Single-molecule Detection in situ

CHATARINA LARSSON
Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Dag Hammarskjölds väg 20, Uppsala, Thursday, April 9, 2009 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

The human body contains a variety of different cell types that share a common genome, but differ in how they use the information encoded therein. Variation in molecular content exists even at the level of individual cells, and to provide deeper insight into complex cellular processes methods that permit analysis of each cell on its own are needed. This thesis presents molecular methods for localized detection of individual nucleic acid molecules. The developed methods are based on padlock probes and target-primed rolling circle amplification. Single-molecule detection sensitivity in combination with single-nucleotide genotyping selectivity enables detection of allelic DNA variants and closely related target sequences directly in cells. Padlock probes further enable multiplex detection of targets, and in combination with image analysis quantitative molecular data for individual cells can be acquired for large cell populations at a resolution that no other in situ detection method can provide at present.

In this thesis, the in situ target-primed rolling circle amplification technique was first used for genotyping of a point mutation in the mitochondrial genome with padlock probes. This displayed mitochondrial DNA heterogeneity in cell populations. Application of the method on comet assay preparations showed that mitochondrial genomes are lost from these samples prior to analysis. Nuclear DNA targets, however, can be efficiently detected in corresponding samples. Padlock probes and rolling circle amplification are thus an attractive alternative to FISH analysis for localized DNA detection in comet assay samples. A method was also developed for localized detection of individual mRNA molecules with padlock probes and rolling circle amplification. This method provides unique possibilities to genotype allelic variants of transcripts in situ. mRNA expression is associated with substantial cell-to-cell variation and our presented method permits simultaneous visualization of multiple transcripts directly in complex tissue samples. Application of the methods presented in this thesis will enable new types of studies of biological samples from both normal and disease states.

Keywords: padlock probes, genotyping, SNP, in situ, rolling circle amplification, RCA, mitochondrial DNA, mRNA, single-molecule, single-cell, fluorescence, microscopy

Chatarina Larsson, Department of Genetics and Pathology, Uppsala University, SE-75185 Uppsala, Sweden

© Chatarina Larsson 2009

ISSN 1651-6206
urn:nbn:se:uu:diva-98538 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-98538)
To my wonderful family
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


III Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. *In situ* detection of individual mRNA transcripts with single nucleotide resolution. *Manuscript*.

Reprints were made with permission from the respective publishers.
Related work by the author

**Peer reviewed research**

**Review articles**


**Book chapters**
Abbreviations

bp  base pair
cDNA  complementary DNA
CNV  copy number variation
DNA  deoxyribonucleic acid
FISH  fluorescence in situ hybridization
ISH  in situ hybridization
LAMP  loop-mediated isothermal amplification
LNA  locked nucleic acid
MELAS  mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes
MIDD  maternally inherited diabetes and deafness
MIP  molecular inversion probe
mRNA  messenger RNA
mtDNA  mitochondrial DNA
ncRNA  non-protein coding RNA
nt  nucleotide
OLA  oligonucleotide ligation assay
PCR  polymerase chain reaction
PNA  peptide nucleic acid
PRINS  primed in situ labelling
RCA  rolling circle amplification
RCP  rolling circle product
RNA  ribonucleic acid
rRNA  ribosomal RNA
RT-PCR  reverse transcription PCR
SNP  single nucleotide polymorphism
tRNA  transfer RNA
qPCR  quantitative PCR
Introduction

Molecular methods for analysis of biomolecules are essential for medical research and diagnostics. In the last decades substantial progress has been made to characterize the collection of biomolecules present in cells of different organisms, with the resulting data made publicly available in databases online, such as Genbank (www.ncbi.nlm.nih.gov) and Ensembl (www.ensembl.org). Also deviations from the normal states are searchable in these and related databases. The public availability of nucleic acid and protein sequences has aided biological and medical research substantially. As the knowledge of different molecules and interactions increases, so does the need of methods able to provide deeper insight into molecular processes.

All molecular methods are associated with characteristics such as sensitivity, specificity, cost and time required to perform an analysis. Method development constantly tries to push the limit of these properties. The optimal method would detect all target molecules in a sample without errors in short time and at low cost. In reality though, compromises typically have to be made regarding these characteristics, keeping in mind the most important properties for the intended application. As method development can be a costly process it is important to consider what a new technology can add to the available repertoire of methods before initiating the development path. Does the method provide more detailed information than available methods? Or does it provide considerable benefits over existing methods providing similar information? To allow for extraction of new types of biological information from samples, we have focused on development of methods that enable discrimination between highly similar target molecules.

The work presented in this thesis describes development and application of two closely related methods for analysis of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The methods both utilize molecular probes known as padlock probes, and signal amplification to enable highly specific detection of targets at the single-molecule level. Analysis takes place at the site of target location in the cell, making it possible to study the molecular content of individual cells. These topics will be discussed further in the following chapters, aiming to provide a comprehensive introduction to the presented work. Finally, the papers included in the thesis will be reviewed and discussed, aiming to demonstrate the advantages of our methods and how they can add unique pieces of information to the scientific exploration of the molecular world of the cell.
Targets for biomolecular analysis

Inheritance of traits was studied long before the nature of the inherited entity was known. Genetic studies were greatly enhanced after the discovery of the double-stranded helical structure of DNA in the 1950’s. The finding that the nucleotides (nt) of two DNA strands hybridize to each other according to specific base pairing rules led to development of methods for identification of specific DNA sequences. The following sections will provide a brief introduction to the nature of DNA and RNA molecules present in a cell, and to some of the variations that can occur in these molecules. As most of the work presented in Papers I-III was carried out on human material, the discussion will be focused on this species unless otherwise stated. Similarly, the discussion will be centered on genetic variation that is interesting in the context of the described methods.

The human genome

The human genome consists of two spatially separated entities, the nuclear genome and the mitochondrial genome. The larger nuclear genome consists of about 3 billion base pairs (bp) of double stranded DNA organised in 23 pairs of chromosomes within the cell nucleus. The nuclear DNA encodes an estimated number of 20 000-25 000 human protein-coding genes. This figure is under constant debate though, with recent reports indicating a lower number of genes. To enable fitting of the genomic DNA within the confined space of the nucleus the DNA is packed into chromatin, a structured organisation of DNA around histones. The chromatin also contains structurally important RNA molecules and further proteins for packing. The histones have a function in regulation of nuclear processes, such as transcription, replication and chromosome condensation. The nuclear DNA is inherited from both parents, with one chromosome copy originating from the mother and the other copy from the father.

The mitochondrial genome consists of a double stranded 16 569 bp circular DNA molecule found associated with proteins in structures called nucleoids within mitochondria. Nucleoids contain up to 10 copies of the mitochondrial genome, and one mitochondrion typically contains several nucleoids, resulting in the genome being present in high copy numbers in individual cells. The mitochondrial genome consists almost exclusively of genes and thus lacks the complexity of coding and non-coding regions found in nuclear DNA. The genes encode 13 of the proteins needed for the respiratory chain pathway as well as two rRNAs and 22 tRNAs needed for protein translation within the mitochondrion. The mitochondrial DNA (mtDNA) is inherited from the mother only, although the transfer of mtDNA also from the father (paternal leakage), has been discovered in one human case. In other animal species paternal contribution to inheritance has been
observed more frequently. It has been argued that paternal leakage may be more prevalent also in humans, however below the detectable level of the methods commonly used. Apart from the possible paternal contribution, the uniparental inheritance of mitochondrial genomes is still complex due to the presence of a mitochondrial genetic bottleneck which causes rapid changes in mitochondrial DNA allele frequency between generations. The reason being that because of the genetic bottleneck only a subset of the maternal mtDNA molecules are transferred to the offspring, where they are replicated to make up a larger number of genomes. The exact mechanism of how the bottleneck works is debated, with several hypotheses proposed recently.

