



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Unraveling virus identity by detection of depleted probes with capillary electrophoresis

Chin-Yu Wang^b, Tzu-Hung Hsiao^a, Liang-Hui Chu^a, Yi-Ling Lin^a, Jau-Ling Huang^c, Chung-Hsuan Chen^{b,*}, Konan Peck^{a,**}

^a Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan

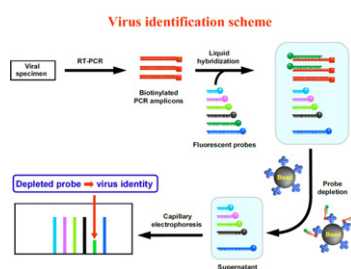
^b The Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan

^c Department of Bioscience Technology, Chang Jung Christian University, Tainan 71101, Taiwan

HIGHLIGHTS

- ▶ The probe depletion approach allowed discriminating from one base for virus identity.
- ▶ The probe depletion method abrogating the need for fluorophore labeled in capillary.
- ▶ The probe depletion approach is applicable to both RNA and DNA viruses.
- ▶ The working principle may be applied in miniaturized portable devices in the future.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 6 March 2012

Received in revised form 26 April 2012

Accepted 28 April 2012

Available online 10 May 2012

Keywords:

Virus identification
Capillary electrophoresis
Liquid hybridization
Signal processing

ABSTRACT

With the emergence of new viral infections and pandemics, there is a need to develop faster methods to unravel the virus identities in a large number of clinical samples. This report describes a virus identification method featuring high throughput, high resolution, and high sensitivity detection of viruses. Identification of virus is based on liquid hybridization of different lengths of virus-specific probes to their corresponding viruses. The probes bound to target sequences are removed by a biotin-streptavidin pull-down mechanism and the supernatant is analyzed by capillary electrophoresis. The probes depleted from the sample appear as diminished peaks in the electropherograms and the remaining probes serve as calibrators to align peaks in different capillaries. The virus identities are unraveled by a signal processing and peak detection algorithm developed in-house. Nine viruses were used in the study to demonstrate how the system works to unravel the virus identity in single and double virus infections. With properly designed probes, the system is able to distinguish closely related viruses. The system takes advantage of the high resolution feature of capillary electrophoresis to resolve probes that differ by length. The method may facilitate virus identity screen from more candidate viruses with an automated 4-color DNA sequencer.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Viral infections have remained major health problems worldwide in spite of all the medical advances in the past century. New viral infections continue to emerge in addition to the existing ones. Consequently, detection and identification of viral pathogens is of critical importance in disease control. Traditionally, virus identification is done by cultivation or serological methods [1,2] in clinical

* Corresponding author. Tel.: +886 2 27871200; fax: +886 2 27899931/27899923.

** Corresponding author. Fax: +886 2 27858594.

E-mail addresses: winschen@gate.sinica.edu.tw (C.-H. Chen), konan@ibms.sinica.edu.tw (K. Peck).

