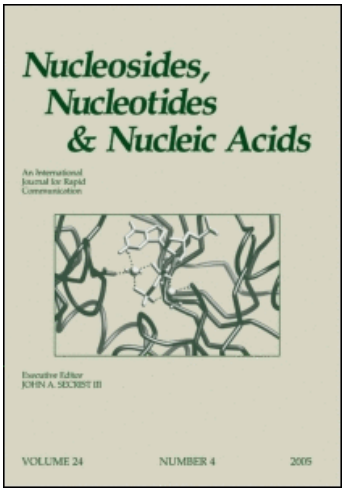


**Nucleosides, Nucleotides and Nucleic Acids**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

# Dna Analysis by Mass Spectrometry: Applications in Dna Sequencing And Dna Diagnostics

H. Köster[a]; D. van den Boom[a]; A. Braun[b]; A. Jacob[a]; C. Jurinke[a]; D. P. Little[b]; K. Tang[b]
[a] Faculty of Chemistry, Department of Biochemistry and Molecular Biology, University of Hamburg, Hamburg, Germany [b] Sequenom, Inc., San Diego, CA, USA

**To cite this Article** Köster, H. , Boom, D. van den , Braun, A. , Jacob, A. , Jurinke, C. , Little, D. P. and Tang, K.(1997) 'Dna Analysis by Mass Spectrometry: Applications in Dna Sequencing And Dna Diagnostics', Nucleosides, Nucleotides and Nucleic Acids, 16: 5, 563 — 571
**To link to this Article: DOI:** 10.1080/07328319708002917
**URL:** http://dx.doi.org/10.1080/07328319708002917

NUCLEOSIDES & NUCLEOTIDES, 16(5&6), 563-571 (1997)

# DNA ANALYSIS BY MASS SPECTROMETRY: APPLICATIONS IN DNA SEQUENCING AND DNA DIAGNOSTICS

H. Köster*(1), D. van den Boom (1), A. Braun (2), A. Jacob (1), C. Jurinke (1), D. P. Little (2) and K. Tang (2)

(1) Faculty of Chemistry, Department of Biochemistry and Molecular Biology, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany
(2) Sequenom, Inc. 11555 Sorrento Valley Road, San Diego, CA 92121, USA



ABSTRACT: Use of mass spectrometry to detect PCR amplified DNA from individuals infected with hepatitis B virus, to derive DNA sequence information through exo-nuclease based degradation or Sanger sequencing and a new format (PROBE) for the efficient determination of mutations is described.

The most frequent way to analyse DNA today is to detect the nucleic molecule by the introduction of fluorescent[1] or chemiluminescent labels[2] and separation of fragment mixtures through gel electrophoresis[3] or using a DNA array format[4]. Fluorescent and chemiluminescent labels especially when coupled with an enzymatic signal amplification step[5] allow a very sensitive detection but have the disadvantage of low information content i. e. the signals so obtained do not supply any specific information regarding the identity of the molecule. Gel electrophoretic methods are hampered by interactions between the analyte molecule and the gel matrix, the influence of secondary structures which in some cases lead to ambiguous or uninterpretable results and additionally are slow, labor-intensive and difficult to automate. DNA array formats used in sequencing-by-hybridization (SBH) applications[6] have the enormous advantage to allow parallel processing of many samples simultaneously and thereby have the potential for very high sample throughput; the drawback here is that sequence information can only be drawn

* to whom correspondence should be addressed

from hybridization events detected by a positive fluorescent signal, i.e. no direct control is possible regarding the identity of the molecule attached through base-pairing. Since a compromise has to be made with respect to the stringency for the hybridization conditions mispairing and the attachment of more than one fragment to the same site on the array cannot be detected. Some of these principle problems linked to SBH have been overcome recently by switching from single-stranded to partially double-stranded probes on the array[8]. This format could be changed in a way that by capturing a target DNA through the single-stranded overhang a primer template system is created suitable for Sanger sequencing[8,9]. Reading Sanger ladders off the DNA array ("DNA Chip") by employing gel electrophoretic separation and fluorescent detection seems to be possible in principle but technically certainly very cumbersome. In contrast detection of DNA fragment mixtures directly off a solid surface is offering a very attractive alternative[10]: Signal acquisition is fast (in the sub millisecond range) and the molecular weight obtained per fragment in the mixture is a very informative electronic signal. The combination of the parallel processing and the two dimensional separation capability of the DNA chip together with the speed and accuracy of DNA fragment detection through the determination of the respective molecular weights is creating a very powerful system for the analysis of nucleic acids.