An introduction to the RNA world
The proportion of the human genome coding for proteins is only about 1.2%. Most of the genomic DNA was previously thought to be non-functional, but the ENCODE (Encyclopedia of DNA elements) project has revealed that a great majority of the nucleotides of the human genome are associated with at least one primary transcript. Transcription results in production of different classes of single stranded RNA molecules. This thesis deals with the class known as messenger RNA (mRNA), which consists of the molecules which are later translated into proteins. Other classes of RNAs include ribosomal RNA (rRNA) and transfer RNA (tRNA), which are needed for translation, and several types of non protein-coding RNA (ncRNA) that have regulatory functions. The transcribed mRNA consists of exons, which constitute the sequences that are translated into proteins, and introns, which are found between the exons and are removed upon formation of the mature mRNA molecule through a process called splicing. Spliced introns sometimes give rise to micro RNA (miRNA) and small nucleolar RNA (snoRNA), providing an example of how sequences initially thought to be “junk” or noise now are known to exhibit important functions. Sequences that direct splicing to intron-exon boundaries were identified early on, and a multitude of other sequence- as well as secondary structure elements that affect splicing have been identified. In addition, different transcription factors binding to promoters have been found to influence the splicing of transcribed RNA. There are still many unknown contributors that influence splicing however, and alternative splicing patterns can be found in a majority of human genes. To increase the stability of mRNA, the 5′-end is protected from degradation by addition of a protein cap, and a sequence consisting of multiple adenosine residues, known as the poly-A tail, is added to the 3′-end. mRNA molecules are then exported from the nucleus into the cytoplasm where they are translated.
Genetic variation

Looking at the human population it is evident that there is great variation in traits among people from different parts of the world. This divergence stems from variation in the human genome that has arisen over time. Also closely related humans have a unique set of traits, and no two human genomes are genetically identical. Even homozygotic twins, the most closely related individuals, have shown to carry copy number differences of large sequence elements in somatic cells and also frequently exhibit epigenetic differences that control how the genome is used in different tissues. The source of genetic variation is mutations in the genome over time. Mutations are changes in the DNA sequence brought about by for example errors in the DNA replication process, or by environmental factors such as radiation or exposure to chemical compounds. DNA mutations can cause proteins to exhibit a gain of function or a loss of function, or they can be silent (lead to no change in amino acid sequence). Silent mutations and variations that occur within exons of genes are said to be synonymous. The opposite can be said for nonsynonymous variations, which lead to a change in the amino acid sequence of the protein. Mutations present in germ cells can be transmitted to the next generation offspring if they are not incompatible with fetal development. When a genetic variant is spread widely enough to be present in 1% of a given population, it is no longer known as a mutation but is called a polymorphism. Mutations that take place in somatic cells are not transmitted to the offspring, but can however lead to disease in the individual. One potential outcome of somatic mutation is cancer. This happens when mutations alter the regulation of cell division. Typically though, many genes need to exhibit changes in their normal function in order for a tumor to develop.

Genetic variation can arise in any part of the genomic DNA, and can be the result of insertions, deletions or exchanges of genetic material. The most common genetic variation in humans is the single nucleotide polymorphism (SNP), as the name suggests being comprised of a change in a single base pair. Other common sources of variation are copy number variations of short repeated sequences (e.g. microsatellites) and larger DNA fragments (known as CNVs). The fact that genetic variation is distributed throughout the human genome in both coding and non-coding regions is utilized for identification and mapping of genes associated with a certain trait or disease. The most common variations used as markers in genetic mapping today are SNPs, and microsatellites. The HapMap project is an initiative to provide a database of common SNPs in the genome. The HapMap database is now a great source of SNP genotype and frequency data for different human populations.

Like nuclear DNA, mtDNA can also be mutated. In fact, mtDNA has a significantly higher mutation rate than nuclear DNA. Reasons for this are suggested to be mutagenic properties of reactive oxygen species formed...
during oxidative phosphorylation, a relatively low fidelity of the mitochondrial DNA polymerase gamma$^{34,35}$, and limitations in some of the mitochondrial DNA repair mechanisms$^{36}$. When mutated mtDNA makes up a certain proportion of the total mtDNA pool in an individual, it can lead to disease in the carrier. A cell that harbours mitochondrial genomes of only one haplotype is called *homoplasmic*, while a cell that has a genome population with mixed haplotypes is called *heteroplasmic*. Most known mitochondrial diseases exhibit heteroplasmy, but also homoplasmic mutations are known to exist$^{37}$. The level of mutated mtDNA in a heteroplasmic individual with a mitochondrial inherited disorder is often proportional to the degree of disease. In addition, symptoms often appear in different organs at different levels of heteroplasmy depending on the nature of the organ and the function of the mutation, a phenomenon known as the *threshold effect*.$^{37}$
Methods for analysis of biomolecules

Genetic variation can exist both between different individuals and between different cells of one individual. To explore genetic variation a number of molecular techniques have been developed. These techniques can generally be divided as those that find unknown variations (e.g. sequencing methods) and those that analyze known variations to find out the status of a given genome (e.g. SNP genotyping methods). This section will introduce some of the properties of biomolecular analysis methods and describe a selection of methods that are of relevance to the work presented in this thesis.

Properties and formats of molecular methods

All assays are associated with different properties which serve to characterize how suitable a method is for different applications. Sensitivity describes the proportion of actual positive samples that are classified as positive by an assay. Sensitivity relates to the limit of detection (LOD) of a molecule, i.e. how many molecules that need to be present in a sample to be able to detect the analyte. Specificity on the other hand describes the ability to correctly identify a particular molecule, by giving the proportion of actual negative samples that are classified as negative. The ability to separate between two highly similar targets is referred to as selectivity, which is of high importance in for example SNP genotyping. In this thesis I refer to the detection efficiency of the method meaning the overall performance of the assay with regards to the number of detected molecules. Another feature of molecular methods is the multiplexing ability, the ability to perform simultaneous detection of several analytes in the same reaction.

Assays can be classified in a number of ways. Analysis of biomolecules in a sample can be carried out either in solution (homogeneous phase) or on a surface (heterogeneous phase). Surface-based analysis can take place with samples either attached on artificial surfaces, such as microarrays, or within cells attached to surfaces. In this thesis, in vitro assays are referred to as assays that analyze biological material that has been removed from the cellular context. This is in contrast to in situ assays, which are carried out within cells or tissues.

In vitro measurements are typically based on the presence of a large population of target molecules, and therefore provide an average result for the
molecules in the sample. Averaging is sufficient for many applications, but rare target variants in for example heterogeneous samples can fail to be detected. Detection with single-molecule resolution can solve this problem but puts demands on the sample preparation and assay technology. Single-molecule analysis methods need to be highly sensitive. Furthermore, it is essential that a signal actually reports the presence of a target molecule and is not created in absence of target recognition. Single-molecule assays are commonly detected by fluorescence to permit multiplexing. As single fluorophores are difficult to detect, amplification methods can be used to create a reporter molecule labeled with multiple fluorophores to increase the signal over the background.

Performing averaging measurements on populations of molecules is not only associated with the risk of failing to detect rare targets, also information on target localization is lost. Even if performing single-molecule measurements, the presence or absence of signal does not indicate the status of individual cells in a sample. Take expression as an example. Quantitative \textit{in vitro} measurements of single β-cells from pancreatic islets have shown large variation in expression levels of selected mRNA molecules\textsuperscript{38}, and it has been shown that cell-to-cell variation in expression can be substantial even in synchronized cell cultures\textsuperscript{39}. From an average measurement it is impossible to determine if all cells in a sample express a detected transcript at similar level, or if there are differences in expression level between different subsets of cells in the sample, as illustrated in Figure 1. Averaging furthermore makes it difficult to correlate the expression of two genes in a cell population.

![Figure 1](image.png)

\textit{Figure 1.} Single-cell based versus population based analysis. Molecular analyses of populations of cells give an average measurement of an investigated property. The true investigated state of each cell in the population can deviate substantially from the average though. Individual cells in 4 populations are illustrated here as small squares and the true state of each cell is shown in black and white. When populations 1-4 are analyzed in bulk, differences between cells do not show and all of the populations give the average result ‘grey’.
since you cannot determine if they are expressed in the same cells or in different cells of the sample in response to a certain stimulus. Single-cell analysis, as the name implies, permits investigation of the molecular status of individual cells. Such analysis can be undertaken by assaying isolated and lysed individual cells, by performing analysis of fixed cells in situ, or by imaging of molecules directly in live cells.

Probing and amplification

For nucleic acid analysis, synthetic oligonucleotides of specified sequence are typically used to “fish” out targets of interest from a population of molecules. Frequently, nucleic acids are detected either by amplification of the molecule of interest, or by amplification of a reporter molecule detecting the presence of a specific target. The reason for this amplification step is to be able to record signals from detected molecules over the background provided by the other molecules in the sample.

