MAX-PLANCK-INSTITUT FÜR MOLECULARE GENETIK

Towards rapid and cost-effective genome characterisation using LNA-modified oligonucleotide DNA hybridisation

Inaugural-Dissertation

Zur Erlangung des akademischen Grades
‘‘Doctor der Naturwissenschaften (Dr. rer. Nat.)’’
am Fachbereich Biologie-Chemie-Pharmazie
dei Freien Universität Berlin

vorgelegt von

Jianping Liu

aus Jiangxi, China

Berlin
June 2007

Diese Arbeit wurde durchgeführt
am Max-Planck-Institut für Molekulare Genetik, Berlin
in der Arbeitspruppe 'Oligofingerprinting/Cell Arrays'
der Abteilung von Prof. Dr. Hans Lehrach.

1. **Gutachter: Prof. Dr. Hans Lehrach**
2. **Gutachter: Prof. Dr. Gerd Multhaup**

Date of defense: October 22nd, 2007

Contents

1. Summary

Oligonucleotide fingerprinting (OFP) is a high-throughput alternative to tag sequencing methods to determine the spectrum and abundance of genes in genomic DNA or cDNA libraries. This method currently relies on the sequential hybridisations of at least 200 short (8-12 mer), radioactively labeled DNA oligonucleotides to arrayed dsDNAs. After image analysis, a sequence fingerprint is generated for each clone based on the hybridisation signals. Then according to the fingerprint similarity, the clones from the same gene can be grouped together using clustering algorithms.

The main problems of the classic OFP method include high oligoprobe number and the use of radioactivity. To resolve these problems, here we have exploited the high affinity and good mismatch discrimination of oligoprobes modified with locked nucleic acid (LNA). In our method, short (hexamer or heptamer) LNA-modified oligoprobes are labelled with fluorescent dyes (e.g. Cy5). Hybridisation results are recorded with a CCD-camera developed for this purpose. The sensitivity of the fluorescence detection permits the convenient use of homogenous liquid assay as well as nylon membrane support. Thus hybridisation data quality is improved, and the process accelerated, simplified, and less expensive.

1.1 Oligonucleotide fingerprinting using iFRET technology

A method for in-solution analysis of DNA libraries by means of hybridisation with short (6- or 7-mer nucleotides long) fluorescent oligoprobes with real-time optical detection by induced fluorescence resonance energy transfer (iFRET) has been established. Only hybridised probe is detected as its fluorescent signal is generated through the excitation of the intercalating dye when a hybridisation event takes place. Thus, elimination of unreacted oligoprobe excess is not necessary. Furthermore immobilisation of the analysed library is avoided. It is essentially independent from the surface binding capacity or DNA immobilisation chemistry, requiring only accurate dispensing of both oligoprobe and target. Therefore, the assay can be performed entirely in solution for conventional OFP analysis, in micro-droplets or nano-well-array format.

We have reduced the length of the hybridisation oligoprobes to 6-mers, which allow the use of smaller total number of oligoprobes and offer better mismatch discrimination. Furthermore, we dramatically increased the throughput of the system in terms of the possible number of clones analysed in a single experiment, by miniaturization of the hybridisation process in a nanowell format, and by optical detection of the hybridisation through iFRET. Miniaturisation of the system and optical signal detection also led to significant reduction of analysis cost per sample. Multiplexing of the assay can be achieved using differentially labelled specific oligoprobes in the same solution. The method is homogeneous, and the generation of a universal set of genotyping reagents is possible because of the use of extremely short LNA-modified oligoprobes.

1.2. Oligonucleotide fingerprinting using nanoporous nylon membrane

Along with further exploitation of non-radioactive technique, we developed a robust method for the detection of hybridisation events using a microarray-based assay on a nanoporous membrane platform. Inherent limitations of standard arrays on glass slides, such as low binding capacity, low signal level, limited re-utilisation of the arrays and long processing time, could be overcome. The application of LNA-modified oligoprobes improved the hybridisation sensitivity, thereby allowing the use of shorter, and therefore reduced total number of oligoprobes for dsDNA characterisation.