A simple application of DNA analysis by mass spectrometry is the detection of hepatitis B virus DNA in blood samples e.g. prior to blood transfusions. In a preliminary report we have demonstrated the principle feasibility to detect hepatitis B virus DNA in blood samples[11]. It is standard practice in a laboratory for clinical microbiology to use immunlogical, dot blot and PCR methods to search for virus contaminations in blood. Patients which were negative in the surface antigen of the virus ($HB_sAg$), negative in a virus specific PCR reaction and also negative in a dot blot assay were classified as HBV negative even when antibodies against the core antigen (anti-$HB_c$) were still present. A patient was classified as cured from a hepatits B virus infection since the presence of anti-$HB_c$ antibodies is believed to be a memory effect of the immune response. It is remarkable that, by using MALDI-TOF mass spectrometry and a nested PCR reaction, we were able to detect unambiguously hepatitis B virus DNA in about 50% of samples from a group of 21 patients having hepatitis B core antibodies but no other detectable

viral markers in the blood. This means that those "cured" patients are potentially still infectious. FIGURE 1 summarizes some of the results[12]. A 67 bp fragment (theoretical molecular mass: 20,839.6 Da) within the core protein coding region of the virus DNA was amplified by nested PCR from a 269 bp first PCR product using standard PCR conditions (*Taq* DNA polymerase, 20 cycles), conditioned employing streptavidin coated magnetic beads and analyzed on a Finnigan MAT Vision 2000 mass spectrometer (measurements were carried out in positive ion reflectron mode using a nitrogen laser and 3-hydroxypicolinic acid as matrix). FIG. 1A shows the analysis of a sample from a patient cured from a hepatitis; it clearly demonstrates the presence of hepatitis B DNA (molecular mass found: 20,837.7) despite a negative $HB_s$-Ag, negative first PCR and a negative dot blot; anti-$HB_c$ was, however, positive. FIG. 1B displays the result from a patient who had no hepatitis B but a hepatitis C virus infection. As expected no 67 bp product is formed under otherwise identical conditions. FIG. 1C depicts the results of a particular interesting case. This patient had a hepatitis B infection in 1991, was treated and determined to be cured in 1992, showing all standard assays to be negative with the exception of anti-$HB_c$. The measurements were repeated in 1995 with the same results. Nevertheless some liver blood values (e.g. transaminases) were unnormal. Analysis by mass spectrometry clearly reveals the presence of hepatitis B virus DNA.



For the detection of mutations we used ligase chain reactions and analyzed the products via mass spectrometry[13]. A very powerful method for the detection of mutations by mass spectrometry uses our PROBE (PRimer Oligo Base Extension) format[14]. It is schematically described in FIGURE 2. It can best be compared with a degenerated Sanger sequencing reaction. Since a mixture of three dNTPs with the fourth base being provided in form of a terminator (ddNTP) leads not to Sanger ladders but to only one termination product, the length of which depends on the DNA sequence between the 3'OH of the primer and the first termination event, the molecular mass of the extension products reveals either the wildtype or a mutated genotype.

The most demanding scheme for mutation detection is certainly DNA sequencing. We are using two different formats. In one approach we are using the generation of fragment ladders through exonuclease degradation. FIGURE 3 represents one time value (25 min) of the enzymatic reaction in which a 60 bp DNA obtained by PCR from the polylinker

