Oligonucleotide Fingerprinting of Arrayed Genomic DNA Sequences Using LNA-Modified Hybridization Probes

Jian-Ping Liu[1,2,§], Mario Drungowski[1,§], Lajos Nyársik[1], Regine Schwartz[1], Hans Lehrach[1], Ralf Herwig[1] and Michal Janitz[*,1]

[1]*Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany*

[2]*Free University Berlin, Department of Biology, Chemistry and Pharmacy, D-14195 Berlin, Germany*

Abstract: Recently, we established a robust method for the detection of hybridization events using a DNA microarray deposited on a nanoporous membrane. Here, in a follow-up study, we demonstrate the performance of this approach on a larger set of LNA-modified oligoprobes and genomic DNA sequences. Twenty-six different LNA-modified 7-mer oligoprobes were hybridized to a set of 66 randomly selected human genomic DNA clones spotted on a nanoporous membrane slide. Subsequently, assay sensitivity analysis was performed using receiver operating characteristic (ROC) curves. Comparison of LNA-modified heptamers and DNA heptamers revealed that the LNA modification clearly improved sensitivity and specificity of hybridization experiment. Clustering analysis was applied in order to test practical performance of hybridization experiments with LNA-modified oligoprobes in recognizing similarity of genomic DNA sequences. Comparing the results with the theoretical sequence clusters, we conclude that the application of LNA-modified oligoprobes allows for reliable clustering of DNA sequences which reflects the underlying sequence homology. Our results show that LNA-modified oligoprobes can be used effectively to unravel sequence similarity of DNA sequences and thus, to characterize the content of unknown DNA libraries.

Keywords: LNA, hybridization, microarray, oligonucleotide fingerprinting.

INTRODUCTION

Microarrays have enabled the realization of low-cost high-throughput hybridization-based assays such as DNA-mapping, sequencing, differential gene expression analysis, polymorphism analysis, and sequencing by hybridization [1]. DNA microarrays are produced by *in situ* synthesis of oligonucleotides [2] or the immobilization of pre-fabricated DNA molecules [3]. Currently, glass slides are mainly used as support media because of their favorable optical characteristics. Since the efficiency of binding DNA to glass slides still limits the sensitivity and the dynamic range of measurements, performance is directly influenced by the amount of DNA on the glass surface. Also, DNA spots of high homogeneity are beneficial, since they simplify image analysis and enhance the accuracy of signal detection. Due to the porous structure of filter membranes, relatively large amounts of DNA can be applied, resulting in strong signal intensities and a good dynamic range. Moreover, filter arrays can be re-used, because the DNA sticks to the nylon surface without significant loss of materials between re-probing [4, 5].

Oligonucleotide fingerprinting (OFP) represents a powerful method of cDNA and genomic DNA library characterization and normalization. It is based on the analysis of arrayed libraries by sequential hybridization of 100-300 oligonucleotides with 7-12 mers in length [6]. Clones are grouped into clusters according to their hybridization fingerprints [7, 8]. The number and the size of the clusters provide information about the spectrum of expressed genes and their relative expression levels, respectively. The power of the approach is the fact that unknown cDNA libraries can be normalized to non- or low-redundant clone sets that represent a large fraction of the unknown genes in that library. These properties have led to a wide range of applications in gene expression analysis [9], gene identification [10] and large-scale genomic sequencing [11]. To further improve the specificity and reduce the total number of probes needed to get adequate sequence information, shorter oligonucleotides (e.g. 7 mers) are desirable [12], although shorter probes lead to a decrease of stability of the hybridization duplex.

Insufficient stability and poor mismatch discrimination of conventional DNA oligonucleotides have resulted in soaring interest in DNA analogues as tools for hybridization. Among the DNA mimics introduced in recent years, a new class of oligomers has become of particular interest. Singh *et al.* [13] described hybridization experiments with a novel nucleotide termed LNA (Locked Nucleic Acid), whose major distinguishing characteristic is the presence of a methylene bridge that connects the 2′-oxygen of ribose with the 4′-carbon of the ribose ring. This bridge results in a locked 3′-endo conformation, reducing the conformational flexibility of the ribose and increasing the local organization of the phosphate backbone. These molecular characteristics allow increased stability of the nucleic acid duplexes formed between LNAs and other nucleic acids. Usually, LNA/DNA duplexes have increased thermal stability (3-8°C per modified base) compared with similar duplexes formed by DNA alone [14]. Due

*Address correspondence to this author at the Max-Planck-Institute for Molecular Genetics, Fabeckstrasse 60-62, D-14195 Berlin, Germany; Tel: +49 (30) 8413 1463; Fax: +49 (30) 8413 1462;
E-mail: janitz@molgen.mpg.de

[§]The authors equally contributed to the study.

1386-2073/07 $50.00+.00

to these features, LNAs have been successfully applied in expression profiling studies [15], fluorescence *in situ* hybridization (FISH) analysis [16], RNA interference [17] and many other applications [for review see 18 and 19].

In a recent proof-of-principle study [20], we showed that modification of oligodeoxynucleotides with LNAs allows for application of heptamer oligoprobes for hybridization with arrayed genomic DNA clones. This in turn leads to reduction of probe number necessary for analysis of large clone sets as well as substantial reduction of sample/reagent consumption. In this follow-up study, we demonstrate the performance of the approach on a larger set of LNA-modified oligoprobes and genomic DNA sequences. We present the application of 26 different LNA-modified 7-mer oligoprobes for hybridization with a set of 66 randomly selected human genomic DNA clones spotted on a nanoporous membrane slide. We demonstrate the performance of the approach by sensitivity analysis using ROC curves. Furthermore, we compare LNA-modified heptamers to DNA heptamers and show that the LNA modification clearly improves sensitivity and specificity of the hybridization. In order to test the clustering performance of LNA-modified probes, we present cluster results with respect to the DNA sequences. Comparing the results with the theoretical sequence clusters we demonstrate that the application of LNA-modified oligoprobes allows for reliable clustering of DNA sequences which reflects the underlying sequence homology.

