On-Line Liquid Chromatography Mass Spectrometry: A Useful Tool for the Detection of DNA Sequence Variation**

Herbert Oberacher, Peter J. Oefner, Walther Parson, and Christian G. Huber*

After completion of the human genome sequence by the end of the year 2001,[1] the discovery of small dissimilarities in the DNA sequences of different individuals, so-called polymorphisms, will gain increasing significance.[2] Information about genetic diversity facilitates valuable insights into inherited disposal to disease as well as into human origin and gene migration. The standard technique for the determination of DNA sequences is the fully automated Sanger sequencing method employing multiplexed capillary electrophoretic analysis.[1] In Sanger sequencing the whole sequence is determined, since polymorphisms between two randomly chosen chromosomes occur only at a frequency of one in 800–62000 base pairs (bp), their discovery by Sanger sequencing would require a lot of time and experimental effort simply for the determination of already known sequences. As a consequence, the development of novel and rapid screening methods for the discovery of variations in DNA sequences is of utmost importance, because they will enable a considerable reduction in the sequencing required and an increase in sample throughput.[3] Moreover, such methods are also mandatory for the quality control of complete genes, prepared by total chemical synthesis, that are gaining importance.[4]

The application of mass spectrometry (MS) to the investigation of high molecular nucleic acids was not feasible before the ionization techniques of matrix-assisted laser desorption/ionization (MALDI) and electospray ionization (ESI) were developed.[5] ESI-MS was applied successfully to the characterization of intact plasmids,[6] polymerase chain reaction (PCR) products,[7, 8] RNA,[9] and synthetic oligonucleotides[10] as well as to the genotyping of single nucleotide polymorphisms (SNPs)[8] and short tandem repeats (STRs).[11, 12] Nevertheless, the success of mass spectrometric assays largely depends on the amount and purity of the nucleic acids samples introduced into the mass spectrometer. Substantial drawbacks of the available purification methods, such as precipitation,[13] solid-phase extraction,[14] and liquid chromatography[9, 15] are the high amounts of sample required and the low recovery. Consequently, these methods are frequently not appropriate for the characterization of real samples or quantitative analysis. Recently, we demonstrated the applicability of ion-pair reversed-phase high-performance liquid chromatography (IP-RP-HPLC) utilizing monolithic, 200 μm internal-diameter capillary columns for the separation and subsequent mass-spectrometric characterization of 180 fmol of DNA restriction fragments up to a length of 267 bp.[16] Nonetheless, sample preparation by PCR typically supplies amounts of nucleic acids merely in the upper amol to lower fmol range. Moreover, the separation column had to be overloaded because of the relatively high detection limit of ESI-MS, which results in poor separation and insufficient characterization of the longer DNA fragments.

Significant improvements in the chromatographic conditions now enable for the first time the separation and subsequent on-line mass-spectrometric investigation of upper amol amounts of DNA. Figure 1 illustrates the total ion

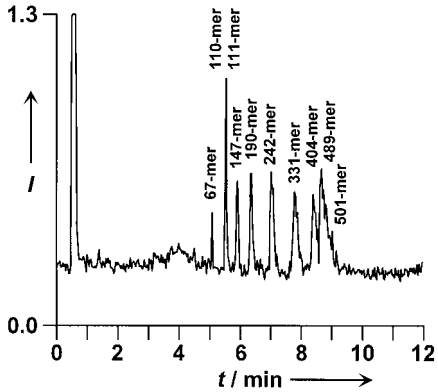

Figure 1. Chromatographic separation and mass analysis of the restriction fragments of a pUC19 MspI digest. Column: PS/DVB monolith, 60 × 0.2 mm internal diameter; mobile phase: A) 25 mM butyldimethylammonium bicarbonate, pH 8.4, B) 25 mM butyldimethylammonium bicarbonate, pH 8.4, 40% acetonitrile; linear gradient: 10–40% B in 3 min, 40–50% B in 12 min; flow rate: 3.0 μL min^{-1}; temperature: 25 °C; scan: 1000–3000 amu; electrospray voltage: 3.40 kV; sheath gas: nitrogen; sheath liquid: acetonitrile, 3.0 μL min^{-1}; sample: 753 amol digest; I = signal intensity × 10^{-6}/counts.