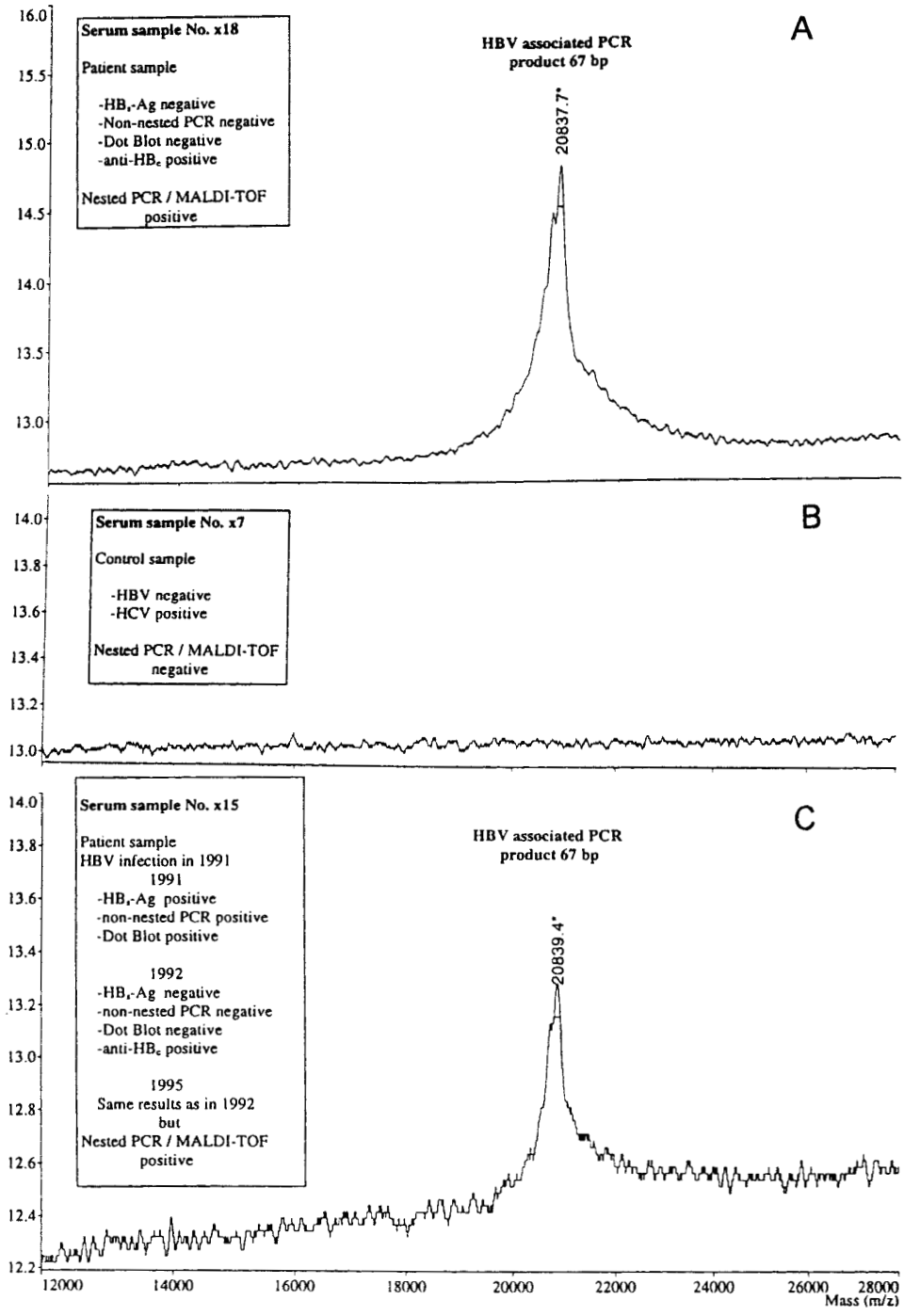


FIG. 1: Detection of hepatitis B virus DNA by MALDI-TOF mass spectrometry.
A: Serum sample from a patient without detectable viral markers but detectable anti-$HB_c$ antibodies; B: Serum sample from a hepatitis C virus infected patient; C: Serum sample from a patient with an acute HBV infection in 1991, cured in 1992 (no detectable viral markers) but with detectable antibodies against hepatitis B core protein until 1995.

