



UPPSALA  
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 532*

# Ligation-mediated Molecular Analysis of Influenza Subtypes, Splicing and Protein Glycosylation

TIM CONZE



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2010

ISSN 1651-6206  
ISBN 978-91-554-7743-1  
urn:nbn:se:uu:diva-120016

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbeck Laboratory, Dag Hammarskjölds Väg 20, Uppsala, Friday, April 23, 2010 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

### **Abstract**

Conze, T. 2010. Ligation-mediated Molecular Analysis of Influenza Subtypes, Splicing and Protein Glycosylation. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 532. 39 pp. Uppsala. ISBN 978-91-554-7743-1.

Binder-based assays are employed throughout the life sciences. Powerful signal amplification techniques have enabled detection of very rare molecule species diluted in simple buffers. Unspecific binding of primary binders leads to increased background in more complex samples. By requiring two recognition events, ligation-based molecular analyses provide highly specific detection of biomolecules in complex samples.

We developed a highly multiplexed padlock-ligation assay targeting signature sequences in the hemagglutinin and neuraminidase genes. From a panel of 77 avian influenza isolates of all major serotypes, 97% were genotyped correctly in accordance with previous classifications by classical diagnostic methods (Paper I).

Alternative splicing is an important mechanism expanding the proteome. Current analysis techniques fail to provide sequences of complete transcripts beyond the read length of sequencing instruments. We devised and implemented a strategy to compress the sequence information contained in the splicing pattern of a transcript into the presence or absence of sequence-blocks. We demonstrate that this assay yields information about the splicing patterns in thousands of transcripts from cellular cDNA (Paper II).

Expression changes of mucin proteins and glycosylation structures are frequently observed from the early stages of cancer development. Expression of mucin 2 and sialyl-Tn are common features of intestinal metaplasia and gastric cancer, and are known to co-locate. Here we have developed an in situ proximity ligation assay (PLA) directed against mucin 2 and sialyl-Tn. Our study on intestinal metaplasia and gastric cancer tissue sections identified mucin 2 as a major carrier of sialyl-Tn in these conditions, and demonstrated how conveniently glycosylation of proteins can be studied by in situ PLA (Paper III).

This thesis shows how the dual recognition requirement of ligation-based assays can be employed to detect target molecules with high specificity, to analyze several sequence features of nucleic acids or to study the proximity of two antigens in situ.

**Keywords:** ligase, proximity ligation, gastric cancer, glycosylation, alternative splicing, avian influenza, padlock probe

*Tim Conze, Department of Genetics and Pathology, Molecular tools, Rudbecklaboratoriet, Uppsala University, SE-75185 Uppsala, Sweden*

© Tim Conze 2010

ISSN 1651-6206

ISBN 978-91-554-7743-1

urn:nbn:se:uu:diva-120016 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-120016>)

Noli turbare circulos meos

Archimedes

*To my family*



# List of papers

This thesis is based on the following publications, which will be referred to in the text by their roman numerals:

- I. Gyarmati, P., **Conze, T.**, Zohari, S., LeBlanc, N., Nilsson, M., Landegren, U., Baner, J. and Belak, S. (2008) *Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses by use of padlock probes*. J Clin Microbiol, 46, 1747-1751.
- II. Conze, T., Göransson, J., Razzaghian, H., Ericsson, O., Öberg, D., Akusjärvi, G., Landegren, U., Nilsson, M., *Single molecule analysis of alternative splicing* manuscript in preparation
- III. **Conze, T.**, Carvalho, A.S., Landegren, U., Almeida, R., Reis, C.A., David, L. and Soderberg, O. (2010) *MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas*. Glycobiology, 20, 199-206.

Reprints of papers I and III were made with the permission of the publishers.

## Related papers

- IV. Nilsson, M., Larsson, C., Stenberg, J., Göransson, J., Grundberg, I., Isaksson, M., **Conze, T.** and Henriksson, S. (2009) *Application of Pad-lock and Selector Probes in Molecular Medicine* In Patrinos, G. and Ansorge, W. (eds), *Molecular Diagnostics*. Academic Press, pp. 117-132.
- V. **Conze, T.**, Shetye, A., Tanaka, Y., Gu, J., Larsson, C., Göransson, J., Tavoosidana, G., Söderberg, O., Nilsson, M. and Landegren, U. (2009) *Analysis of Genes, Transcripts, and Proteins via DNA Ligation*. *Annual Review of Analytical Chemistry*, 2, 215-239.
- VI. Wang, X., Hu, B., Zieba, A., Neumann, N.G., Kasper-Sonnenberg, M., Honsbein, A., Hultqvist, G., **Conze, T.**, Witt, W., Limbach, C., et al. (2009) *A protein interaction node at the neurotransmitter release site: domains of Aczonin/Piccolo, Bassoon, CAST, and rim converge on the N-terminal domain of Munc13-1*. *J Neurosci*, 29, 12584-12596.
- VII. Weibrecht, I., Leuchowius, K-J., Clausson, C-M., **Conze, T.**, Jarvius, M., Howell, WM., Kamali-Moghaddam, M., Söderberg, O. (2010) *Proximity Ligation Assays: a recent addition to the proteomics toolbox*, accepted in *Expert Reviews in Proteomics*
- VIII. Tavoosidana, G., Ronquist, G., Darmanis, S., Carlsson, L., Wu, D., **Conze, T.**, Ek, P., Semjonow, A., Larsson, A., Landegren, U., Kamali-Moghaddam, M. *Prostasomes Assessed in Blood Plasma by 4PLA Are Biomarkers Reflecting Disease Aggressiveness in prostate cancer* Manuscript in preparation

Supervisor: Ulf Landegren, Professor  
Molecular Medicine  
Department of Genetics and Pathology  
Uppsala University  
Sweden

Faculty opponent: Jörn Glökler, Docent  
Department of Vertebrate Genomics  
Max Planck Institute for Molecular Genetics, Berlin  
Germany

Review board: Stefan Schwartz, Docent  
Molecular Virology/Viral Model Systems  
Department of Medical Biochemistry and Microbiology  
Uppsala University  
Sweden

Marion Kusche-Gullberg, Professor  
Matrix Biology  
Department of Biomedicine  
University of Bergen  
Norway

Berndt Klingebom, Professor emeritus  
Department of Virology  
National Veterinary Institute, Uppsala  
Sweden





# Contents

Introduction .....	11
Ligation-mediated molecular analyses .....	13
Definitions .....	13
Sensitivity .....	13
Specificity .....	13
Affinity .....	13
Selectivity .....	14
DNA ligation .....	14
Enzymatic ligation .....	14
Chemical ligation .....	15
Binders .....	16
Amplification methods .....	16
Polymerase chain reaction (PCR) .....	17
Rolling circle amplification (RCA) .....	17
Requirements of highly sensitive assays .....	18
Padlock ligation .....	18
Proximity ligation assay (PLA) .....	19
PLA in solution .....	20
PLA in situ .....	20
Present investigations .....	23
Paper I – Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses by use of padlock probes .....	23
Results .....	24
Future perspectives .....	24
Paper II – Analysis of full transcript splice patterns at single molecule resolution .....	25
Results .....	27
Future perspectives .....	27
Paper III – MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinoma .....	28
Results .....	29
Future perspectives .....	30
Future perspectives .....	31
Acknowledgments .....	33

# Abbreviations

AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
cDNA	Complementary DNA, reverse transcribed mRNA
DNA	Deoxyribonucleic acid
E1A	Human adenovirus type 5 early region 1A
EIA	Enzyme linked immunoassay
ELISA	Enzyme linked immunosorbent assay
EST	Expressed sequence tag
FAB	Fast atom bombardment
GC	Gastric cancer
HPAI	Highly pathogenic avian influenza
HPLC	High performance liquid chromatography
IHC	Immunohistochemistry
IM	Intestinal metaplasia
MIP	Molecular inversion probes, synonym for padlock probes
MS	Mass spectrometry
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PLA	Proximity ligation assay
qPCR	Quantitative real-time PCR
RCA	Rolling circle amplification
RCP	Rolling circle product
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism

# Introduction

Life science research has made great progress in the last 30 years due to the availability of molecular biological methods. New biological findings and development of new analytical methods have propelled each other forward. Cloning and DNA sequencing have enabled sequencing of many organisms' genomes, arguably changing the way that life science research is done. Development of the polymerase chain reaction (PCR) gave rise to an entire field of applications, from cloning to extremely sensitive detection of nucleic acids for diagnostics. Microarray strategies made possible very high throughput searches for signaling pathways, candidate genes in diseases and previously unknown degrees of multiplexed analyses.

Clinical samples archived in biobanks linked to well documented clinical data are a valuable resource for life science. To take advantage of these large collections of samples optimally, users must accept strict limitations on sample consumption. Thus, current life science research increasingly requires methods that analyze minute amounts of often complex samples, measure multiple parameters at the same time, and collect information down to single molecule levels.

