Genotyping RNA and DNA Using Padlock Probes

BY

DAN-OSCAR ANTSON
ABSTRACT

Novel techniques are needed to investigate the genetic variation revealed in the first draft of the human genome sequence. Padlock probes are recently developed reagents, suitable for detecting single-nucleotide variations of DNA and RNA in situ or in solution. The probes are oligonucleotides of about 70-140 nucleotides that can be circularized by ligation in the presence of a correct target sequence. Standard chemical synthesis of padlock probes is difficult due to the requirement for intact 5’ and 3’ ends of these long oligonucleotides.

A novel PCR-based method is presented in this thesis, whereby longer, densely labeled padlock probes can be made as compared to conventional chemical synthesis. PCR-generated padlock probes produced a stronger signal and a more resolved staining pattern, compared to chemically synthesized probes in fluorescence in situ analysis of an alpha-satellite sequence variant present in human chromosomes 13 and 21. Padlock probes used for in situ analysis of metaphase chromosomes had an optimal length of 140 nucleotides. They were used to identify individual chromosomes 7 and 15, and to follow the transmission of chromosome homologues for two consecutive generations. The specificity of the padlock probes to detect single copy genes in genomic DNA samples was demonstrated by detecting a single-nucleotide mutation in the ATP7B gene.

It has not previously been known if T4 DNA ligase can be used for RNA sequence analysis. In this thesis, it is demonstrated that T4 DNA ligase can be used for distinguishing single-nucleotide RNA sequence variants. Reaction conditions were defined where most mismatches could be discriminated by a factor of 80 and all mismatches by a factor of at least 20. Under these conditions padlock probes could detect and distinguish RNA sequence variants with ligation efficiency almost as high as on the corresponding DNA sequence.

A detailed study of the parameters influencing RNA-templated DNA ligation revealed that DNA ligation on RNA templates proceeds at a much slower rate compared to the same reaction on DNA, and that a molar excess of enzyme is required. Furthermore, the ligation reaction is inhibited by high concentrations of the cofactor ATP and NaCl.

The work presented in this thesis demonstrates that PCR-generated padlock probes can detect and distinguish single-nucleotide variation in both RNA and DNA.

Key words: Padlock probe, enzymatic synthesis, PCR, in situ hybridization, RNA-templated DNA ligation.

Dan-Oscar Antson, Department of Genetics and Pathology, Section of Medical Genetics, Rudbeck laboratory, Uppsala University, SE-751 85 Uppsala, Sweden

© Dan-Oscar Antson 2001

ISSN 0282-7476
ISBN 91-554-5057-1

Printed in Sweden by Lindbergs Grafiska HB, Uppsala 2001
This is the book I never read
These are the words I never said
This is the path I’ll never tread
These are the dreams I’ll dream instead
This is the joy that’s seldom spread
These are the tears...
The tears we shed
This is the fear
This is the dread
These are the contents of my head
And these are the years that we have spent
And this is what they represent
And this is how I feel
Do you know how I feel?
‘Cause I don’t think you know how I feel

Excerpt from Why by Annie Lennox 1992
MAIN REFERENCES

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


* Authors contributed equally to the work.
Reprints were made with permission from the publisher.
# TABLE OF CONTENTS

## ABBREVIATIONS .................................................................................................................. 6

## BACKGROUND .................................................................................................................... 7

RNA DETECTION METHODS ........................................................................................................ 7
Northern blot ............................................................................................................................ 8
Real-time PCR ........................................................................................................................... 8
The invader technique ............................................................................................................... 9
DNA microarrays ..................................................................................................................... 9
SAGE ..................................................................................................................................... 10

THE HISTORY OF CYTOGENETICS ...................................................................................... 10

IN SITU HYBRIDIZATION ....................................................................................................... 11

TECHNIQUES FOR IN SITU SNP ANALYSIS ........................................................................ 12
Allele-specific hybridization ................................................................................................. 12
Primed in situ labeling ........................................................................................................... 13
In situ PCR ............................................................................................................................. 13

## INTRODUCTION TO PRESENT INVESTIGATIONS ......................................................... 14

PADLOCK PROBES ............................................................................................................... 14
SYNTHESIS OF PADLOCK PROBES .................................................................................. 16
Chemical synthesis ................................................................................................................. 16

LIGASES AND THE LIGATION MECHANISM ...................................................................... 17

CENTROMERE FUNCTION AND STRUCTURE ................................................................... 18

## PRESENT INVESTIGATIONS ............................................................................................. 20

I. PCR-GENERATED PADLOCK PROBES DETECT SINGLE-NUCLEOTIDE VARIATION IN GENOMIC DNA .......................................................................................................................... 20
II. ENHANCED DETECTION AND DISTINCTION OF RNA BY ENZYMATIC PROBE LIGATION ................................................................................................................................. 22
III. RNA-TEMPLATED DNA LIGATION FOR TRANSCRIPT ANALYSIS .......................... 24
IV. PCR-GENERATED PADLOCK PROBES DISTINGUISH HOMOLOGOUS CHROMOSOMES AT A SINGLE-NUCLEOTIDE RESOLUTION ENABLING QUANTITATIVE IN SITU GENOTYPING .................................................. 26

## DISCUSSION .......................................................................................................................... 28

IN SITU GENOTYPING USING PCR-GENERATED PADLOCK PROBES ....................... 28
RNA TEMPLATED DNA LIGATION ...................................................................................... 29

## CONCLUDING REMARKS ................................................................................................ 29

## SWEDISH SUMMARY ......................................................................................................... 30

## ACKNOWLEDGEMENTS ...................................................................................................... 31

## REFERENCES ....................................................................................................................... 33
ABBREVIATIONS

ATP  adenosine triphosphate
cDNA complementary DNA
DNA  deoxyribonucleic acid
DMT  dimethoxytrityl
FISH fluorescence in situ hybridization
FRET fluorescence resonance energy transfer
HPLC high pressure liquid chromatography
ISH  in situ hybridization
kb   kilo bases
NaCl sodium chloride
NAD  nicotinamide adenine dinucleotide
NMN  nicotin mononucleotide
nt   nucleotides
PCR  polymerase chain reaction
RCR  rolling circle replication
RNA  ribonucleic acid
RT   reverse transcription
SNP  single-nucleotide polymorphism
BACKGROUND

With the first draft of the human genome sequence completed, the scientific community must confront new challenges. Apparently, humans only have about 30,000 to 40,000 protein-coding genes, roughly twice as many as a fly or a worm (1). Despite the surprisingly small number, studying the intricate expression pattern of all of our genes, identifying their function, and investigating the complex genetic variation present within them will be an arduous task. To overcome these challenges novel methods are needed for accurate, efficient, and simultaneous detection of genetic variation and expression patterns among a large number of genes. Several methods that have been developed to meet these challenges are reviewed herein, but the focus of this thesis will be on a recently developed method for detecting single-nucleotide variations in DNA and RNA using so-called padlock probes. Padlock probes are particularly suitable for the detection of single-nucleotide variations in situ, such as for the in situ genotyping of centromeres based upon sequence variations in alpha-satellite sequences, which is presented herein. Furthermore, several improvements enhancing the utility of this method will be described and discussed.