Virus identification scheme

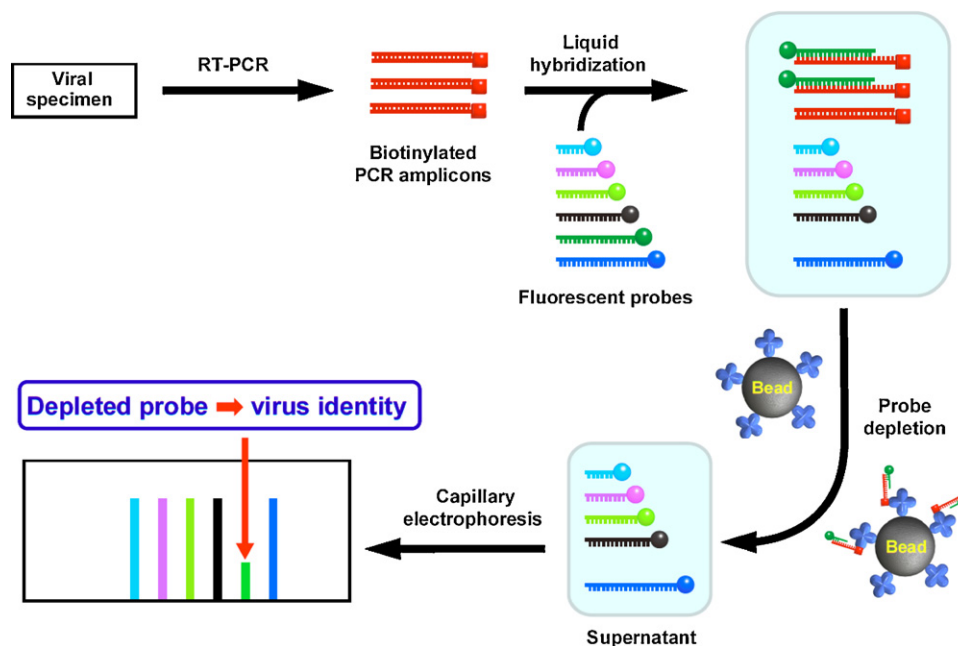


Fig. 1. A schematic diagram illustrating the strategy for rapid identification of multiple viruses.

laboratories. Most methods for viral disease diagnosis are geared toward detecting the presence of a particular virus. When a symptomatic patient of unknown etiology is presented, a number of different methods are often executed in parallel to reach a final diagnosis.

Quantitative real-time PCR (qPCR) is a very sensitive and specific method for detecting target viruses but is not suitable for unraveling the identity of an unknown virus in the sample. On the other hand, DNA microarrays with probes capable of interrogating hundreds of viruses have been developed to unravel the identities of viruses in a sample [3–5]. The DNA microarray approach is based on the slow and low efficiency solid phase hybridization process. Unraveling the virus identities for a large number of samples by DNA microarray is still a costly and time consuming approach. There is a growing interest in making virus identification faster, less costly, and more automated.

Through years of development, capillary electrophoresis is an established method for biomolecule analysis [6–8] and an emerging method for clinical applications [9,10]. With laser induced fluorescence detection method, capillary gel electrophoresis based DNA sequencer has reached the detection limit of 10^{-21} mol of DNA molecules [11] and single base resolution.

We have developed a rapid method for unraveling virus identity by using capillary gel electrophoresis based DNA sequencer. The working principle of the method is shown in Fig. 1. RT-PCR amplification with multiple primer pairs for virus genetic element amplification is employed to enrich viral DNA targets and liquid hybridization facilitates efficient binding of the probes to their cognate viral targets. The probes corresponding to the genetic elements of viruses present in the specimen bind the viral DNA targets and are retained on magnetic beads by a biotin–streptavidin pull-down mechanism. With capillary electrophoresis separation, the identities of the viruses in the specimens are unraveled by the retention time of the diminished peaks in the electropherogram. The high resolution DNA sequencer allows discrimination of probes differing by one base in length to facilitate identification of the pathogenic viruses from a multitude of candidate viruses. In order to automatically detect the diminished peaks to unravel the virus identities, an

integrated signal processing algorithm is developed to align peaks, measure peak areas, and determine the identity of the viruses. This technical note demonstrates successful implementation of the above strategy with 9 viral isolates using widely available capillary electrophoresis instrument.