Complex samples like serum or tissues commonly stored in biobanks are studied either with hypothesis-free methods or methods specifically targeting desired analytes. The former result in an overview image of the molecular contents, usually dominated by abundant species and with less information on rare molecules. The latter provide more detailed information about a chosen aspect of the sample. Often, analyte molecules are targeted by a specific binder, which is in turn quantified by a sensitive detection method. While the selectivity of antibodies and other common binders is sufficient to detect abundant targets, even the best antibodies exhibit some unspecific binding to other molecules. The detectable signal from rare target molecules is thus frequently overshadowed by signal from unspecific binding events.

To abolish signal generated from spurious binding, many analysis methods for liquid samples produce signal only from target molecules that share at least two orthogonal features with the intended analyte: Southern, northern and western blots resolve molecules by their electrophoretic mobility and require binding of an analyte-specific probe. Target molecules can be distinguished by chromatographic mobility and molecular mass in samples analyzed by liquid chromatography coupled to mass spectrometry. In PCR, sandwich ELISA, padlock probing and proximity ligation two probe binding events to the target molecule are required to yield a detectable signal, leading

to exquisite specificity. Only the latter group of methods where signal amplification and detection strictly depend on double specificity has the potential to detect very rare analytes from unenriched samples.

In my doctoral research presented herein, I have applied both padlock probing and *in situ* proximity ligation. In both methods, two primary binding events are required to generate signals, thereby providing good assay selectivity. Binding reagents in these two methods are oligonucleotides and antibodies conjugated to oligonucleotides, respectively. When binding brings these oligonucleotides into close proximity, a ligase can create a single-stranded DNA circle. This newly generated sequence can be amplified by PCR, or rolling circle amplification (RCA) can be used to generate long single-stranded DNA molecules consisting of concatemeric copies of the DNA circle. These concatemers provide numerous binding sites for labeled oligonucleotides, so that local enrichment of the label makes them easily distinguishable from any non-specifically bound detection oligonucleotides.

# Ligation-mediated molecular analyses

This chapter gives an overview of DNA ligation in biomolecular assays. After defining a few key terms, I begin by pointing out the integrating function of ligation reactions. I briefly discuss the properties and reaction mechanisms of DNA ligase enzymes, and of chemical ligation.

Probe-based biomolecular assays involve recognition of target molecules by some type of binder, so I describe how several binder-classes can be combined with ligation-based assays. Thereafter, two DNA amplification reactions are introduced that can be used to simplify detection of the ligation products. I close by pointing out how the high specificity throughout ligation-based biomolecular assays and amplified detection makes them particularly suitable for detection of low concentrations of target molecules in complex samples.

To conclude this chapter, I describe padlock probe ligation and proximity ligation, the two ligation-mediated techniques applied in this work.

## Definitions

### Sensitivity

Sensitivity describes how efficient the assay is in detecting the target molecules.

### Specificity

Specificity is a measurement of how efficient the assay is in detecting the correct target molecules. A specific assay correctly measures low concentrations of target molecules despite high concentrations of other substances in the sample.

### Affinity

The affinity describes the equilibrium binding of one molecular species to another. A high affinity binder binds tightly to the target, and even low concentrations of binder are sufficient to bind the majority of the target molecules. While researchers consider antibodies to be aimed against a target species, in principle any antibody has a finite affinity for any other molecular

species. Antibodies are useful when they have very high affinity towards their antigen and undetectable affinity towards other molecules in the sample.

## Selectivity

The ratio of a binder's affinity for the target versus its affinity for other molecules is known as selectivity. A selective binder binds only the intended target molecule despite the presence of high concentrations of other molecules.

## DNA ligation

While most life scientists know DNA ligases for their applications for construction purposes in *e.g.* molecular cloning, the ligation of two DNA ends is also an attractive reaction for molecular analyses, as it functions as a molecular AND gate (Figure 1). Just like the electronic logical gates, a new molecular signal in the form of the joint sequence can only be generated when the ends of two suitable oligonucleotides are located near each other. Thus two signals can be integrated to either increase the specificity of an assay, or to simultaneously interrogate the combination of two separate features of a target.

DNA ligation is the process of joining of the 5' end of one oligonucleotide to the juxtaposed 3' end of another oligonucleotide, usually in a nicked duplex DNA structure, by either enzymatic or chemical methods.

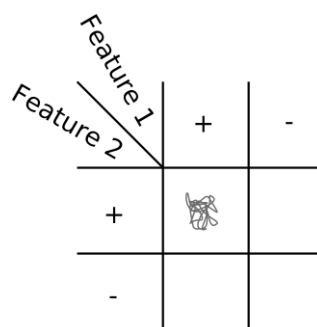


Figure 1: DNA ligation and amplification as a molecular AND gate. A signal is only generated if two features are recognized in close proximity.

## Enzymatic ligation

DNA ligases were first isolated in the late 1960s (1-3) and applied for molecular cloning (4, 5) and synthesis of long oligonucleotides (6). In their cellular environment DNA ligases catalyze the joining of Okazaki fragments produced during the replication of genomic DNA, forming the lagging strand, and repair nicks that arise during DNA damage repair and DNA recombination. DNA ligases have been isolated from a number of organisms, and share general reaction mechanism, but differ in substrate requirements and temperature optima.

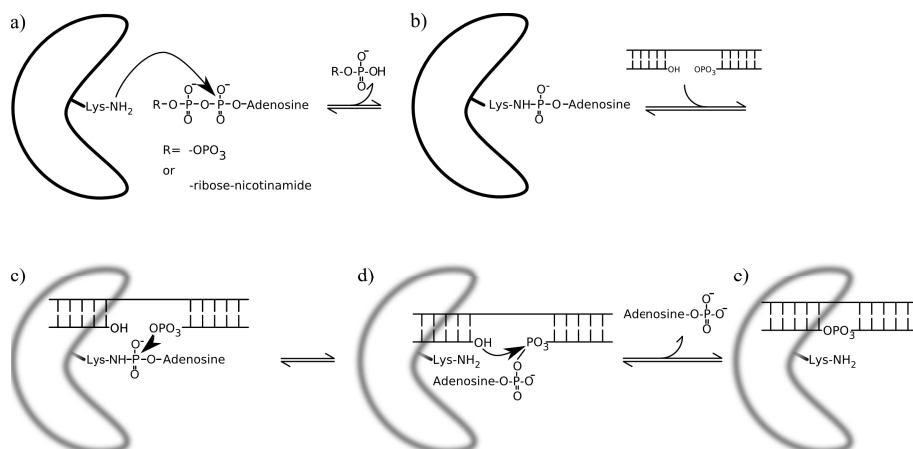


Figure 2: Mechanism of enzymatic DNA ligation. In a first step a primary amino group at the active site of the ligase is adenylated, preserving the energy of a phosphodiester bond from a cofactor, either NAD<sup>+</sup> or ATP, depending on if the enzyme is of eubacterial origin or not (a, b). On binding to a nicked DNA duplex, this adenylyl group is transferred to the 5'-phosphate at the nicked site (c). Finally, the enzyme catalyzes an attack of the activated 5' end by the nearby 3'-hydroxyl group (d), resulting in the formation of a phosphodiester bond between the 5' and 3' ends, and releasing AMP (e).

All ligases contain a lysine residue in their active site. Depending on the origin of the enzyme, this lysine is adenylated by reaction with either ATP or NAD<sup>+</sup> (Figure 2a, b). The ligation reaction begins when the adenylated ligase enzyme binds to a single-stranded break (with a phosphorylated 5' end) in double-stranded DNA (Figure 2c). The adenylyl group is transferred to the 5' phosphate, which is in turn attacked by the 3' hydroxyl group (Figure 2d), releasing the adenylate. The enzyme is released and is ready for another ligation after acquiring a new adenylyl charge (7). Ligases isolated from thermophilic organisms have higher optimal reaction temperatures, and can be used for thermocycled ligation reactions (8). Higher reaction temperatures also destabilize transiently hybridized imperfectly matched DNA complexes, further decreasing the risk of chance ligation events.

The substrate specificity of DNA ligases ensures that only nicks with correctly matched basepairs are closed. This mismatch discrimination is due to the structure of DNA ligases, whose footprint covers several bases upstream and downstream of the nick position (9). Substrates for ligation that are mismatched at a single position at the nick are ligated 10<sup>4</sup> times slower than correctly matched complexes (10).

## Chemical ligation

A number of investigators have developed non-enzymatic ways to target-specifically join two oligonucleotides. Reactive groups introduced at the

ends of oligonucleotides can be joined when brought in close proximity by hybridizing in juxtaposition. As this reaction is not catalyzed, the kinetics of the reactive groups must be a compromise between ligation speed and specificity – chance encounters of highly reactive oligonucleotides in solution are very hard to distinguish from target-dependent ligations. Chemical ligation is less useful for single nucleotide typing, as mismatched complexes reduce the uncatalyzed reaction speed much less (typically ca. 15 times (11, 12)) than the enzymatic ligations.