RNA detection methods

Messenger RNA transcripts, mRNA, play an important role in living cells as an intermediate step in the protein synthesis process. The presence of specific transcripts or changes in their expression levels may provide information about cellular processes at the protein level. RNA detection methods span from the specific quantitative detection of a few transcripts of interest to the large-scale parallel detection of thousands of different RNA molecules. Methods detecting specific RNA transcripts such as northern blot, real-time PCR or the invader technique can be used to detect and quantify transcripts present in different tissues or cell types. DNA microarrays (2) and the SAGE technique (3) can be used to study complex differences in expression patterns of thousands of transcripts in tissues or cells undergoing changes in their cellular program.
**Northern blot**

The starting point for the development of RNA detection methods was in 1977 when Alwine et al. described a method for the sensitive detection of specific RNA molecules. In this method, total RNA or mRNA is isolated and size-separated on an agarose gel before being transferred to a membrane. The RNA molecule of interest can then be detected by the hybridization of a labeled target-specific probe (4). This method is an extension of the popular DNA analysis method called Southern blot (5), and became known by the name of northern blot. Not only can RNA molecules be detected in a specific and sensitive manner, but changes in RNA levels can be monitored as well. In addition, alternative RNA splice forms can be detected due to size differences from insertions or deletions of exons. One drawback with the northern blot technique is that single-nucleotide differences are difficult to detect. Similar genes can therefore be difficult to distinguish with this method. The number of genes that can be investigated in one experiment is also limited.

**Real-time PCR**

Highly sensitive detection of specific RNA species in a complex RNA sample can be achieved if the reverse transcription (RT) reaction is followed by PCR. Real-time detection of the ongoing PCR via the TaqMan assay (6,7) exploits the nuclease activity of the Taq-polymerase (8). A probe complementary to the amplified target sequence, labeled with a reporter fluorophore at the 5’-end and a quencher fluorophore at the 3’-end (9,10), is added to the PCR. During the extension step Taq-polymerase degrades hybridized probes, releasing the 5’ fluorophore. The fluorescence from released fluorophores is proportional to the amount of degraded probes, and hence the amplification cycle at which the fluorescence exceeds a threshold value can be taken as a measure of the amount of target present in the reaction. Alternative probe designs have been developed for real-time detection, such as molecular beacons (11) and scorpion probes (12). These probes contain reporter fluorophores and quencher molecules as in the TaqMan probes previously described, but are constructed so that degradation of the probes is not necessary. Furthermore, they can be used to distinguish transcript variants that differ in a single-nucleotide position (6,7). The PCR step renders the Taqman method sensitive to false positive results due to contamination of the samples, and multiplexing is difficult, demanding labor-intensive optimizations. Biased results may arise from the reverse transcription step. This is avoided in the invader technique, since RNA is used as the target in the reaction.
The invader technique

A recently developed method for the quantitative detection of SNPs in DNA or RNA is the invader technique. Two oligonucleotides are hybridized to the targets so that the 5’ end of the downstream probe is prevented from hybridizing by the 3’ end of the upstream probe. The displaced 5’ ends can be cleaved by FLAP endonucleases, so-called cleavases. The cleaved 5’ part of the molecule can then be detected, provided it is labeled. The reaction is cyclic because it is performed at a temperature that allows dissociation of cleaved oligonucleotides and hybridization of new uncleaved oligonucleotides. The cleavases are sensitive to mismatches at the cleavage junction between the target sequence and oligonucleotide with the 5’ overhang. A correctly matched 5’ overhang oligonucleotide is cleaved while a mismatched oligonucleotide remains intact (13). The method has been developed to detect RNA molecules as well, using a two-step FRET-based invader assay (14). Due to the discriminating power of the cleavase enzyme, highly homologous genes can be accurately detected and quantified. The method does not allow for multiplexing to any high degree, thus limiting the number of genes that can be studied in an experiment.

DNA microarrays

High-throughput parallel analyses of many transcripts or target sequences are possible with DNA microarray technology. There are two principally different approaches to array manufacturing. The first approach is to synthesize oligonucleotides in situ on the array (15). The other approach is to construct microarrays by spotting oligonucleotides or cDNAs onto the surface (2,16). DNA microarrays are used to analyze patterns of gene expression level differences by comparing two subsets of RNA transcripts. The mRNA transcripts are converted into fluorescence-labeled cDNA in a RT reaction. The labeled cDNA is thereafter hybridized to the DNA microarray. Fluorescent signals of hybridized DNA can be detected after theremoval of unspecifically bound DNA using stringent washes. Although the transcription of a large number of genes can be studied in parallel using DNA microarrays, the results obtained must be evaluated with caution. The RT reaction may skew the ratios of different transcripts by favoring reverse transcription of some mRNA variants while others are poorly transcribed. The expression levels of transcripts from individual members of a gene family may be difficult to determine accurately using DNA microarrays due to cross-hybridization of similar cDNA molecules. Kane et al. have showed that a high level of cross-hybridization occurs between DNA molecules that are >75% homologous over a stretch of 50 nucleotides (17).
**SAGE**

Serial analysis of gene expression (SAGE) is another approach used to study expression levels of genes (3). In this method, 5' biotinylated reverse transcription reaction products are isolated by using streptavidin-coated beads. After several restriction and ligation steps, a final product of concatenated short sequences characteristic of specific mRNA molecules are cloned and sequenced. By enumerating the short sequence tags, the expression of individual gene levels can be calculated from the frequency of occurrence in the concatenated products. A similar approach has recently been developed for gene expression analysis whereby short sequence tags immobilized on a microbead array are sequenced (18). A cDNA library is constructed by cloning cDNA templates into a cloning vector containing a 32-mer oligonucleotide tag sequence. The inserts are PCR amplified and rendered single-stranded. The single-stranded tagged PCR products are hybridized to a population of microbeads, each with a different tag complementary sequence attached to the surface. The beads are immobilized in a random two dimensional array. Finally, a short part of the cDNA product present on each bead is sequenced through repeated cycles of restriction, ligation of linker molecules with a zip code sequence, and detection of ligated linker through hybridization of fluorescently-labeled probes complementary to the zip code sequence.

With the SAGE and microbead array approaches, expression levels of both known and unknown genes can be studied. The expression levels obtained by SAGE correlate well with those obtained screening cDNA libraries for the same genes. One drawback with the method is that the discrimination of similar genes with this method is difficult since only a short motif is used to identify individual genes.

