International Journal of Mass Spectrometry 304 (2011) 124–129

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Tandem mass spectrometric de novo sequencing of oligonucleotides using simulated annealing for stochastic optimization

Herbert Oberacher*, Florian Pitterl

Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria

ARTICLE INFO

Article history:
Received 29 March 2010
Received in revised form 7 June 2010
Accepted 7 June 2010
Available online 15 June 2010

Keywords:
Oligonucleotide
Electrospray ionization
Tandem mass spectrometry
De novo sequencing
Simulated annealing

ABSTRACT

A global strategy for tandem mass spectrometric de novo sequencing of oligonucleotides is presented which is based on finding among all possible sequences the sequence, whose simulated tandem mass spectrum shows the highest degree of similarity to the measured spectrum. Global de novo sequencing can become a time-consuming task due the exponential increase of the number of possible sequences with increasing oligonucleotide length. For the reduction of sequencing effort and time, simulated annealing was introduced as stochastic optimization technique that is capable to find the optimal sequence by testing only a subset of all possible configurations. We have evaluated the performance of the de novo sequencing approach by sequencing 13 synthetic oligonucleotides ranging in length from 5 to 24 nucleotides. Mass spectrometric and tandem mass spectrometric experiments were performed on a quadrupole–quadrupole–time-of-flight instrument. Accurate molecular mass measurements were used to determine the nucleotide compositions of the oligonucleotides. For the majority of tested oligonucleotides two different charge states were selected as precursor ions. Each precursor ion was fragmented applying several different collision voltages. Overall 107 fragment ion mass spectra were acquired. The nucleotide compositions together with the tandem mass spectral information represented the input for the de novo sequencing algorithm. Sequences of oligonucleotides as large as 22-mers were correctly determined in more than 90% of cases. The optimal sequence was retrieved within 1–2 min. Fragment ion mass spectra of larger oligonucleotides were inappropriate for de novo sequencing. The occurrence of extensive internal fragmentation causing low sequence coverage paired with a high probability of assigning fragment ions to wrong sequences gave rise to incorrect results.

1. Introduction

Synthetic oligonucleotides have become important tools in molecular biology and molecular diagnostics. They are used as primers and probes in many different molecular biological assays, the most prominent ones being the polymerase chain reaction [1] and sequencing [2]. Moreover, synthetic oligonucleotides are used as antisense reagents. This type of therapeutics comprises antisense oligonucleotides, ribozymes, DNAzymes, as well as small interfering RNAs (siRNAs) [3–6]. All these species are designed to specifically modulate or even prevent the transfer of the genetic information to the protein.

Synthetic oligonucleotides are usually single-stranded oligo(deoxy)nucleotides consisting of 10–50 nucleotides. Verification of the sequence of the synthetic oligonucleotides is an essential part of quality control. Modifications and short lengths hamper the sequence determination with conventional techniques. For the characterization of nucleic acid species hardly amenable to Sanger sequencing, mass spectrometry (MS) has become the method of choice [7–10]. Principally, mass spectrometric sequencing is based on the measurement of characteristic mass differences between sequence ladders. These informative ladders can either be produced prior to the mass spectrometric analysis via chemical or enzymatic reactions [11–21] or in the course of tandem mass spectrometric experiments via gas-phase fragmentation of oligonucleotide ions [22–32]. The principles of tandem mass spectrometry (MS/MS) of oligonucleotides have been extensively studied and reviewed [33]. Collision induced dissociation (CID) is the most commonly applied technique to produce fragment ions. For oligodeoxyribonucleotides, CID typically produces a_n–B_n- and w_n-type fragment ions whereas for oligoribonucleotides c_n- and y_n-type ions dominate [34–36].

A distinct advantage of tandem mass spectrometric sequencing is its inherent speed of data generation. Fragment ion mass spectra can be obtained very rapidly in a time frame of several seconds upon fragmentation and subsequent mass analysis of the

* Corresponding author at: Institute of Legal Medicine, Innsbruck Medical University, Müllerstrasse 44, 6020 Innsbruck, Austria. Tel.: +43 512 9003 70639; fax: +43 512 9003 73600.
E-mail address: herbert.oberacher@i-med.ac.at (H. Oberacher).

1387-3806/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2010.06.008

Table 1
Summary of oligodeoxynucleotides analyzed in this study.

	Sequence	Length (nt)	Charge state: collision voltage range (V)
5-mer_1	5′-ACGTA-3′	5	2−: 20–45
5-mer_2	5′-CCGAT-3′	5	2−: 20–45
5-mer_3	5′-TTAGC-3′	5	2−: 20–45
10-mer_1	5′-CGTATTAGCC-3′	10	3−: 20–45
10-mer_2	5′-ATTTGTACGT-3′	10	3−: 20–45
10-mer_3	5′-GCTCGGAATC-3′	10	3−: 20–45
15-mer_1	5′-CGTATTAGCCACGTA-3′	15	3−: 20–45
			4−: 20–45
15-mer_2	5′-ATAGCAGTCCGATTC-3′	15	3−: 20–45
			4−: 20–45
15-mer_3	5′-ACGCATTACGGCGGT-3′	15	3−: 20–45
			4−: 20–45
20-mer_1	5′-GCACCCATTACCCGAATAAA-3′	20	4−: 20–45
			5−: 20–45
20-mer_2	5′-TGCACTCCAGCCTGGGCAAC-3′	20	4−: 20–45
			5−: 20–45
22-mer	5′-TTGGTGCACCCATTACCCGAAT-3′	22	5−: 20–45
24-mer	5′-TCCAGAGACAGACTAATAGGAGGT-3′	24	5−: 20–45

fragments directly in a mass spectrometer. Data interpretation, however, still represents a bottleneck. The complexity of fragment ion mass spectrum interpretation increases with the length of the precursor ion rendering "manual" interpretation of MS/MS spectra a difficult and time-consuming task [37]. To overcome this problem, computational data interpretation routines have been developed [30,38–42].