2. Experimental

2.1. Materials and reagents

Nine virus isolates, SARS coronavirus (SARS), Dengue type 1–4 viruses (DENV1, DENV2, DENV3, DENV4), Influenza A virus (FluA), Human adenovirus C serotype 5 virus (HAdV-C), Japanese encephalitis RP-9 virus (JEVRP-9), and transmissible gastroenteritis coronavirus (TGEV) were used in the study. The PCR primers and 6-FAM fluorescence labeled probes were synthesized by Prisma Biotech Corporation (Taipei, Taiwan). Streptavidin coated magnetic beads with particle size of $0.86 \mu\text{m}$ were from Merck (Darmstadt, Germany). The Platinum *Taq* DNA polymerase, RNaseOUT™, and Superscript III reverse transcriptase were from Invitrogen (Carlsbad, CA). The Microcon molecular weight cut-off spun column (YM-10) was from Millipore (Billerica, MA). The electrophoretic sieving medium POP-5 gel and 3700 running buffer ($10\times$ with EDTA) were from ABI (Applied Biosystems, Foster City, CA).

2.2. Design of primers and probes

To design virus specific primers and probes, fully sequenced viral genomes covering 46 viral families, 160 genera, and 1120 species were downloaded from the GenBank database (Release 152). PCR primers were designed by the Primer 3 program for each individual virus and then aligned with the 9 viral genomes to avoid cross hybridizations. Viral species-specific probes were designed with the previously developed algorithms [4]. The primers and probes were further analyzed by the established rules [12] to ensure the primers and probes do not share sequence similarities to result in cross hybridization between them. Totals of 15 probes for 15 different viruses with an incremental 1–3 base length difference

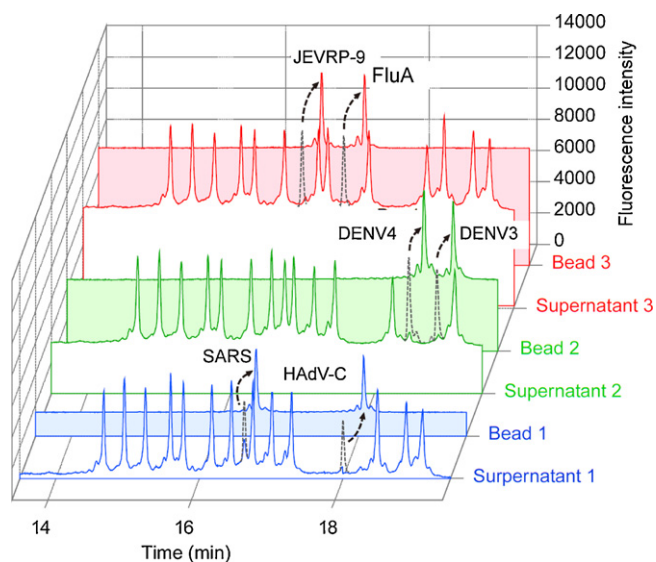


Fig. 2. Electropherograms of both the supernatant and the eluate from beads in solutions of liquid hybridization followed by magnetic bead pull-down of the hybrids. Three sample sets each containing two viruses are shown in this figure. The vanished peaks in the supernatant electropherograms are indicated by dashed peaks and lines to point to the peaks that appear in the complementary bead electropherograms.

between each probe were used in the study. The primer and probe sequences are listed in Tables S1 and S2 in the Supplementary data.

2.3. One-step reverse transcription PCR

The one-step reverse transcription PCR (RT-PCR) reaction was done with forward and biotinylated reverse primers. Each reaction mixture (100 μ L) contained: 1 μ L viral RNA derived from ca. 10^4 to 10^5 plaque-forming units (pfu) of viruses, 20 μ L of 5 μ M primers, 10 μ L of $10\times$ PCR buffer, 2 μ L of 10 mM dNTP, 5 μ L of 50 mM $MgCl_2$, 20 μ L of 100 mM betainine, 0.25 μ L of 100 mM DTT, 1.0 μ L of hot start *Taq* DNA polymerase, 8 unit of RNaseOUTTM, 4 μ L of DMSO, and 0.25 μ L of Superscript III reverse transcriptase. The one-step RT-PCR was carried out with a PTC-225 thermal cycler (MJ Research, Waltham, MA) with the reaction condition of 55 $^\circ$ C for 30 min, 95 $^\circ$ C for 6 min, and followed by 40 cycles of 95 $^\circ$ C for 20 s, 58 $^\circ$ C for 30 s, and 72 $^\circ$ C for 20 s.