**The history of cytogenetics**

In 1879, Alexander Fleming named the substance present in the nucleus of cells chromatin, a substance which he observed was organized in distinct structures. These structures were later christened chromosomes by Waldenayer in 1888, who also described the behavior of the chromosomes during mitosis. The distribution of the chromosomes during meiosis was described by Strasbourg the same year (19). The next milestone in the history of cytogenetics was in 1915 when Thomas H. Morgan postulated from genetic crossing experiments that genes were arranged linearly on chromosomes (20). Although chromosomes could be studied, the quality of the human cell preparations during the first half of the twentieth century was of such poor quality that the exact number of human chromosomes remained uncertain until 1956. That year Tjio and Levan established that the human normal diploid number of chromosomes was 46. This discovery was a consequence of the development of
better techniques to obtain metaphase cell preparations, such as the hypotonic
treatment of cells (21) and the addition of colchicine to cell cultures to arrest
cells in metaphase (22-24), making it possible to produce abundant numbers of
metaphase cells of high quality. This development led to the discovery of a
number of chromosomal aberrations. In 1959, trisomy 21 was shown to be
associated with Downs syndrome (25). The same year a missing and
supernumerary X-chromosome was discovered to be associated with Turner’s
(26) and Klinefelter’s (27) syndromes, respectively. At the end of the 1960s, the
ability of certain dyes to preferably stain AT or GC rich zones within
chromosomes was discovered (28-30). These chromosome banding techniques
allowed scientists to identify all the normal chromosomes and to detect
relatively small chromosomal aberrations.

**In situ hybridization**

In 1969, the same year as the first banding technique was discovered, the
first in situ hybridization attempts were made. RNA molecules labeled with
radioisotopes were hybridized to DNA in cytological preparations (31-33). The
experimental procedures were cumbersome but resulted in relatively sensitive
results and a resolution far exceeding that of any banding technique.
Radiolabeled probes dominated the in situ hybridization field during the 1970s.

This dominance was broken during the 1980s with the introduction of the
fluorescence in situ hybridization technique, FISH. RNA and DNA probes were
labeled by incorporation of either fluorescence reporter molecules or haptens,
allowing the probes to be detected directly or via fluorescence-labeled
antibodies, respectively. Bauman et al. showed in 1980 that RNA probes 3’-
labeled with a fluorophore could be used to detect specific DNA sequences
in situ (34). A year later, Langer et al. introduced probes labeled with biotin-dUTP
(35). This nucleotide was compatible with the enzymatic techniques used to
label radioactive probes, and could be detected by several different methods, e.
g. via colloidal gold, immunocytochemical procedures or by enzymatic
deposition of colored end products (19). New haptens, and fluorophore-
conjugated antibodies targeting them, further expanded the scope of FISH. In
1986, two colors were detected simultaneously in situ (36), and with the
introduction of AMCA, three colors could be detected by 1989 (37).

In recent years, several multicolor FISH methods have been developed. One
means to achieve multiple FISH colors is by combinatorial labeling of the
probes with several fluorophores. Two different methods to detect probes with
combinatorial labeling emerged in 1996. In the multi-FISH (M-FISH) method
the probe signals are detected using a fluorescence microscope with optimized
filters having narrow band passes to minimize spectral overlap and crosstalk.
between different fluorophores (38). Another method to detect combinatorially-labeled probes is by measuring the emission spectra simultaneously in every sample pixel e.g. spectral karyotyping (SKY) (39). With these methods, all 24 different chromosomes can be individually labeled using only five fluorophores, facilitating swift karyotyping of tumor cells with complex numerical and/or structural aberrations. Multiple FISH colors can also be generated by ratio labeling of the probes (40). The individual labelings represent combinations of fluorophores that are further varied by using different ratios of the two fluorophores in the probe labeling reaction. A further development of multicolor FISH is the addition of a combined binary label to the ratio labeling, combined binary ratio labeling (COBRA). A set of 12 colors can be achieved using three fluorophores in ratio labeling. The number of colors can then be doubled to 24 by binary addition of a fourth fluorophore. The binary addition of a fifth fluorophore increases the number of colors to 48 (41). This not only enabled for the discrimination of individual chromosomes, but also allowed for the p- and q-arms of individual chromosomes to be distinguished.

Techniques for in situ SNP analysis

Allele-specific hybridization

There are only a few techniques available for the detection and discrimination of single-nucleotide polymorphisms in situ. Short variant-specific oligonucleotides have been used to detect a G/T variation in the alpha-satellite repeat sequence on human chromosome 15 (42). The 23 nt long fluorescence-labeled probes used in the study could distinguish between the two sequence variants due to the fact that probes destabilized by a single-base mismatch could be removed under suitably stringent washing conditions. The drawback with this technique is that the probes have to be relatively short in order to be sensitive to mismatches. This restricts the amount of label that can be incorporated into the probes, and in the end the signal that can be obtained from them, limiting the sensitivity to studies of highly repeated sequences. Another limitation is that these probes give rise to a qualitative rather than quantitative signal. Small variations in the washing conditions can have a large impact on the hybridization efficiencies of the probes resulting in biased signals. The utility of obtaining quantitative results from genotyping of highly repeated sequences is discussed in present investigation IV.
**Primed in situ labeling**

The selectivity of an in situ hybridization reaction with short oligonucleotide probes can be improved by utilizing the ability of a DNA polymerase to discriminate between matched and mismatched 3’ probe ends in an extension reaction. This mechanism is used in the primed in situ labeling procedure, PRINS (43). Single-nucleotide variants can be distinguished by hybridizing short unlabeled oligonucleotides with the 3’ end, opposite the variable position. Correctly hybridized probes are extended by the polymerase, and become labeled via incorporation of labeled nucleotides in the extension product, while mismatched probes are not extended by the polymerase. Dual color PRINS can be achieved by two consecutive hybridizations and extension reactions using differently labeled nucleotides. To avoid cross-reactions between the two different extension reactions, the 3’ ends of the first extension products can be terminated in an extension reaction with dideoxynucleotides before the second hybridization reaction (44). This method can be used to distinguish single-nucleotide variations in repetitive genomic DNA, but quantitative signals are probably difficult to obtain, as the consecutive reactions needed to obtain dual colors may skew signal proportions.