MATERIALS AND METHODS

Resources of Oligoprobes and Targets

The oligoprobes used in this investigation were randomly selected from a set of computationally generated heptamers for the human genome based on a superior algorithm with Shannon entropy as a quality criterion, as described previously [8]. Impractical oligoprobes were removed from the starting set by considering the G/C content and palindromic and other problematic sequences. As suggested previously [21], in order to achieve greater stability, most of the oligoprobes were LNA-modified at pyrimidine positions. Moreover, building of continuous stretches of LNAs was avoided for most of the probes, so that particular LNAs were separated by at least one non-modified nucleotide. Due to the difficulties of LNA incorporation, the nucleotides at the 3'-end were not subjected to modification.

DNA only and LNA-modified oligoprobes were synthesized by MWG Biotech (Ebersberg, Germany) and TIB MolBiol (Berlin, Germany). Cy5 was coupled to the 5'-terminus using amidite chemistry. The sequences of probes used in this work are shown in Table **1**. The concentration of the probe was 10 pmol/µl. Since it is difficult to precisely discriminate between true positives and unspecific binding for 7-mer oligoprobes with more than two LNA modifications [20], we used only one to three LNA modifications per LNA-modified oligoprobe.

The melting temperatures (Tm) of DNA and LNA-modified oligoprobes (listed in Table **1**) were determined using a previously described thermodynamic model [22, 23]. The model was modified for prediction of the melting temperatures of DNA/LNA mixmers (more details can be found at http://lna-tm.com). The conditions used for Tm prediction were the same as used for hybridization (100 mM salt and 0.01 µM oligoprobe).

Table 1. Individual Characteristics of Oligoprobes

Oligo ID	Sequence	GC-Content (%)	Matching Rate (%)	Tm (°C)
LNA Probes				
OP-01L	Cy5-tcAgaAg	43	37.9	6
OP-02L	Cy5-CtgaaGc	57	18.2	7
OP-03L	Cy5-aTgAgGa	43	19.7	13
OP-04L	Cy5-CtgaaGg	57	37.9	9
OP-05L	Cy5-TgcTgGg	71	45.5	25
OP-06L	Cy5-tTcctcc	57	60.6	1
OP-07L	Cy5-caggaCa	57	16.7	8
OP-08L	Cy5-TccTgct	57	16.7	6
OP-09L	Cy5-aAgTgct	43	18.2	10
OP-10L	Cy5-tCacTgt	43	18.2	8
OP-11L	Cy5-cacTgca	57	28.8	10
OP-12L	Cy5-tcCtgGa	57	37.9	9
OP-13L	Cy5-cctccTg	71	42.4	9
OP-14L	Cy5-agcTgag	57	10.6	7
OP-15L	Cy5-cTcctCc	57	15.2	12
OP-16L	Cy5-cagccTc	71	9.1	10
OP-17L	Cy5-tGCtgGt	57	18.2	16
OP-18L	Cy5-cTgCcct	86	30.3	19
OP-19L	Cy5-ttGccAa	43	28.8	9
OP-20L	Cy5-ttcTctg	43	37.9	-1
OP-21L	Cy5-agTccTc	57	4.5	7
OP-22L	Cy5-aaTgaGg	43	15.2	11
OP-23L	Cy5-aGcTcag	57	18.2	10
OP-24L	Cy5-gTcTgga	57	12.1	10
OP-25L	Cy5-gTcTggc	71	13.6	19
OP-26L	Cy5-TgagcTg	57	28.8	9
DNA Probes				
OP-02	Cy5-ctgaagc	57	18.2	3
OP-03	Cy5-atgagga	43	19.7	-7
OP-13	Cy5-cctcctg	71	42.4	4
OP-15	Cy5-ctcctcc	57	15.2	3
OP-17	Cy5-tgctggt	57	18.2	3
OP-19	Cy5-ttgccaa	43	28.8	0

Modifications with LNA bases are indicated with capital letter in the oligo sequences.

Sixty-six sequence-verified double-stranded DNA (dsDNA) clones were used with a size range between 288bp and 1719bp. All the clones were derived from the human chromosome Xq28 cosmid clone (GeneBank accession number:

AL034384). The dsDNA target sequences matched the cosmid fragment at the position 89182-137664.

Targets Preparation

Plasmid inserts were amplified as described previously [24]. Concentrations of the ethanol-precipitated PCR products were measured using a Bio-Photometer (Eppendorf, Hamburg, Germany) and adjusted to 0.1 μM. Samples were transferred into 384-well plates, which were sealed and stored at -20°C until further use for spotting.

Microarray Construction

Briefly, target DNA samples dissolved in TE buffer were spotted onto the Vivid Pall nanoporous membrane slides (Pall Corporation, East Hills, NY) using a piezoelectric pipette spotting robot (SciFlexArrayer, Scienion, Berlin, Germany). A single nozzle was used to avoid differences between pins. DNA solution with desired volume for each DNA spot was released and each DNA target was printed in quadruplicate at a 900 μm centre-to-centre spacing. In total, 1,152 spots were arrayed in a printable area of 60 mm × 20 mm. Slides were left at room temperature to dry the water and then heat-treated on a metal block at 95°C for 5 min. The DNA was crosslinked to the support by UV irradiation with a total energy of 1200 mJ in the Stratagen Crosslinker (Stratagene, La Jolla, CA). Blocking of the non-reacted surface to minimize nonspecific binding of probes to the slide was achieved by using a solution containing 3% (w/v) casein-hammerstein followed by a washing step [20]. The slides were then air-dried and stored at 4°C until hybridization.