Automated de novo sequencing of oligonucleotides can follow either local or global search paradigms. Rosenski and McCloskey have introduced a local search algorithm called "Simple Oligonucleotide Sequencer" (SOS) [38], which represents an efficient sequencing algorithm for the *ab initio* determination of unknown oligonucleotide sequences at approximately the 20-mer level and below. We have presented a global sequencing strategy [43]. Starting from a known nucleotide composition, all possible nucleic acid sequences and corresponding theoretical tandem mass spectra are generated and compared to the experimental spectrum. The correct sequence is indicated by showing the highest similarity between the simulated spectrum and the measured spectrum. Using this concept, sequences of 5–12-mer oligonucleotides were successfully *de novo* determined. A major limitation for the widespread use of this sequencing strategy arose from the exponential increase of the number of possible sequences with increasing oligonucleotide length. Accordingly, calculation times become unacceptably long even for rather small oligonucleotides (9–12-mers). The sequencing effort and time can be reduced via the application of some kind of stochastic optimization technique which is capable to find the optimal sequence by testing only a subset of all possible configurations. Herein, we introduce simulated annealing [44] as stochastic optimization method for fast and automated *de novo* sequencing of oligonucleotides. The possibilities and limitations of the *de novo* sequencing strategy have been evaluated with a set of 107 tandem mass spectra obtained from CID of synthetic oligonucleotides consisting of 5–24 nts on a quadrupole–quadrupole–time-of-flight (QqTOF) mass spectrometer.

2. Materials and methods

2.1. Chemicals and oligodeoxynucleotides

Acetonitrile (HPLC gradient-grade) was obtained from Sigma–Aldrich (St. Louis, MO, USA). A stock solution (1.0 M) of cyclohexyldimethylammonium acetate (CycHDMAA) was prepared by the titration of a cyclohexyldimethylamine solution (Fluka, Buchs, Switzerland) with acetic acid (Fluka) at 5 °C until pH 8.3 was reached. HPLC-grade water was from Merck (Darmstadt, Germany). Synthetic oligodeoxynucleotides were obtained from Sigma–Aldrich. The sequences were randomly selected and are summarized in Table 1.

2.2. Mass spectrometric measurements

ESI-MS and -MS/MS experiments were performed on a QSTAR XL mass spectrometer (AB Sciex, Foster City, CA, USA). A modified TurboIonSpray source was used for all experiments. The modifications included the replacements of the Peek tubing transfer line and of the stainless steel sprayer capillary by fused silica capillaries (transfer line: 375 μm o.d., 20 μm i.d., sprayer capillary: 90 μm o.d., 20 μm i.d., Polymicro Technologies, Phoenix, AZ, USA) [45,46]. Oligodeoxynucleotides were analyzed in the negative ion mode by continuous infusion of 5–20 pmol/μL solutions in 25 mM aqueous CycHDMAA containing 50% acetonitrile (v/v) at a flow rate of 2.0 μL/min. Cations present in the sample solutions were removed by on-line cation-exchange using a 20 mm × 0.50 mm i.d. cation-exchange microcolumn packed with 38–75 μm Dowex 50 WX8 particles (Serva, Heidelberg, Germany) [47]. The spray voltage was set to 3.9 kV. Gas flows of 10–15 arbitrary units for the nebulizer gas and 40 arbitrary units for the turbo gas were employed. The temperature of the turbo gas was adjusted to 200 °C. For MS/MS, the Q1 resolution was set to unit resolution. The collision gas (N_2) flow was set to 5 arbitrary units. The collision energy was changed by varying the voltage applied to the collision cell (Table 1). The accumulation time was set to 1 s. Mass spectra collected over a period of 1.0–2.0 min were averaged. Mass spectra were recorded on a personal computer with the Analyst QS software (service pack 8 and Bioanalyst extension, Applied Biosystems).

2.3. Data processing

Measured MS/MS spectra were exported from the Analyst software as text files. Each file contained a list of the centroided fragment ion *m/z*-values and the corresponding relative signal intensities. To improve sequencing efficiency, files were processed [42]. Data treatment was accomplished fully automatically using a program written in ActivePerl 5.6.1® (Active State Corporation, Vancouver, BC, Canada) and started with the deletion of all signals detected within a range of ±5 mass units around the *m/z* of the precursor ion. Furthermore, all signals having either a signal intensity smaller than 1.0% of the intensity of the most intense fragment ion or which were identified as isotopic peaks were deleted.