**In situ PCR**

Single- or low-copy number DNA sequences can be detected in situ through PCR amplification (45-48). Cycles of denaturation, hybridization and extension of two target-specific primers are performed in cells fixed in solution or spread on glass. The amplification products are either detected directly by the incorporation of labeled nucleotides or indirectly by hybridization of an amplicon-specific labeled probe (49). This method has been used to discriminate between two rat albumin gene variants differing by a seven bp deletion (50). Although the technique is claimed to be sensitive and specific enough to detect single-copy sequences (51), the usefulness of the method is limited by several problems. The amplification of the sequences of interest is not exponential as in ordinary PCR, but appears rather to progress linearly (49,52,53). Furthermore, false positive signals are common. In the direct PCR approach, false positive signals probably arise from unspecific incorporation of labeled nucleotides into fragmented genomic DNA or priming of nonspecific PCR products by fragmented DNA. Another source of false positive signals is the diffusion of amplified products from cells harboring the target sequence into the surrounding reaction solution. Such PCR products are then efficiently amplified, and may subsequently become lodged in target sequence-deficient cells (49).
INTRODUCTION TO PRESENT INVESTIGATIONS

A way to detect single-nucleotide polymorphisms quantitatively in situ is to use oligonucleotide probes that become circularized by a ligase after hybridization to the correct target sequence. The properties of these circularizable probes, padlock probes, are described in detail hereafter, as well as some of the properties of ligase enzymes, which are an essential component in the method. Finally, the structure and function of human chromosomal centromeres, which are used as a model system in the in situ experiments, is described.

Padlock probes

Padlock probes are linear oligonucleotides of 90 or so nucleotides in length, consisting of two target-complementary segments, one at each end of the probe. These target-complementary segments are joined together by a target-noncomplementary linker segment (54). Upon hybridization of the probe to the correct target sequence, the 5' and 3' target-complementary ends of the probe are brought into juxtaposition, thus creating a double helix with a nick. This nick can be covalently sealed by a DNA ligase, thereby circularizing the probe. Due to the length of the target-complementary segments, usually 20 nt per segment, and the helical nature of double-stranded DNA, the probe is typically wound around the target strand approximately four times (Figure 1).

Figure 1. An illustration of a padlock probe interacting with the target sequence.
Padlock probes are excellent tools to detect single-nucleotide variants in RNA and DNA. This is because the simultaneous hybridization of the two target-complementary segments ensures a high specificity in recognizing a target sequence (54-56). The discriminating power of the padlock probes is increased by the fact that the ligation reaction is strongly inhibited by any mismatches at the ligation junction, especially at the 3’ ends of the hybridized probes (57-60). Distinction of single-nucleotide sequence variants is therefore possible with padlock probes. The catenation of the probes to the correct target sequence after hybridization and ligation renders the probe resistant to stringent washes if the probes are bound between two points of the target molecule that are attached to a solid phase, reducing the background signal from non-specifically hybridized probes (54).

Highly sensitive detection of circularized padlock probes is possible, e.g. by amplifying reacted probes using a rolling circle replication (RCR) mechanism (61-63) or by PCR (64,65). The RCR reaction produces concatenated complementary sequences, and can be initiated by using a primer complementary to the linker segment of the padlock probe (61-63). If a DNA polymerase with 3’ to 5’ exonuclease activity is used, then the target sequence can be used to prime RCR, since the DNA polymerase can degrade the target DNA-strand, by removing any 3’ nonprobe-complementary nucleotides from a nearby end before starting to extend from the 3’ end of the target strand (62). The RCR products can then be detected either by hybridization of labeled probes or by incorporation of labeled nucleotides during RCR. By using a pair of linking segment-specific primers, PCR can be used to specifically amplify circularized padlock probes across the ligation junction (64,65).
Synthesis of Padlock probes

Padlock probes can be synthesized using standard solid-phase chemical synthesis or via a novel PCR-based method. The probes are demanding to synthesize because of the length of the probes and the strict requirement of intact 5' and 3' ends. The PCR-based synthesis method is described in present investigation I, whereas the chemical synthesis method and the problems inherent with it are described below.

Chemical synthesis

In the first step of the oligonucleotide synthesis cycle, the 5'-protecting dimethoxytrityl (DMT) group of the nucleoside phosphoramidite is removed with trichloroacetic acid (TCA). This is followed by addition of nucleoside phosphoramidites (66), which in the presence of tetrazol, reacts with free 5' hydroxyl groups, to form a phosphotriester bond to the added nucleoside. The phosphotriesters are oxidized with iodine before the last step in which unreacted 5' hydroxyls are capped using acetic anhydride, preventing further extension of these molecules. These steps are repeated until the last nucleotide has been added to the oligonucleotide. The synthesis products are then released from the solid phase by incubation in ammonia (NH₃ (aq.)) over night at 56°C. Base-protecting groups are also removed during the incubation in ammonia (67).

Many types of imperfect molecules accumulate during standard automated solid-phase oligonucleotide synthesis. The main portion of the impurities consist of 5' truncated, capped molecules that failed to couple a nucleotide. Oligonucleotides with intact 5' ends can be conveniently isolated on a reverse phase chromatography column, if the hydrophobic DMT group is left on after the final coupling cycle. Another mechanism yielding imperfect molecules is the creation of abasic deoxyribose residues formed mainly from purines during every acidic detritylation step. These residues will be cleaved during the ammonia deprotection step yielding shorter fragments (68). The 5' end fragments will be equipped with a DMT, and are notoriously difficult to separate from full-length products using HPLC. They are therefore the major impurity in chromatographically purified oligonucleotides. In a further mechanism, 3' truncated products can be produced if synthesis is primed on a hydroxyl other than the one at the 5' end of the nucleotides bound to the support. Finally, inefficient capping or detritylation, or reversible capping, results in internally deleted molecules (69,70).

The functionality of the padlock probe is completely hampered by missing nucleotides at the 3' or 5' ends, while internal deletions may affect the probe function less. Intact 5' and 3' ends can be selected for by using a novel solid
support containing a disiloxyl bond. This bond allows for the cleavage of abasic deoxyribose residues under mild basic conditions after the last coupling step. These mild basic washes remove the 5' DMT-protected part of the molecule while the 3' part remains intact. These truncated fragments are easily removed from the full-length trityl-containing molecules using a reverse phase chromatography column after the release of the synthesized products by cleavage of the disiloxyl bond with tetrabutylammoniumfluoride (TBAF) (71). Padlock probes of lengths greater than what can be conveniently synthesized chemically can be constructed by enzymatic synthesis. Most of the problems with solid-phase oligonucleotide synthesis can be avoided using the enzymatic procedure, which is described later in present investigation 1.

**Ligases and the ligation mechanism**

DNA ligases catalyze the formation of phosphodiester bonds between the adjacent 3'-hydroxyl and 5'-phosphate termini at breaks in one or both strands of a DNA duplex. DNA ligases can be divided according to the cofactor requirements of the enzymes. The known eukaryotic, archebacterial and viral ligases all require ATP as a cofactor and are monomeric proteins ranging from 30 to >100 kDa. The eubacterial ligases are a more homogenous group of proteins in the range of 70-80 kDa, and they require NAD⁺ as a cofactor (72-74). ATP- and NAD⁺-dependent DNA ligases share six conserved motifs involved in building up the active site of the enzymes (75). Motif I, closest to the N-terminal, most motif I contains the conserved KXDG sequence which is involved in the formation of an adenylated enzyme intermediate (76-78).