2.4. Liquid hybridization, probe depletion, and detection

After the one-step RT-PCR, the biotinylated amplicons were denatured at 95 $^\circ$ C for 10 min and immediately chilled on ice for 5 min. The liquid hybridization was carried out by adding 15 different 6-FAM labeled virus probes to the biotinylated amplicons and incubated at 52 $^\circ$ C in a rotating oven for 30 min. The total volume of liquid hybridization was 60 μ L and the concentration of each 6-FAM labeled virus probe was 1 nM. After liquid hybridization, the streptavidin coated magnetic beads were added to pull down the hybrids at room temperature for 30 min. The beads were washed with $3\times$ SSC at 52 $^\circ$ C for two times. The supernatant was desalted with a spun column and had a final volume of 10 μ L. To strip the probes off the hybrids retained on the beads, the beads were washed with formamide at 95 $^\circ$ C. The capillary electrophoretic separation was performed with an Applied Biosystems 3100 genetic analyzer. Before electrokinetic injection, Hi-Di formamide solution (Applied Biosystems, Foster City, CA) was added to the solution to a final volume of 60 μ L. The capillaries were pre-run for 5 min before the samples were introduced by electrokinetic injection for 15 s at 15 kV. The electrophoretic separations were run for 20 min with the electric field strength set at 250 V cm^{-1} . The entire process starting