Membrane Slide Hybridization

A total of 32 short oligoprobes, listed in Table **1**, were consecutively hybridized with the filters, as previously described [20]. A total of 200 pmol of particular Cy5-labeled oligoprobe was added to 20 mL of hybridization buffer [20] in a plastic or glass box and then denatured at 98°C for 5 min. The membrane slides were incubated in the boiling hybridization solution for 5 min with gentle shaking and then cooled down on ice. Subsequently, slides were washed with hybridization buffer at 4°C for 2 min. For the oligoprobes OP-05L, OP-16L and OP-25L washing time was extended to 5 min whereas for the OP-18L to 10 min due to the stronger affinity to target DNA. All four probes have elevated GC content (71% or 86%).

Stripping Procedure

The oligoprobe was removed by incubation of the membrane in stripping buffer [20] at 80°C for 30 min when the LNA-modified oligoprobes were used. In cases in which non-modified oligoprobes were used, a 10 min incubation in stripping buffer was applied. Subsequently, the slides submerged in hybridization buffer were cooled down to room temperature. The membranes were verified using a charge-coupled device (CCD) camera to ensure that all Cy5 fluorescence had been removed. Once the fluorescent intensity had been eliminated, dried membranes were stored at room temperature until re-use.

Microarray Image Acquisition and Quantification

For hybridization signal imaging, a 16-bit CCD-camera type CH350 (Photometrix, Tucson, AZ, USA) was used with a 512 × 512 pixel chip size cooled down to -40°C. The camera allowed the measurement of pixel values in the range from 0 to 65,535. The acquired pixel values were corrected for background and presented as arbitrary units of fluorescence intensity. The CCD-camera was operated by software that had been developed in-house. An object field of 20 × 20 mm² could be detected using this approach. The excitation light was generated by a halogen light source with an interference filter. The emission filter was placed directly on the front side of the camera objective. For excitation, a 620 nm filter (AHF, Tübingen, Germany) was used; for emission, a 670 nm filter (Andover Corporation, Salem, NH, USA) was used. For signal quantification, the images were evaluated using AlphaEaseFC Software version 4.0.0 (Alpha Innotech Corporation, San Leandro, CA, USA). The data file was imported into a custom Microsoft Excel worksheet. The values of the mean signal intensity from each spot were subtracted from the local mean background intensity. Deviating spots (e.g., high background, dust, irregularities, etc.) were filtered out from the analysis. The mean background-corrected spot intensity of each sample was used in subsequent data analysis, as described previously [25].

Data Normalization

For each hybridization with an oligoprobe, a median normalization was used, i.e. raw signals were divided by the median signal of that experiment. Replicate signals of the same genomic DNA clone in the same experiment were averaged.

Sensitivity Analysis

To test the performance of the hybridization, receiver operating characteristic (ROC) curves were computed. A ROC curve is a graphical plot of sensitivity *vs* specificity. It can also be represented by plotting the fraction of true positive (TP) *vs* the fraction of false positives (FP) signals. Sensitivity/specificity was judged with the theoretical sequence matches of the oligoprobes and the genomic DNA sequences and a continuous discrimination threshold across the signal range.

Clustering Oligonucleotide Fingerprints

For clustering the hybridization fingerprints of the 66 clones across the set of oligoprobes, a hierarchical clustering method with Pearson correlation was applied as a similarity measure and average-linkage as an update rule using J-Express Pro V 2.7 software (http://www.molmine.com). Pairwise similarity was calculated based on median-row-centered hybridization profiles with 26 LNA-modified oligoprobes.

RESULTS AND DISCUSSION

Microarray Hybridization

Sixty-six sequence-verified genomic DNA clones were randomly selected from a previously constructed library [11]. The selected clones have overlapping sequences and

can thus be grouped into different clusters. The nylon membrane was divided into 3 fields as Field-1, -2 and -3 (shown in Fig. **1A**, left panel). Different volume of PCR products produced from the randomly chosen 66 genomic DNA clones could be dispensed onto the three fields, so that we could investigate at how low concentration the hybridization signals could be detected. In this case, the three fields (from Field 1 to 3) correspond to 10 nL, 20 nL and 40 nL of ethanol-precipitated PCR products. In each field, 96 samples were spotted in quadruplicates to produce 384 spots (16 columns by 24 rows) in each field with 1,152 spots on the whole membrane.

The purified dsDNAs from the 66 genomic DNA clones with the concentration of 0.1 μM were spotted in column 1-11 (shown in Fig. **1A**). After dispensing, the nylon membrane was dried, cross-linked and blocked before being applied to the hybridization with the 26 selected 7-mer LNA-modified oligoprobes. In our experimental set-up, we found that comparing with 10 nL (field 1 in Fig. **1A**) and 40 nL (Field 3 in Fig. **1A**), when 20 nL DNA materials were dispensed (Field 2 in Fig. **1A**), reliable significant and cost-effective hybridization signals could be achieved. As shown in the images of the right panel in Fig. **1A**, the hybridization signals from the matching targets are clearly distinguishable from the background, and specific hybridization signals from different spots were observed for oligoprobes as an example OP-15L, -25L and -26L with different sequences (Table **1**).