The polymerase chain reaction

The method that has had one of the largest impacts on biological research is the polymerase chain reaction (PCR). PCR is a sensitive method used for amplification of specific stretches of DNA. The method consists in its simplest form of cycles of incubations of the sample together with a pair of sequence specific PCR primers at different temperatures. The method has multiplexing abilities, but it has been shown that as the number of primer pairs included in the reaction increases, so does the risk of amplification artifacts such as false and truncated products. Another limitation with conventional PCR is that the amount of product formed in a reaction is typically not indicative of the initial concentration of the target DNA in the sample. Numerous variants of PCR exist today. One particularly interesting PCR variant is real-time PCR. Real-time PCR has a fluorescent read-out stemming from either hybridization of sequence specific fluorescence-labeled probes (e.g. molecular beacons or Taqman probes), or from the interaction between the PCR product and a dye that fluoresces upon duplex DNA binding (e.g. SYBR green I). The resulting fluorescence is proportional to the amount of PCR product that has been formed in the assay and can be tracked in real-time as the analysis is carried out. This permits use of real-time PCR for quantitative measurements of specific target sequences and the method is therefore often referred to as quantitative PCR (qPCR). PCR can be combined with other methods, such as reverse transcription or proximity ligation to enable analysis also of RNA and protein, respectively.
Padlock probes

Padlock probes are synthetic oligonucleotide probes which enable sensitive, specific, multiplex and localized detection of nucleic acids. The padlock probe technology is a refinement of the oligonucleotide ligation (OLA) assay48, which utilizes enzymatic ligation of two oligonucleotide probes on a target DNA molecule to detect and genotype DNA. A padlock probe typically consists of a 70-100 nt long single stranded DNA oligonucleotide. The two end segments of the padlock probe are complementary to the target DNA and are designed to hybridize next to each other on the target site so that the 3'- and 5'-ends of the probe become positioned immediately next to each other. This enables the two ends of the probe to be covalently joined together through DNA ligation, forming a circular molecule49, as shown in Figure 2. The DNA ligation reaction is sensitive for mismatches in the probe-target duplex, a fact which is used for discriminating between highly similar targets. Sequences deviating from the padlock probe target sequence with as little as one nt can be distinguished, making ligation based assays highly suitable for SNP or point mutation analysis. The mismatch-dependent reduction in ligation efficiency is greatest when padlock probes are designed to hybridize to the nucleotide position to be genotyped with the 3'-end of the probe50, 51. The non-target complementary part (in this thesis referred to as the back piece) of a padlock probe consists of sequences for amplification and/or identification of circularized probes.

Circularized padlock probes are typically detected following some sort of amplification reaction such as PCR or rolling circle amplification (RCA) (described in next section). PCR is carried out over the ligation site so that enrichment takes place for circularized padlock probes only52. To further reduce the risk of amplification artifacts, linear padlock probes and probes that have undergone intermolecular ligation to other probe molecules can be removed by exonucleases prior to amplification53. The circular template amplified in the padlock probe assay can only be present if two recognition events have taken place at the correct location in the sample. This is in contrast to PCR, where primers hybridize to the target DNA independent of each other. The strict requirement of correct hybridization makes padlock probes very suitable for multiplex reactions since amplification products from false priming sites cannot be formed. Multiplexing is also facilitated by the ability to encode specific tag sequences in the back piece of different padlock probes for read-out and identification.

Padlock probes are also known under other names, i.e. circularizable oligodeoxyribonucleotide probes (c-probes) and molecular inversion probes (MIPs). MIPs differ from the original padlock probe concept by not being based solely on ligation for allele recognition. Instead of hybridizing to the nucleotide to be genotyped, allele distinction is made by a polymerase filling in the appropriate nucleotide at this position, allowing for a ligase to then
Figure 2. SNP genotyping with padlock probes. Padlock probes hybridize to target sequences and perfectly base-paired ends are joined through ligation, creating a circular molecule. Padlock probes that do not make a perfect match to the target sequence with the 3'-end terminal nucleotide cannot undergo ligation and remain linear. Circular padlock probes are detected by a method of choice whereas linear probes do not give rise to a detection signal. 

a) Padlock probe T matched to target A, 
b) Padlock probe T mismatch hybridized to target G, 
c) Padlock probe C matched to target G. Tag-sites on padlock probes for A and G detection are shown in grey and white, respectively, and matched targets are color coded accordingly. The image is not drawn to scale.

close the extended probe into a circle. This step is carried out in four parallel reactions, one for each nucleotide. MIPs further differ from regular padlock probes by being opened at a location in the back piece of the circular probe prior to amplification. This inversion of the probe increases the PCR efficiency by releasing the probe from the target strand. The MIP assay has been used for highly multiplexed genotyping of more than 50 000 SNPs in order to detect allele copy number aberrations.

DNA ligases for padlock probe analyses

DNA ligases are critical enzymes for DNA replication and repair that exist as two principal types, ATP-dependent and NAD⁺-dependent ligases. There are many commercially available ligases for molecular biology, but mainly two DNA ligases are used for padlock probe ligation. The first is the bacteriophage T4 DNA ligase, which is an ATP-dependent ligase most active around 37 °C. The ligation specificity of T4 DNA ligase is dependent on the NaCl concentration. The second ligase is the thermally stable Thermus thermophilus (Tth) ligase, also known as Ampligase, which joins DNA nicks in double stranded DNA through an NAD⁺-dependent reaction. Am-
pligase is active at high temperatures, enabling stringent hybridization of oligonucleotides, and has been found to have significantly higher ligation specificity than T4 DNA ligase\textsuperscript{50}. Although specificity is generally higher with Ampligase, T4 DNA ligase has the advantage of working at low temperatures which makes it suitable for some padlock probe applications.

**Rolling circle amplification**

RCA (also referred to in the literature as rolling circle replication (RCR) or rolling circle DNA synthesis) is an isothermal amplification technique. The components necessary to perform an RCA reaction is a circular DNA molecule and a polymerase with ability for synthesis of long stretches of DNA using circular DNA as template\textsuperscript{58}. Due to their circular nature, padlock probes are excellent templates for RCA. The amplification reaction can be carried out in either a linear fashion\textsuperscript{59}, or in a hyperbranched fashion\textsuperscript{60}. The DNA synthesis is initiated from either an added RCA primer which is complementary to part of the padlock probe, or from the 3’-end of the target molecule\textsuperscript{59}, as also shown herein in Paper I. The RCA mechanism has been used for signal amplification for both nucleic acid and protein detection.

The product of an RCA reaction is a long single stranded DNA molecule consisting of multiple repeats of the complementary sequence of the template molecule. The presence or absence of an amplification product can be detected by for example gel electrophoresis\textsuperscript{59, 60}, on arrays\textsuperscript{61} or in real time\textsuperscript{62}. A considerable strength with the rolling circle amplification method is the possibility to take advantage of the coiling of the amplification product into a small micro-meter sized bundle of DNA. This phenomenon occurs to rolling circle products (RCPs) formed in solution and on surfaces\textsuperscript{60, 63}. RCPs can be visualized fluorescently as bright spots by hybridization of fluorophore-tagged oligonucleotides to sequences incorporated in the padlock probe back piece\textsuperscript{60}.

**The \(\Phi\text{29} \) DNA polymerase for efficient RCA**

There are a number of commercially available DNA polymerases capable of replicating circular molecules. The most widely used for RCA today is the \(\Phi\text{29} \) DNA polymerase, which is also the polymerase used for the work presented in this thesis. \(\Phi\text{29} \) DNA polymerase is a DNA dependent polymerase which synthesizes DNA with high fidelity due to efficient proofreading\textsuperscript{64}. There are several properties that make \(\Phi\text{29} \) DNA polymerase our enzyme of choice for performing RCA \textit{in situ}. First, both the DNA synthesis and the 3’-5’ exonuclease activities of the enzyme are associated with high processivity. During polymerization, one polymerase molecule can synthesize more than 70 kb DNA in length\textsuperscript{65}. Second, the polymerase has a strand displacement activity which could be of importance when replicating a circular molecule. It has been argued however that it is more likely the bending force associated
with replication of small circular molecules such as padlock probes that cause the strand displacement needed for efficient replication of these molecules with RCA\textsuperscript{59}. \(\Phi 29\) DNA polymerase can further also use RNA as primer for DNA synthesis\textsuperscript{66, 67}, which is of special interest in RNA analysis with padlock probes.