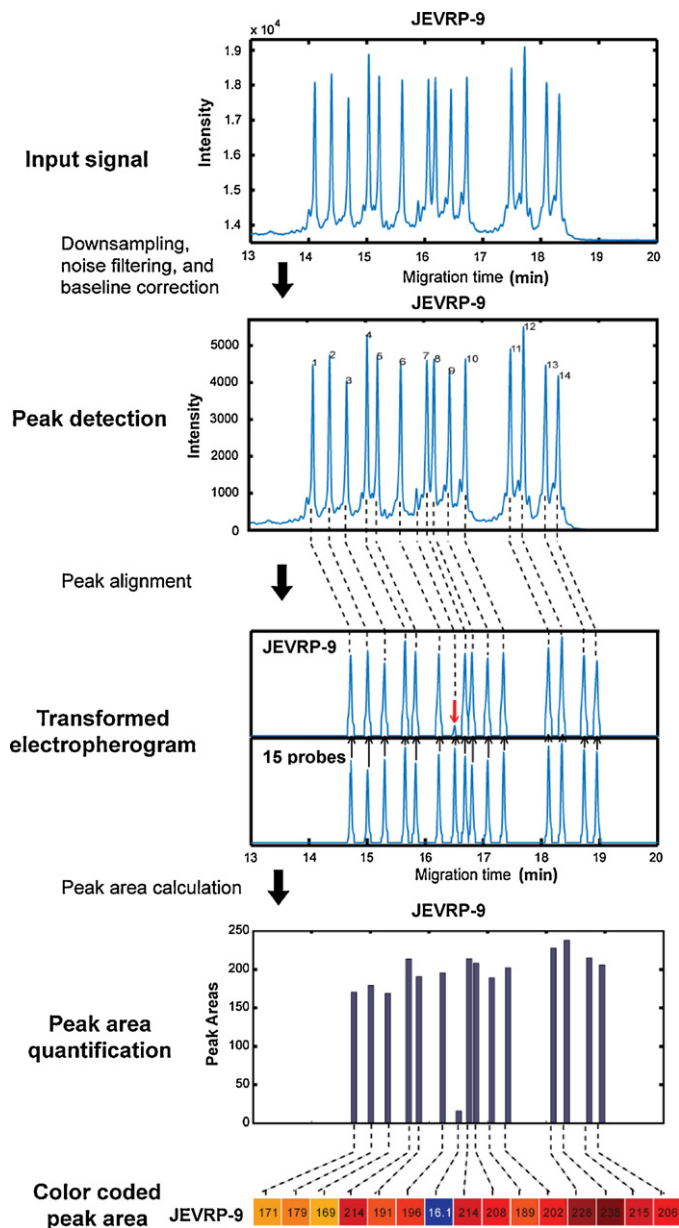


Fig. 3. A flowchart of peak detection and virus identification process. The procedural pipeline starts from raw signal output obtained from capillary electrophoresis instrument through signal processing algorithms to yield the identity of the virus in the sample.

from one-step RT-PCR to completion of capillary electrophoresis analysis was done in 6 h.

3. Results and discussion

Three sets of experimental results are shown in Fig. 2. Each set consists of two samples, one contains probes left behind in the supernatant after pull-down of hybrids by magnetic beads and the other contains probes eluted from hybrids on the magnetic beads. The blue electropherogram designated 'Supernatant 1' shows that the 8th and the 12th peaks are diminished. These two peaks show up in the complementary electropherogram designated 'Bead 1'. The green electropherograms show that the 13th and the 14th probes missing in the 'Supernatant 2' show up in the 'Bead 2' electropherogram. Likewise, the diminished 7th and 10th peaks in the 'Supernatant 3' electropherogram show up in the 'Bead 3' electropherogram.

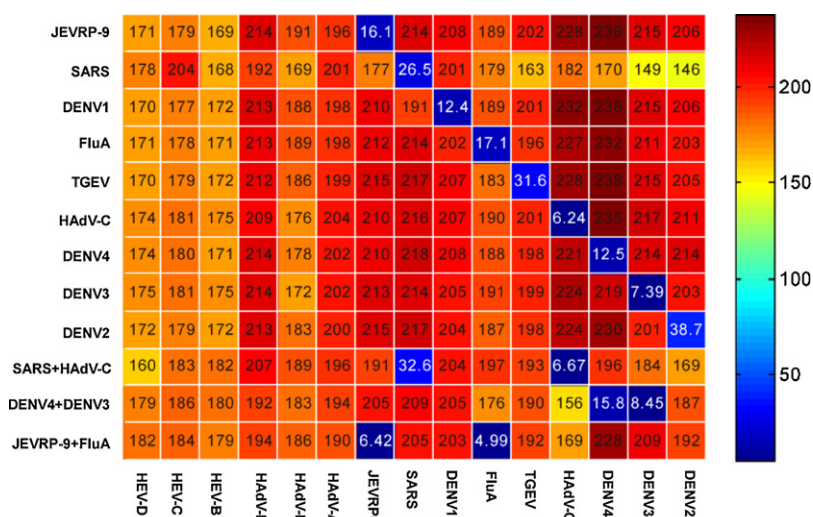


Fig. 4. Heatmap of 180 peak areas resulted from 12 viral samples and 15 virus probes. The peak area is indicated in the map and the values are pseudocolor encoded for visualization of diminished probes and the corresponding viruses.

By looking up Table S2 for virus probes which are listed by the order of probe length, we identified that the first sample set contained SARS and HAdV-C viruses, the second sample set contained DENV4 and DENV3 viruses, and the third sample set contained JEVRP-9 and FluA viruses. These results were in accordance with the viruses we spiked in the test samples. In this experiment, we used the ‘supernatant and bead’ complementary electropherogram pairs to cross exam the experimental results of the probe depletion approach and correctly identified the viruses from both the bead and supernatant electropherograms. The peaks in the supernatant electropherogram serve as size markers to unravel the identity of the diminished peaks. On the other hand, the two standalone peaks in the bead electropherograms would require additional size markers labeled with a different fluorophore and run in the same capillary to unravel the identity of the probes if the bead electropherograms were to be used without the complementary supernatant electropherogram. For this reason, the probe depletion approach provides an advantage for automated identification of viruses present in the test sample.