The technique is characterized by a hybridisation time of less than 1 hour and the use of Cy5-labelled, 7-mer oligoprobes modified with LNA nucleotides. The volume of the DNA spotted onto a nanoporous membrane could be reduced to ~4 nl with detectable signal intensity. Moreover, the amount of the DNA target could be reduced to 4 fmol. The slides could be successfully stripped and re-probed for at least 19 times without significant loss of signal intensity.

In a proof of principle study 26 different LNA-modified 7-mer oligoprobes were hybridised with a set of 66 randomly selected human genomic DNA clones spotted on a nanoporous membrane slide. The performance of the approach was evaluated by sensitivity analysis using Receiver Operating Characteristic (ROC) curves. The hybridisation performances of LNA-modified and non-modified heptamers were compared. It was shown that the LNA modification clearly improves sensitivity and

specificity of the hybridisation. In order to test the clustering performance of LNA-modified oligoprobes de novo clustering based on hybridisation patterns was performed and compared with in silico clustering. It was demonstrated that the application of LNA-modified oligoprobes allows for reliable clustering of DNA sequences.

The described approach might dramatically increase the throughput of techniques based on OFP, by decreasing the total number of oligoprobes requested for analysis of large clone sets and reduction of the sample/reagent consumption. The method is particularly advantageous when numerous hybridisation-based assays must be performed for characterisation of sample sets of 100,000 or more.

2. Zusammenfassung

Oligonukleotid Fingerprinting (OFP) ist eine attraktive Alternative zu TAG-Sequenzierungsmethoden zur Bestimmung vorhandener Gene in genomischer DNA und cDNA Bibliotheken im Hochdurchsatz. Diese Methode basiert im Moment auf der sequentiellen Hybridisierung von mindestens 200 kurzen (8–12 Nukleotide) radioaktiv markierten Oligonukleotiden, um dsDNA aufzureihen. Basierend auf den Hybridisierungssignalen wird nach einer Bildanalyse für jeden Klon ein Sequenz-Fingerprint erstellt. Mit Hilfe von Cluster-Algorithmen können dann Fingerprints mit großer Ähnlichkeit zusammengefaßt werden.

Ein Problem der aktuellen Methodik stellt die hohe Anzahl benötigter Oligonukleotide und die verwendete Radioaktivität dar. In dieser Arbeit wurden die hohe Affinität und zugleich gute Diskrimination von sogenannten Locked Nucleic Acid (LNA) untersucht, um die Anzahl der benötigten Oligos und deren Länge zu reduzieren. Zusätzlich wurden zur Vermeidung von Radioaktivität kurze (Hexamere oder Heptamere), modifizierte LNA's benutzt, die einen fluoreszierenden Farbstoff trugen (z.B. Cy5). Die Ergebnisse wurden mit einer für diesen Zweck entwickelten CCD-Kamera aufgenommen. Die Empfindlichkeit der Fluoreszenzbestimmung diese Methode sorgt dafür, dass sowohl in Lösung befindliche Assays als auch solche auf Nylon-Membranen verwendet werden können. Dadurch ist die Datenqualität verbessert, der Prozeß ist einfacher, schneller und damit auch billiger.