**RCA for single molecule detection**

A beauty of the rolling circle mechanism is that it can be used in any assay that is based upon target dependent DNA circularization. Examples of this are padlock probe-, selector probe\textsuperscript{68} or proximity ligation\textsuperscript{69} assays, which enable studies of both known and unknown nucleic acid sequences, as well as proteins. Each RCP represents the detection of one molecule. By digital counting of RCPs it is thus possible to create a detailed overview of the molecules present in a sample. Such quantification can be done either by scanning of RCPs in solution flown through a microfluidic channel\textsuperscript{63, 70}, or by depositing the RCP solution on microscope slides, followed by imaging\textsuperscript{71}. RCPs can also be created on microarray surfaces, providing quantitative microarray data\textsuperscript{72}, or \textit{in situ}, enabling single molecule detection well above background in both cell and tissue samples as described in Papers I-III. Notably, RCA also permits visualization of endogenous proteins, protein complexes and protein modifications with single-molecule/interaction resolution \textit{in situ}, taking advantage of the highly specific proximity ligation mechanism for target recognition\textsuperscript{69, 73}.

By depositing RCPs on a surface, the multiplexing ability of the detection assay can be increased by applying serial and combinatorial labeling of the RCPs\textsuperscript{71}. This approach involves multiple rounds of hybridization and stripping of detection oligonucleotides to extract information about the identity of molecules present in the original sample. The presented design enabled identification of 31 genomic loci represented as RCPs deposited on a microscope slide, and it was argued that >1500 targets should be decodable with only minor modifications to the procedure\textsuperscript{71}.

**In situ** detection of nucleic acids

The methods described so far in this thesis all share the common characteristic that they do not provide any information on the status of individual cells in a sample. For single-cell analysis, technologies must be very sensitive. Expression analysis of individual cells can be performed with reverse transcription quantitative PCR (RT-qPCR)\textsuperscript{38}, but this is associated with noise problems when very rare mRNA targets are investigated\textsuperscript{74}. The method is also dependent on physical isolation of single cells. Microfluidic approaches to single cell biology are also attracting interest\textsuperscript{75}. These technologies currently demand specialized equipment and it will probably take time before
they become widespread in the research community. Instead of going through the trouble of isolating single cells, analysis can be carried out in situ, providing not only information on individual cell content but also physical location of cells in a tissue and of target molecules within cells.

In situ hybridization

The detection of specific nucleic acid molecules by hybridization within cells is known as in situ hybridization (ISH) and was first described in 1969\textsuperscript{76}. Detection for ISH was initially done by radioactive labelling of probes, but is now performed through antigen labelling (for subsequent immunostaining) or fluorescence labelling of the probes. Based on fluorescence, the ISH technique is known as fluorescence in situ hybridization (FISH), permitting simultaneous analysis of multiple targets in the same sample. FISH is routinely applied for both diagnosis and research and there is a large collection of different FISH methods available for different applications\textsuperscript{77}.

FISH to DNA in metaphase chromosome spreads (2D-FISH) is limited to the detection of large (several Mb) fragments. FISH on interphase nuclei (3D-FISH) and chromatin fibers (Fiber-FISH) has higher resolution (down to 50 kb and 1 kb respectively), but detection of short targets is generally associated with a decrease in detection efficiency\textsuperscript{77, 78}. FISH for detection of nuclear DNA has been applied for highly multiplexed detection, identifying all human chromosomes in individual metaphase spreads\textsuperscript{79, 80} and interphase nuclei\textsuperscript{81}. Increased resolution of multicolour-FISH even allows for staining of individual chromosome arms\textsuperscript{82}, enabling close investigation of chromosome stability in cell lines\textsuperscript{83}.

Although fluorescent detection enables multiplexing and closer investigations of the sub-cellular localization of transcripts, RNA in situ hybridization is often carried out using other detection methods than fluorescence. RNA FISH using oligonucleotide probes tagged with multiple fluorophores can be used for detection of single-copy mRNA molecules in fixed cultured cells\textsuperscript{84}. Using probes labelled with a different fluorophore combination for each mRNA, multiple different transcripts can be detected simultaneously, as demonstrated with detection of 10 different active sites of transcription in cultured human fibroblasts\textsuperscript{85}. Single-molecule FISH has also been performed with multiple probes carrying one fluorophore label each, requiring hybridization of many probes to the same mRNA to give rise to a signal\textsuperscript{86}.

Modified nucleotides for increased sensitivity and specificity in ISH

The specificity of hybridization in ISH is dependent on the hybridization probe length. A long probe generally has higher affinity to the target, but as the length increases, the ability to distinguish the target from similar sequences decreases\textsuperscript{87}. SNPs, the most common genetic variants, are currently
difficult to study in the context of the cell due to the low target discrimination abilities. Recently efforts have been made to increase FISH resolution by incorporating modified nucleotides in hybridization probes. Locked nucleic acid (LNA) is an RNA analog which can be incorporated in synthetic oligonucleotides, providing increased hybridization affinity compared to normal nucleotides. The increased thermal stability of DNA-LNA heteroduplexes enables high specificity and increased mismatch discrimination abilities through raised hybridization temperatures. LNA-modified primers have been used for enhanced SNP genotyping in, for example, allele specific PCR, but so far only repetitive non-polymorphic DNA sequences have been detected using FISH with LNA-modified probes. LNA modified probes have proven very useful for imaging of small non-coding RNAs through normalization of small non-coding RNAs in situ though, where normal oligonucleotide probes lack the sensitivity needed for detection.

Similar to LNA, peptide nucleic acid (PNA) oligonucleotides have been proposed to increase selectivity in situ. This was demonstrated by discrimination of two centromeric repeats differing in a single base pair. PNA is a synthetic DNA analog that has a non-charged peptide backbone instead of a deoxyribose phosphate backbone. This makes PNA-containing duplexes more stable than DNA-DNA or DNA-RNA duplexes. FISH with PNA probes for DNA detection has so far been limited to repeated sequences due to low detection efficiency. One additional interesting feature of PNA is that it can form triplex structures with double stranded DNA, and PNA openers have successfully been used for local denaturation of DNA in combination with padlock probing in situ in bacteria.

Techniques based on in situ polymerization

Since ISH based techniques lack the sensitivity needed for short single copy sequence detection, other in situ techniques such as primed in situ labelling (PRINS) and in situ PCR have aimed at solving this problem. These two techniques both build on DNA dependent DNA polymerization, but differ somewhat in how they are carried out. In PRINS a sequence-specific unlabelled oligonucleotide primer is elongated by Taq polymerase, incorporating labelled nucleotides in the elongation product. Multiple targets can be detected on the same slide by carrying out sequential elongation of primers for the different targets in different steps, and the specificity of the primer elongation thus enables single nucleotide differences to be distinguished. PRINS has mainly been applied for identification of chromosomes in metaphase spreads. Although, using multiple primers for the same gene and tyramide signal amplification (TSA), a method commonly used for signal amplification in immunoassays, single copy genes have been detected. In situ PCR also relies on elongation of unlabelled sequence specific primers, but for this technique the temperature is cycled as in solution phase PCR,
leading to accumulation of PCR amplicons in cells. In situ PCR is mainly used for detection of RNA, and is then coupled to reverse transcription which can be performed in the same reaction as the PCR (in situ RT-PCR)\textsuperscript{103}. Detection is either carried out by incorporation of labelled nucleotides in the PCR reaction, or by performing a separate ISH to detect the amplicon.

Although put forward as quick and cheap alternatives to FISH for some applications, PRINS and in situ PCR have gained limited success. Both of these polymerization-based techniques are associated with characteristics that make them insufficient for visualization of single-copy base pair variants in situ. PRINS has specificity to genotype SNPs, but is not sensitive enough to visualize single-copy sequences unless a cocktail of primers is used. Similarly, in situ PCR is limited by detection sensitivity, and also needs to be carefully optimized to avoid background staining\textsuperscript{103}. Padlock probes on the other hand are highly specific and offer single molecule sensitivity of detection, making them potentially suitable for in situ detection, as will be described next.