2.4. De novo sequencing

The nucleotide compositions of the oligonucleotides were deduced from accurate molecular mass measurements and were used to generate starting sequences as input for the optimization process.

The similarity between any sequence and the sequence of the fragmented oligonucleotide was specified using routines part of the comparative sequencing algorithm (COMPAS) [40–42,48]. Briefly, the algorithm involves the comparison of a measured MS/MS spectrum to a set of fragment ion *m/z*-values predicted from a given sequence. The closeness of matching between the measured spectrum and the predicted set of ions is characterized by a value called fitness (FS). FS determination starts with the generation of a list of monoisotopic *m/z*-values representing all theoretically possible a_n–B_n- and w_n-ions for the given reference sequence. Then, the predicted *m/z*-values and those obtained from the experimental spectrum are compared and FS is calculated. The maximum tolerable mass deviation for matching predicted and calculated fragment ions was fixed at 0.1. FS takes into account the difference between measured and predicted *m/z*-values, the relative intensity of the fragment ions, the number of matched fragment ions, and the number of nucleotide positions not covered by fragment ions in the experimental spectrum. The larger the value for FS, the closer the match between the measured and the predicted spectra.

For finding the sequence showing the highest correlation between the predicted fragmentation pattern and the measured spectrum (= optimal sequence), a simulated annealing algorithm was applied. The iterative improvement of the sequence involves random generation of new sequences and consecutive calculation of the corresponding FS-value. The sequence generator interchanges nucleotides at two randomly selected positions within the sequence. The permutation is accepted if the FS-value of the new sequence (FS2) is larger than the FS-value of the previous sequence (FS1). The new sequence is the starting point of the next step. The case $\Delta FS < 0$, meaning decline of the similarity, is treated probabilistically: the probability that the configuration is accepted is $P(\Delta FS) = \exp(\Delta FS/FS1/T)$. If $P(\Delta FS)$ is larger than a random number in the interval {0,1} the new sequence is retained; if not, the original sequence is used to start the next step. Within the probability function *T* is a global parameter which is called temperature. Initially *T* is set to 0.5 and is decreased after every 50 cycles by 10%. After 800 cycles and after 1600 cycles *T* is set back to 0.5. Overall 2400 optimization steps are performed to find the best matching sequence.

All calculations were performed on a personal computer under Windows XP™ operating system (1.7 GHz Pentium, 1.0 GB RAM). Automated *de novo* sequencing was performed with a program written in ActivePerl 5.6.1® (Active State Corporation). A copy of the program for academic use is available upon request from the corresponding author.

3. Results and discussion

3.1. The de novo sequencing strategy

The developed *de novo* sequencing approach consists of three steps (Fig. 1). Firstly, the molecular mass of the oligonucleotide is determined (Fig. 1a). Depending on the accuracy of the mass spectrometric measurement a certain number of nucleotide compositions will be retrieved [49,50]. Secondly, tandem mass spectrometric experiments are performed (Fig. 1b). Ions representing a charge state of the oligonucleotide are isolated and fragmented. Thirdly, the measured tandem mass spectrum together with all possible nucleotide compositions are used as input for computer-assisted *de novo* sequencing to find the sequence showing the highest probability to represent the true sequence of the oligonucleotide (Fig. 1c). The similarity between any sequence and the sequence of the fragmented oligonucleotide is specified using routines part of COMPAS [40–42,48]. The true sequence should retrieve the highest FS-value.

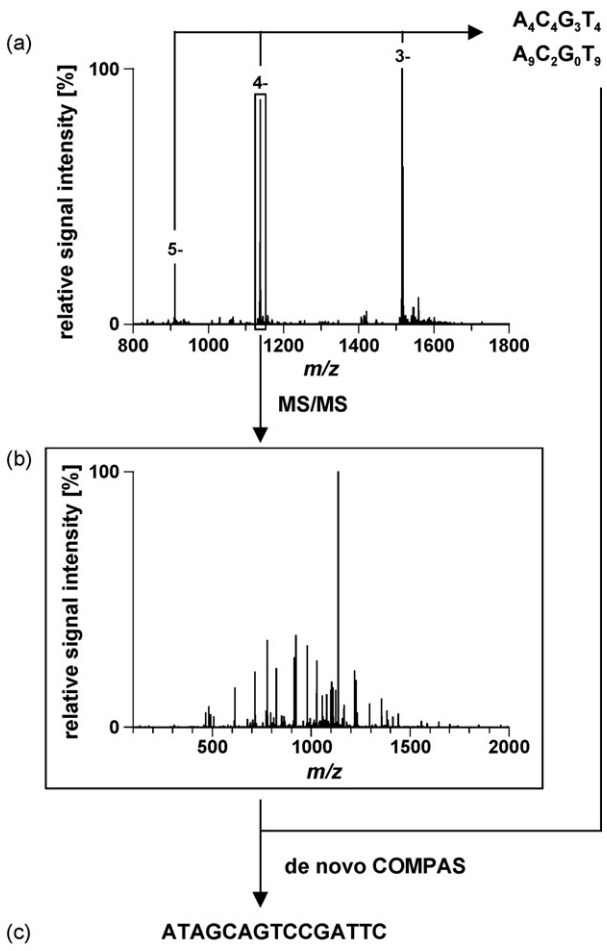

Fig. 1. Outline of the steps involved in de novo sequencing of oligonucleotides.