Assays that employ non-enzymatic ligation of oligonucleotides to detect intracellular targets can dispense with the addition of ligases, and can in principle measure events without fixing the cells.

## Binders

Ligation-based assays can be used to measure a wide range of targets. Only a pair of target-specific binders that each contain an oligonucleotide are required to perform ligation-based assays, and many classes of binders either consist entirely of DNA or can be modified to contain oligonucleotides.

Ligation of linear oligonucleotide binders has been used to detect single nucleotide differences in genomic DNA (13), and the probe design has been further refined into padlock probes – linear oligonucleotide probes with target specific sequences at both ends that become circularized upon ligation at the target sequence. Aptamers (14) are oligonucleotides that have been selected to contain target-binding domains for such targets as small molecules, proteins and protein modifications. Aptamers can be easily extended to include a ligatable sequence extension attached to the binding portion (15). In this form, aptamers constitute very stable, easily produced and often highly specific binders for ligation-based assays. Antibodies are binding reagents routinely used and commercially available for thousands of targets. The huge repertoire of antibodies can be converted into ligatable probes by conjugation of oligonucleotide arms (16). Other protein-based binders could also be conjugated, such as affibodies (17), ankyrin repeats (18), trinectins (19) or single chain antibodies. Thus, almost every class of available binders can be used in ligation-dependent assays with unified readout formats for DNA.

## Amplification methods

Once the DNA portions of the binders have been ligated, the newly formed joint sequence can be amplified with nucleic acid techniques for easier detection and quantification.



## Polymerase chain reaction (PCR)

Ligation products can be exponentially amplified by PCR. With primer sites located on both sides of the ligation site, only successfully ligated probes can act as a template for PCR (Figure 3a, b). As the amplification products are soluble, PCR is suitable for bulk quantification of ligations in a sample. Combined with real-time quantitative PCR (qPCR), up to 20 protein markers have been measured from a 10  $\mu$ l serum sample (20).

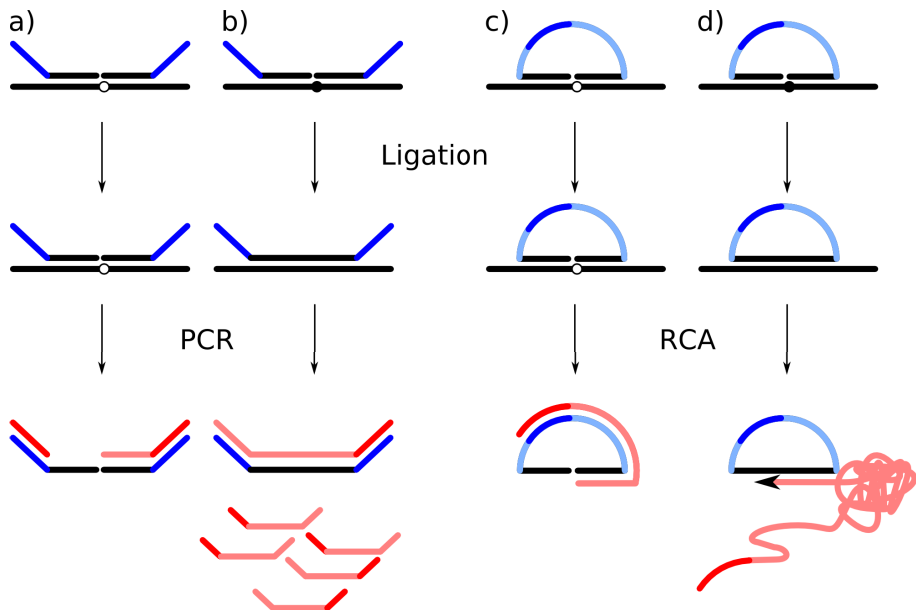


Figure 3: Oligonucleotide probes in ligation assays: a) Two probes hybridize to the position of a single nucleotide polymorphism (SNP) so that the SNP is positioned at the nick. The mismatched SNP (white circle) inhibits ligation so that PCR with primers (red) against primer target sites in the probes (blue) can only extend (light red) up to the end of the probe. No exponential amplification occurs. b) When the target DNA contains the matching variant of the SNP (black circle), ligation forms a new DNA species which can act as a template for PCR. The amplification can be detected *e.g.* by qPCR. c) A padlock probe with target-complementary regions (black) designed to bind at a SNP site contains a primer site and further synthetic sequences (light blue). The mismatched SNP inhibits ligation into a circle, so that RCA only results in a short extension of the primer. d) The padlock probe becomes circularized on the matched SNP variant and acts as a template for RCA generating a long concatemeric single stranded amplification product.

## Rolling circle amplification (RCA)

The ligation step of both DNA and protein detection assays can be designed so that only recognition events result in formation of single-stranded DNA circles.

The highly processive phi29 DNA polymerase can elongate a primer templated by a DNA circle, producing a long DNA strand composed of hundreds of concatemeric copies of the circle complementary sequence from each circular DNA template (Figure 3d). These threads of single stranded DNA typically fold into a ball of DNA (a rolling circle product, RCP) that is attached to the assay target via the binder and primer. Labeled oligonucleotides directed against the circle sequence can be hybridized to the RCP, so that the high local concentration of label can be detected by appropriate microscopic methods. The strong, point-like signals from RCPs are easily counted and can be re-probed to interrogate multiple features encoded in the DNA circle.

## Requirements of highly sensitive assays

Any assay that aims to detect a very rare species in complex samples needs to maintain specificity in all steps from sample binding to signal quantification. This chain of highly specific steps can be realized with ligation-based assays. During recognition, the requirement for two independent binding events drastically decreases false positive signals. Amplification by PCR requires the template sequence, but in the absence of template jumping PCR artifacts can result in amplification from non-contiguous templates (21, 22). Target-specific quantification of PCR amplification with molecular beacons or TaqMan probes ensures that amplification artifacts are not mistaken for true signal.

RCA produces one amplification product per template, and (at least in the absence of appropriate primers) the amplification product cannot act as template. Thus RCPs originate only from correctly ligated DNA circles. The intense, point-like signal from RCPs is so distinct from the surrounding background signal that they are easily counted in liquid samples (23) and tissue sections (24).

## Padlock ligation

Padlock probes – also referred to as molecular inversion probes (MIPs) – are linear, single stranded oligonucleotides comprising an artificial, non-target-complementary tag sequence flanked by two target-complementary sequences (Figure 3c). After their juxtaposition on the target sequences, these can be joined by ligation into DNA circles. These circles become topologically linked to the target strand, hence locked in place. The specificity of the double hybridization combined with the single nucleotide mismatch discrimination of DNA ligases is sufficient for detection of single nucleotide polymorphisms (SNPs) in genomic DNA (25). The non-target-complementary portion can be chosen as binding sites for amplification primers, hybridiza-

tion sites for labeled oligonucleotides or for identification and quantification of amplification products of reacted probes on tag microarrays (26, 27).

Padlock probes have been used *in vitro* for multiplexed SNP genotyping (27), studies of copy-number variations (28), detection of rare pathogens and highly sensitive quantification of cDNA (29). The small hybridization footprint of padlock probes enables studies on short or degraded targets, like formalin-fixed, paraffin-embedded tissues (28) and forensic or ancient DNA samples. In an *in situ* application of padlock ligation, Larsson and coworkers detected a SNP in mitochondrial genomes (30).

## Proximity ligation assay (PLA)

Padlock ligation is restricted to detection of nucleic acid targets. To detect, quantify or locate *e.g.* protein targets with a ligation-based assay, binder-oligonucleotide conjugates (proximity probes) are used to translate the binding event into a nucleic acid signal. In the basic form of the proximity ligation assay, two proximity probes bind to separate sites on the target protein molecules, bringing their oligonucleotide portions close to each other.