Padlock probes for in situ detection and genotyping

The first application of padlock probes in the in situ setting was for detection and genotyping of centromeric sequences in chromosomes 13 and 21\textsuperscript{104}. In this study padlock probes labelled with biotin and digoxigenin were detected using labelled streptavidin and antibodies. Detection of ligated padlock probes using antibodies suffered from a lack of sensitivity, preventing studies of rare targets due to background generated by unspecific binding of the detection reagents. Introduction of the rolling circle amplification mechanism increased the specificity of detection by increasing the signal strength and decreasing the unspecific background\textsuperscript{60}. This was possible since RCA signals only can appear if a DNA circle, and thus a target sequence, is present in the sample. In this first publication, RCA in situ was reported to be very inefficient for detection. In 2001 two further papers demonstrating RCA of circular probes for SNP detection in situ were published\textsuperscript{105, 106}, aiming to raise the detection efficiency by different approaches. These two papers will be discussed further in the background for Paper I.

The publications mentioned above comprised the state of the art for padlock probe analyses in situ when the work resulting in this thesis was initiated. Following the publication of Paper I, the field has become more active and a number of papers on applications and adaptations of the technique have appeared and will be discussed in the following chapter.
Present investigation

The aim of the present investigation was to develop molecular methods for detection and genotyping of specific target DNA and RNA molecules in situ. The work resulted in the three papers presented below. The developed methods all share several key elements regarding the generation of the read out signal, showing the potential of the underlying padlock probe technology for use in a wide variety of research areas. This section will discuss the presented work, introduce some recent developments in the field, and provide future perspectives on the use of padlock probes and target-primed RCA in molecular biology.

Paper I: In situ genotyping individual DNA molecules with padlock probes and target-primed RCA.

This paper describes the first demonstration of target-primed RCA of circularized padlock probes for the detection and genotyping of DNA in situ. The paper has a technological focus, with discussion of several important reaction parameters and design issues, and provides comparisons to related in situ detection methods based on padlock probes. The paper illustrates the usefulness of single-molecule imaging through genotyping of a point mutation in the mitochondrial genome and quantification of the proportion of mutated mtDNA in individual cells in a population.

Background

Already in the first publication on padlock probes in situ detection of nucleic acids was suggested as an application where these probes could provide unique benefits in molecular biology. As the specific probing was combined with RCA, sensitivity was increased to detection of individual molecules. Previously published protocols for RCA-mediated single-molecule genotyping in situ did not take full advantage of the combination of specific probing and RCA, as illustrated in Figure 3. ISH and OLA have been used in combination with RCA for visualization of single-copy genes. This was accomplished by hybridization and, in some instances, ligation of single stranded oligonucleotide probes at the target site. Detection of oli-
gonucleotides interacting with the target sequence was performed by amplification of an in vitro circularized probe hybridized to a non-target complementary part of the linear oligonucleotide\textsuperscript{106}. Unspecific binding of hybridization probes is common in ISH if reaction conditions are not optimized carefully. Furthermore, RCA can be unintentionally primed by ligation templates carried over from the in vitro probe circularization reaction. Using pre-formed DNA circles as template for RCA in situ to detect hybridized oligonucleotide probes is therefore associated with a risk of generating signals in absence of target DNA. It has been shown in vitro that Φ29 DNA polymerase-mediated RCA of circular DNA that can not release itself from the detected target strand is inhibited and results in only a few copies of the template circle\textsuperscript{59}. To avoid this inhibition an in situ DNA denaturation method based on enzymes was used to prepare target DNA for padlock probe recognition in another publication\textsuperscript{105}. Through restriction digestion the enzymatic denaturation procedure allows for precise determination of where target DNA should be fragmented. Performing the DNA digestion close to the target site enables the padlock probe to slide off the target strand when RCA is initiated, and thus escape the inhibitory effect caused by the presence of the target strand. Padlock probes were circularized by specific target-dependent ligation in this publication. However, RCA was initiated by addition of an external primer, leading to the problem of RCPs losing association with the target molecule.

We made adjustments to the enzymatic denaturation procedure to create a 3'-end close to our target sequence. This 3'-end was then used for priming of the RCA reaction, omitting the need of any external RCA primer to be added. The use of the target molecule as primer leads to the method being referred to as ‘target-primed’ RCA. The procedure is illustrated in Figure 3c. Benefits of our proposed reaction scheme are:

- We take advantage of the gentle denaturation conditions provided by enzymatic digestion, preserving the morphology of the sample.
- We ensure specificity of detection by performing padlock probe ligation in the sample with the target DNA as ligation template.
- We prevent inhibition of the RCA reaction by enabling the circular padlock probe to be displaced from the target sequence as the polymerase reads through the circle to synthesize DNA.
- By priming the RCA from the target strand, the resulting RCP becomes covalently linked to the target molecule. This ensures association between the signal and the detected target.

The objective of this study was to demonstrate these properties and to provide detailed mechanistic information. The mitochondrial genome was an ideal target for the method development due to its presence in high copy numbers and the absence of tight packing, as compared to nuclear DNA. The mutation chosen as genotyping target was an A to G transition at position
3243 of the mitochondrial genome. This mutation in a tRNA\(^{(Leu)}\) gene is one of the most well studied human mitochondrial mutations, and is associated with the mitochondrial disorders MELAS (Mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes) and MIDD (Maternally inherited diabetes and deafness). The mutation is also prevalent among diabetic patients and patients with certain types of hearing loss\(^{107}\).

*Figure 3.* Different strategies for single-molecule genotyping of DNA with RCA *in situ*. **a)** Pre-formed DNA circles are amplified by RCA to detect linear oligonucleotides hybridized to heat denatured target DNA *in situ*\(^{106}\). Ligation of an allele-specific probe (black) to an anchor probe (blue) enables SNP genotyping through stringent washing to remove unligated probes. **b)** Target DNA is denatured by restriction digestion on the 5’-side of the target followed by exonucleolytic removal of the non-target strand. Padlock probes are circularized *in situ* and as amplification is initiated, the padlock probes are displaced from the target via the nearby 5’-end\(^{105}\). This allows for efficient RCA, but leads to loss of association with the target molecule. **c)** Target-primed RCA for padlock probe detection *in situ*, as described in Paper I. The target DNA is denatured by restriction digestion at the 3’-side of the target followed by exonucleolyis of the non-target strand. Padlock probes are circularized *in situ* and RCA is primed from the target DNA strand, covalently linking the amplification product to the target molecule. Grey - cellular DNA, red - detection probes and detection probe sites. 3’-ends are indicated by arrowheads, 5’-ends are indicated by blocks. The image is not drawn to scale.
Results

Cell lines wild type, mutant, or heteroplasmic for the mitochondrial A3243G mutation were used as samples in the first part of the paper. First the enzymatic target preparation procedure was demonstrated to work in an RCA-free system by antibody-mediated detection of bound padlock probes. The immunohistochemical detection resulted in high non-specific background originating from unspecifically bound labelled antibodies. When target-primed RCA was combined with the denaturation procedure, the background from unspecific signals disappeared. Advantages in specificity and efficiency of our method compared to previously published RCA in situ protocols were demonstrated. Furthermore, different criteria for successful detection were examined. The overall detection efficiency of the presented method was estimated to 10% by comparison of typical numbers of counted RCA signals in situ, with qPCR measurements of purified DNA samples from one of the cell lines. The specificity and selectivity of the method was demonstrated by genotyping of the A3243G point mutation in two co-cultured cell lines homoplasmic for each of the two different genotypes. A heteroplasmic cell line was then used to demonstrate production of quantitative data by counting of individual RCA products. Using target-primed RCA, the average level of heteroplasmy in this cell line was determined to 66% ± 8% (mean ± s.d.). This was in good agreement with the measurement produced by the gold standard genotyping method based on gel quantification, which gave a result of 64% ± 7%. Additionally, the in situ genotyping visualized the inter-cellular variation of heteroplasmy in the cell population. Finally, mtDNA detection in tissue was demonstrated by genotyping of the A3243G mutation in a fresh frozen tonsil section.