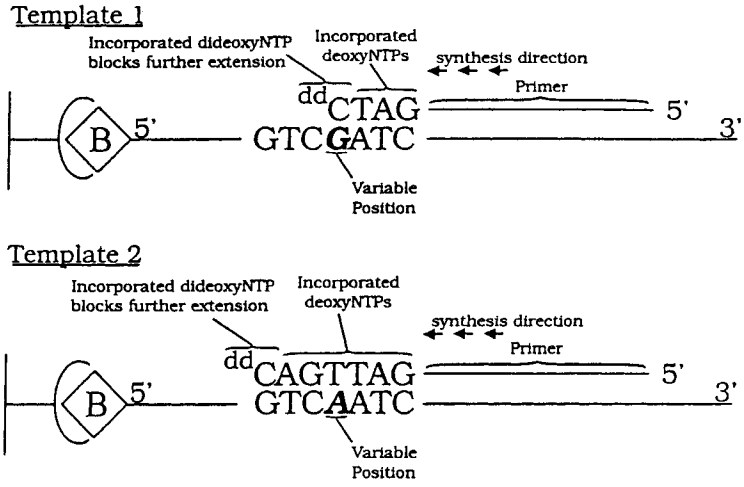


FIG. 2: Schematic representation of the PROBE mutation detection method.

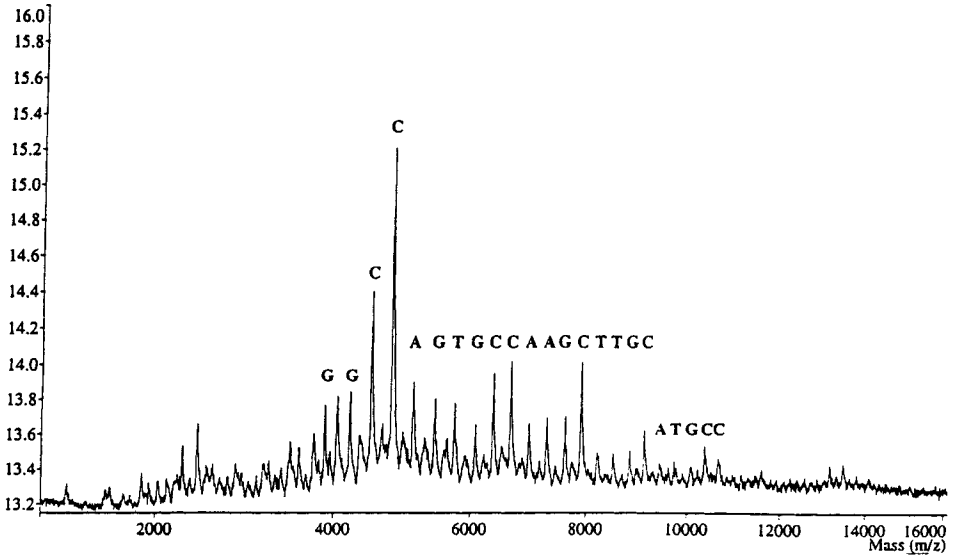


FIG. 3: Exonucleolytic degradation of an immobilized 60-mer using snake venom phosphodiesterase.

region of M13mp18 using one 5'-biotinylated primer, N7-deaza-dGTP and N7-deaza-dATP was exonucleolytically degraded. The double-stranded amplicon was immobilized on streptavidin coated magnetic beads, the antisense strand denatured and the remaining immobilized single strand treated with snake venom phosphodiesterase. 23 Bases of the sequence d(bio-GTA AAA CGA CGG CCA GTG CCA AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCC CGG) could be easily read from one time value. This is a very robust and economical method to rapidly check the puritity and identity of synthetic oligoncleotides as well as for the detection of mutations.

In a simulation of Sanger DNA sequencing reactions a mixture of synthetic oligonucleotides terminating in dT were analyzed by MALDI-TOF mass spectrometry. A mixture of a 7-mer, 10-, 11-, 18-, 19-, 20-, 24-, 26-, 33-, 37-, 38-, 42-, 46- and 50-mer were prepared in almost equal amounts (with the exception of the 33-mer and 37-mer, which were present in 20 and 40% of the amounts of the other fragments, respectively; the 7-mer apparently was lost during sample preparation). The raw spectrum of the mixture of these 14 fragments is shown in FIGURE 4.

Whereas in the low mass range peaks originating from the matrix molecules are seen, the spectrum also reveals a number of unexpected peaks. In contrast to gel electrophoretic methods the interpretation and assigment of all peaks is straightforward and summarized in TABLE 1.