The ligation reaction mechanism can be divided into three steps. First, the ligase is activated through the formation of a covalent protein-AMP intermediate with the concomitant release of PPI or NMN, depending on the cofactor. The AMP molecule is bound to the ε-amino group on the lysine residue in the conserved KXDG motif. After the recognition of and binding to a nick, the AMP molecule is transferred to the phosphorylated 5' end of the nick. In the third and final step, the enzyme catalyzes a nucleophilic attack of the adjacent 3' hydroxyl group on the pyrophosphate bond between the AMP and the 5' phosphate. The subsequent formation of a phosphodiester bond seals the nick, and releases the AMP and the ligase (72,74,79).

The two most commonly used ligases are the T4 DNA ligase and Tth DNA ligase. T4 DNA ligase is derived from bacteriophage T4. Tth ligase originates from the eubacterium Thermus thermophilus. T4 DNA ligase has two temperature optima for ligation, 28°C and 37°C, while Tth ligase is thermostable and has an optimal temperature range between 65-72°C for nick closure. The
optimal pH for nick closure ranges between 7.2-7.8 for T4 DNA ligase, and 8.5 for Tth ligases (60,73,80). Both of these enzymes require divalent cations present in the active site during the ligation reaction, probably in a similar configuration that has been proposed for T7 and Tfi ligase (74). In general, Mg\(^{2+}\) is used as the divalent metal ion for ATP-dependent and NAD\(^+\)-dependent ligases. Slightly increased ligation rates have been observed with the T4 DNA ligase when Mn\(^{2+}\) was used (73). Tth ligase exhibited about 6 fold higher mismatch ligation rate with Mn\(^{2+}\) compared to Mg\(^{2+}\) (80). DNA ligases need a certain length of double-stranded DNA, referred to as a footprint, for nick ligation. For the T7 ligase, the footprint has been shown to be asymmetrical, spanning 7-9 and 3-5 bases at the 5' and 3' ends of the nick, respectively (75). T4 DNA ligase and Tth ligase are more discriminating toward mismatches at the 3' end than the 5' end of the ligation junction (57,60,81). T4 DNA ligase mismatch discrimination can be enhanced further by a NaCl concentration of 200 mM in the ligation reaction (57,60). Some ATP-dependent ligases, such as T4 DNA ligase, also have the ability to ligate DNA hybridized to RNA (82,83). This property of T4 DNA ligase has been exploited in a ligase-dependent PCR (LD-PCR) assay. Sensitive detection of hepatitis C virus RNA sequences present in serum was achieved with this method (64). Probes that become circularized when ligated have also been used to detect hepatitis C virus RNA (64). The RNA-templated DNA ligation ability of T4 DNA ligase has been explored and optimized for RNA sequence analysis in the present investigations II and III. Recently, Xu et al. presented chemical autoligation of DNA hybridized to RNA, combined with a fluorescence resonance energy transfer (FRET) detection system as an alternative to ligase-based assays (84). The ligation efficiency on RNA equaled that on DNA with chemical ligation, but the chemical autoligation reaction had a 2.5 to 36 times higher t\(_{1/2}\) compared to results obtained in present investigation III.

**Centromere function and structure**

The genetic information needed for the normal development and viability of a human cell is divided into 46 chapters, called chromosomes. It is crucial that the two daughter cells receive a complete and identical set of chromosomes during cell division. The proper segregation of chromosomes during the two types of cell division, mitosis and meiosis, is ensured by a cis-acting DNA element, the centromere. This structure is visible on condensed metaphase chromosomes as a primary constriction, and must perform an entirely different task during the two different cell divisions. In mitosis and meiosis II, the centromere must preserve the two sister chromatids as an entity until the appropriate point in cell division at which time the mitotic spindles separate
pulling the two sister chromatids to the opposite poles of the dividing cell. By contrast, in the first reductional meiotic division, meiosis I, the centromere ensures that the two sister chromatids are not separated by the meiotic spindles. Instead, the number of chromatids is reduced to half by the segregation of the paired homologous chromosomes.

The human centromeres consist of highly condensed heterochromatin, built up from at least five different classes of repeat sequences (85). Of these five classes, the alpha satellite repeat class is considered to be the necessary DNA component of the centromere, since it is present on all human chromosomes at the primary constriction. The basic unit of alpha satellite DNA is an approximately 171 bp long monomer sequence. A variable number of these monomer units are in turn tandemly repeated in arrays organized in higher order repeat structures ranging from 200 kb up to 9000 kb (85). Although most of these higher order repeat structures are chromosome-specific, some of them are present on several chromosomes, probably as a result of rare events of interchromosomal exchange between non-homologous chromosomes (86). The higher order repeat units on homologous chromosomes from different individuals differ in both sequence and repeat number. This may be explained by unequal crossing-over events between sister chromatids or homologous chromosomes (87-89).

Alpha satellite sequences are conserved among primates, but there is little or no conservation between all mammalian centromeric sequences. Still, centromeres from one species are functional inside cells of another species, as illustrated by the relative stability of somatic cell hybrids. The function of centromeres is probably not dictated by the DNA sequence as such but, rather by structural features present in the repeated DNA sequence. There is evidence that a conserved region in the 171 bp monomer unit of primate alpha satellite DNA has a dyad symmetry potentially allowing for the formation of hairpin structures or H-formed DNA (90). Centromere-binding proteins such as HMGI (alpha protein) and topoisomerase II have been reported to bind such structures (91-93).
PRESENT INVESTIGATIONS

I. PCR-generated padlock probes detect single-nucleotide variation in genomic DNA

This paper describes a fast and flexible small-scale method for the synthesis of long padlock probes for detecting single-nucleotide sequence variants in situ. It is difficult to chemically synthesize probes of more than 100 nucleotides. This is because the proportion of imperfect oligonucleotides increases with oligonucleotide length. We have solved this problem by developing a PCR-based method to synthesize padlock probes. By using two primers, an amplification template, and modified nucleotide triphosphates, longer probes can be made that are more densely labeled than those constructed by conventional chemical synthesis. One primer equipped with a 5'-phosphate forms the 5' target complementary sequence of the padlock probe. The 3' target-complementary part of the probe is formed by the complement of the 5' biotinylated primer. The target non-complementary linker segment is derived from the PCR-template sequence. The probes can be labeled during PCR by incorporation of modified nucleotide triphosphates. The double-stranded PCR products are immobilized on streptavidin-coated paramagnetic beads via the 5' biotin, and the complementary strand is released under denaturing conditions. After neutralization, precipitation, and quantification, the single-stranded amplification product is ready to be used as a padlock probe (Figure 2).