Discussion

This paper was the first report of padlock probes and target-primed RCA in situ, and we visualized for the first time single-nucleotide differences in individual mitochondrial genomes in individual cells. In situ genotyping of co-cultured cell lines of different mtDNA genotype demonstrated the high specificity and selectivity of the method. We also illustrated that the method can be used for relative quantification of target molecules, which is of essence for a single-molecule detection method. The impact of the modifications in the protocol compared to previous in situ RCA techniques was considerable with regards to specificity and detection efficiency. The target-dependent ligation and the covalent linkage of the RCP to the target molecule confined signals to the cytoplasm of cells. In addition, the target-priming was superior in detection efficiency compared to priming of RCA with an external primer.

In this paper we put great effort in presenting a robust protocol for detection of mtDNA molecules. The method has potential to be used for detection
of any cytoplasmic DNA molecules that carry, or can be manipulated to carry, a nearby free 3'-end. The method quickly attracted interest from groups involved in mitochondrial research, and collaborations were initiated to investigate properties of mtDNA inheritance and segregation. Being applicable on both cell and tissue samples, mechanisms of the mitochondrial bottleneck should be possible to investigate with target-primed RCA directly in biological samples. This has so far resulted in one paper describing assays for segregation analysis\textsuperscript{108}. As the threshold effect causes disease to present differently depending on mutation load in different tissues and organs of the body\textsuperscript{37}, visualization of disease-causing mutations in tissue is another application within mitochondrial biology that could be of interest. mtDNA mutations in cancer have recently gained attention, although first discovered long ago. Most mtDNA mutations found associated with cancer are somatic mutations, however a few examples of inherited tumour-predisposing mutations exist\textsuperscript{109}. The biological relevance of mtDNA mutations in cancer is unclear and investigation of tumour tissue to look at mtDNA genotype levels in individual cancer cells could be an interesting and relatively straight-forward application of the technique.

In a collaborative study, we applied the method on comet assay preparations, resulting in the work presented in Paper II. The method was later applied for detection of DNA on metaphase chromosomes by a co-author on Paper I, but due to low detection efficiency only highly repetitive sequences could be stably detected\textsuperscript{110}. An independent research group recently applied the assay for detection of Anaplasma subspecies bacteria in cultured cells, combining specific DNA detection with immunofluorescent staining of host cell antigens\textsuperscript{111}. The fact that an independent group could adapt the protocol for a completely different application, with only relatively minor target-dependent modifications, indicates that we have provided a stable protocol that can be easily adapted to suit the needs of the user. The target-priming strategy has furthermore been used in assays for SNP genotyping with branched RCA in homogeneous phase\textsuperscript{112}, and for SNP genotyping with chemiluminescent detection of RCPs on a solid phase surface\textsuperscript{113}. 
Paper II: Detection of Alu sequences and mtDNA in comets using padlock probes

In this paper the method described in Paper I was applied for DNA detection in the Comet assay. As in Paper I, mtDNA was detected and the method was then applied also for detection of repeated Alu sequences in nuclear DNA using the same reaction scheme.

Background

Single-cell gel electrophoresis, also referred to as the comet assay, is a method for detecting and measuring DNA damage and repair in individual cells\textsuperscript{114}. In this assay cells are embedded in agarose, lysed in the gel and subjected to electrophoresis. Lysis of the cells leads to removal of most histones and other nuclear proteins, leaving the DNA as nucleoids with supercoiled DNA loops in the gel. If breaks are present in the DNA, the supercoiling will be relaxed and the broken DNA will move away from the nucleoid core during electrophoresis. This causes the nucleoid to adapt an extended shape resembling a comet, with the head-part consisting of the remaining DNA in the nucleoid, and the tail-part consisting of DNA associated with strand breaks. The amount of DNA breaks in the cell determines the proportion of DNA in the comet tail relative to the head. The assay is used in many research areas, including toxicology studies of compounds and basic investigations of DNA repair mechanisms.

Fluorescence \textit{in situ} hybridization has been applied to study the location of specific DNA sequences within comet preparations\textsuperscript{115}. Since hybridization needs to take place in agarose gels, standard protocols for FISH have had to be modified to keep the gel from being destroyed. For example, high temperatures have to be avoided in order to prevent the agarose from melting. Therefore chemicals rather than heat are used for denaturation and for increasing the stringency in hybridization and washing steps. Also, long incubation times for hybridization of labelled probes are used, up to several days\textsuperscript{116}. Faster and gentler techniques for investigation of the localization of specific DNA sequences in comet preparations are wanted. Furthermore, it was unknown what happened to mitochondrial genomes within comet assay preparations. In this study we evaluated the method from Paper I for use in the comet assay, aiming to demonstrate:

- That gentle enzymatic target denaturation can be used in these sensitive preparations.
- The suitability of using padlock probes and RCA within the preparations with regards to signal strength, background and efficiency of detection.
- The localization of mtDNA within the preparations.
Results

To investigate the fate of mitochondrial DNA in comet preparations, we carried out target-primed RCA of an mtDNA specific padlock probe at different stages of the assay, using in essence our previously published protocol (Paper I\textsuperscript{117}). mtDNA detection was first performed in cells embedded in agarose gels without carrying out lysis or electrophoresis. This demonstrated that the different steps of the protocol function within the gel. The mtDNA-specific padlock probe was known from previous work to be detected \textit{in situ} with high efficiency, and distinct RCA products could be seen surrounding the cell nuclei in the gel embedded cells. Then the effect of exposure of cells to lysis solution and electrophoresis was examined. The results showed that the mtDNA started to drift away from the cell nuclei in the gel directly after addition of lysis solution. After electrophoresis had been carried out, association between the mtDNA and the cellular DNA had been completely lost. The results indicated that rather than just migrating away from the nucleoids, mtDNA is most likely washed out of the gel after release from the cellular membranes.

We also detected sequences in nuclear DNA to show that the target-primed RCA method was not inhibited by cells having gone through different steps of the assay. The Alu element, a repetitive unit spread over the genome in primates, representing approximately 10\% of the human genomic DNA\textsuperscript{3} was chosen as target. Due to the vast number and variability of Alu sequences in the human genome it would be impossible to design specific probes to detect all of these simultaneously. We therefore chose the conserved 26 bp Alu core sequence as target for our padlock probe. This sequence was reported to be present in a large proportion of Alu elements in an early version of GenBank\textsuperscript{118}. For restriction digestion we used the restriction enzyme \textit{AluI}, which has given name to the element and whose restriction site is present in the approximately 300 bp long consensus sequence for Alu\textsuperscript{119}. Using this enzyme for digestion, we hoped that we would be able to detect at least some of the available sequences. Results showed detection of Alu sequences in all stages of the comet assay, and signals were present in both heads and tails of comets. In contrast to the mitochondrial DNA, the signals from Alu sequences did not show signs of dispersal out of the gel, further supporting our finding that mtDNA is lost from comet assay preparations.

Discussion

In this paper we demonstrated that padlock probes and target-primed RCA can be used for specific DNA detection within agarose gels. Targeting the core sequence of the highly repetitive Alu element, we showed that our method provides a comparatively gentle and quick way of performing the same type of analysis as FISH with low background in comet preparations.
Since mtDNA was found to disperse from cells and possibly be washed out from gels during preparation procedures, padlock probe detection is however not suitable for investigation of damage and repair in the mitochondrial genome in the comet assay.

Trials to detect low-copy sequences with target-primed RCA on regular fixed cell preparations and metaphase chromosome spreads have reported low efficiency of detection\textsuperscript{60, 110}. This low efficiency is probably linked to poor target availability due to the very tight and complex arrangement of DNA in the cell nucleus. Subjecting cells to harsher denaturation conditions to make the target DNA more available easily leads to loss of genetic material from the sample. This delicate target-preparation balance has so far prevented stable \textit{in situ} detection of nuclear single-copy DNA sequences with padlock probes. Conversely, lysis in the comet preparation protocol renders DNA almost “naked” in the gel, readily available for analysis. With the supporting gel matrix that keeps the DNA from dispersing, detection of single copy genes with padlock probes could be possible within these preparations. It would then be possible to look at damage and repair to specific genes and even regions of genes in the assay, subjects currently being investigated with FISH. A limitation of FISH analysis in the comet assay is that the resolution is low, making it necessary to use very large probes. This causes localization of signals in the comet tail to actually only be indicative of that DNA damage has occurred in the vicinity of the investigated gene\textsuperscript{115}. With higher resolution, as offered by padlock probing, true gene-specific effects could be studied by targeting multiple sites within single genes.
Paper III: *In situ* detection of individual mRNA transcripts with single nucleotide resolution.