The signal-to-noise ratios varied across the three hybridizations of OP-15L, -25L and -26L (Fig. **1A**, right panel). This was mostly due to the different hybridization performance of these three oligos with difference sequences and LNA modifications. Other signal fluctuations could occur as a result of target and array preparation procedure, hybridization process itself, and image processing. Nevertheless, the background remained homogeneous across the images (Fig. **1A**, right panel), thus not influencing subsequent data analysis.

Sensitivity of Hybridization Experiments

In order to prove the performance of the hybridization experiments with LNA-modified oligoprobes, we compared the experimentally derived signals from those genomic DNAs that contain the probe sequence (or its reverse complementary sequence) with those genomic DNAs that do not contain the probe sequences. For our study, it was important that the probe sequences could be considered approximately randomly selected, since we wanted to avoid individual probe bias resulting from unusually high (or low) matching rates. The matching rate of an oligoprobe is dependent on the length of the probe and the average length of the genomic DNAs under consideration. The practical performance of the hybridization experiment is a trade-off between probe length (shorter probes match to more sequences) and stability of the match (shorter probes show less stability). The 66 human genomic DNA sequences have a mean length of 1,322 bp (standard deviation 239 bp). Theoretically, the expected number of matches of a specific 7-mer probe within a set of N sequences of that length is equal to N*0.15, i.e. the probe will match to 15% of the clones [8]. The hit rates of the selected LNA-modified oligoprobes in our study ranged from 4.5-60.6% with a median hit rate of 19% what comes close to a randomly selected probe. Thus, we conclude that there was no sequence bias in our data set. Other individual characteristics of the LNA-modified oligoprobes are shown in Table **1**.

ROC curves are shown in Fig. **1B**. For each hybridization experiment, we computed the average signal for each genomic DNA across its four replicate values. This signal was used as a threshold to determine the amount of false positive (X-axis) and true positive (Y-axis) genomic DNAs. True positive signals were defined as signals above the threshold derived from a genomic DNA containing the probe sequence; false positive signals were those above the threshold derived from a genomic DNA that did not contain the probe sequence. Most LNA-modified oligoprobes showed a very good performance in separating signal from noise indicated by the value for the integral of the ROC curve (green line, maximum equal to one).

However, for some LNA-modified oligoprobes, the performance dropped below the 0.9 limit (OP-04L, OP-09L, OP-14L, and OP-18L). The performance measure used here was the area under the ROC curve (AUC) indicated by the green lines. In classification theory, this measure is usually known as discrimination and can be interpreted as the ability of the classifier to correctly classify an unknown sample. In our case, the AUC measured the ability of the individual oligoprobe to distinct signal from noise. The lack of performance of these oligoprobes cannot be explained straightforwardly by the individual probe statistics (Table **1**), neither by an unusually high (or low) matching rate nor any differences with respect to the melting temperature or the GC content. The discrepancy between the number of LNA modifications in probes and the hybridization specificity could also be observed. Liu, *et al.* [20], previously demonstrated that more than two LNA modifications result in decreased probe specificity. However, in the present study, the oligoprobes with three LNA modifications, as for example oligoprobes OP-03L, -05L and -17L in Fig. **1B**, were sufficiently specific. From this observation, we conclude that either two or three LNA modifications in heptamer probes will lead to improvement in hybridization specificity.

Comparison of DNA and LNA-Modified Oligoprobes

In order to demonstrate the gain in performance of separating signal from noise with respect to LNA modification of the oligoprobe, we compared the same ROC curves for selected DNA probe sequences. Fig. **2A** shows the ROC curves of six different DNA and LNA-modified probe pairs. All LNA-modified oligoprobes showed an increase in performance compared to the respective DNA sequence. Moreover, only one of the DNA probes (OP-15) exceeded the quality level (ROC curve integral, green line) of 0.9, whereas all LNA-modified oligoprobes did. A possible reason for this increase of performance is the increase in melting temperature when switching from DNA to LNA (comparison in Table **1**). On average, the increase was 11.0 °C for the six oligoprobes under consideration (range from 4 to 20°C). Higher stability of sequence matching between oligoprobe and genomic DNA sequence was also demonstrated by the reproducibility of the hybridization signals. Fig. **2B** shows the histograms of the CVs (coefficients of variation)