In global *de novo* sequencing principally all possible sequences need to be tested to find the optimal sequence. Starting from a given nucleotide composition oligonucleotide sequences can be assembled by changing the nucleotide order. The length and the nucleotide composition define the number of possible permutations. For an oligonucleotide having the composition $A_aC_cG_gT_t$ the

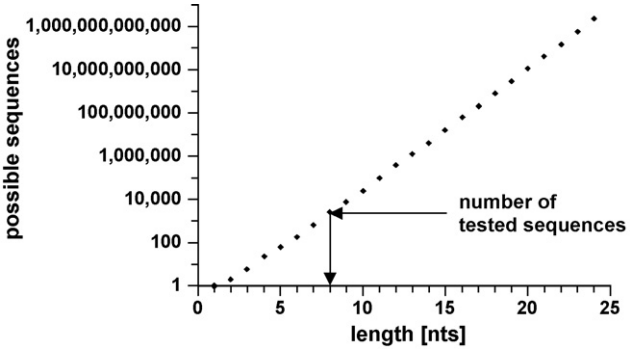

Fig. 2. Estimation of the maximum number of possible sequences for oligonucleotides consisting of different numbers of nucleotides.

Table 2
De novo sequencing of 15-mer_2. The tandem mass spectrum was obtained from fragmentation of the quadruply charged ion (CE = 140 eV).

Composition	Cycles	Best matching sequence	Fitness value (FS)
$A_9C_2G_0T_4$	800	5′-ATATAAAAAACATTC-3′	510.8
	1600	5′-ATATAAAAAACATTC-3′	510.8
	2400	5′-ATATAAAAAACATTC-3′	510.8
	800	5′-ATATAAAAAACATTC-3′	510.8
	1600	5′-ATATAAAAAACATTC-3′	510.8
	2400	5′-ATATAAAAAACATTC-3′	510.8
	800	5′-ATATAAAAAACATTC-3′	510.8
	1600	5′-ATATAAAAAACATTC-3′	510.8
	2400	5′-ATATAAAAAACATTC-3′	510.8
$A_4C_4G_3T_4$	800	5′-ATAGCAGTCCGATTC-3′	5462.9
	1600	5′-ATAGCAGTCCGATTC-3′	5462.9
	2400	5′-ATAGCAGTCCGATTC-3′	5462.9
	800	5′-ATAGCAGTCCGATTC-3′	5462.9
	1600	5′-ATAGCAGTCCGATTC-3′	5462.9
	2400	5′-ATAGCAGTCCGATTC-3′	5462.9
	800	5′-ATAGCAGTCCGATTC-3′	5462.9
	1600	5′-ATAGCAGTCCGATTC-3′	5462.9
	2400	5′-ATAGCAGTCCGATTC-3′	5462.9

number of permutations can be calculated using the following equation:

$$\text{number of permutations} = \frac{(a+c+g+t)!}{a!\,c!\,g!\,t!}$$

The number of possible sequences increases exponentially with increasing oligonucleotide length (Fig. 2). The exponential explosion has a severe impact on computing effort; only for very small oligonucleotides all possible permutations can be tested within seconds or minutes [41]. To overcome this problem and to extend the size range for the applicability of the global *de novo* sequencing strategy, we used a stochastic search approach to come to a sequence that is hoped to be the true sequence. Stochastic optimization refers to minimization or maximization of a function in the presence of randomness in the optimization process. The goal is to find the global minimum/maximum of a function that may possess several local minima/maxima by testing only a subset of all possible configurations. Stochastic search techniques are used when the structure of a space is not well understood or is not smooth. The stochastic optimization method that we have selected for *de novo* sequencing of oligonucleotides was simulated annealing [44]. Simulated annealing works by emulating the physical process whereby a solid, initially at high temperature and disordered, is slowly cooled so that when eventually its structure is frozen a state of minimum internal energy is reached. The role of temperature is to allow the configurations to reach higher energy states with a probability given by the Boltzmann's exponential law, so that they can overcome energy barriers that would otherwise force them into local minima. Any simulated annealing algorithm needs four ingredients: (1) a concise description of a configuration of the system; (2) a random generator of rearrangements of the elements in a configuration; (3) a quantitative objective function; and (4) an annealing schedule of the temperature and length of time for which the system is to be evolved. We have adapted the simulated annealing procedure to find a sequence whose simulated tandem mass spectrum shows the highest degree of similarity to the measured spectrum. The start configuration 5′-a(A)c(C)g(G)t(T)-3′ was generated from a given nucleotide composition $A_aC_cG_gT_t$. Rearrangements were introduced into the sequence by randomly changing the positions of two nucleotides. For each new configuration the FS-value was determined. The permutation was accepted if the FS-value of the new sequence (FS2) was larger than the FS-value of the previous sequence (FS1) and the new sequence was the starting point of the next step. The case $\Delta FS < 0$, meaning decline of the similarity, was treated probabilistically: the probability that