This paper describes the application of padlock probes and target-primed RCA for detection of mRNA *in situ*. The transfer of the padlock probe technique from DNA to RNA detection has not been straight-forward. In this paper we present a robust and reproducible method for visualization of transcripts with single-molecule resolution. We apply the method for detection of protein isoform mRNAs and transcripts from different cancer-related genes.

Background

Detection of RNA targets using padlock probes has been shown to be less efficient and less specific than detection of DNA targets\(^\text{120, 121}\). The reduced ligation rate is due to a drop in ligation efficiency when using RNA rather than DNA as ligation template. Attempts have none-the-less been made to use padlock probes for *in situ* detection of RNA\(^\text{105}\). Although, this protocol has not been reproduced in any subsequent publications, which may be an indication of that the detection efficiency is low. Inefficient padlock probe detection was recently documented in a paper describing RNA *in situ* detection based on circularizeable turtle probes and RCA\(^\text{122}\). In this study, turtle probes had far higher detection efficiency than padlock probes when directed at the same RNA target. Turtle probes are padlock-like probes that undergo ligation to form circular molecules using an intramolecular hairpin structure as ligation template\(^\text{122}\). Since the ligation takes place on an internal hairpin in the turtle probe, small sequence differences in the target can not be assayed. The self-templated ligation can furthermore cause specificity problems if washing steps are not sufficiently stringent, as discussed for RCA using pre-formed DNA templates\(^\text{106}\) in Paper I. As RNA can work as primer for RCA with Φ29 DNA polymerase\(^\text{66, 67, 122}\), target-primed RCA can be used for detection of RNA molecules with padlock- or turtle probes. A considerable limitation however, is that since the RNase activity of Φ29 DNA polymerase seems to be inhibited by double stranded RNA formed in secondary structures\(^\text{122}\), priming can only be done from RNA ends located very near the target site. This limits detection to sequences in the 3’-end of non-polyadenylated RNA, which excludes analysis of the majority of mammalian mRNAs.

We developed a method in which we first perform reverse transcription *in situ* to create a cDNA molecule. Using cDNA as ligation template ensures high specificity and efficiency of padlock probe circularization. This way, the SNP-genotyping resolution offered by padlock probes can be utilized for RNA detection. RCA is primed by the 3’-end of the cDNA molecule, keeping the RCP located at the site of the target molecule. A schematic compar-
son of the cDNA detection strategy with direct RNA detection with padlock- and turtle probes is shown in Figure 4. The aim of the study was to investigate important reaction parameters and then apply the method to demonstrate unique features of the technique. We further wanted to exemplify research areas where the technique could offer important contributions.

Results

Optimizations were mainly carried out on steps that were new for this method or were most likely to differ from detection of mtDNA. Reverse transcription has been carried out in situ previously, but then detection of the formed cDNA molecules was done by real-time PCR following laser microdissection of individual cells. This publication reported that a majority of the reverse transcribed cDNAs were of full length. Reverse transcription is also carried out for RNA detection with in situ RT-PCR, recommending

![Diagram of target-primed RCA for detection of RNA molecules in situ](image)

*Figure 4. Target-primed RCA for detection of RNA molecules in situ. a) The method used in Paper III. RNA is reverse transcribed into cDNA, which is targeted by padlock probes that are circularized upon recognition of the target sequence. b) Direct targeting of RNA with padlock probes in situ. c) RNA is targeted by turtle probes that hybridize to the target sequence and then undergo ligation on an intramolecular hairpin structure. Target-primed RCA detects circularized probes in a-c. Note that the target sequence must be located next to an RNA 3'-end for RCA priming in b and c, whereas target location in a does not have this limitation. Grey - RNA, black - DNA, red - detection probes and detection probe sites. 3'-ends are indicated by arrowheads, 5'-ends are indicated by blocks. The image is not drawn to scale.*
cDNA synthesis length with specific primers to be kept below 1500 bp to be efficient\(^{103}\). In contrast to these two reports we saw that the cDNA molecules detected \textit{in situ} with padlock probes were mainly short. LNA has been proposed as a great enhancer for ISH efficiency\(^{91}\), and has made \textit{in situ} detection of small RNA molecules possible\(^{92}\). We found that incorporating LNA bases in the cDNA primer greatly enhanced transcript detection \textit{in situ}. In contrast to turtle probes that are confined to detection of RNA 3’-ends\(^{122}\), we could detect several different sites located on the same transcript with sustained detection efficiency. Detection of a single nucleotide difference in mRNA molecules in two co-cultured cell lines demonstrated the high selectivity of the method. Furthermore, we showed that multiplexed detection can be carried out for expression analysis of both tissue and cultured cells. Different transcript isoforms were shown to have distinct mRNA localization patterns in fresh frozen sectioned tissue. Furthermore, different cell lines where analyzed for the presence of cancer-related transcripts, showing that the mRNA expression profile differed in cancer cell lines compared to a primary fibroblast cell line without known malignancy.

\section*{Discussion}

In this paper we demonstrated a highly specific method for expression analysis with single-molecule resolution at the single-cell level. Method development and application based on single-molecule visualization has been quite active recently, with publications applying methods based on both hybridization\(^{84, 86, 124, 125}\) and on signal amplification\(^{122}\). Hybridization-based methods have a great advantage in that they are very sensitive, detecting in essence all targeted mRNAs\(^{86, 125}\). In contrast, padlock probes and target-primed RCA is limited to detection of rather highly transcribed genes. Conversely, padlock probe detection is highly specific. In contrast, hybridization, in particular with heavily labeled probes\(^{84, 124, 125}\), is associated with a risk of generating false positive signals due to probes binding unspecifically to the sample. Single-molecule visualization through hybridization of multiple probes carrying single fluorophore-labels is less sensitive to false positives due to the requirement of several probes to bind to the same transcript\(^{86}\). Due to the necessity of having long sequences available for hybridization, this strategy instead suffers from a lower ability to distinguish between similar transcripts. Single-molecule detection with amplification of turtle probes, on the other hand, requires only enough dissimilarity between two sequences to allow for specific hybridization of 30-40 bp of probe\(^{122}\). However, ultimate target selectivity among single-RNA molecule visualization methods thus far is only provided by padlock probing, being able to separate sequences differing only in one nucleotide position.

Point mutation detection in mRNA \textit{in situ} was recently reported also in an assay based on loop-mediated isothermal amplification (LAMP)\(^{126}\). LAMP is
a complex but rapid amplification technique requiring binding of several primers to ensure high specificity\textsuperscript{127}. For \textit{in situ} detection of mRNA, the LAMP method was combined with an allele-specific reverse transcription, where primer extension was inhibited if mismatched to the target. The \textit{in situ} LAMP method does not offer data on single molecules, but instead scores cells as positive or negative for a mutation. The short assay time and being applicable on routine paraffin embedded samples are advantageous for the \textit{in situ} LAMP method, but authors reported need for improvement of specificity as false positive samples were detected \textsuperscript{126}. Padlock probes and RCA for point mutation genotyping is so far restricted for use on fresh frozen tissue sections. Benefits in comparison with LAMP are instead the high selectivity and the possibility to generate quantitative data for multiple transcripts in individual cells.