2.1 Oligonukleotid Fingerprinting in Kombination mit der iFRET Technologie

Es wurde eine Methode zur Analyse von DNA Bibliotheken durch Hybridisierung von kurzen (6-7 Nukleotide) fluoreszenzmarkierten Oligonukleotiden entwickelt. Dabei kann die Hybridisierung in Echtzeit optisch mit Hilfe der Induzierten Fluoreszenz Resonanz Energie Transfer (iFRET) Technologie verfolgt werden. Bei Hybridisierung wird durch den Fluoreszenzfarbstoff ein vorhandener interkalierter Farbstoff zum Leuchten angeregt. Auf diese Weise ist es nicht nötig, den Überschuss von Oligoproben, die nicht reagiert haben, zu entfernen. Des weiteren wird die Immobilisierung der analysierten Bibliothek vermieden. Sie ist damit unabhängig von der

Kapazität zur Oberflächenbindung oder der chemischen DNA-Immobilisierung und benötigt einzig ein präzises Dispensieren der Oligoprobe und des Targets. Somit kann das Assay vollständig in einer Lösung für konventionelle OFP-Analysen durchgeführt werden, in Micro-Tropfen oder Nano-Well-Anordnung („nano-well array“) Format.

Die Länge der Hybridisierungs-Oligoproben wurde auf Hexamere reduziert, was die Verwendung einer geringere Gesamtzahl von Oligoproben und bessere Unterscheidungen bei der Fehleranpassung ermöglicht. Ferner wurde durch die Miniaturisierung des Hybridisationsprozessen in ein Nano-Well Format und durch den optischen Nachweis der Hybridisation durch iFRET der Durchlauf des Systems in Hinblick auf die mögliche Anzahl der Klone, die in einem Einzelexperiment analysiert werden können, deutlich gesteigert. Die Miniaturisierung des Systems und der optische Signalnachweis führt außerdem zu einer signifikanten Reduzierung der Kosten für die Analysierung der Proben. Die Datenübertragung des Assays kann durch die Verwendung unterschiedlich gekennzeichneter spezifischer Oligoproben in der gleichen Lösung erreicht werden. Die Herstellung eines Universalsets von Genotypisierungsreagenzien wird ermöglicht durch die Verwendung von extrem kurzen LNA-modifizierten Oligoproben.

2.2 Oligonukleotid Fingerprinting durch Verwendung einer Nanoporen-Nylonmembran.

Neben des weiteren Ausbaus der nicht-radioaktiven Technik wurde eine widerstandsfähige Methode für den Nachweis von Hybridisationsereignissen entwickelt, die ein Microarray-basiertes Assay auf der Plattform einer Nanoporenmembran verwendet. Grundlegende Grenzen von Standardarrays auf Glasobjektträgern wie geringe Bindungskapazität, geringe Signalstärke, begrenzte Wiederverwendung der Arrays und lange Fertigungszeiten konnten überwunden werden. Der Einsatz von LNA-modifizierten Oligoproben verbesserte die Hybridisierungs-Sensitivität und erlaubt dadurch den Gebrauch von kurzen Oligoproben, was die Gesamtanzahl von Oligoproben für dsDNA-Charakterisierung reduziert.

Die Technik ist gekennzeichnet durch eine Hybridisierungszeit von weniger als einer Stunde und dem Gebrauch von Cy5-markierten Heptamer-Oligoproben, modifiziert

mit sogenannten Locked Nucleid Oligos (LNA) Nukleotiden. Die Menge an DNA, die auf eine Nanoporen-Membran gespottet wird, konnte bei nachweisbarer Signalstärke auf ca. 4 nl reduziert werden. Darüber hinaus konnte die Menge des DNA Targets auf 4 fmol reduziert werden. Die Objektträger konnten für mindestens 19 Mal erfolgreich gereinigt und erneut mit Proben beladen werden, ohne signifikanten Verlust der Signalintensität.