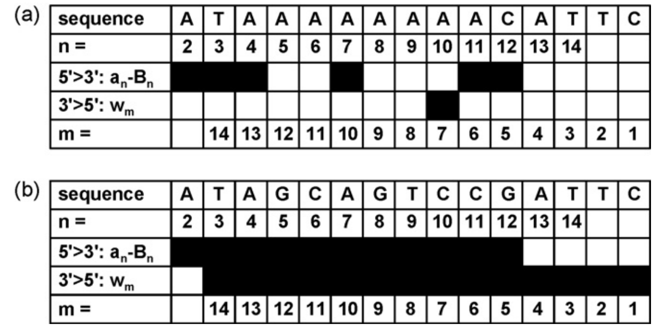

Fig. 3. Sequence coverage diagrams for two sequences matched to 15-mer_2. The tandem mass spectrum was obtained from fragmentation of the quadruply charged ion (CE = 140 eV).

the configuration was accepted was $P(\Delta FS) = \exp(\Delta FS/FS1/T)$. If $P(\Delta FS)$ was larger than a random number in the interval [0,1] the new sequence was retained; if not, the original sequence was used to start the next step. Within the probability function T is a global parameter which is called temperature. The temperature distinguishes classes of rearrangements, so that rearrangements causing large changes in FS occur at high temperatures, while the small changes are deferred until low temperatures. Sufficient efficiency in finding the best matching sequence was obtained by initially setting T to 0.5 and decreasing it after every 50 cycles by 10%. After 800 cycles and after 1600 cycles T was set back to 0.5. Overall 2400 optimization steps were performed to find the best matching sequence, and these were accomplished within 1–2 min. The number of optimization steps was nearly equal to the maximum number of permutations for an 8-mer oligonucleotide (2520, Fig. 2). Thus, for very small oligonucleotides (nts $\leq$ 8) no improvement in calculation time with the simulated annealing approach was obtained. For larger oligonucleotides, however, the simulated annealing approach significantly reduced calculation time and effort.

3.2. *De novo sequencing of a 15-mer oligonucleotide*

De novo sequencing of an oligonucleotide is exemplified by sequencing of 15-mer_2 (Fig. 1). The monoisotopic mass of 15-mer_2 was 4549.80. The mass deviation for a QqTOF instrument is typically below 50 ppm [46]. Under these constraints only two nucleotide compositions are possible: $A_4C_4G_3T_4$ and $A_9C_2G_0T_4$. The tandem mass spectrum was obtained from the quadruply charged ion (CE = 140 eV). Using 5′-AAAAAAAAACCTTTT-3′ as starting

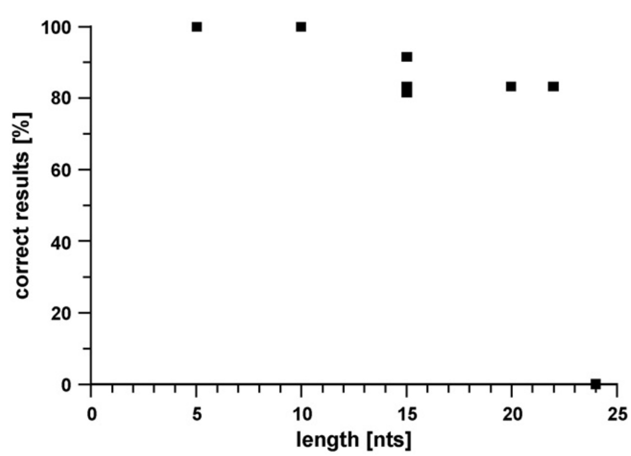

Fig. 4. Oligonucleotide length dependence of correct sequencing results.

Table 3
Comparison of the true and the optimal sequence for incorrect de novo sequencing results.

	True sequence	Charge state, Collision voltage (V)	Optimal sequence
15-mer_1	5′-CGTATTAGCCACGTA-3′	4-, 40	5′-**GC**TATTAGCCACGTA-3′
		4-, 45	5′-**TATTACG**GCCACGTA-3′
15-mer_2	5′-ATAGCAGTCCGATTC-3′	4-, 40	5′-**TA**AGCAGTCCGATTC-3′
		4-, 45	5′-**TA**AGCAGTCCGATTC-3′
15-mer_3	5′-ACGCATTACGGCGGT-3′	4-, 45	5′-**GCA**CATTACGGCGGT-3′
20-mer_1	5′-GCACCCATTACCCGAATAAA-3′	4-, 20	5′-GCACCC**TTA**ACCCGAATAAA-3′
		5-, 20	5′-GCACCCATTACCCGAA**AT**AA-3′
20-mer_2	5′-TGCACTCCAGCCTGGGCAAC-3′	5-, 25	5′-TGCACTCCAGCCTGG**CG**AAC-3′
		5-, 45	5′-T**CCGA**TCCAGCCTGGGCAAC-3′
22-mer	5′-TTGGTGCACCCATTACCCGAAT-3′	5-, 45	5′-T**GT**GTGCACCCATTACCCGAAT-3′

configuration, optimization resulted in 5′-ATATAAAAAACATTC-3′ (Table 2). The corresponding FS-value was 510.8. In Fig. 3a the sequence coverage diagram is shown. The sequence coverage (26%) is rather low which suggests that there might be an alternative sequence corresponding to the second composition that might fit best. For 5′-AAAACCCCGGGTTTT-3′, 5′-ATAGCAGTCCGATTC-3′ was obtained as sequencing result (Table 2). The corresponding FS-value was 5462.9. The sequence coverage was 93% (Fig. 3b). *De novo* sequencing was repeated two more times. In all cases the same sequence was retrieved as sequencing result (Table 2), which in fact was equal to the true sequence of 15-mer_2 (5′-ATAGCAGTCCGATTC-3′).