DNA polymerases deficient of proofreading activity are known to leave extra, untemplated nucleotides at the 3' end of PCR products. DNA polymerases with a processive 3' → 5' exonucleolytic activity, on the other hand, can produce PCR products missing one or more nucleotides at the 3'-terminus. When using Taq polymerase, which lacks proofreading activity, the problem of additional nucleotides at the 3' end of the PCR product can be avoided by decreasing the nucleotide concentration in the amplification reaction. While this lowers the probe synthesis yield, it is compensated for by the reduced consumption of expensive modified nucleotide triphosphates. If a polymerase possessing proofreading activity, such as the Pfu-polymerase, is used at low nucleotide concentration, then the problem of missing 3' nucleotides can be avoided by increasing the nucleotide concentration during the last extension step in the PCR.
Figure 2. A schematic outline of the method for enzymatic synthesis of padlock probes.
When the in situ signal of an enzymatically synthesized 90-mer padlock probe was compared to a chemically synthesized probe of the same size, a three-fold stronger signal was achieved. The two probes detected an alpha-satellite sequence variant on human chromosome 21, and the signals were weighed against signals from a probe specific for an alpha-satellite repeat sequence present on chromosome 12. The PCR-generated probes targeting a single-nucleotide sequence variant on chromosomes 13 and 21 also distinguished the staining pattern of homologous chromosomes better than chemically synthesized probes. Furthermore, a probe length of 140 nt was found to be optimal when used in situ on metaphase chromosomes. It is probable that longer probes gain access to the target sequence less efficiently, accounting for the weaker signal. This is probably due to difficulties for probes longer than 140 nucleotides to access the target sequence. Myer and Darren concluded recently that PCR-generated probes up to a length of 800 nt can be efficiently ligated in solution with PCR products as target (94).

Finally, PCR-generated padlock probes were used to analyze sequence variants of a single-copy gene in genomic DNA. Padlock probes detected a homozygous C to A mutation present in the ATP7B gene. In homozygous form, this mutation causes a copper transport deficiency called Wilson’s disease. The probes were ligated in solution and amplified across the ligation junction by PCR. The PCR products were subsequently detected on a 2% agarose gel. This was the first demonstration that padlock probes have the required specificity to detect single copy-genes in total genomic DNA.

II. Enhanced detection and distinction of RNA by enzymatic probe ligation

DNA sequence variants can be accurately distinguished through ligase-mediated gene detection. However, it has not previously been known if ligases can also distinguish between RNA sequence variants in a similar manner. A model system was designed to investigate the ligation of pairs of DNA oligonucleotides hybridizing juxtaposed on RNA strands. Four in vitro RNA target transcripts of amplified synthetic oligonucleotides, differing in one centrally located position, were synthesized. Furthermore, four oligonucleotides targeting these in vitro transcripts were synthesized. Each oligonucleotide had a variable base at the 3’-terminal position, and was size-coded by a 5’-sequence addition. Finally, a 5’ phosphorylated 3’ fluorescently-labeled oligonucleotide complementary to all the in vitro transcripts was synthesized (Figure 3). Nicked heteroduplex molecules serving as ligation templates were preformed by mixing the 3’ fluorescently-labeled oligonucleotide, RNA target transcripts and variable oligonucleotides at a molar ratio of 1:2:4. This was done to ensure that
every labeled oligonucleotide participated as substrates for the ligation reaction, and that the four size-coded oligonucleotides competed in the reaction. The ligation products were separated, detected, and quantified using a fluorescence-sequencing instrument. Using this ligase-based assay, we defined one set of reaction conditions where most mismatches were discriminated by a factor greater than 80, compared to the corresponding matched substrate (G-G, T-G, A-G, T-U, C-U, G-A, C-C). The remaining mismatches (G-U, C-A, T-C, A-C, A-A) were all discriminated by a factor larger than 20.

Figure 3. The experimental system used in the ligation studies. The four 5' probe molecules differ at their 3'-most position, and can be identified by the different numbers of T residues added at their 5' ends. The 3' probe is labeled with a fluorophore, facilitating detection of ligated products. The variable base in the RNA template is denoted with the corresponding capital letter.

A padlock probe was constructed targeting the RNA transcript with an A in the variable position. Efficient circularization of the padlock probe was only obtained in the ligation reaction in the presence of the correct RNA or DNA sequence (Figure 4). Ligation of the probes hybridized to RNA was almost as efficient as on DNA. These results indicate that ligase-mediated detection and distinction of RNA sequence variants can be done highly sensitively and accurately using padlock probes in solution or in situ.
Figure 4. Ligation of a padlock probe targeting the A variant RNA transcript. The negative control (-) contained no target sequence.

III. RNA-templated DNA ligation for transcript analysis

The different parameters influencing RNA-templated DNA ligation were studied in detail using the model system previously described in present investigation II. A time-course study of the ligation of DNA probes correctly hybridized to four RNA targets was performed. The time-course study revealed that DNA ligation by T4-DNA ligase proceeded at a much slower rate compared to the same reaction on DNA template and required large amounts of enzyme. The accumulation of processed end products in the ligation reaction was saturated only when a molar excess of ligase was added to the reaction. A large fraction of adenylated products was observed in the reaction, probably as a result of the ligase detaching from the template before the final nucleophilic attack of the 3' hydroxyl group at the nick, leaving an AMP group attached to the 5' end of the nick and blocking further ligation attempts. This is most likely caused by premature AMP reloading of the ligase enzyme. The ligation kinetics also differed among RNA target sequences that differed in only one position (Figure 5). The reason for this discrepancy is not understood. The accumulation of adenylated and ligated products follows first order kinetics, but the completed ligation reaction alone does not. The rate limiting step in the reaction is probably the joining step, since 5' adenylated reaction intermediates are accumulated during the course of the reaction.
Figure 5. A time-course study of the ligation of DNA probes, correctly base-paired to corresponding RNA targets. The RNA targets used in each experiment are indicated with capital letters in the lower right corner of the graph. Squares represent adenylated and ligated probes, while circles denote ligated probes only. The $t_{1/2}$ estimates of the reactions are shown adjacent to the reaction curves.