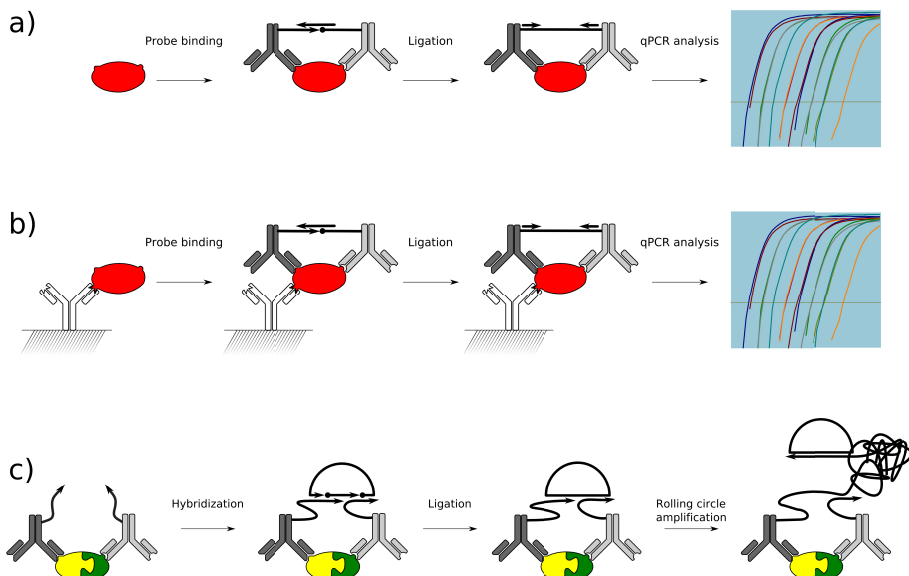


Figure 4: Proximity ligation. a) In PLA in solution, binding of two proximity probes to the target (red) places their oligonucleotide arms close to each other. Upon hybridization of a short ligation template oligonucleotide, the arms are joined by ligase. The newly formed joint sequence is quantified by qPCR. b) In the solid phase version of PLA, the analyte is precipitated onto a solid phase by target-specific binders. Unbound sample components or excess reagents are washed away before ligation and quantification by qPCR. c) In *in situ* PLA, the proximity probes are designed to act as ligation templates for formation of a DNA circle from two additional oligonucleotides, and to act as primers for RCA.

Depending on whether the assay detects targets in solution or *in situ*, the proximity of the binders is converted into reporter molecules by different means:

### PLA in solution

Primer sites are integrated in the oligonucleotide arms of the proximity probes. When two proximity probes bind close to each other they can be ligated by jointly hybridizing to a short synthetic ligation template oligonucleotide (Figure 4a). The resulting ligation product can be quantified as a surrogate marker via PCR amplification (15, 31).

PLA in solution is a highly sensitive assay amenable to automation, as it requires no manipulations except addition of reagents. While buffer optimization in a solution phase assay must consider the requirements of all the consecutive assay steps, buffers can easily be changed when the target protein is immobilized on a solid support by binders (Figure 4b), and optimization can concentrate on each assay step separately. To improve resilience of PLA to PCR inhibitors present in complex samples the immobilized target protein can be washed before addition of proximity probes (32). While the removal of non-target sample components eliminates a source of false-positive signals, extra care must be taken to block unspecific adhesion of proximity probes to the solid phase. Washing also constitutes a source of variability, *e.g.* when liquid is removed incompletely or part of a bead solid phase is accidentally removed. Solid phase PLA carried out on functionalized tube surfaces is limited by diffusion kinetics of targets and reagents, while the kinetics of PLA using bead solid phases approach those of a solution phase assay (33). Target molecules can also be captured on beads from a large volume of low concentrated samples, to continue the assay in a smaller volume.

### PLA *in situ*

Macromolecular complexes (24) and post-translational modifications (34) can be located in cultured cells and sectioned tissues using *in situ* PLA. In solution PLA the proximity between probes results in ligation of the oligonucleotides attached to the probes. By contrast, in *in situ* PLA proximity probes act as ligation templates for the joining of two additional oligonucleotides (Figure 4c). Only when both proximity probes template these ligation reactions is a DNA circle formed. The distance requirement is set by the size of the binders and the length of the oligonucleotides, usually in the range between 10 to 50 nm. This circle then acts as the reporter molecule when phi29 polymerase amplifies the circle into an RCP, primed by one of the oligonucleotide arms. The RCP folds into a ball of DNA and can be stained by fluorescent oligonucleotides with the tag sequence derived from the DNA circle. The high local concentration of fluorophores is easily discerned from background fluorescence and tissue auto-fluorescence, so that the intense, point-like signals can be visualized and counted (Figure 5).

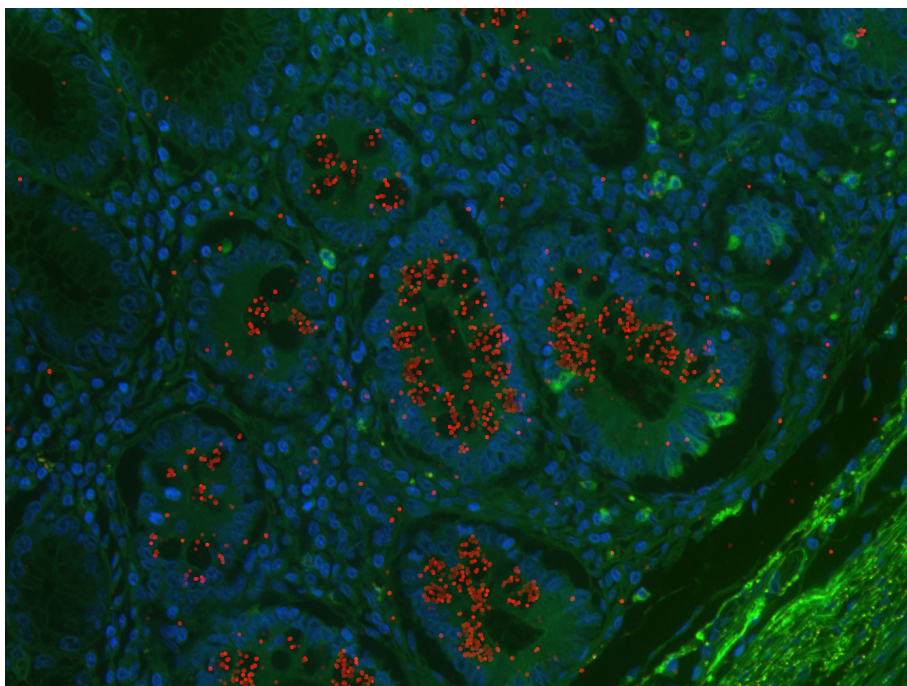


Figure 5: *in situ* PLA with proximity probes against mucin 2 and sialyl-Tn. Staining the RCPs with fluorescent oligonucleotides (red) makes their location visible in the microscope.

To reduce the number of conjugation reactions and testing of proximity probes required to set up a new *in situ* PLA, non-modified primary antibodies can be used together with secondary proximity probes, directed against species (34) or isotype-specific antibodies. This permits the direct application of thousands of commercially available antibodies with PLA, omitting the need to conjugate each new antibody.

Strong amplification of a signal with limited specificity leads to background signals and impairs the limit of detection. The improved specificity provided by the requirement for double recognition permits full use of powerful signal amplification techniques. Application of PLA to proteins transferred to a western blot membrane can significantly improve the specificity and thereby the sensitivity by decreasing the effects of cross-reactive antibodies; alternatively PLA introduces the unique ability to directly study post-translational modifications of a specific protein on a membrane.



# Present investigations

This chapter briefly outlines the three investigations included in this thesis, discusses their novelty and future perspectives.

## Paper I – Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses by use of padlock probes

Highly pathogenic avian influenza H5N1 and H1N1 “swine flu” have posed serious health problem for animal and human populations around the globe. Since the first outbreak of highly pathogenic avian influenza (HPAI) of serotype H5N1 in 1997, wholesale slaughter of poultry herds in outbreak areas made clear the need for sensitive, specific and faster methods for serotyping influenza. If the infectious agent in non-pandemic influenza infections can be identified within hours, animal lives can be spared and the economic impact of culling can be minimized. Fast diagnostics of human influenza have concentrated on confirming influenza infection in order to efficiently use antiviral drugs – these drugs need to be applied shortly after onset of symptoms in order to be effective (35). Therefore simple and fast, but relatively insensitive methods as immunochromatography have been developed for bedside diagnostics. With turnaround times between 10 and 30 minutes these tests can be performed by relatively untrained personnel, albeit with limited sensitivity and at considerable cost (36, 37). Highly sensitive methods as virus culture in eggs and cell culture combined with immunofluorescence or enzyme linked immunoassays (EIA) complete the repertoire of the virologist. The circulating influenza viruses differ widely in species specificity, virulence and genotype. Reactivity to antisera (serotyping) is used to classify influenza isolates systematically. Serotypes are groups of isolates recognized by the same anti-hemagglutinin and anti-neuraminidase antisera. To serotype a virus isolate, sufficient amounts of virus must be produced, currently by biological replication in an infectable tissue (38) – resulting in long turnaround times for these assays. Despite three days incubation during the biological amplification, these tests are inexpensive and highly sensitive when fresh virus isolates are used.

Reverse transcriptase PCR (RT-PCR) assays are used for sensitive detection of viral nucleic acids (39, 40), but can only detect one serotype at a time (41) and suffer from the high mutation rate of RNA viruses like influenza.

## Results

The aim of this work (42) was to genotype all relevant influenza hemagglutinin and neuraminidase antigens, in an assay that is superior to state-of-the-art methods with respect to handling, turnaround time and multiplexing. Genotyping the virus isolate could make use of the quick amplification mechanisms available to nucleic acid analysis. However the mutation rate of the influenza RNA genome and the small sequence differences between serotypes complicate attempts to genotype influenza by PCR. Padlock probes can reliably genotype hundreds or more short sequences per sample, enabling redundant testing of several characteristic sequences to identify a serotype.