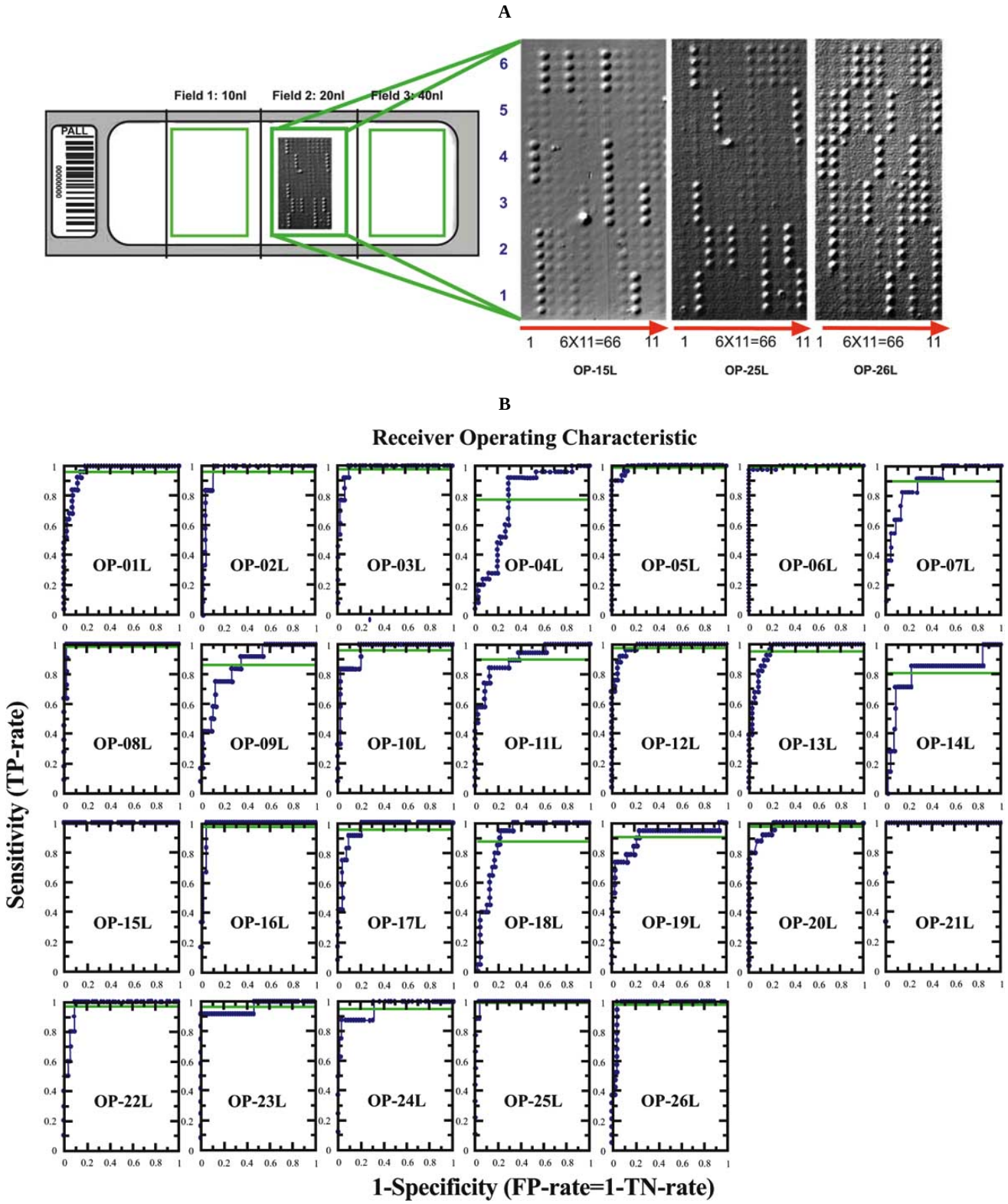
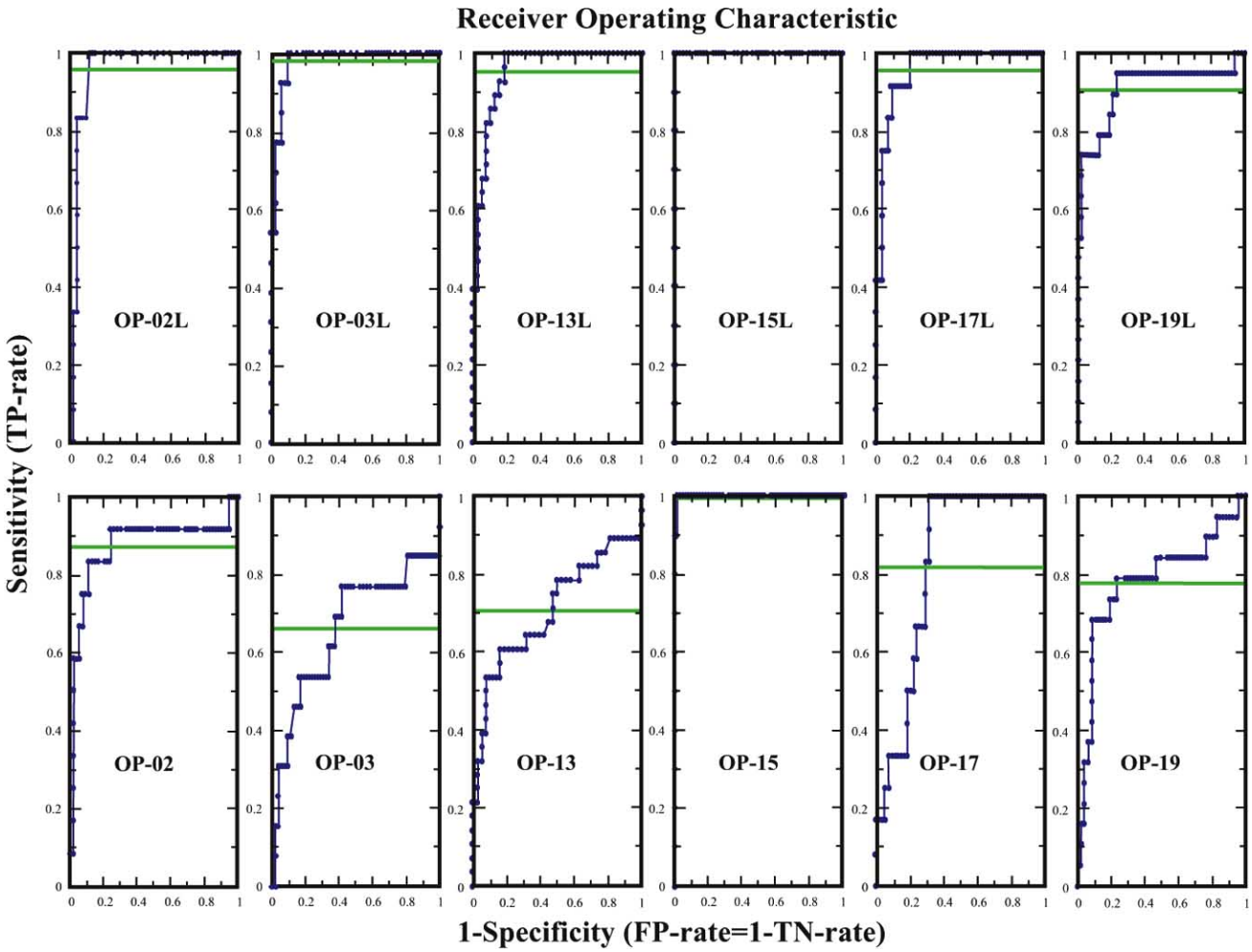