chromatogram of 753 amol of a digest of the cloning vector pUC19 with the restriction enzyme MspI. Because of the high separation efficiency and the avoidance of overloading the ten fragments of the digest, ranging in size from 67–501 bp, could be resolved into eight peaks. Only the 110- and 111-mer as well as the 489- and 501-mer fragments coeluted, a consequence of the small difference in their lengths. Mass spectra were extracted from the total ion chromatogram and deconvoluted by means of a computer-aided algorithm,[17] which yielded the molecular masses of the DNA fragments with high accuracy. As indicated in Table 1, mass spectra were obtained for all the fragments upon analyzing 15 fmol of material, from which the molecular masses were measured with an average accuracy of ±0.025%. This value correlates well with the mass accuracy reported for an ion-trap mass spectrometer (0.02% for a single-stranded 81-mer PCR product).[18] Even the coeluting DNA fragments were unambiguously identified

[*] Ao. Univ.-Prof. Mag. Dr. C. G. Huber, Mag. H. Oberacher
Institute of Analytical Chemistry and Radiochemistry
Leopold-Franzens University
Innrain 52a, 6020 Innsbruck (Austria)
Fax: (+43) 512-507-2767
E-mail: christian.huber@uibk.ac.at
Univ.-Doz. Mag. Dr. W. Parson
Institute of Legal Medicine, Leopold-Franzens University
Innsbruck (Austria)
Univ.-Doz. Dr. P. J. Oefner
Genome Technology Center, Stanford University
Palo Alto, CA (USA)

[**] This work was financially supported by the Austrian Science Fund (P-14133-PHY) and the National Institutes of Health (HG01932). The suggestion of Dr. Jeff van Ness from the Keck Graduate Institute of Applied Life Sciences in Claremont, CA, USA, to utilize butyldimethylamine for IP-RP-HPLC separations is gratefully acknowledged.

  1433-7851/01/4020-3828 $ 17.50+.50/0

Table 1. Masses of the DNA restriction fragments from the pUC19 cloning vector digested with MspI; n.d. = not detected.

Length [bp]	M_{calcd}	M_{found} (753 amol)	Relative deviation [%]	M_{found} (15 fmol)	Relative deviation [%]
67	41438.1	41413	−0.061	41423	−0.036
110	68005.6	67986	−0.029	67977	−0.042
111	68623.0	68664	0.060	68623	0.000
147	90883.4	90841	−0.047	90862	−0.024
190	117438.0	117413	−0.021	117424	−0.012
242	149566.0	149583	0.011	149596	0.020
331	204588.8	204520	−0.034	204611	0.011
404	249643.9	n.d.	n.d.	249750	0.043
489	302222.7	n.d.	n.d.	302391	0.056
501	309574.2	n.d.	n.d.	309567	−0.002

in the extracted mass spectra on the basis of their different masses (Table 1). After a reduction in the amount of sample to 753 amol, molecular masses were still obtained with equivalent mass accuracy for fragments ranging in size up to 331 bp (Table 1).

The attainment of low detection limits was made possible by the use of butyldimethylammonium bicarbonate as the ion-pair reagent instead of triethylammonium bicarbonate. A direct comparison of 25 mM butyldimethylammonium bicarbonate and 25 mM triethylammonium bicarbonate in 80:20 (v/v) water:acetonitrile as solvents for the ESI-MS of nucleic acids in direct-infusion experiments revealed an improvement in signal-to-noise ratio by a factor of 1.5 with butyldimethylammonium bicarbonate. This tendency correlates well with the observation that spectrum quality is enhanced by the use of less basic amines[19] (pK_a of triethylamine: 10.72, pK_a of butyldimethylamine: 10.06). Moreover, the concentration of acetonitrile necessary to elute the nucleic acids from the column is higher with butyldimethylammonium ions (16.0–17.6%) compared to triethylammonium ions (6.5–7.9% acetonitrile), a result of the higher affinity of the former to the stationary phase. The concomitant decrease in surface tension and the increase in volatility of the eluent cause an additional improvement in electrospray ionization efficiency.[20]