Signature genomic sequences for each serotype were identified by aligning the hemagglutinin and neuraminidase genes of thousands of influenza isolates. Padlock probes were directed to target sequences that include characteristic sequence variants, so that the probes could be ligated and amplified in parallel with fluorescent PCR primers. The fluorescent PCR products were separately identified and quantified by hybridizing to tag sequences on a DNA-microarray with oligonucleotide tags complementary to a probe-specific sequence in each padlock probe. Of the 77 virus strain samples available for this study (combinations of 9 neuraminidase and 16 hemagglutinin serotypes) 97% were correctly identified (75/77) by thresholding of the fluorescence signal scanned on the microarrays. The only sample of serotype H10N8 didn't yield any signal for hemagglutinin or neuraminidase, and one of two samples of H12N5 serotype was correctly genotyped for H12 hemagglutinin, but no neuraminidase genotype could be determined.

## Future perspectives

This paper demonstrated that a padlock probe genotyping test with multiplexed readout can be useful for screening of pre-pandemic cases. Rapid, inexpensive and reliable influenza subtyping has the potential to markedly alleviate the strain on health care systems during potential influenza epidemics in populations where widespread vaccination is not an option. To make a difference, the tests need to be quick enough to be reliably performed in the doctor's office by healthcare staff without special training, while a positive/negative result for the serotype in question would be sufficient.

To transfer the padlock test into a doctor's office test, several issues need to be addressed: In order to keep handling to a minimum, blood samples would need to be tested with very little purification. A microfluidic RNA preparation could be used to enrich RNA from a sufficient volume to capture



enough pathogen material (43). To save time and simplify temperature requirements, reverse transcription of the RNA could be omitted in favor of direct padlock ligation on the RNA target (44). Amplification would ideally take place at constant or room temperature, arguing for RCA instead of PCR. In principle, ligation, RCA and hybridization of labeled oligonucleotides can be carried out in one or two steps, further reducing handling. Depending on the label used, a fluorescence reader could directly quantify the result, or the operator could read out the result manually after addition of a chromogenic substrate (45) (requiring enzyme-labeled detection oligonucleotides).

Virus serotyping is just one application of differentiation of closely related RNA and DNA targets. Other bacterial, fungal or viral pathogens could be distinguished in a similar test. Determination of antibiotic resistance marker genes in bacterial infections could speed up informed treatment decisions and delay the appearance of multiply resistant strains.

The de-multiplexing of amplification products on DNA tag microarrays permits the readout of many tests on the same microarrays – a major cost-factor for the establishment of new tests in a research environment. Even tests in a clinical environment can be produced more flexibly: To add a target sequence for *e.g.* a new influenza serotype to the tested panel, one only needs to design padlock probes for their characteristic sequences so that their amplification products become de-multiplexed to unused DNA tag features. Continuous updating of the serotype panel thus entails mixing of padlock probes, rather than re-designing the entire assay and producing new microarrays.

## Paper II – Analysis of full transcript splice patterns at single molecule resolution

Before transport to the cytoplasm and translation into protein, primary RNA transcripts are processed in the nucleus of eukaryotic organisms. Long stretches of non-coding RNA (introns) are removed and the remaining coding regions (exons) are spliced together. A protein-RNA complex – the spliceosome – catalyzes the removal of intron sequences, followed by the joining of the remaining exons. The RNA components of the spliceosome recognize the exon-intron borders by consensus sequence elements. The majority of introns bear canonical 5'-GT and 3'-AG dinucleotides, which directly interact with spliceosome components to determine the splice site (46).

Nucleosomes preferentially bind at genomic sequences that are transcribed into exons (47), which decreases transcription speed (48) and appears to favor assembly of the spliceosome during transcription. Thus chromatin structure – as well as RNA sequence – influences which splice sites are used by the spliceosome.

Regulatory sequences in transcripts are bound by factors that favor or disfavor recruitment of the initial spliceosome components. In addition to the

sequences marking the exon-intron borders, exonic and intronic regulatory sequences are required for regulated assembly of the spliceosome and correct splicing (49-51).

Once the complete spliceosome is assembled, splicing begins with the attack of a 2'-OH group of an intronic adenosine on the phosphate of the 5'-splice site, forming a lariat structure and releasing a free 3'-OH at the 5'-splice site. This 3'-OH in turn attacks the phosphate at the 3'-splice site, joining the two exons (52).

The macromolecules involved in assembling the spliceosome are regulated by signal transduction events (53, 54), though the mechanisms are yet poorly understood (55). Modifications by *e.g.* phosphorylation of splicing factors change their binding preferences to the primary transcript, to spliceosome components, or make the factor unavailable by sequestration (56). This results in the selection of alternative splice sites. Depending on the position of the splice sites within the transcript, the result can be exclusion of a cassette exon (exon-skipping, most frequently), retention of an intron, mutually exclusive splicing of an internal exon (57) or polyadenylation site or alternative 3'- or 5'-splice sites (58).

Differently spliced mRNAs code for protein isoforms with different amino acid sequences. These isoforms often differ in their biological function, as they bind to different partners, localize differently in the cell or possess different catalytic properties or pharmacological effects (59).

The human genome contains between 20000 and 25000 protein-coding genes (60), barely four times more than in budding yeast. Alternative splicing is an important mechanism adding to the complexity of higher organisms. Most genes consist of several exons and the proportion alternatively spliced is estimated between 95% and 100% (61, 62). mRNA isoforms generated by alternative splicing are differentially expressed in development stages (63), tissues (64) and individuals. Changed splicing patterns are involved in disease (51, 58, 64-66).

Sequencing of cloned RT-PCR products provides information about the entire transcript (as far as it is captured by the PCR primers), one gene at a time (ESTs (67)). Due to the massive effort of cloning and sequencing, this method is useful as a gold-standard for defining splice sites but it is unsuitable to investigate more complex splicing patterns and rare splice variants. Known alternatively spliced isoforms can often be differentiated by RT-PCR followed by gel electrophoresis, via differences in sizes. This simple method requires prior knowledge about the splice variants, and it yields data about one gene/splice event at a time. Setting up a new assay is easy, but analysis of multiple genes in many samples is labor-intensive and expensive.

More parallel strategies include whole genome exon arrays, where variable splicing can be detected as *e.g.* tissue-dependent expression of an exon. This genome-wide approach is suitable for detection of clear-cut changes in splicing patterns, where transcripts are consistently spliced in each sample, but in a manner that differs between samples (68). Johnson *et al.* analyzed

splicing of 10000 multi-exon genes in several tissues with microarrays containing probes for all exon-exon junctions (69). Such a high-throughput strategy is useful to screen for dominant use of splice sites in several samples. Less strongly expressed, functional splice variants are submerged under the signal from the majority variant.

To investigate more complex alternative splicing patterns, Zhu *et al.* propose use of PCR colonies (colonies) in combination with mini-sequencing. This technique provides splicing information from thousands of single transcripts, which permits studies of rare, yet important alternative transcripts (70).

## Results

In this paper, we develop and demonstrate a ligase-based method to analyze the complete splice patterns of a gene in thousands of individual transcripts to detect rare spliced transcript isoforms. Instead of cloning and sequencing transcripts, the exon information from individual cDNA molecules is encoded as synthetic tag chains, used to determine the presence or absence of each exon in individual transcripts. The design of the probes permits parallel analysis of several genes from the same 1  $\mu$ l of cDNA.

We demonstrated the specificity of this method with pure artificial cDNA of known exon composition. Probes were hybridized to the cDNA and ligated to form an encoded representation of the cDNA sequence. These DNA representations were circularized and amplified by RCA. The RCPs were deposited on a microscope slide and decoded by hybridization with one fluorescent detection oligonucleotide per exon. The collected images were analyzed in a customized Matlab script. Between 96 and 99.9% of all transcripts were decoded correctly.

We further analyzed alternative splicing during the well-studied infection cycle of adenovirus early region 1-RNA (E1A, (71, 72)) and host  $\beta$ -actin RNA in the same reaction. Total RNA was prepared from HeLa cells infected with adenovirus at 6, 12, 24, 36 and 48 hours post infection. Cellular RNA was reverse transcribed into cDNA, and analyzed as describe above. At 24h post-infection, exons B to E had been spliced out from 90% of the transcripts, while earlier cDNA preparations contained higher proportions of isoforms ABCDF and ABCF.

## Future perspectives

High throughput sequencers have become an efficient way to refine splice site maps, but complete sequences from a single transcript molecule cannot be collected due to the limited read lengths, and detection of rare splice events still requires very deep sequencing. Alternatively spliced transcripts can be present at very different concentrations in a sample.

With the present technique, we confirmed the known adenovirus E1A splicing patterns (71, 72), demonstrating that splicing can be simultaneously analyzed in thousands of transcripts over the entire transcript. To our knowledge, no functional minority splice variants have been observed so far, possibly due to the lack of efficient methods to discover them. Now that the spliceotyping method is well established, it is time to study genes with more complicated splice patterns like *KCNMA* (73, 74) or *mod(mdg4)* (75), to observe whether the postulated linked splicing events occur in nature.