Most additional peaks are dependent ion peaks from the unfragmented parent ion peaks, i.e. the two peaks around 7800 to 8000 m/z corresponding to the parent ions of the unfragmented 24- und 26-mer have a double charged dependent ion peak around 3900 to 4000 m/z. Other peaks are the result of a depurination due to the presence of the acidic matrix molecules. This can effectively be reduced by using N7-deazapurine modified DNA[15, 16]. Unfortunately deoxynucleoside triphosphates of N7-deaza-adenine and N7-deaza-guanine are poor substrates for most DNA polymerases. We recently were able, however, to completely substitute these valuable purine deoxynucleoside triphosphates by their corresponding N7-deaza analogs during PCR reactions by using the *Pfu*(exo-) DNA polymerase[17]. The advantage of combining solid phase methodologies for sample conditioning and the incorporation of N7-deazapurine desoxynucleoside triphosphates

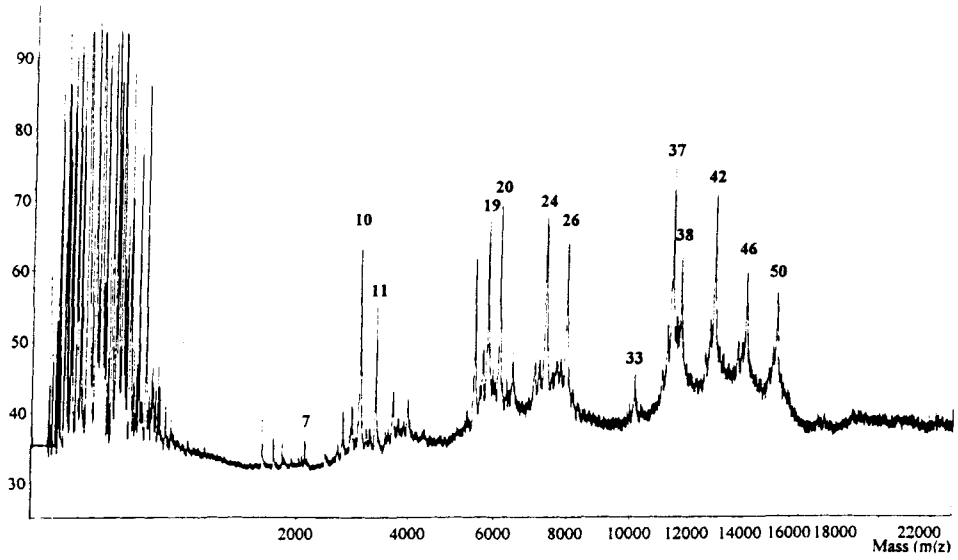


FIG. 4: MALDI-TOF spectrum of a mixture of 14 fragments ranging from 7-mer to 50-mer.



TABLE 1: Interpretation and assigment of peaks corresponding to the 14 fragment mixture of FIG. 4.

| | Sequence | $[M+H]^+_{cal.}$ | $[M+H]^+_{peak}$ | Remarks |
|---|---|---|---|---|
| | | | 1514.0 | $[10mer+2H]^{2+}$ |
| | | | 1665.8 | $[11mer+2H]^{2+}$ |
| | | | 1791.4 | $[7mer-329.6+H]^+$ |
| 7 mer | 5′-TAACGGT | 2121.4 | 2121.0 | Δ = −0.4 |
| | | | 2738.5 | $[19mer-313.4+2H]^{2+}$ |
| | | | 2894.5 | $[19mer+2H]^{2+}$ |
| 10 mer | 5′-TAACGGTCAT | 3028.0 | 3027.9 | Δ = −0.1 |
| 11 mer | 5′-TAACGGTCATT | 3332.2 | 3332.3 | Δ = +0.1 |
| | | | 3670.3 | $[24mer+2H]^{2+}$ |
| | | | 3977.7 | $[26mer+2H]^{2+}$ |
| | | | 5475.6 | $[19mer-313.4+H]^+$ |
| 19 mer | 5′-TAACGGTCATTACGGCCAT | 5788.8 | 5789.0 | Δ = +0.2 |
| 20 mer | 5′-TAACGGTCATTACGGCCATT | 6093.0 | 6092.8 | Δ = −0.2 |
| | | | 6444.6 | $[42mer+2H]^{2+}$ |
| 24 mer | 5′-TAACGGTCATTACGGCCATTGACT | 7328.8 | 7339.0 | Δ = +10.2 |
| 26 mer | 5′-TAACGGTCATTACGGCCATTGACTGT | 7962.2 | 7957.5 | Δ = −4.7 |
| 33 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCT | 10129.6 | 10129.7 | Δ = +0.1 |
| 37 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCTGCAT | 11365.4 | 11365.3 | Δ = −0.1 |
| 38 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCTGCATT | 11669.6 | 11669.5 | Δ = −0.1 |
| 42 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCTGCATTACAT | 12889.4 | 12889.2 | Δ = −0.2 |
| 46 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCTGCATTACATGACT | 14125.2 | 14122.7 | Δ = −2.5 |
| 50 mer | 5′-TAACGGTCATTACGGCCATTGACTGTAGGACCTGCATTACATGACTAGCT | 15361.0 | 15360.9 | Δ = −0.1 |