3.3. Performance of the de novo sequencing approach

The performance of the developed *de novo* sequencing approach was evaluated by sequencing 13 synthetic oligodeoxynucleotides ranging in length from 5 to 24 nts (Table 1). For the majority of tested oligonucleotides two different charge states were selected as precursor ions. Each precursor ion was fragmented at several different collision voltages. According to a recently published rule of thumb [42], collision voltages ranging from 20 to 45 V were applied for the fragmentation of precursor ions carrying at least one charged nucleotide per 7 nts to ensure the generation of a sufficient amount of sequence-specific fragment ions. Utmost sequence coverage is of importance for unequivocal and correct *de novo* sequencing. Altogether 107 tandem mass spectra were acquired. The filtered spectra together with the nucleotide compositions were used as inputs for automated *de novo* sequencing. Optimized sequences were compared with the true sequences to evaluate the performance of the *de novo* sequencing algorithm (Fig. 4). Up to the 22-mer the true sequence was retrieved as optimal sequence in the majority of cases. On average correct results were obtained in 90.1% of cases. The sequences of the 5- and 10-mer oligonucleotides were correctly determined in 100% of cases. For the other oligonucleotides the percentage of correct results was always larger than 82%. The incorrect sequencing results are summarized in Table 3. Difficulties were mainly encountered at high collision energies, where more extensive fragmentation leading to smaller fragment ions was observed. Extensive fragmentation is connected with a reduction of sequence coverage particularly at the ends of the oligonucleotide sequences. Interestingly, at high collision energy settings sequencing of the 5′-end was more affected than sequencing of the 3′-end, which suggests that loss of coverage with long-chain w_n-type ions was more likely than loss of coverage with long-chain a_n–B_n-type ions. For 20-mer_1, a collision voltage of 20 V was too low to enable sufficient sequence coverage. Thus, permutations of the true sequence showing reversed orders of A and T were obtained as sequencing results. For the 24-mer all sequencing results were incorrect. This was due, presumably, to an increased amount of unspecific fragmentation for long-chain oligonucleotides (Fig. 5 in Ref. [42]). As demonstrated recently (Fig. 6b in Ref. [42]), the amount of internal fragments seems to rise for oligonucleotides exceeding a length of 20–22 nts by using a QqTOF instrument for CID. Unspecific fragmentation has a negative effect on sequence coverage and therefore reduces the FS of the true sequence. The probability that an alternative sequence might become the optimal sequence is further increased due to the fact that unspecific fragment ions can match to simulated ions corresponding to incorrect sequences. These unwanted matches give rise to FS-values exceeding the FS of the true sequence.

4. Conclusions

Tandem mass spectrometric sequencing represents a valuable tool for the sequence elucidation of small oligonucleotides. By applying CID on a QqTOF instrument, tandem mass spectra exhibiting sufficiently high sequence coverage for *de novo* sequence are obtained up to approximately the 20-mer level. Tandem mass spectrometric sequencing of larger oligonucleotides is severely hampered by secondary fragmentation. Extension of the size range to or beyond the 40-mer level might be accomplished by using alternative instrumental platforms or activation methods for fragmentation. Particularly ion trap instruments in combination with high-performance orbitrap or Fourier-transform ion cyclotron resonance mass analyzers seem to represent appropriate platforms for nucleic acids sequencing.

For *de novo* sequencing a global search paradigm was used which is based on finding the sequence, whose simulated tandem mass spectrum shows the highest degree of similarity to the measured spectrum. A major limitation for the application of this strategy to the sequencing of larger oligonucleotides arises from the exponential increase of the number of possible sequences with increasing oligonucleotide length. To reduce sequencing effort and time, simulated annealing was introduced as stochastic optimization technique, which has been shown to represent a reliable, fast and automated approach for tandem mass spectrometric *de novo* sequencing of small oligodeoxyribonucleotides consisting of the four naturally occurring DNA bases. By allowing the occurrence of modifications particularly at the bases as well as alternative fragmentation pathways (e.g. c- and y-type fragmentation) the applicability of the *de novo* sequencing approach can be extended to the sequencing of oligoribonucleotides, the characterization of DNA adduct formation, or the metabolic profiling of nucleic acids used as pharmaceuticals.

Acknowledgments

The authors wish to thank AB Sciex for the generous provision of the mass spectrometer and the associated equipment.

References

[1] R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, Enzymatic amplification of b-globin genomic sequences and restriction site analysis for Diagnosis of Sickle Cell anemia, Science 230 (1985) 1350–1354.

[2] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 5463–5467.

[3] J.B. Opalinska, A.M. Gewirtz, Nucleic-acid therapeutics: basic principles and recent applications, Nat. Rev. Drug Discov. 1 (2002) 503–514.

[4] W.H. Pan, G.A. Clawson, Antisense applications for biological control, J. Cell. Biochem. 98 (2006) 14–35.