Fig. (1). A) Hybridization signal images using LNA-modified oligoprobes OP-15L, -25L and -26L for hybridization with the spotted 66 dsDNA templates. The hybridization signals from the matching targets can be clearly recognized from the background. **B**) Receiver Operating Characteristic (ROC) of all 26 LNA-modified oligoprobe hybridization experiments. X-axis shows the false positive rate, Y-axis the true positive rate. The green lines indicate the area under the ROC curve (maximum equal to one).

derived from the four replicated signals for each genomic DNA in each hybridization experiment. The blue line shows the CVs for hybridizations using all 26 LNA-modified oligoprobes, and the green histogram shows the CVs for all six DNA hybridizations. Whereas the peak of the LNA histo-

A



B

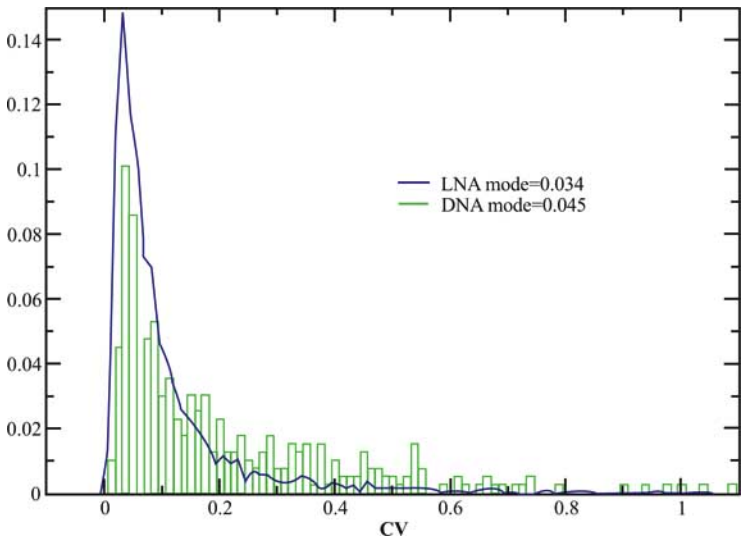
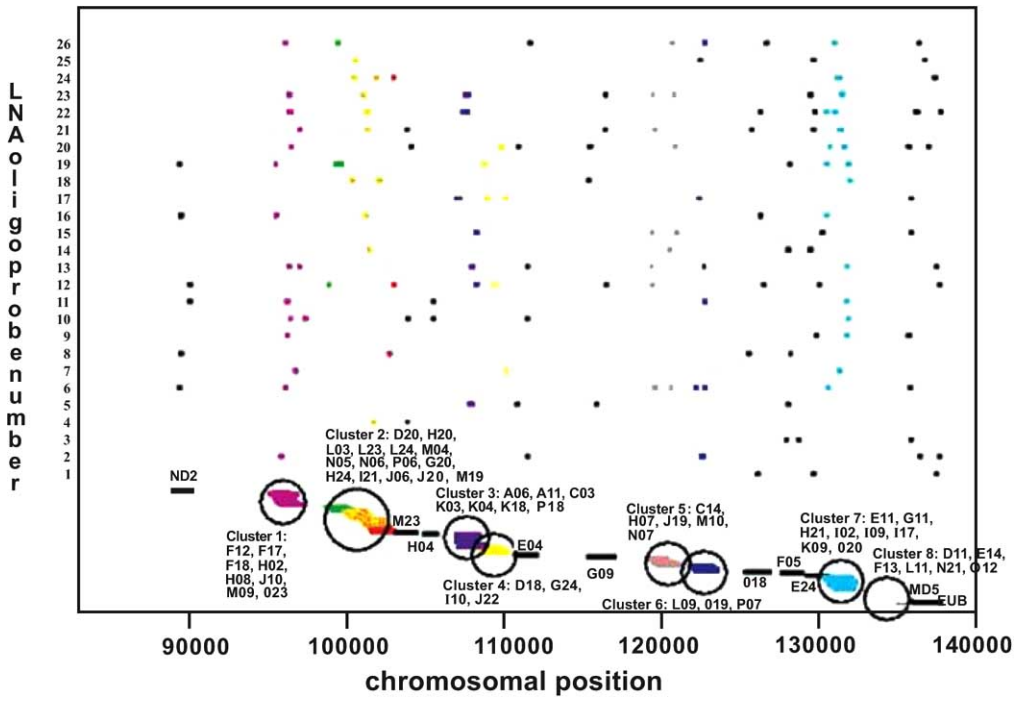


Fig. (2). A) Comparison of LNA-modified (upper panel) with DNA-oligoprobe hybridizations (lower panel) using dsDNAs as templates. The graphs show ROC curves for each single experiment. **B**) Histogram of coefficients of variations (CVs). Blue line shows the histogram of CVs derived from replicate genomic DNA signals of all hybridization experiments with LNA-modified oligoprobes, the green histogram shows the CVs derived from all DNA hybridization experiments.