**Perspectives on Papers I-III**

The papers presented in this thesis describe development and application of the \textit{in situ} target-primed RCA technique for detection of padlock probes circularized in a target-dependent reaction. So far, the target-primed RCA technique is most suitable for detection of molecules in the cell cytoplasm. This may be considered a big limitation, but as it is becoming increasingly apparent that the biological processes going on within cells are much more complex than originally thought, it becomes more of a detail in the margin. RNA molecules for example are assigned more and more functions in the maintenance and regulation of cells\textsuperscript{20}, and the known complexity of the transcriptome is constantly increased. Alternative splicing is common for human genes, creating genetic diversity beyond the limited number of genes present in the genome. It has been shown that a large proportion of the gene expression variation in one of the populations characterized in the HapMap project is based on transcript isoforms\textsuperscript{128}. Disruption of splicing patterns has furthermore been linked to a number of diseases\textsuperscript{129}. Apart from splicing, random monoallelic expression has been found to be widespread for genes on human autosomes\textsuperscript{130}. This adds a class of monoallelically expressed genes to the X-inactivated and imprinted gene classes. The above mentioned example studies have all been undertaken \textit{in vitro}, as average measurements of what is going on in an investigated cell population. With the notion that gene expression is highly variable between cells, occurring in “bursts”\textsuperscript{124}, it would obviously be interesting to explore this transcriptional variation on a single-cell level. This provides just a few examples of where the specificity of padlock probes could be utilized and where other \textit{in situ} detection methods are limited by low target selectivity. Similarly, also mitochondrial DNA heteroplasmy levels can be highly variable between cells, which can be visualized by padlock probing as demonstrated in Paper I. Another application area
where individual cell status is of interest is in infection by pathogens. Single-molecule visualization with RCA in this area has been exemplified by detection of the Epstein-Barr virus (EBV) in specific cells in tissue with turtle probes\textsuperscript{122} and detection of Anaplasma bacteria in infected cells by padlock probes\textsuperscript{111}.

Image analysis tools provide important contributions to the use of single molecule methods since they permit more unbiased analysis of the generated data. The quantitative data presented in Paper I was counted by hand in a time-consuming fashion. Image analysis software for analysis of data from \textit{in situ} assays using RCA has since then been developed (BlobFinder\textsuperscript{131}). Automated RCP counting reduces bias and allows for larger populations to be analyzed in a relatively short time. Image analysis furthermore facilitates analysis of multiplexed detection assays\textsuperscript{71}. The BlobFinder software was used for RCP quantitation in Paper III. As average measurements based on many cells were sufficient to determine the overall efficiency of optimization procedures, we did not take into account cell-to-cell variation in this paper. Nevertheless we did observe substantial differences in expression level between cells at times, demonstrating the need of suitable image analysis to fully take advantage of the strength with this and other single-molecule \textit{in situ} detection methods.
Summary

The target-primed RCA method for detection of padlock probes *in situ* was first described in Paper I, where we genotyped point mutations in mtDNA molecules. In Paper II we detected Alu repeat sequences and mtDNA in individual cells in comet assay preparations. Finally, in Paper III we developed target-primed RCA for padlock probe based detection of *in situ* reverse transcribed cDNA molecules, demonstrating single nucleotide variation in mRNA transcripts. New types of *in situ* analyses are now made possible by the combination of excellent genotyping selectivity and single-molecule detection sensitivity offered by padlock probes and target-primed RCA.

The findings of the work included in this thesis can be summarized as follows:

- Padlock probes can genotype single-nucleotide variants of DNA sequences *in situ*, providing visualization of genetic variants directly in biological samples.

- Target-primed RCA is an efficient amplification method for padlock probe detection *in situ*, enabling highly specific single-molecule detection within cells.

- The priming of RCA from the target strand enables localized detection by creating the RCP as an extension of the target molecule.

- Single RCPs can be counted, enabling relative quantification of target molecules in individual cells.

- Padlock probes and target-primed RCA can be used for analysis of cultured cells and sections of fresh frozen tissue.

- Padlock probes and target-primed RCA provide increased detection resolution and a gentle and quick alternative to FISH for analysis of specific nuclear DNA sequences in comet assay samples.
Acknowledgements

The work presented in this thesis was performed at the Department of Genetics and Pathology at Uppsala University. Financial support was provided by Knut och Alice Wallenbergs stiftelse, Beijerstiftelsen, Stiftelsen för Medicinsk Forskning, Svenska Läkarsällskapet, Linnéstiftelsen, Borgströms stiftelse, Göran Gustafssons stiftelse, Erik K Fernströms stiftelse, Vetenskapsrådet, the EU Framework Program 6 projects MolTools and COMICS, and the EU Framework Program 7 project READNA.

I would like to express my gratitude to everyone that has contributed in any way to support me on my road to finishing this thesis. Especially I would like to thank the following persons in a more informal way:

My supervisor Mats Nilsson for taking me on as a PhD student and teaching me everything there was to know about in situ analysis with padlock probes, both theoretically and in the lab. Many thanks for your support and enthusiasm during these years.

My co-supervisor Ulf Landegren for interesting discussions and for providing valuable input on the work.

All the collaborators that I have worked with during these years, who have helped identifying important areas of application for our methods. Thanks also to our collaborators at CBA at Uppsala University for providing us with image analysis tools.

Present and former members of the research groups for Molecular diagnostics and Molecular medicine. Thanks for providing good company at the lab! - The girls of the “old” office for many interesting discussions about everything and research. Jenny, my pretend twin, for endless advice on the quality of detection probes, company at various conferences and courses, and for always getting yourself into funny situations. Sigrun, for fun times at the Santorini conference and fun parties in the old days. I hope that you will have a great time in Boston with the family. Edith, for many and long discussions. Malin, for finishing after me 😊. Good luck with the expanding family and what comes afterwards.
- Ida and Sara, my fellow *in situ* padlockers. Thanks for good collaborations and for taking care of my cells when I don’t have the time/energy to do it myself. A special thanks to Ida for all the hard work put into the cDNA project and the figures for Paper III. Good luck with the future work and your theses.

- The post-docs for sharing your wisdom. Mathias for being the best office neighbor a PhD student could have (you know everything!), for answering all my PCR questions and for helping me proofread this thesis. Carolina W for all the help with image analysis. Ola, for helpful advice on cells and methods. Masood, Yuki, Katerina, Reza and Ling for your positive attitudes and challenging questions. Jonas J, Johan S and Olle, for setting good examples of how to defend a thesis. A special thanks to Jonas J (and Malin) for helping a young PhD student with an intensifying panic attack to submit the figures for her manuscript late at night many years ago!

- The PhD students. Tim, my lab neighbor, for your never-ending helpfulness with everything. And for running ICA Tim of course. Henrik, for keeping the level of irony at a constant high in the lab. Kalle, for hosting a much needed workshop on Illustrator together with Mathias. Magnus, Irene & Spyros, for nice company and interesting discussions at group meetings and other get-togethers. The (relative) newcomers, Rachel, Gucci, Agata, Johan V, Maria, Camilla and Rongqin. Good luck with your research everyone!

- Lena and Carla for organizing the lab and keeping track of everything. Johan O and everyone else that has helped me with various computer issues during these years. Anders for agreeing to function as my taxi to and from various events and for chairing the thesis defense. Elin, for reminding me of Silsoe-times.

- The former group members; Johan B, Fredrik, Mia, Carolina R and Jonas M.

- All the students that I have supervised who have helped out in the projects. Pernilla, without you we (and certainly not the cells) would not have survived the last few months. Lotta and Ylva; many thanks for your hard work during your SOFOSKO projects. I will let you know when the project finally finishes (because it will!). Pulkit, you have an alternative view on things and I hope that your ambitions take you as far as you want. Tabish, keep up the hard work and the commitment. Björn, you did an excellent job as my first ever PRAO-student.

Researchers, students and the administration at IGP and Rudbeck Laboratory. A special thanks to my new office neighbors Lore Zech and Karl-Henrik Gustavson for providing a quiet and relaxed space to write this thesis. I very much enjoy your inspirational stories about old times. Thank you also Lore for valuable input on the method development.
Friends. Thank you for support provided outside of the lab. I especially want to thank:

- *Cecilia, Anna, Rebecka, Lisa & Nina* with families. Thanks for more than a decade of fun parties, trips and get-togethers. Although the road to Stockholm (and certainly Växjö) feels like a million miles sometimes it’s always worth the trip to see you!

- *Miriam (& Hasse)* for many fun and late nights “when we were young”, and fika times and dinners at work and with the families.

- *Sofia & Daniel* for being such nice people. Thank you for babysitting and giving us a few moments of peace now and then.

- *Emma* for providing some grown-up contact and even some talk about research during walks with Valle and Ida.

- *Jon*, for helping me entertain a restless soul.

- The Nyköping people for providing a refreshing environment 100% free of talk about ligases, polymerases and probes. You are the perfect vacation company.

If I for some reason have forgotten to mention you above, I apologize. Please come see me and I will thank you in person.

Finally, my family and relatives. Tack för allt!

- The new Larsson family. *Anne, Thomas, Andreas, Johan & Gunni.*


References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 431

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-98538