The applicability of IP-RP-HPLC-ESI-MS to the detection of insertions and deletions was evaluated through the characterization of restriction fragments of the pUC18 cloning vector. After digestion with HaeIII, all fragment masses were determined with a maximum mass deviation of ±0.09%. The only exception was the 257 bp fragment showing a mass deviation of −0.39%, which differed significantly from the average value for the mass deviation given above (±0.025%). The measured mass difference of 620 mass units (found: 158220, calcd: 158840.1, Figure 2) corresponded almost exactly to the mass of one base pair. In fact, Sanger sequencing revealed the deletion of a C–G base pair at position 184 proving that the fragment is actually a 256-mer (calcd mass: 158221.7, Figure 2). Assuming that for the detection of an insertion or deletion the mass has to be measured to ±200 mass units and that the maximum mass deviation is 0.09%, the method is generally applicable for nucleic acid fragments up to a mass of approximately 222000 (360 bp).

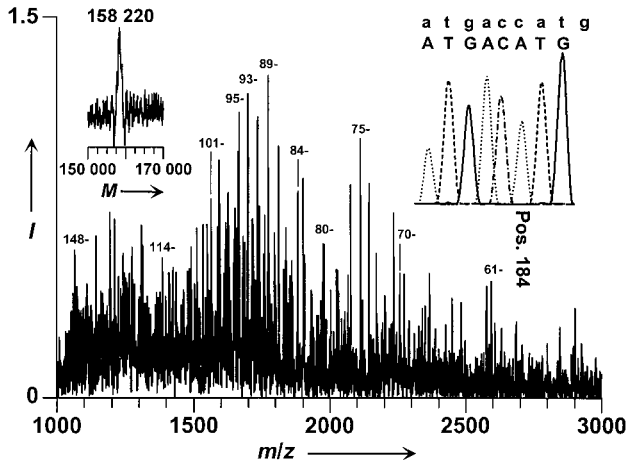

Figure 2. Mass spectrum for the detection of a deletion in a double-stranded DNA fragment from the cloning vector pUC18 by means of IP-RP-HPLC-ESI-MS. Mobile phase: A) 25 mM triethylammonium bicarbonate, pH 8.4, B) 25 mM triethylammonium bicarbonate, pH 8.4, 20% acetonitrile; linear gradient: 5–16% B in 3 min, 16–25% B in 12 min; flow rate: 3.0 μL min^{-1}; sample: 12 fmol pUC18 DNA HaeIII digest; other conditions as in Figure 1. The inset on the left side shows the deconvoluted mass spectrum, the one on the right side the sequence determined by Sanger DNA sequencing (in uppercase letters), which confirms the deletion suspected by mass spectrometry as a missing G–C base pair at position 184 in the published sequence (in lowercase letters); I = signal intensity × 10^{-3}/counts.

IP-RP-HPLC-ESI-MS analysis is also suitable for the characterization of SNPs. The distinction of the two alleles of an SNP both in homozygous and heterozygous individuals is possible by means of completely denaturing HPLC.[21] For this method short DNA double strands (<100 bp) are amplified by PCR, then completely denatured into single strands, chromatographically separated by IP-RP-HPLC, and finally detected by UV-absorbance. The retention behavior of the single strands is characteristic for each allele. However, differences in retention are typically rather subtle, and thus, unequivocal identification of the alleles can sometimes be difficult or even impossible, especially in the case of a G/C polymorphism.[21] Detection by means of ESI-MS can substantially alleviate the problem, since coeluting fragments having different masses are readily distinguished. An example for the recognition of an A/C polymorphism in a heterozygous individual is illustrated in Figure 3. Although all four single strands coeluted because of the application of a steep gradient (Figure 3a), the presence of the alleles A and C both in the forward and reverse strand (identified by the indices f and r) was effortlessly deduced from the mass spectrum (Figure 3b). Based on the high accuracy of mass measurements and a comparison of the measured and theoretical masses it is evident from Table 2 that from the four possible alleles only the alleles A and C fit the experimental data. As expected, the masses of the forward strands were shifted by 312 mass units to higher mass, a result of the non-template addition of an additional deoxyadenosine unit by Taq-polymerase.