To implement the automatic virus identification strategy with the probe depletion approach, we developed a signal processing and data analysis algorithm. The algorithm was implemented with MATLAB software using built-in functions and commands. Fig. 3 illustrates the flow of the procedure. Shannon–Nyquist sampling theorem [13] and Savitzky–Golay low-pass filtering [14] were employed to down-sample the electropherogram. Baseline correction was employed to have better quantification of the peak area, peak height, and for peak alignment. Parameters for peak detection were defined prior to the detection and included minimum peak height, minimum peak width, and sampling frequency. After peak detection, the peaks of the electropherogram were aligned with the blank control electropherogram containing peaks of the 15 virus probes. The aligned peaks were further attended to fit the first-order Gaussian distribution. By calculating the peak areas of the fitted curves, we are able to tell which peaks are diminished and unravel the identity of the corresponding viruses (refer to Supplementary Data for more details). To facilitate visual identification, the peak area is pseudocolor encoded.

To verify the automated virus identification algorithm, we used nine samples each containing one virus and three samples each containing two viruses to test virus identification in single and double viral infections. The 12 samples were denoted by x_1 to x_{12} and the 15 virus probes were denoted by y_1 to y_{15} , respectively. Altogether, 180 peak areas are shown in Fig. 4. By

excluding the statistical outlier, i.e., the diminished peak, the software program calculated the mean and standard deviation of each electropherogram. The program identified the diminished probe and its corresponding virus by 95% confidence interval. The outputs are displayed in a numerical format as x_i and y_j coordinated values, where $i = 1, \dots, 12$ and $j = 1, \dots, 15$. The diminished peaks: (1,7), (2,8), (3,9), (4,10), (5,11), (6,12), (7,13), (8,14), (9,15), (10,8), (10,12), (11,13), (11,14), (12,7), and (12,10) are presented by the heatmap in Fig. 3. With the pseudocolor heatmap, one can easily identify the diminished peaks which appear in blue.

All 12 samples were processed with one-step RT-PCR and capillary gel electrophoresis to demonstrate that the method works for both DNA and RNA virus identification. Depending on the coverage of viruses needed for an identity screen, the number of probes may be extended by a multi-channel capillary electrophoresis instrument to have more virus probes in one run. The number of probes can be increased by two ways, by labeling the probes with up to 4 different fluorophores and/or increase the number of probes of different lengths. The high resolution of capillary gel electrophoresis allowed discriminating probes difference in length by one base, e.g., the 64mer probe for SARS and the 65mer probe for DENV1. Therefore, it permits using tens of fluorescently labeled probes for virus identification. The number may go beyond a hundred if more than one fluorophores are used. The number of sample per run is limited by the capacity of the DNA sequencer and is up to 96 samples at a time with an ABI 3730xl DNA analyzer.

It is known that samples run in different capillaries vary in migration time and signal intensity. With the aforementioned algorithm, we were able to align and adjust the retention time of the peaks. To assess the variation of peak area in different capillaries, we used a blank control and 7 different quantities of SARS virus ranging from 10 to 5000 copies of virus genome to go through the one-step RT-PCR process. The 8 samples were run in separate capillaries by the ABI 3100 genetic analyzer. The results are shown in Fig. S1 in the Supporting Information. The peak area data fit virus quantity data on logarithmic scale with a coefficient of determination, $r^2 = 0.97$. The data present in the plot were performed in triplicates to determine the standard deviation. For 10 copies, the mean peak area was 175 and the standard deviation was 14; for 50 copies, the mean was 140 and the standard deviation was 6. By 95% confidence interval, i.e., with 1.96-fold standard deviation, the method is able to detect virus down to the order of 10 copies and resolve virus quantity by order of magnitude. Based on this quantification study and the PCR conditions described in Section

2, it took 10^5 copies of SARS virus to have completely diminished peak.

