidyanathan, G. Charge State Effects in the Decompositions of Single-Nucleobase Oligonucleotide Polyanions. *Int. J. Mass Spectrom. Ion Processes* **1997,** *162,* 1–16.

32. Luo, H.; Lipton, M. S.; Smith, R. D. Charge Effects for Differentiation of Oligodeoxynucleotide Isomers Containing 8-oxo-dG Residues. *J. Am. Soc. Mass Spectrom.* **2002,** *13,* 195–199.

33. Little, D. P.; Thannhauser, T. W.; McLafferty, F. W. Verification of 50- to 100-mer DNA and RNA Sequences with High-Resolution Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *95,* 2318–2322.

34. Barry, J. P.; Vouros, P.; Van Schepdael, A.; Lay, S.-J. Mass and Sequence Verification of Modified Oligonucleotides Using Electrospray Tandem Mass Spectrometry. *J. Mass Spectrom.* **1995,** *30,* 993–1006.

35. Premstaller, A.; Ongania, K.-H.; Huber, C. G. Factors Determining the Performance of Triple Quadrupole, Quadrupole Iion Trap, and Sector Field Mass Spectrometers in Electrospray Ionization Tandem Mass Spectrometry of Oligonucleotides. 1. Comparison of Performance Characteristics. *Rapid Commun. Mass Spectrom.* **2001,** *15,* 1045–1052.

36. Premstaller, A.; Huber, C. G. Factors Determining the Performance of Triple Quadrupole, Quadrupole Ion Trap, and Sector Field Mass Spectrometers in Electrospray Ionization Tandem Mass Spectrometry of Oligonucleotides. 2. Suitability for de Novo Sequencing. *Rapid Commun. Mass Spectrom.* **2001,** *15,* 1053–1060.

37. Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N. Enzymatic Amplification of b-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **1985,** *230,* 1350–1354.