The ligation reaction was inhibited by concentrations of the cofactor ATP normally used in ligation reactions. ATP concentrations exceeding 14 $\mu$M, the $K_m$ for ATP binding, (79) resulted in a high yield of 5’ adenylated end products and a low yield of ligated end products (Figure 6). This can be explained by a premature ATP-reloading of the ligase enzyme after the 5’ adenylation step in the ligation reaction, resulting in the dissociation of the enzyme-substrate complex. Increased mismatch discrimination can be achieved in DNA-templated DNA ligation by increasing the NaCl concentration to 200 mM. This effect does not apply for RNA-templated DNA ligation, since ligation was completely abolished at NaCl concentrations at or above 150 mM. We further investigated the ability of magnesium and manganese ions to support DNA ligation reactions templated by RNA strands having an A or a C at the diagnostic position. These templates were chosen for the experiment because
they exhibited the fastest and the slowest reaction kinetics, respectively. There were no detectable differences in the adenylation rate in the presence of either of the two ions, but the ligation reaction proceeded twice as fast when manganese was used.

Figure 6. ATP-dependence of probe adenylation and ligation reactions in RNA-templated ligation. A 60 min reaction was performed with a ligation probe having A at the 3’-most position and an RNA target with a U at the variable position. Averages and ranges of adenylated (squares) and ligated (circles) products from triplicate reactions are shown.

IV. PCR-generated padlock probes distinguish homologous chromosomes at a single-nucleotide resolution enabling quantitative in situ genotyping

There are a few methods available to detect single-nucleotide polymorphisms in situ. The methods are based either on the hybridization of short variant-specific oligonucleotides, or on the hybridization of oligonucleotides followed by a discriminating polymerization step, the so-called primed in situ labeling (PRINS) technique. PCR has also been employed in situ to distinguish between single-nucleotide variants. To date, these methods have only yielded qualitative data, which may decrease the informativity of the markers used, whereas padlock probes can produce quantitative data, enhancing the informativity of the markers. Padlock probes targeting two alpha-satellite sequence variants present at the centromere of human
chromosome 7 were designed and enzymatically synthesized. These sequences have not previously been used in this manner. Probes were also synthesized to target an A/G alpha-satellite sequence variation present on chromosome 15. This variation has previously been shown to be polymorphic when short variant-specific oligonucleotides were used to detect the variation in situ, but only a restricted number of allelic variants could be distinguished with the method (42). The probes were co-hybridized and ligated simultaneously on metaphase spreads prepared from lymphoblast cell lines established from a CEPH family. The probes were visualized using two layers of fluorescently-labeled antibodies. Homologues of chromosomes 7 and 15 segregating in the family were distinguished by quantifying the red and green staining pattern of at least 20 metaphases from each individual. We successfully followed the transmission of the centromeres 7 and 15 during two consecutive generations in a pedigree. The centromeres seem to cluster into allelic variants, when mean signal intensities are plotted for all of the 14 centromeres present in the 7 individuals in the pedigree. At least four and three allelic variants may be detected for chromosome 7 and 15, respectively (figure 7).

**Figure 7.** Red and green mean signal intensities from 20 metaphase images, where the background was subtracted and the strongest signal was set to 100%, are shown with the standard errors of the means. The different allelic variants are enclosed by a dotted line.
DISCUSSION

In situ genotyping using PCR-generated padlock probes

The novel PCR-based method for the enzymatic synthesis of padlock probes presented in this thesis offers several advantages over standard solid-phase chemical synthesis. Enzymatic synthesis enables for the synthesis of long probes that are densely labeled. Signals detected from densely labeled enzymatic probes allow for a more precise quantitation, compared to signals from less strongly labeled chemically synthesized probes of a corresponding length. Flexible labeling of probes by selecting differently labeled nucleotides in the PCR reaction permits radical redesign of multicolor probe systems. The ability of the padlock probes to distinguish single-nucleotide variations, rendering them suitable for detection and discrimination of closely similar sequences, has been utilized in the present investigation.

In this study, homologous chromosomes were identified and genotyped for the presence of alpha-satellite repeat sequence variants in the centromeres. The chromosome 7 and 15 centromeres present in the 7 individuals used in the study seem to clustered into allelic variants. Some centromeres present only once in the pedigree were not assigned to any of the allelic variants. Instead, they may represent additional allelic variants present in the population (Figure 7). A larger sample size is needed to confirm whether or not these centromeres truly belong to additional allelic variants. These results indicate that it should be possible to identify human chromosomes and to distinguish homologous chromosomes using polymorphic padlock probes. Padlock probes could therefore be used as a cytogentical technique to establish the parental origin in dicentric chromosomes or chromosomes with a balanced translocation, or to determine if nondisjunction has occurred in meiosis I or II in cases of trisomia.

The heterochromatin part of the human genome will be the last part sequenced in the human genome sequencing effort, temporarily leaving the centromeric regions uncharted. The physical context of sequence variation was lost with previous methods used to study sequence variation in centromeres. Mapping of single-nucleotide variation can be performed on DNA fibers in situ with padlock probes, resolving the physical distribution of sequence variations in repeated sequences. Due to the present detection limit of padlock probes of around 100 probes in situ, low- or single-copy sequence variations cannot be detected. This problem may be overcome by amplifying the signal from padlock probes using RCR (61-63), enabling sensitive detection of single-copy sequences in situ in the future.
Results obtained in present investigations II and III, allow for novel studies where accurate and quantitative detection and distinction of specific RNA sequence variants may be performed in solution or in situ. A benefit of the localized signal produced in in situ experiments is that additional biological information about the transcripts studied can be obtained, such as identification of cells containing the specific transcripts and the subcellular localization of the RNA molecules. Potential uses for padlock probes include, for instance, monitoring transcription of imprinted genes or genes belonging to the same gene family, and detection of rare transcripts in a complex population of transcripts, such as viral or mutated endogenous transcripts in biological samples.

Current methods for the parallel investigation of expression patterns for a large number of genes lack the ability to accurately distinguish between transcripts from closely similar genes. Moreover, most of the techniques involve an RT step which can result in skewed representation of the sequences. The RT step is avoided with padlock probes because RNA can be used as a template in the highly specific ligation reaction. Multiplexing can be achieved by simultaneous amplification over the ligation junction of several different ligated padlock probes using a standard primer pair targeting the linker segments of the padlock probes. The amplified products can thereafter be sorted on a DNA array on the basis of a zip code sequence present in the amplified part of the linker segment. This technique might in the future be the method of choice for studying the expression patterns of closely related genes.

CONCLUDING REMARKS

The first draft sequence of the human genome has revealed more than 1.4 million single-nucleotide polymorphisms (1). Most of these SNPs are probably physiologically insignificant variants, but some of them are probably involved, in combination with environmental factors, in causing complex diseases such as arteriosclerosis, psychiatric disorders and diabetes. Novel methods are needed in the forthcoming years to study the complex patterns of genetic variations behind these and other multifactorial diseases. Although padlock probes as a technique is still somewhat in its infancy, the method has already proven to be versatile and flexible. The findings presented in this thesis have enabled the expansion of padlock probes into new areas of use, further enhancing the usefulness of the method. I believe that these findings and others to come will
help the technique to evolve and mature during the next years and to secure a prominent position among techniques used to study genomic variation.