Spliceotyping analyses of genes with complicated splicing patterns require many exon probes, and thus many hybridization cycles using our current decoding scheme. Instead of repeatedly staining, microscoping and stripping RCPs, and analyzing image data, the tag strings' sequence could simply be read by high throughput sequencers. Transcript classification would become much more certain, and scaling to read hundreds of thousands of transcripts – abbreviated by our ligation approach – is easily accommodated in current sequencers.

### Paper III – MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinoma

Mucin 2 is a secreted member of the mucin family of highly glycosylated proteins that cover epithelia of the reproductive, gastrointestinal and respiratory tracts. The family members display tissue-specific expression patterns. Mucin 2 is not expressed in healthy gastric mucosa (76-81). During the course of the gastric carcinogenesis pathway (*Helicobacter pylori* infection, gastritis, intestinal metaplasia and dysplasia (82, 83)) gastric expression of mucins changes and the tissue undergoes a morphological change towards intestinal phenotype, including expression of mucin 2. The expression of mucin 2 is thought to be part of the defense against *H. pylori* infection.

The transition from a normal epithelium towards gastric cancer is also accompanied by changes in glycosylation patterns. The carbohydrate structure sialyl-Tn (Neu5Ac $\alpha$ -6GalNAc-O-Ser/Thr) is an important cancer antigen not normally expressed in human tissues (84, 85) and a serum marker of poor prognosis in *e.g.* ovarian cancer (86) and gastric cancer (87). Antibodies against sialyl-Tn are used in diagnostic kits for measuring circulating mucins bearing sialyl-Tn in cancer patients. Sialyl-Tn was also a component of Theratope, a trial vaccine that was used to treat *e.g.* breast cancer (88). So far the carrier protein of sialyl-Tn in gastric cancer and intestinal metaplasia is unknown.

Due to the complexity of glycosylation structures, analysis of the sugar structure of proteins has lagged behind other protein and nucleic acid analyses. Classical chemical approaches have used large quantities of antigens

to analyze chemical degradation products by chromatographic methods and mass spectrometry (MS). More modern approaches include the following:

1. Fast atom bombardment (FAB) mass spectrometers coupled to HPLC can provide information about sugar branching points and linkage (89, 90). Typically a protein of interest is purified and the sugar cleaved from the apoprotein. The sugar portion is then analyzed by FAB-MS. This method requires moderate amounts of biological material, but it cannot be used for localized detection of apoprotein glycosylation. Altered glycosylation in a subpopulation of cells (*e.g.* cancer surrounded by healthy tissue) is not easily detected.
2. Nuclear magnetic resonance spectrometry (NMR) methods can provide complete macromolecule solution structures, but require large amounts of sample (91). Interpretation of the NMR spectra is difficult and time-consuming. Due to the large amounts of material required, this method is not useful for localized detection of glycosylation in patient samples.
3. Immunohistochemical and immunofluorescent analyses of tissue sections complement the analytical MS and NMR methods. While the former are used to delineate sugar structures without previous information, the immunological methods are used to locate the expression of a sugar epitope in tissues. Analyses of the cellular location of the sugar epitope fail to reveal the nature of its apoprotein, but co-localization of the sugar epitope and a protein indicates that the protein may carry the sugar residue. However, due to the relatively low resolution of immunohistochemical and immunofluorescent staining methods, co-localization cannot serve as proof of chemical linkage (92).
4. Samples that have been immunoprecipitated with binders against sugar epitopes or apoproteins can be resolved by gel-electrophoresis, and further analyzed with blotting or MS methods. Without prior hypotheses this approach can identify the carrier of *e.g.* a sugar epitope by MS-sequencing. The sample amounts required to perform such analyses (of the order of 1 mg) precludes analyses in patient material, much less tissue localization of apoprotein-sugar pairs (93).
5. Change of immunohistochemical (IHC) staining intensities of specific proteins in tissues before and after treatment with glycosidases is an indication of apoprotein epitope obstruction by a substrate of the glycosidase. The results are difficult to evaluate due to the low resolution of IHC staining methods and the difficulty of comparing staining intensities between different sections.

## Results

Here we determined whether mucin 2 is a carrier of the sialyl-Tn antigen in intestinal metaplasia (IM) and gastric carcinoma (GC), and visualize the tissue location of sialyl-Tn on mucin 2 *in situ*.

Mucin 2 is a secreted glycoprotein that is expressed in intestinal metaplasia and gastric carcinoma. Sialyl-Tn is a sugar epitope that is used as a serum cancer marker. Based on the co-location of immunofluorescence of mucin 2 and sialyl-Tn (92) we hypothesized that mucin 2 is the carrier of sialyl-Tn in gastric cancer. To test this hypothesis we developed isotype-specific secondary proximity probes against mouse IgG and IgM and applied them in an *in situ* PLA (24) against mucin 2 and sialyl-Tn, using primary monoclonal IgG and IgM antibodies.

Tissue sections were immunohistochemically stained for sialyl-Tn and mucin 2 in separate reactions and counterstained with haematoxylin eosin. Microscope images were recorded from sites with IM or GC morphology. PLA against mucin 2 and sialyl-Tn was performed on adjacent sections. As expected, staining occurred in IM and GC areas of the tissue where immunohistochemical staining indicated the expression of both mucin 2 and sialyl-Tn. PLA signals were found in 11 of the 12 tissue sections containing IM and GC, demonstrating that mucin 2 is a major carrier of sialyl-Tn in these tissue sections.

## Future perspectives

Due to the limited patient material included in this study, we didn't attempt to correlate clinical progress of the cases to the glycosylation status of mucin 2. To investigate whether sialyl-Tn on mucin 2 has any prognostic value, a larger number of archived GC and IM tissues with documented clinical history should be analyzed.

Further mucins and sugar epitopes can be readily studied with the available molecular tools. The combinations of mucins 1, 2, 5AC and 6 with the sugar epitopes sialyl, Tn, sialyl-Tn, sialyl-Lewis A and sialyl-Lewis X are relevant candidates due to their expression in the gastrointestinal tract.

Establishing an early test for precursors of gastric cancer would enable studies of earlier GC stages in patients, and could ultimately lead to earlier intervention. As tissue section analyses are out of the question for a screening test, we would like to test whether glycosylated mucins are leaked into the blood stream and can be detected there. As observed in one of the GC cases in this study, the lymph system can be invaded by cells expressing mucin 2 bearing sialyl-Tn. With the sensitivity of PLA in solution, early detection of small amounts of glycosylated mucins leaked from small lesions becomes a possibility.

## Future perspectives

Ligation-mediated molecular assays make it possible to analyze DNA, RNA and protein targets with highly sensitive and specific nucleic acid amplification techniques. Together with the potential to quantify or locate several analytes within the same reaction, studies of the interactions of these three classes of biological macromolecules are now within our reach. Imagine to genotype a somatic single nucleotide mutation within a tumor that influences mRNA transport, detect an mRNA of interest as well as its protein product – all in the same tissue section, with utmost specificity.

The need to study biological processes in all their detail pushes the limits of analytical methods. An ultimate goal for analysis is to generate a detailed model of the location and interactions of all molecules in a system. Binder-dependent assays offer a narrowed view that concentrates on some chosen molecule species, but binding and detection steps never reach completion. Like other multistep detection methods, ligation-mediated assays comprise several consecutive steps to convert the presence of the analyte into a signal. As each step is less than completely effective, the conversion rate of analyte to signal is insufficient to detect every single analyte molecule. While *e.g.* binding constants set theoretical limits to what sensitivity can be achieved, further optimization of efficiencies will bring the goal of amplified detection of every single-molecule closer.

While biomarkers having sufficient specificity to determine the donor status (*e.g.* hCG in the pregnancy test) are rare and limited to conditions with one common cause, diagnosing complex diseases like cancer from a blood sample is less straightforward.

Cancers can arise via several routes. For cells to become cancerous, several cellular functions have to be disabled or up-regulated, but alterations of components at different steps of *e.g.* signaling cascades can result in the same inactivating effect. Thus it is only to be expected that biomarker panels are more useful than single markers to detect all the underlying changes that lead to this same phenotype we call cancer. To identify the most promising panels of biomarkers, large numbers of valuable, limited biobank samples need to be screened. Multiplexed assays for multiple biomarkers can reduce the sample consumption, but are challenging to set up, as the concentrations of different candidate markers can vary over many orders of magnitude. The ability to represent proteins, protein-complexes or post-translational modifi-

ations of protein biomarkers as DNA surrogate markers by PLA enables quantification by qPCR over a range of approximately 6 orders of magnitude, so that tens of biomarkers can be quantified with one unified readout format. Studies that systematically screen large numbers of promising cancer biomarker candidates in hundreds or thousands of small samples have the potential to provide doctors with powerful differential diagnostic tools to improve our understanding of early stages of cancer development and guide treatment decisions.