[5] A. de Fougerolles, H.P. Vornlocher, J. Maraganore, J. Lieberman, Interfering with disease: a progress report on siRNA-based therapeutics, Nat. Rev. Drug Discov. 6 (2007) 443–453.

[6] R. Bhindi, R.G. Fahmy, H.C. Lowe, C.N. Chesterman, C.R. Dass, M.J. Cairns, E.G. Saravolac, L.Q. Sun, L.M. Khachigian, Brothers in arms: DNA enzymes, short interfering RNA, and the emerging wave of small-molecule nucleic acid-based gene-silencing strategies, Am. J. Pathol. 171 (2007) 1079–1088.

[7] E. Nordhoff, F. Kirpekar, P. Roepstorff, Mass spectrometry of nucleic acids, Mass Spectrom. Rev. 15 (1996) 76–138.

[8] P.A. Limbach, P.F. Crain, J.A. McCloskey, Characterization of oligonucleotides and nucleic acids by mass spectrometry, Curr. Opin. Biotechnol. 6 (1995) 96–102.

[9] J.A. McCloskey, A.B. Whitehill, J. Rozenski, F. Qiu, P.F. Crain, New techniques for the rapid characterization of oligonucleotides by mass spectrometry, Nucleosides Nucleotides 18 (1999) 1549–1553.

[10] C.G. Huber, H. Oberacher, Analysis of nucleic acids by on-line liquid chromatography–mass spectrometry, Mass Spectrom. Rev. 20 (2001) 310–343.

[11] U. Pieles, W. Zürcher, M. Schär, H.E. Moser, Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides, Nucleic Acids Res. 21 (1993) 3191–3196.

[12] R.P. Glover, G.M.A. Sweetman, P.B. Farmer, G.C.K. Roberts, Sequencing of oligonucleotides using high performance liquid chromatography and electrospray mass spectrometry, Rapid Commun. Mass Spectrom. 9 (1995) 897–901.

[13] K. Faulstich, K. Wörner, H. Brill, J.W. Engels, A sequencing method for RNA oligonucleotides based on mass spectrometry, Anal. Chem. 69 (1997) 4349–4353.

[14] N.I. Taranenko, V.V. Golovlev, S.L. Allman, N.V. Taraneko, C.H. Chen, J. Hong, L.Y. Chang, Matrix-assisted laser desorption/ionization for short tandem repeat loci, Rapid Commun. Mass Spectrom. 12 (1998) 412–418.

[15] H. Wu, H. Aboleneen, Improved oligonucleotide sequencing by alkaline phosphatase and exonuclease digestions with mass spectrometry, Anal. Biochem. 290 (2001) 347–352.

[16] D. Alazard, M. Filipowsky, J. Raeside, M. Clarke, M. Majlessi, J. Russell, W. Weisburg, Sequencing of production-scale synthetic oligonucleotides by enriching for coupling failures using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Anal. Biochem. 301 (2002) 57–64.

[17] P.M. Vallone, J.M. Devaney, M.A. Marino, J.M. Butler, A strategy for examining complex mixtures of deoxyoligonucleotides using ion-pair-reverse-phase high-performance liquid chromatography, matrix-assisted laser desorption ionization time-of-flight mass spectrometry, and informatics, Anal. Biochem. 304 (2002) 257–265.

[18] A.P. Null, L.M. Benson, D.C. Muddiman, Enzymatic strategies for the characterization of nucleic acids by electrospray ionization mass spectrometry, Rapid Commun. Mass Spectrom. 17 (2003) 2699–2706.

[19] J. Farand, M. Beverly, Sequence confirmation of modified oligonucleotides using chemical degradation, electrospray ionization, time-of-flight, and tandem mass spectrometry, Anal. Chem. 80 (2008) 7414–7421.

[20] J. Farand, F. Gosselin, De novo sequence determination of modified oligonucleotides, Anal. Chem. 81 (2009) 3723–3730.

[21] U. Bahr, H. Aygun, M. Karas, Sequencing of single and double stranded RNA oligonucleotides by acid hydrolysis and MALDI mass spectrometry, Anal. Chem. 81 (2009) 3173–3179.

[22] D.P. Little, R.A. Chorush, J.P. Speir, M.W. Senko, N.L. Kelleher, F.W. McLafferty, Rapid sequencing of oligonucleotides by high-resolution mass spectrometry, J. Am. Chem. Soc. 116 (1994) 4893–4897.

[23] D.P. Little, F.W. McLafferty, Sequencing 50-mer DNAs using electrospray tandem mass spectrometry and complementary fragmentation methods, J. Am. Chem. Soc. 117 (1995) 6783–6784.

[24] Z. Wang, K.X. Wan, R. Ramanathan, J.S. Taylor, M.L. Gross, Structure and fragmentation mechanisms of isomeric T-rich oligodesoxynucleotides: a comparison of four tandem mass spectrometry methods, J. Am. Soc. Mass Spectrom. 9 (1998) 683–691.

[25] A. Premstaller, C.G. Huber, 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. 15 (2001) 1053–1060.

[26] S. Schurch, E. Bernal-Mendez, C.J. Leumann, Electrospray tandem mass spectrometry of mixed-sequence RNA/DNA oligonucleotides, J. Am. Soc. Mass Spectrom. 13 (2002) 936–945.