4. Conclusions

We have developed a rapid virus identification method with one-step RT-PCR, liquid hybridization, probe depletion, capillary electrophoresis, and signal processing algorithm for unraveling the identity of unknown viruses in test samples. The probe depletion approach allows more probes to be used for virus identity screen by abrogating the need for an additional fluorophore labeled size markers in each capillary. The method is applicable to both RNA viruses and DNA viruses, e.g., HAdV-C. Although the method is for unraveling virus identity rather than quantification, we determined that as low as 10 copies of virus may be sufficient for identification and the method has 4 orders of magnitude in dynamic range for SARS virus. The high resolution 4-color DNA sequencer and the signal processing algorithm allow discriminating probes that differ by one base in length and may permit virus identification from more candidate viruses than this research simultaneously in a time span of less than 1 h. With wide availability of capillary electrophoresis instruments, the technology described in this report is readily applicable in clinical laboratories for virus identification. The working principle and the procedural pipeline of the technology may be applied and integrated in miniaturized devices [15] for portable uses in the future.

Acknowledgement

We acknowledge Academia Sinica and National Science Council of Taiwan for the financial support under grant number: NSC 92-2751-B-001-018-Y

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.aca.2012.04.040>.

References

- [1] T. Briese, C.G. Hatalski, S. Kliche, Y.S. Park, W.I. Lipkin, *J. Clin. Microbiol.* 33 (1995) 348–351.
- [2] L. Cao, C.M. Chan, C. Lee, S.S. Wong, K.Y. Yuen, *Infect. Immun.* 66 (1998) 966–973.
- [3] D. Wang, L. Coscoy, M. Zylberberg, P.C. Avila, H.A. Boushey, D. Ganem, J.L. DeRisi, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 15687–15692.
- [4] C.C. Chou, T.T. Lee, C.H. Chen, H.Y. Hsiao, Y.L. Lin, M.S. Ho, P.C. Yang, K. Peck, *BMC Bioinformatics* 7 (2006) 232.
- [5] O.J. Jabado, S. Conlan, P.L. Quan, J. Hui, G. Palacios, M. Hornig, T. Briese, W.I. Lipkin, *BMC Bioinformatics* 11 (2010) 354.
- [6] T.N. Chiesl, W.K. Chu, A.M. Stockton, X. Amashukeli, F. Grunthaner, R.A. Mathies, *Anal. Chem.* 81 (2009) 2537–2544.
- [7] S.D. Woodward, I. Urbanova, D. Nurok, F. Svec, *Anal. Chem.* 82 (2010) 3445–3448.
- [8] D. Brambilla, R. Verpillot, M. Taverna, L. De Kimpe, B. Le Droumaguet, J. Nicolas, M. Canovi, M. Gobbi, F. Mantegazza, M. Salmona, V. Nicolas, W. Scheper, P. Couvreur, K. Andrieux, *Anal. Chem.* 82 (2010) 10083–10089.
- [9] P.L. Chang, Y.S. Chang, J.H. Chen, S.J. Chen, H.C. Chen, *Anal. Chem.* 80 (2008) 8554–8560.
- [10] Y. Jiang, Y. Ma, *Anal. Chem.* 81 (2009) 6474–6480.
- [11] H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S.L. Wu, N.J. Dovichi, C. Fuller, *Anal. Chem.* 63 (1991) 2835–2841.
- [12] P.C. Chang, K. Peck, *Bioinformatics* 19 (2003) 1311–1317.
- [13] R.J. Marks, *II Handbook of Fourier Analysis and Its Applications*, Oxford, New York, 2009.
- [14] A. Savitzky, M.J.E. Golay, *Anal. Chem.* 36 (1964) 1627–1639.
- [15] N. Thaitrong, P. Liu, T. Briese, W.I. Lipkin, T.N. Chiesl, Y. Higa, R.A. Mathies, *Anal. Chem.* 82 (2010) 10102–10109.