SWEDISH SUMMARY

Nu när i stort sett hela det mänskliga genomet är sekvenserat står den biologiska och medicinska vetenskapen inför nya utmaningar. Sekvenseringen har uppdagat att människan har omkring 30.000 – 40.000 gener, ungefär dubbelt så många som hos bananflugan. Det har dessutom visat sig att det finns ett stort antal singel-nukleotidvariationer i DNA-sekvensen. Dessa är i de flesta fall harmlösa, men vissa variationer kan, i kombination med andra variationer och olika miljöfaktorer, orsaka sjukdomar.

Sjukdomar som anses bero på flera olika faktorer är vanliga inom befolkningen; till dessa räknas bland annat åderförkalkning, fetma, diabetes, olika typer av cancer och ett flertal psykiska sjukdomar. För att försöka förstå variationen i gener och i den mänskliga arvsmassan och vilka kombinationer av variationer som ger upphov till sjukdomar behövs det nya metoder. En ny metod att hitta och studera just variationer i DNA och RNA i celler är hänglåsprober, padlock probes på engelska.

En hänglåsprob är ett sökfragment av enkelsträngat DNA, konstruerat på så sätt att när den hittat rätt målsekvensen av DNA så placerar sig båda dess ändar bredvid varandra (se figur 1 på sidan 14). Glappet mellan ändarna kan förslutas med hjälp av ett enzym, ett så kallat ligas, vilket leder till att sökfragmentet blir cirkulärt. Eftersom dubbelsträngat DNA är spiralvridet kommer det cirkulariserade sökfragmentet att sitta fasttvinnat runt målsekvensen.

Ordlista

DNA kemisk molekyl som bygger upp arvsmassan
gen arvsanlag
genom arvsmassa
kromosom bärare av arvsmassa
nukleotid bokstav i det genetiska språket
probe sökfragment
RNA kemisk molekyl som styr proteinsyntesen
sekvensera avläsa arvsmassan

Hänglåsprober kan framställas på kemisk väg, men det är svårt att få deras kvalitet så god att de fungerar väl. I min avhandling presenterar jag en alternativ metod att enkelt tillverka dessa sökfragment med hjälp av ett enzym. Det enzymatiskt framställda proberna användes sedan till att studera variationer...
i repeterat DNA på kromosom 7 och 15. Individuella kromosomer kunde identifieras med hänglåsproberna och nedärvningen av dessa kromosomer kunde följas i en familj på tre generationer. Att kunna identifiera specifika kromosomer med sökfragmentet kan avslöja hos vilken av föräldrarna det skett ett misstag när ett barn har fått en kromosom för mycket eller saknar en kromosom.

Det har inte tidigare varit känt om hänglåsprober kan detektera singelnukleotidvariationer på RNA. Jag har tillsammans med mina kollegor visat att ligasenzymet kan skilja mellan olika nukleotider och detektera RNA-molekyler som skiljer sig åt med endast en nukleotid. Vi studerade dessutom vilka faktorer som påverkade enzymet, för att hitta de reaktionsbetingelser som gav det bästa ligeringsresultatet och den största diskrimineringen mellan olika nukleotider. Resultaten i min avhandling pekar mot att det är möjligt att hitta och särskilja närbesläktade RNA-molekyler med hjälp av hänglåsprober. Detta är av vikt när man vill studera uttrycket, mängden RNA och i slutändan mängden protein från närbesläktade gener i en cell.

ACKNOWLEDGEMENTS

Yes, you have come to the right page. This is the list of who’s hot and who’s not, the acknowledgments. Well, here we go! I would like to thank:

First and foremost, my supervisor professor Ulf Landegren for letting me roam free in the lab, and allowing me to do science under his supervision. As a guinea pig for your different ways of supervision, I must admit I preferred the last one to the other methods.

Dr. Mats Nilsson, my co-supervisor, without your practical assistance in and out of the lab, endless enthusiasm, good humor and encouragement, this thesis would not have been possible.

Past and present colleagues in Ulf’s group, Johan, Simon, Mats (the bad boy bunch), Sigrun (lopapeysa, see I now know some icelandic), Anders for bringing the samba into the lab, Marek for trying to teach me DNA chemistry, Maritha for all the fun we have had on conferences (I won’t tell if you won’t tell.), and Jonas (Järven) for helping me with silly computer problems and for actually understanding how a microscope works (We’ll always have Paris....).

Lotta and Anette for all their support during the first struggling years at BMC. I would never have pulled through without you, besides I learned a lot about fashion, make-up, interior design and what goes on inside the heads of girls when we shared an office in the old days.
Kicki for all the fun, friendship and syncopated dancing we have had since you joined the coca-cola light break club. You’re the tops!

I am also especially indebted to Jeanette (the weekend club) for all of her help during these years, without your expertise and organizational skills I would have been lost forever in university beaurocracy. Elisabeth, Anette, William & Viktor, Stig, Eddie, Marcus, and Eva are not to be forgotten, thank you for all of your help during these last years.

Professor Lore Zech for always being interested and impressed by my FISH work (NO, I did not use DAPI to color my hair). Anna and Kerstin for helpful hints and suggestions whenever my FISH experiments did not work.

Chris Ninalga, the queen of hyphens and ruler of commas for actually reading this thesis and crossing the t’s and dotting the i’s.

All of the other people at the department during the years at BMC and now at Rudbeck laboratory (none mentioned, none forgotten) for all the high spirited parties and for company during the long evenings and weekends in the lab.

Sorken and Emelie, a.k.a. the sandwich sisters, for dragging me out into the real world from the lab. I will always remember the three little kittens....tour 2000 (tre små fjantar, tappa sina vantar), and all the Seinfeld-like discussions at cafe Linné (I’m not as selfish as Elaine). A special thanks to Emelie for scrutinizing the language in the swedish summary.

Fredrik (Björe) macho man for all the hiking tours into the mountains (South-America next year?). Nisse and Linda for accompanying some of these expeditions and for beeing good friends, Lena (Heffan) for your encouragement, Marie (Mr. Queen) and Carro for simply being you, and Fredrik my oldest friend. The party bunch, Malin-Darling, Chris, Stefan, Marlena, Kristina, Niclas and many more for all the fun parties. Lars for being a ray of light during the dark and stormy thesis writing nights.

My family for all the long boring Sunday afternoons at the soccer field, the starting point for my interest in the natural sciences. Who could have imagined that the many hours I spent studying bugs and weeds beside the soccer field would amount to this thesis.

Oh, are you not on the list?? Well, write you’re name on the dotted line and underline the appropriate word, and hey presto, you’ve made it....

for all the fun we had.

Finally, I would especially like to thank ............................................ for all of the support.

you’re the best!
REFERENCES