A



B

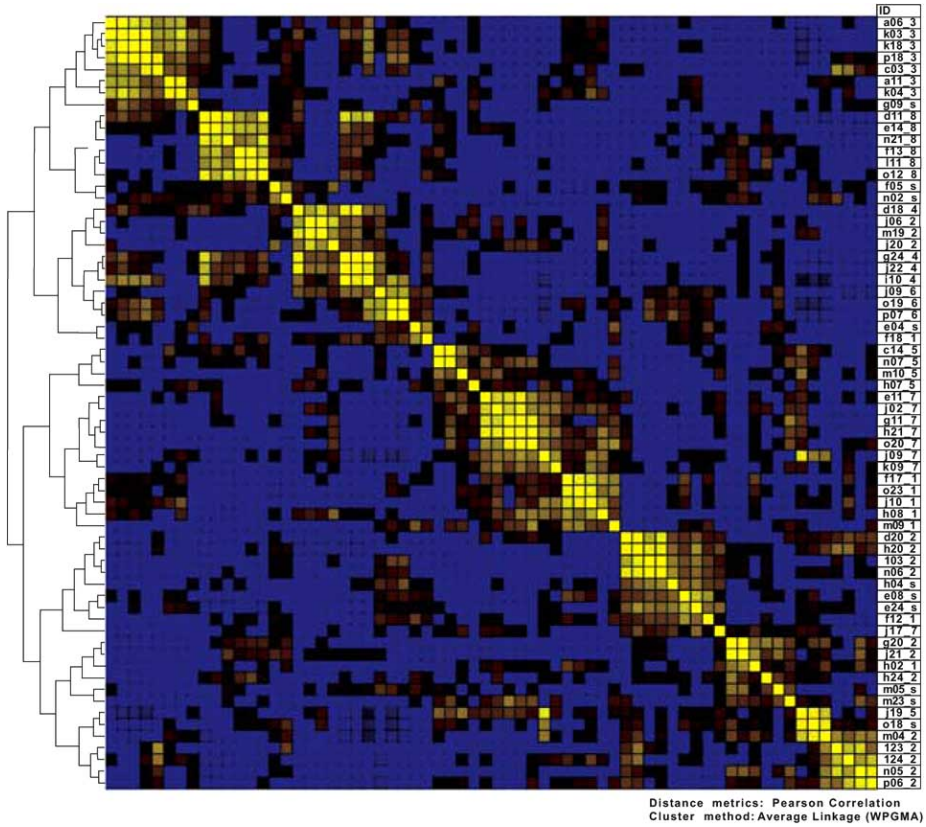


Fig. (3). A) Oligoprobe positions in the human chromosome Xq28 cosmid clone range 89182-137664 which comprises of all the 66 genomic DNA clones. Clusters of overlapping sequences are marked with circles. **B**) Hierarchical clustering of genomic DNA clones using Pearson correlation as similarity measure and average linkage as an update rule (J-Express Pro V2.7, www.molmine.com). Similarity of the genomic DNA clones is visualized with colored boxes (yellow = high similarity, blue = low similarity). The panel on the right shows the genomic DNA clone identifier and the sequence cluster they are associated with (s = singletons).

gram (0.034), and the DNA histogram (0.045) are comparable, the DNA histogram shows a much broader tail indicating a high proportion of signals with low reproducibility. Since the CVs compute the ratios of the standard deviations and the means across the replicates, they can be used to compare the reproducibility in the different hybridization experiments. CVs range from zero (perfect reproducibility) to higher values. CVs above one indicate complete irreproducibility of hybridization signals. In the hybridization experiments using the LNA-modified and DNA oligoprobes, there was no difference of the CVs when distinguishing true positive signals, i.e. CVs derived from replicate signals of a genomic DNA matching the respective probe sequences from true negative signals (data not shown), so that both samples were combined for hybridization experiments with LNA-modified and DNA oligoprobes. Thus, we conclude that the LNA approach compared to DNA leads to a higher reproducibility and stability of the experiment.

Clustering of Genomic DNA Clones

In order to test the practical performance of hybridization experiments with LNA-modified oligoprobes in recognizing similarity of genomic DNA sequences, we used a clustering approach. Clustering by DNA oligoprobe hybridizations has been applied previously in order to identify the set of genes and their abundance in large cDNA libraries [7, 26]. The ratio behind this approach is that the pattern of hybridization of oligoprobes to the cDNA reflects its sequence and can therefore be used as a 'fingerprint' for its identification. The overall distribution of LNA-modified oligoprobes (or their reverse complementary sequences) in the genomic DNA clones is given in Fig. **3A**. It may be observed that the genomic DNA sequences partition in several fragments of overlapping sequence clusters of different sizes ranging from 1 to 14 members (indicated by black circles). The goal of the clustering approach was to reconstruct this partition by the genomic DNAs fingerprints of signal intensities across the LNA-modified oligprobes. For clustering, we used a hierarchical clustering method with Pearson Correlation as a pairwise similarity measure and average linkage as an update rule (Fig. 3B) using the J-Express Pro V2.7 software package. The result shows that the oligofingerprint approach reproduced the sequence clusters very well. Separated and highly overlapping clusters, such as sequence clusters 1, 3, 4, 5, 6, 7 and 8, are recovered accurately. Sequence clusters that are more distinct, such as cluster 2, are spread in subgroups within the histogram.