into enzymatically generated DNA lead to significant improvements in generating DNA sequence information by mass spectrometry[18]. Since this interdisciplinary field is still underdeveloped the potential for further improvements is large.

## REFERENCES


1. Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughs, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; and Hood, L. E. *Nature* **1986**, *321*, 674-679.
2. Beck, S.; and Köster, H. *Anal. Chem.* **1990**, *62*, 2258-2270.
3. Hunkapiller, T.; Kaiser, R. J.; Koop, B. F.; Hood, L. *Science* **1991**, *254*, 59-67.
4. Drmanac, R.; Drmanac, S.; Strezoska, Z.; Paunesku, T.; Labat, I.; Zeremski, M. et al. *Science* **1993**, *260*, 1649-1652.
5. Köster, H.; Beck, S.; Coull, J. M.; Dunne, T.; Gildae, B. D.; Kissinger, C.; and O'Keefe, T. *Nucleic Acids Res. Symposium Ser.* **1991**, *No 24*, 318-322.
6. Pease, C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. B.; and Fodor, S. P. A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5022-5026.
7. Broude, N. E.; Sano, T.; Smith, C. L.; and Cantor, C.R. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3072-3076.
8. Fu, D.; Broude, N. E.; Köster, H.; Smith, C. L.; and Cantor, C. R. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10162-10166.
9. Fu, D.; Broude, N. E.; Köster, H.; Smith, C. L.; and Cantor, C.R. *Genet. Anal.* **1996**, *12*, 137-142.
10. Tang, K.; Fu, D.; Kötter, S.; Cotter, R. J.; Cantor, C. R.; and Köster, H. *Nucleic Acids Res.* **1995**, *23*, 3126-3131.
11. Jurinke, C.; Zöllner, B.; Feucht, H.-H.; Jacob, A.; Kirchhübel, J.; Lüchow, A.; van den Boom, D.; Laufs, R.; and Köster, H. *Genet. Anal.* **1996**, *13*, 67-71.
12. Jurinke, C.; Zöllner, B.; Feucht, H.-H.; van den Boom, D.; Polywka, S.; Laufs, R.; and Köster, H. in preparation.
13. Jurinke, C.; van den Boom, D.; Jacob, A.; Tang, K.; Wörl, R.; and Köster, H. *Anal. Biochem.* **1996**, *237*, 174-181.
14. Braun, A.; Little, D. P.; and Köster, H. **1997** *Clin. Chem.* in press.

15. Schneider, K.; and Chait, B. *Nucleic Acids Res.* **1995**, *23*, 1570.

16. Kirpekar, F.; Nordhoff, E.; Kristiansen, K.; Roepstorff, P.; Hahner, S.; and Hillenkamp, F. *Rapid Comm. Mass Spectrom.* **1995**, *9*, 525-531.

17. Siegert, C.; Jacob, A.; and Köster, H. *Anal. Biochem.* **1996**, *243*, 55-65.

18. Köster, H.; Tang, K.; Fu, D.-J.; Braun, A.; van den Boom, D.; Smith, C. L.; Cotter, R. J.; and Cantor, C. R. *Nature Biotech.* **1996**, *14*, 1123-1128.