The possibility to genotype an SNP in the form of a deletion was examined by means of a T/−T polymorphism in an 82 and 81 bp PCR product. Both alleles were unambiguously identified by the molecular masses of the single strands

  1433-7851/01/4020-3829 $ 17.50+.50/0

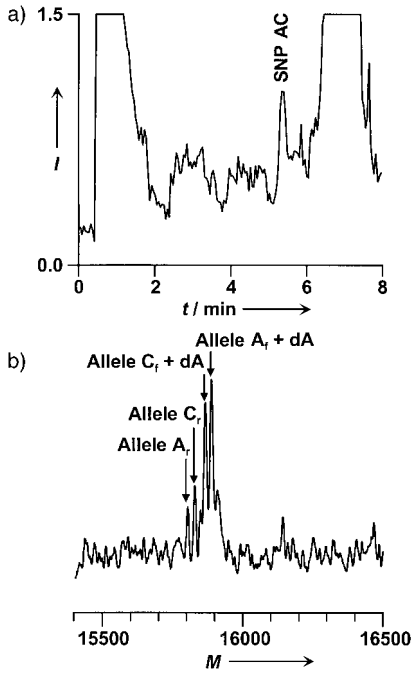

Figure 3. Identification of a single nucleotide polymorphism by IP-RP-HPLC-ESI-MS. Mobile phase: A) 25 mM triethylammonium bicarbonate, pH 8.4, B) 25 mM triethylammonium bicarbonate, pH 8.4, 20% acetonitrile; linear gradient: 10–60% B in 10 min; flow rate: 3.0 μL min^{-1}; temperature: 70 °C; scan: 500–2000 amu; sample: 500 nL PCR-amplicon; other conditions as in Figure 2; I = signal intensity $\times 10^{-6}$/counts.

Table 2. Allele identification through comparison of the eight theoretically possible masses with the measured masses.

Allele	M_{calcd}	M_{found}
$A + dA_f$	15893.54	15890
A_r	15806.42	15806
$T + dA_f$	15884.52	–
T_r	15797.40	–
$G + dA_f$	15909.54	–
G_r	15871.47	–
$C + dA_f$	15869.51	15869
C_r	15831.44	15831

(found/calcd T_f: 25277/25278.55; T_r: 25261/25262.75; $-T_f$: 24973/24974.36; $-T_r$: 24945/24949.54).

The on-line coupling to liquid chromatography represents an important step towards the application of electrospray ionization mass spectrometry for the detection of variations in DNA sequences in real molecular biological and biochemical samples. Highly accurate mass measurements enable the characterization of nucleic acids by their molecular mass as an intrinsic property of these molecules. Future developments will focus on the possibility of nucleic acid sequencing by means of fragment ion mass spectra, and it must be emphasized that the on-line separation and purification using liquid chromatography will be essential to obtain mass spectra of high quality and information content.