# Acknowledgments

These studies were performed at the Department of Genetics and Pathology at Uppsala Universitet.

Many people made this work possible and exciting. These deserve special mention:

My principal supervisor **Ulf Landegren** for building the creative environment that makes the Molecular Medicine group so different, for sharing your vision of molecular diagnostics and many fruitful discussions. **Mats**, for always leaving the door to your office open, and the endless knowledge about everything that ligates. **Ola**, for your unwavering optimism, for many opportunities and your great spirit. **Anders** and **Carolina R** for keeping track of all our projects reporting.

**Kalle**, **Irene**, **Ida** for being great office-mates, stealing berries in deserted gardens and many fun evenings. **Yuki** for Japanese culinary delights and for initiating almost all our hotpot parties. **Jenny** for always having a few minutes to discuss our latest images. To **Olle** and **Jonas J** for my first steps in ligationology, arrays and weightlifting. To **Jörg** for German Thursdays. To **Edith** for showing me that you can get everything when you have the right attitude. To **Mathias** for your positive approach to life and having fun while getting the work done. To **Carolina W** for getting me hooked on MatLab and image analysis. To **Sara** for lots of paper napkins, joining me in the police station and fingerprinting together. For **Chatarina** who got her RNase-free experiments to work despite working next to my bench. To **David** for a new perspective on our work, and long discussions about fairness and ethics. **Henrik** for consolation about needle-in-the-haystack projects and showing me that 5 is less than 4.

**Fredrik**, **Johan B**, **Mats G**, **Jonas M**, **Sigrun** and **Johan S**, **Katerina**, **Reza**, **Mia**, **Agata**, **Camilla**, **Carl Magnus**, **Gucci**, **Johan V**, **Magnus**, **Malin**, **Marco**, **Maria**, **Rachel**, **Rongqin**, **Spyros**, **Wu Di**, **Lotte**, **Andries**, **Masood**, **Mikaela**, **Pier**, **Yanling** in short the entire Molecular Medicine lab for free discussions in the corridor, swimming, parties, running, teaching me ligationology and working with me in different stages of my career as a PhD student.

**Elin**, **Johan O**, **Lena** and **Carla** who make the lab go round.

To my students **Hamid** and **Alysha** for good questions, inspiration and teaching me to explain.

To my collaborators **Leonor**, **Ana**, **Francoise**, **Mariam**, **Milan** and **Peter** for the great time working with you.

To all the customers of my infamous “Snickers bar” for the excuse to go shopping like in a little boy's dream. To my friends from ICM and Döbeln for my first contacts with Swedish language, student life and explaining “lagom”.

My friends in Uppsala. **Anna** for teaching the meaning of reverse psychology in Puerto Rico. **Natalia** and **Andrey** for epic battles in Carcassonne, **Christoph** and **Nicole** for good company and great breakfast on Midsummer, **Babett** and **Jens** for being our farming neighbors and good advice.

**My family** – spread around the globe as we are – for giving me confidence to “make it”, and giving me curiosity and uncountable opportunities to see the world.

Endless thanks to my wife **Lei** for standing by me even when I'm stressed and totally intolerable. Last but not at all least to my son **Jan Louis** for your great smile and impressing me with the little things in life.

This work was supported by grants from the Karl & Anders Wallenberg foundation and the European Commission.

# References

1. Weiss B, Jacquemin-Sablon A, Live TR, Fareed GC, Richardson CC. 1968. Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from *Escherichia coli* infected with bacteriophage T4. *J Biol Chem* 243: 4543-55
2. Zimmerman SB, Little JW, Oshinsky CK, Gellert M. 1967. Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proc Natl Acad Sci U S A* 57: 1841-8
3. Gellert M. 1967. Formation of covalent circles of lambda DNA by *E. coli* extracts. *Proc Natl Acad Sci U S A* 57: 148-55
4. Jackson DA, Symons RH, Berg P. 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc Natl Acad Sci U S A* 69: 2904-9
5. Maniatis T, Sambrook J, Fritsch EF. 1982. *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y. :: Cold Spring Harbor Laboratory. 545 pp.
6. Agarwal KL, Buchi H, Caruthers MH, Gupta N, Khorana HG, et al. 1970. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* 227: 27-34
7. Lehman IR. 1974. DNA ligase: structure, mechanism, and function. *Science* 186: 790-7
8. Abravaya K, Carrino JJ, Muldoon S, Lee HH. 1995. Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res* 23: 675-82
9. Doherty AJ, Dafforn TR. 2000. Nick recognition by DNA ligases. *J Mol Biol* 296: 43-56
10. Luo J, Bergstrom DE, Barany F. 1996. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res* 24: 3071-8
11. Gryaznov SM, Schultz R, Chaturvedi SK, Letsinger RL. 1994. Enhancement of selectivity in recognition of nucleic acids via chemical autoligation. *Nucleic Acids Res* 22: 2366-9
12. Gryaznov SM, Letsinger RL. 1993. Chemical ligation of oligonucleotides in the presence and absence of a template. *J. Am. Chem. Soc.* 115: 3808-09
13. Landegren U, Kaiser R, Sanders J, Hood L. 1988. A ligase-mediated gene detection technique. *Science* 241: 1077-80
14. Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249: 505-10
15. Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, et al. 2002. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* 20: 473-7
16. Kozlov IA, Melnyk PC, Stromborg KE, Chee MS, Barker DL, Zhao C. 2004. Efficient strategies for the conjugation of oligonucleotides to antibodies enabling highly sensitive protein detection. *Biopolymers* 73: 621-30

17. Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlen M, Nygren PA. 1997. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol* 15: 772-7
18. Stumpp MT, Binz HK, Amstutz P. 2008. DARPin: a new generation of protein therapeutics. *Drug Discov Today* 13: 695-701
19. Koide A, Bailey CW, Huang X, Koide S. 1998. The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* 284: 1141-51
20. Fredriksson S, Horecka J, Brustugun OT, Schlingemann J, Koong AC, et al. 2008. Multiplexed proximity ligation assays to profile putative plasma biomarkers relevant to pancreatic and ovarian cancer. *Clin Chem* 54: 582-9
21. Meyerhans A, Vartanian JP, Wain-Hobson S. 1990. DNA recombination during PCR. *Nucleic Acids Res* 18: 1687-91
22. Paabo S, Irwin DM, Wilson AC. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265: 4718-21
23. Melin J, Jarvius J, Goransson J, Nilsson M. 2007. Homogeneous amplified single-molecule detection: Characterization of key parameters. *Anal Biochem* 368: 230-8
24. Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius KJ, et al. 2006. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* 3: 995-1000
25. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* 19: 225-32
26. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, et al. 2005. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* 15: 269-75
27. Baner J, Isaksson A, Waldenstrom E, Jarvius J, Landegren U, Nilsson M. 2003. Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids Res* 31: e103
28. Wang Y, Moorhead M, Karlin-Neumann G, Falkowski M, Chen C, et al. 2005. Allele quantification using molecular inversion probes (MIP). *Nucleic Acids Res* 33: e183
29. Ericsson O, Jarvius J, Schallmeiner E, Howell M, Nong RY, et al. 2008. A dual-tag microarray platform for high-performance nucleic acid and protein analyses. *Nucleic Acids Res* 36: e45
30. Larsson C, Koch J, Nygren A, Janssen G, Raap AK, et al. 2004. In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat Methods* 1: 227-32
31. Gullberg M, Gustafsdottir SM, Schallmeiner E, Jarvius J, Bjarnegard M, et al. 2004. Cytokine detection by antibody-based proximity ligation. *Proc Natl Acad Sci U S A* 101: 8420-4
32. Gustafsdottir SM, Nordengrahn A, Fredriksson S, Wallgren P, Rivera E, et al. 2006. Detection of individual microbial pathogens by proximity ligation. *Clin Chem* 52: 1152-60
33. Darmanis S, Yuan Nong R, Hammond M, Gu J, Alderborn A, et al. 2009. Sensitive plasma protein analysis by microparticle-based proximity ligation assays. *Mol Cell Proteomics*
34. Jarvius M, Paulsson J, Weibrecht I, Leuchowius KJ, Andersson AC, et al. 2007. In situ detection of phosphorylated platelet-derived growth factor receptor beta using a generalized proximity ligation method. *Mol Cell Proteomics* 6: 1500-9
35. WHO. 2005. WHO recommendations on the use of rapid testing for influenza diagnosis World Health Organization