[27] K.A. Sannes-Lowery, S.A. Hofstadler, Sequence confirmation of modified oligonucleotides using IRMPD in the external ion reservoir of an electrospray ionization Fourier transform ion cyclotron mass spectrometer, J. Am. Soc. Mass Spectrom. 14 (2003) 825–833.

[28] K. Hakansson, R.R. Hudgins, A.G. Marshall, R.A. O'Hair, Electron capture dissociation and infrared multiphoton dissociation of oligodeoxynucleotide dications, J. Am. Soc. Mass Spectrom. 14 (2003) 23–41.

[29] K.M. Keller, J.S. Brodbelt, Collisionally activated dissociation and infrared multiphoton dissociation of oligonucleotides in a quadrupole ion trap, Anal. Biochem. 326 (2004) 200–210.

[30] K.A. Kellersberger, E. Yu, G.H. Kruppa, M.M. Young, D. Fabris, Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation, Anal. Chem. 76 (2004) 2438–2445.

[31] T.Y. Huang, J. Liu, X. Liang, B.D. Hodges, S.A. McLuckey, Collision-induced dissociation of intact duplex and single-stranded siRNA anions, Anal. Chem. 80 (2008) 8501–8508.

[32] M. Taucher, U. Rieder, K. Breuker, Minimizing base loss and internal fragmentation in collisionally activated dissociation of multiply deprotonated RNA, J. Am. Soc. Mass Spectrom. 21 (2010) 278–285.

[33] J. Wu, S.A. McLuckey, Gas-phase fragmentation of oligonucleotide ions, Int. J. Mass Spectrom. 237 (2004) 197–241.

[34] S.A. McLuckey, G.J. Berkel, G.L. Glish, Tandem mass spectrometry of small, multiply charged oligonucleotides, J. Am. Soc. Mass Spectrom. 3 (1992) 60–70.

[35] S.A. McLuckey, S. Habibi-Goudarzi, Decompositions of multiply charged oligonucleotide anions, J. Am. Chem. Soc. 115 (1993) 12085–12095.

[36] G. Holzl, H. Oberacher, S. Pitsch, A. Stutz, C.G. Huber, Analysis of biological and synthetic ribonucleic acids by liquid chromatography–mass spectrometry using monolithic capillary columns, Anal. Chem. 77 (2005) 673–680.

[37] J. Ni, S.C. Pomerantz, J. Rozenski, Y. Zhang, J.A. McCloskey, Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry, Anal. Chem. 68 (1996) 1989–1999.

[38] J. Rozenski, J.A. McCloskey, SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry, J. Am. Soc. Mass Spectrom. 13 (2002) 200–203.

[39] W.T. Muhammad, D.L. Tabb, K.F. Fox, A. Fox, Automated discrimination of polymerase chain reaction products with closely related sequences by software-based detection of characteristic peaks in product ion spectra, Rapid Commun. Mass Spectrom. 17 (2003) 2755–2762.

[40] H. Oberacher, B. Wellenzohn, C.G. Huber, Comparative sequencing of nucleic acids by liquid chromatography–tandem mass spectrometry, Anal. Chem. 74 (2002) 211–218.

[41] H. Oberacher, W. Parson, P.J. Oefner, B.M. Mayr, C.G. Huber, Applicability of tandem mass spectrometry to the automated comparative sequencing of long-chain oligonucleotides, J. Am. Soc. Mass Spectrom. 15 (2004) 510–522.

[42] H. Oberacher, F. Pitterl, On the use of ESI-QqTOF-MS/MS for the comparative sequencing of nucleic acids, Biopolymers 91 (2009) 401–409.

[43] H. Oberacher, B.M. Mayr, C.G. Huber, Automated de novo sequencing of nucleic acids by liquid chromatography–tandem mass spectrometry, J. Am. Soc. Mass Spectrom. 15 (2004) 32–42.

[44] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.

[45] H. Oberacher, H. Niederstätter, B. Casetta, W. Parson, Detection of DNA sequence variations in homo- and heterozygous samples via molecular mass measurements by electrospray ionization time-of-flight mass spectrometry, Anal. Chem. 77 (2005) 4999–5008.

[46] H. Oberacher, H. Niederstatter, W. Parson, Characterization of synthetic nucleic acids by electrospray ionization quadrupole time-of-flight mass spectrometry, J. Mass Spectrom. 40 (2005) 932–945.

[47] C.G. Huber, M.R. Buchmeiser, On-line cation-exchange for suppression of adduct formation in negative-ion electrospray mass spectrometry of nucleic acids, Anal. Chem. 70 (1998) 5288–5295.

[48] H. Oberacher, P.J. Oefner, G. Hölzl, A. Premstaller, C.G. Huber, Re-sequencing of multiple single nucleotide polymorphisms by liquid chromatography–electrospray ionization mass spectrometry, Nucleic Acids Res. 30 (2002) e67.

[49] S.C. Pomerantz, J.A. Kowalak, J.A. McCloskey, Determination of oligonucleotide composition from mass spectrometrically measured molecular weight, J. Am. Soc. Mass Spectrom. 4 (1993) 204–209.

[50] D.C. Muddiman, G.A. Anderson, S.A. Hofstadler, R.D. Smith, Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry, Anal. Chem. 69 (1997) 1543–1549.