Experimental Section

All DNA separations were performed in monolithic, divinylbenzene cross-linked polystyrene (PS/DVD) polymer-based capillary columns (60 × 0.2 mm internal diameter).[16] Separation of nucleic acids was accomplished by IP-RP-HPLC using 25 mM triethylammonium bicarbonate or butyldimethylammonium bicarbonate as the ion-pair reagent. Reproducible gradients at a flow rate of 1–2 μL min^{-1} were formed upon splitting a primary flow of 100–200 μL min^{-1} by means of a fused silica capillary (1 m length, 375 μm external diameter, 50 μm internal diameter) and a tee-piece. Separated nucleic acids were detected by ESI-MS in an ion trap mass spectrometer (LCQ, Finnigan, San Jose, CA). A mixture of caffeine, methionyl–argininyl–phenylalanyl–alanine and Utramark 1621 was used for external mass calibration according the manufacturer's protocol. Subsequently, the parameters of the ion optics were tuned at m/z 1808.7 for maximum ion transmission in the negative ion mode with a solution of 20 pmol μL^{-1} $(dT)_{24}$ and 25 mM triethylammonium bicarbonate in 80:20 (v/v) water:acetonitrile. Acetonitrile at a flow rate of 3 μL min^{-1} was added as sheath liquid to improve detection sensitivity.[22] The digests of the cloning vectors were purchased from Sigma (St. Louis, MO). The sequence and PCR amplification of the 51 bp biallelic A/C polymorphism is described in ref.[21] that of the 82/81 bp T/–T polymorphism in the SNP database of the NCBI at http://www.ncbi.nlm.nih.gov/SNP, No. rs769382.

Received: April 17, 2001 [Z16947]

[1] N. J. Dovici, J. Zhang, *Angew. Chem.* **2000**, *112*, 4635–4640; *Angew. Chem. Int. Ed.* **2000**, *39*, 4463–4468.
[2] D. G. Wang, J. B. Fan, C. J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M. S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. J. Hudson, R. Lipshutz, M. Chee, E. S. Lander, *Science* **1998**, *280*, 1077–1082.
[3] R. G. H. Cotton, *Trends Genet.* **1997**, *13*, 43–46.
[4] K. L. Beattie, R. F. Fowler, *Nature* **1991**, *352*, 548–549.
[5] E. Nordhoff, F. Kirpekar, P. Roepstorff, *Mass Spectrom. Rev.* **1996**, *15*, 76–138.
[6] X. Cheng, D. G. Camp, Q. Wu, R. Bakhtiar, D. L. Springer, B. J. Morris, J. E. Bruce, G. A. Anderson, C. G. Edmonds, R. D. Smith, *Nucleic Acids Res.* **1996**, *24*, 2183–2189.
[7] D. C. Muddiman, A. P. Null, J. C. Hannis, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1201–1204.
[8] M. T. Krahmer, Y. A. Johnson, J. J. Walters, K. F. Fox, A. Fox, M. Nagpal, *Anal. Chem.* **1999**, *71*, 2893–2900.
[9] D. P. Little, T. W. Thannhauer, F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2318–2322.
[10] C. G. Huber, M. R. Buchmeiser, *Anal. Chem.* **1998**, *70*, 5288–5295.
[11] J. C. Hannis, D. C. Muddiman, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 954–962.
[12] H. Oberacher, W. Parson, C. G. Huber, *Anal. Chem.* **2001**, *73*, in press.
[13] N. Portier, A. Van Dorsselaer, Y. Cordier, O. Roch, R. Bischoff, *Nucleic Acids Res.* **1994**, *22*, 3895–3903.
[14] A. Deroussent, J.-P. Le Caer, A. Gouyette, *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1–4.
[15] A. Apffel, J. A. Chakel, S. Fischer, K. Lichtenwalter, W. S. Hancock, *J. Chromatogr. A* **1997**, *777*, 3–21.
[16] A. Premstaller, H. Oberacher, C. G. Huber, *Anal. Chem.* **2000**, *72*, 4386–4393.
[17] M. Mann, C. K. Meng, J. B. Fenn, *Anal. Chem.* **1989**, *61*, 1702–1708.
[18] S. Hahner, A. Schneider, A. Ingendoh, J. Mosner, *Nucleic Acids Res.* **2000**, *28*, e82.
[19] M. J. Greig, R. H. Griffey, *Rapid Commun. Mass Spectrom.* **1995**, *9*, 97–102.
[20] K. Bleicher, E. Bayer, *Biol. Mass Spectrom.* **1994**, *23*, 320–322.
[21] P. J. Oefner, *J. Chromatogr. B* **2000**, *739*, 345–355.
[22] C. G. Huber, A. Krajete, *J. Chromatogr. A* **2000**, *870*, 413–424.

  1433-7851/01/4020-3830 $ 17.50+.50/0