36. WHO. 2002. WHO Manual on Animal Influenza Diagnosis and Surveillance. World Health Organization
37. Dominguez EA, Taber LH, Couch RB. 1993. Comparison of rapid diagnostic techniques for respiratory syncytial and influenza A virus respiratory infections in young children. *J Clin Microbiol* 31: 2286-90
38. Alexander D. 2008. Avian Influenza. In *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*, pp. 465-81: Office Intl Des Epizooties
39. Steininger C, Kundi M, Aberle SW, Aberle JH, Popow-Kraupp T. 2002. Effectiveness of reverse transcription-PCR, virus isolation, and enzyme-linked immunosorbent assay for diagnosis of influenza A virus infection in different age groups. *J Clin Microbiol* 40: 2051-6
40. Ng EK, Cheng PK, Ng AY, Hoang TL, Lim WW. 2005. Influenza A H5N1 detection. *Emerg Infect Dis* 11: 1303-5
41. Allwinn R, Preiser W, Rabenau H, Buxbaum S, Sturmer M, Doerr HW. 2002. Laboratory diagnosis of influenza--virology or serology? *Med Microbiol Immunol* 191: 157-60
42. Gyarmati P, Conze T, Zohari S, LeBlanc N, Nilsson M, et al. 2008. Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses by use of padlock probes. *J Clin Microbiol* 46: 1747-51
43. Hong JW, Studer V, Hang G, Anderson WF, Quake SR. 2004. A nanoliter-scale nucleic acid processor with parallel architecture. *Nat Biotechnol* 22: 435-9
44. Jonstrup SP, Koch J, Kjems J. 2006. A microRNA detection system based on padlock probes and rolling circle amplification. *RNA* 12: 1747-52
45. Zieba A, Wahlby C, Hjelm F, Jordan L, Berg J, et al. 2010. Bright-field microscopy visualization of proteins and protein complexes by in situ proximity ligation with peroxidase detection. *Clin Chem* 56: 99-110
46. Johnson PJ. 2002. Spliceosomal introns in a deep-branching eukaryote: the splice of life. *Proc Natl Acad Sci U S A* 99: 3359-61
47. Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16: 990-5
48. Batsche E, Yaniv M, Muchardt C. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* 13: 22-9
49. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297: 1007-13
50. Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25: 106-10
51. Caceres JF, Kornblihtt AR. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet* 18: 186-93
52. Michel F, Ferat JL. 1995. Structure and activities of group II introns. *Annu Rev Biochem* 64: 435-61
53. Stamm S. 2008. Regulation of Alternative Splicing by Reversible Protein Phosphorylation. *J. Biol. Chem.* 283: 1223-27
54. Xie J, Black DL. 2001. A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. *Nature* 410: 936-39
55. Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457-63
56. Markus MA, Heinrich B, Raitskin O, Adams DJ, Mangs H, et al. 2006. WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Experimental Cell Research* 312: 3379-88
57. Gooding C, Smith CW. 2008. Tropomyosin exons as models for alternative splicing. *Adv Exp Med Biol* 644: 27-42

58. Tazi J, Bakkour N, Stamm S. 2009. Alternative splicing and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1792: 14-26
59. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, et al. 2004. Novel RNAs Identified From an In-Depth Analysis of the Transcriptome of Human Chromosomes 21 and 22. *Genome Research* 14: 331-42
60. Stein LD. 2004. Human genome: end of the beginning. *Nature* 431: 915-6
61. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, et al. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-6
62. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413-5
63. Lopez AJ. 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* 32: 279-305
64. Nagao K, Togawa N, Fujii K, Uchikawa H, Kohno Y, et al. 2005. Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the PTCH gene with exon junction microarrays. *Hum Mol Genet* 14: 3379-88
65. Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR. 2006. Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, SMN1 and SMN2. *The American Journal of Human Genetics* 78: 63-77
66. Byers PH, Duvic M, Atkinson M, Robinow M, Smith LT, et al. 1997. Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the COL1A1 and COL1A2 genes of type I collagen. *Am J Med Genet* 72: 94-105
67. Zavolan M, Kondo S, Schonbach C, Adachi J, Hume DA, et al. 2003. Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res* 13: 1290-300
68. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, et al. 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 7: 325
69. Johnson JM, Castle J, Garrett-Engle P, Kan Z, Loerch PM, et al. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141-4
70. Zhu J, Shendure J, Mitra RD, Church GM. 2003. Single molecule profiling of alternative pre-mRNA splicing. *Science* 301: 836-8
71. Wang J, Manley JL. 1995. Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. *RNA* 1: 335-46
72. Caceres JF, Stamm S, Helfman DM, Krainer AR. 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265: 1706-9
73. Rosenblatt KP, Sun Z-P, Heller S, Hudspeth AJ. 1997. Distribution of Ca<sup>2+</sup>-Activated K<sup>+</sup> Channel Isoforms along the Tonotopic Gradient of the Chicken's Cochlea. *Neuron* 19: 1061-75
74. Navaratnam DS, Bell TJ, Tu TD, Cohen EL, Oberholtzer JC. 1997. Differential Distribution of Ca<sup>2+</sup>-Activated K<sup>+</sup> Channel Splice Variants among Hair Cells along the Tonotopic Axis of the Chick Cochlea. *Neuron* 19: 1077-85
75. Dorn R, Reuter G, Loewendorf A. 2001. Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. *Proc Natl Acad Sci USA* 98: 9724-9
76. Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, et al. 1993. Heterogeneity of mucin gene expression in normal and neoplastic tissues. *Cancer Res* 53: 641-51

77. Gambus G, de Bolos C, Andreu D, Franci C, Egea G, Real FX. 1993. Detection of the MUC2 apomucin tandem repeat with a mouse monoclonal antibody. *Gastroenterology* 104: 93-102
78. Carrato C, Balague C, de Bolos C, Gonzalez E, Gambus G, et al. 1994. Differential apomucin expression in normal and neoplastic human gastrointestinal tissues. *Gastroenterology* 107: 160-72
79. Reis CA, David L, Seixas M, Burchell J, Sobrinho-Simoes M. 1998. Expression of fully and under-glycosylated forms of MUC1 mucin in gastric carcinoma. *Int J Cancer* 79: 402-10
80. Ho JJ, Siddiki B, Kim YS. 1995. Association of sialyl-Lewis(a) and sialyl-Lewis(x) with MUC-1 apomucin in pancreatic cancer cell line. *Cancer Res* 55: 3659-63
81. Audie JP, Janin A, Porchet N, Copin MC, Gosselin B, Aubert JP. 1993. Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by in situ hybridization. *J Histochem Cytochem* 41: 1479-85
82. Correa P. 1992. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 52: 6735-40
83. Correa P. 1988. A human model of gastric carcinogenesis. *Cancer Res* 48: 3554-60
84. Reddish MA, Jackson L, Koganty RR, Qiu DX, Hong W, Longenecker BM. 1997. Specificities of anti-sialyl-Tn and anti-Tn monoclonal antibodies generated using novel clustered synthetic glycopeptide epitopes. *Glycoconjugate Journal* 14: 549-60
85. Gero EJ, Colcher D, Ferroni P, Melsheimer R, Giani S, et al. 1989. CA 72-4 radioimmunoassay for the detection of the TAG-72 carcinoma-associated antigen in serum of patients. *J Clin Lab Anal* 3: 360-9
86. Kobayashi H, Terao T, Kawashima Y. 1992. Serum sialyl Tn as an independent predictor of poor prognosis in patients with epithelial ovarian cancer. *Journal of Clinical Oncology* 10: 95-101
87. Werther JL, Tatematsu M, Klein R, Kurihara M, Kumagai K, et al. 1996. Sialosyl-Tn antigen as a marker of gastric cancer progression: An international study. *International Journal of Cancer* 69: 193-99
88. MacLean GD, Reddish MA, Koganty RR, Longenecker BM. 1996. Antibodies against mucin-associated sialyl-Tn epitopes correlate with survival of metastatic adenocarcinoma patients undergoing active specific immunotherapy with synthetic STn vaccine. *J Immunother Emphasis Tumor Immunol* 19: 59-68
89. Butters TD, Neville DC. 2008. Glycoprotein Analysis. pp. 495-513
90. Silverman HS, Parry S, Sutton-Smith M, Burdick MD, McDermott K, et al. 2001. In vivo glycosylation of mucin tandem repeats. *Glycobiology* 11: 459-71
91. Dwek RA, Edge CJ, Harvey DJ, Wormald MR, Parekh RB. 1993. Analysis of Glycoprotein-Associated Oligosaccharides. *Annual Review of Biochemistry* 62: 65-100
92. Ferreira B, Marcos NT, David L, Nakayama J, Reis CA. 2006. Terminal {alpha}1,4-linked N-acetylglucosamine in Helicobacter pylori-associated Intestinal Metaplasia of the Human Stomach and Gastric Carcinoma Cell Lines. *J. Histochem. Cytochem.* 54: 585-91
93. Bäckström M, Thomsson KA, Karlsson H, Hansson GC. 2009. Sensitive Liquid Chromatography-Electrospray Mass Spectrometry Allows for the Analysis of the O-Glycosylation of Immunoprecipitated Proteins from Cells or Tissues: Application to MUC1 Glycosylation in Cancer. *Journal of Proteome Research* 8: 538-45

# Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 532*

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](http://publications.uu.se)  
urn:nbn:se:uu:diva-120016



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2010