Our results show that LNA-modified oligoprobes can be used effectively to unravel sequence similarity of genomic DNA sequences and thus, to characterize the content of unknown genomic DNA libraries. In practice, the number of oligoprobes is typically much smaller compared to the number of genomic DNAs as in this pilot experiment. There, the number of genomic DNAs is on the order of hundreds of thousands of sequences. The number of LNA-modified probes needed to characterize such libraries can be extrapolated from oligofingerprint experiments using DNA probes. In a recent study, we used approximately 250 octamer probes in order to derive a low-redundant set of sequences from approximately 250,000 cDNAs from different libraries of the sugar beet plant [26]. With this number of probes, we were able to identify even small sequence clusters (corresponding to lowly expressed genes) with high accuracy. Since heptamer probes hybridize with much larger fractions of sequences than octamers and since the stability of LNA-modified oligoprobes is much higher than that of DNA probes, this number is rather an upper bound, and the effective number of LNA-modified probes needed to characterize such large numbers of sequences would rather be in the order of 100-150.

ACKNOWLEDGEMENTS

Sequence-verified DNA clones were kindly provided by the RZPD Deutsches Ressourcenzentrum für Genomforschung, Germany. This work was supported by the European Commission project MolTools, the National Genome Research Network (NGFN-2) and the Max-Planck-Society.

REFERENCES

[1] Hoheisel, J.D. *Nat Rev Genet.*, **2006**, *7*, 200.
[2] Maskos, U.; Southern, E.M. *Nucleic Acids Res.*, **1992**, *20*, 1679.
[3] Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. *Science*, **1995**, *270*, 467.
[4] Hauser, N.C.; Vingron. M.; Scheideler, M.; Krems, B.; Hellmuth, K.; Entian, K.D.; Hoheisel, J.D. *Yeast*, **1998**, *14*, 1209.
[5] Southern, E.M.; Maskos, U.; Elder, J.K. *Genomics*, **1992**, *13*, 1008.
[6] Beier, M.; Hoheisel, J.D. *Nucleic Acids Res.*, **1999**, *27*, 1970.
[7] Herwig. R.; Poustka, A.J.; Muller, C.; Bull, C.; Lehrach, H.; O'Brien, J. *Genome Res.*, **1999**, *9*, 1093.
[8] Herwig, R.; Schmitt, A.O.; Steinfath, M.; O'Brien, J.; Seidel, H.; Meier-Ewert, S.; Lehrach, H.; Radelof, U. *Bioinformatics*, **2000**, *16*, 890.
[9] Lennon, G.; Lehrach, H. *Trends Genet.*, **1991**, *7*, 314.
[10] Dramanac, S.; Stavropoulos, N.A.; Labat, I.; Vonau, J.; Hauser, B.; Soares, M.B.; Dramanac, R. *Genomics*, **1996**, *37*, 29.
[11] Radelof, U.; Hennig, S.; Seranski, P.; Steinfath, M.; Ramser, J.; Reinhard, R.; Poustka, A.; Francis, F.; Lehrach, H, *Nucleic Acids Res.*, **1998**, *26*, 5358.
[12] Drmanac, R.; Strezoska, Z.; Labat, I.; Drmanac, Sand.; Crkvenjakov, R. *DNA Cell Biol.*, **1990**, *9*, 527.
[13] Singh, S.K.; Nielsen, P.; Koshkin, A.A.; Wengel, J. *Chem Commun.*, **1998**, *4*, 455.
[14] Dwaine, A.B.; David, R.C. *Biochemistry & Biology*, **2001**, *8*, 1.
[15] Dahlgren, C.; Wahlestedt, C.; Thonberg, H. *Biochem Biophys Res Commun.*, **2006**, *341*, 1211.
[16] Petersen, M.; Wengel, J. *Trends Biotechnol.*, **2003**, *21*, 74.
[17] Jepsen, J.S.; Sorensen, M.D.; Wengel, J. *Oligonucleotides*, **2004**, *14*, 130.
[18] Tolstrup, N.; Nielsen, P.S.; Kolberg, J.G.; Frankel, A.M.; Vissing, H.; Kauppinen, S. *Nucleic Acids Res.*, **2003**, *31*, 3758.
[19] Silahtaroglu, A.N.; Tommerup, N.; Vissing, H. Mol. *Cell Probes*, **2003**, *17*,165.
[20] Liu, J.P.; Guerasimova, A.; Schwartz, R.; Lange, M.; Lehrach, H.; Nyarsik, L.; Janitz, M. *Comb. Chem. High Throughput Screen.*, **2006**, *9*, 591.
[21] McTigue, P.M., Peterson, R.J. and Kahn, J.D. *Biochemistry*, **2004**, *43*, 5388.
[22] SantaLucia, J.J. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 1460.
[23] Allawi, H.T.; SantaLucia, J.J. *Biochemistry*, **1997**, *36*, 10581.
[24] Bauer, O.; Guerasimova, A.; Sauer, S.; Thamm, S.; Steinfath, M.; Herwig, R.; Janitz, M.; Lehrach, H.; Radelof, U. *Rapid Commun. Mass Spectrom.*, **2004**, *18*, 1821.
[25] Guerasimova, A.; Nyarsik, L.; Liu, J.P.; Schwartz, R.; Lange, M.; Lehrach, H.; Janitz, M. *Biomol. Eng.*, **2006**, *23*, 35.
[26] Herwig, R.; Schulz, B.; Weisshaar, B.; Hennig, S.; Steinfath, S.; Drungowski, M.; Stahl, D.; Wruck, W.; Menze, A.; O'Brien, J.; Lehrach, H.; Radelof, U. The *Plant J.*, **2002**, *32*, 845.

Received: February 19, 2007 **Revised: March 20, 2007** **Accepted: March 22, 2007**