In einer Studie zur prinzipiellen Beweisführung wurden 26 unterschiedliche LNA-modifizierte Heptamer-Oligoproben mit einem Set von 66 zufällig ausgewählten auf einem Nanoporen-Membran-Objektträger gespotteten, menschlichen, genomischen DNA-Klonen hybridisiert. Die Leistungsfähigkeit der Methode wurde ausgewertet durch eine Sensitivitätsanalyse unter Verwendung von sogenannten „Receiver Operating Characteristic“ (ROC) -Kurven. Die Effizienz der Hybridisierung der LNA-modifizierten und nicht-modifizierten Heptamere wurde verglichen. Es konnte gezeigt werden, dass LNA-Modifikationen die Empfindlichkeit und Spezifität der Hybridisierung deutlich verbessern. Um die Effizienz der Anhäufung („Clustering“) LNA-modifizierten Oligoproben de novo zu testen wurden Hybridisierungsmuster basierend auf Clustering, hergestellt und mit in silico Clustering verglichen. Es konnte gezeigt werden, dass die Anwendung von LNA-modifizierten Oligoproben das zuverlässige Clustern von DNA-Sequenzen ermöglicht.

Die beschriebene Anwendung dürfte dafür sorgen, den Durchlauf von Techniken, basierend auf OFP durch Herabsetzen der Anzahl von Oligoproben, die für die Analyse von großen Klonsets nötig sind, und durch Reduzierung des Verbrauchs an Proben und Reagenzien, deutlich zu verstärken. Die Methode ist vor allem von Vorteil wenn große Mengen hybridisierungsbasierter Assays für die Charakterisierung von Probensets von 100.000 und mehr erstellt werden müssen.

10.1 Acknowledgement

This dissertation represents the efforts of about three years of work, and would not have been possible to be finished without the effective cooperation and great assistance of so many people who deserve my acknowledgement.

I would like to express my highest gratitude to Professor Dr. Hans Lehrach for his providing me the opportunity to carry out my Ph.D. program in his department at the Max-Planck-Institute for Molecular Genetics, for the guidance and for reviewing this thesis. I am indebted to Professor Dr. Gerd Multhaup for being my adviser in Free University Berlin and for reviewing this thesis.

I am exceptionally grateful to my group leader, Dr. Michal Janitz, for inviting me to be here, for all his supervision, discussion and supports on my Ph.D. program during the last four years, for his correction of this dissertation, for his additional help with my family, as well as for his advice on my career development which is of vital importance for my future.

A special thank-you goes to Dr. Anna Guerasimova and Dr. Lajos Nyarsik for their stimulating discussions and valuable suggestions, which helped me with understanding the technical details, and for giving critical comments on my thesis.

Biotechnology engineer Regine Schwartz from the group of Dr. Nyarsik deserves my thanks for her professional spotting skills, technical assistance and experimental work, and for her warm help to find the comfortable apartment for my family and the excellent babysitter for my little son.

Thanks are also due to Sabine Thamm for her daily support and warmth on my numerous requests for German translation during my staying in the lab, to Dr. Andrea Fiebitz for her kindly translating the summary from English into German and for a lot of very generous help, to Andreas Dahl for his encouraging smile, sparkling discussion and strong support for my private life, to Irina Girnus for her assistance during the RNA preparation step.

Particular thanks go to Dr. Ralf Herwig and Mario Drungowski for their bioinformatic discussion and help during the paper preparation.

I am indebted to my colleagues, Dr. Dominique Vanhecke, Dr. Yuhui Hu, Xi Cheng, Matthias Lange and some other members for their creating a pleasant working atmosphere throughout my project, for their technical supports, valuable discussion on projects, for their reagent supply and for their allowing me to use their advanced computers.

Also I would like to thank Hilke Wegwerth, Jeanette Brylla, Johanna Belart, Ulrich Barenbrock, and Anna Skotarczyk for their administrative helps on my life here in Germany as a foreigner.

Finally, I want to express my deep appreciation to my families: my parents have been providing me with their continuous support over all these years which enable me to get this far; my wife Chang Yin, with her love and care, has made my life in the lab full of joy and interests as at home; my son Yuchong Liu, with his admiring smile as well as the bothering cry during the two-year period of growing up, clears up and fuels my mind every time when I stay with him.

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.

10.2 List of figures (to be continued)

10.2 List of figures (continued)

10.3 List of tables