DNA-Mediated Detection and Profiling of Protein Complexes

MARIA HAMMOND
Proteins are the effector molecules of life. They are encoded in DNA that is inherited from generation to generation, but most cellular functions are executed by proteins. Proteins rarely act on their own – most actions are carried out through an interplay of tens of proteins and other biomolecules.

Here I describe how synthetic DNA can be used to study proteins and protein complexes. Variants of proximity ligation assays (PLA) are used to generate DNA reporter molecules upon proximal binding by pairs of DNA oligonucleotide-modified affinity reagents. In Paper I, a robust protocol was set up for PLA on paramagnetic microparticles, and we demonstrated that this solid phase PLA had superior performance for detecting nine candidate cancer biomarkers compared to other immunoassays. Based on the protocol described in Paper I I then developed further variants of PLA that allows detection of protein aggregates and protein interactions. I sensitively detected aggregated amyloid protofibrils of prion proteins in paper II, and in paper III I studied binary interactions between several proteins of the NFκB family. For all immunoassays the selection of high quality affinity binders represents a major challenge. I have therefore established a protocol where a large set of protein binders can be simultaneously validated to identify optimal pairs for dual recognition immunoassays (Paper IV).

Keywords: Proximity ligation assay, Protein complexes, Protein interactions, Biomarkers, Prions, Antibodies

Maria Hammond, Uppsala University, Department of Immunology, Genetics and Pathology, Molecular tools, Rudbecklaboratoriet, SE-751 85 Uppsala, Sweden.

© Maria Hammond 2013

ISSN 1651-6206
ISBN 978-91-554-8718-8
urn:nbn:se:uu:diva-204861 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-204861)
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Related Work by the Author


# Contents

Introduction ..................................................................................................... 7  
Detecting and profiling biomolecules – terminology ...................................... 7  
DNA – a well-defined carrier of information ............................................... 8  
Tools for studying DNA ................................................................................. 8  
Proteins – life’s effector molecules .............................................................. 9  
Tools for studying proteins ............................................................................ 10  
Immunoads ..................................................................................................... 11  
The proximity ligation assay ......................................................................... 13  
Functionalizing affinity binders with reporter molecules .......................... 14  
Large scale initiatives to produce binders to all human proteins .......... 15  
Finding the monospecific antibody – an impossible task? ..................... 16  
Developing highly specific assays with promiscuous binders .............. 17  
Complexes of proteins generate biological function in health and disease ................................................................................................. 18  
Detecting complexes of biomolecules ......................................................... 19  

Investigations ................................................................................................. 21  
Paper I – Sensitive plasma protein analysis by microparticle-based proximity ligation assays ................................................................. 21  
Paper II – Sensitive detection of aggregated prion proteins by proximity ligation assay .................................................................................. 23  
Paper III – Profiling cellular protein complexes by proximity ligation with dual tag microarray readout ..................................................... 25  
Paper IV – A DNA-mediated search for optimal combinations of protein binders ................................................................. 28  

Conclusion .................................................................................................... 33  

Acknowledgements ....................................................................................... 34  
References ..................................................................................................... 37
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSE</td>
<td>Bovine spongiform encephalopathy</td>
</tr>
<tr>
<td>BRET</td>
<td>Bioluminescence resonance energy transfer</td>
</tr>
<tr>
<td>CJD</td>
<td>Creutzfeldt-Jakob disease</td>
</tr>
<tr>
<td>CoIP</td>
<td>Co-immunoprecipitation</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTM</td>
<td>Dual tag microarray</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>FCS</td>
<td>Fluorescence correlation spectroscopy</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>HGP</td>
<td>Human Genome Project</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IPCR</td>
<td>Immuno polymerase chain reaction</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of detection</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PLA</td>
<td>Proximity ligation assay</td>
</tr>
<tr>
<td>PMCA</td>
<td>Protein misfolding cyclic amplification</td>
</tr>
<tr>
<td>PrEST</td>
<td>Protein epitope signature tag</td>
</tr>
<tr>
<td>PrPC</td>
<td>Prion protein cellular</td>
</tr>
<tr>
<td>PrPSc</td>
<td>Prion protein scrapie</td>
</tr>
<tr>
<td>PSA</td>
<td>Prostate specific antigen</td>
</tr>
<tr>
<td>PTM</td>
<td>Post translational modification</td>
</tr>
<tr>
<td>QuIC</td>
<td>Quaking-induced conversion</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>RCA</td>
<td>Rolling circle amplification</td>
</tr>
<tr>
<td>RIA</td>
<td>Radioimmunoassay</td>
</tr>
<tr>
<td>SELEX</td>
<td>Systematic evolution of ligand by exponential enrichment</td>
</tr>
<tr>
<td>SOMAmers</td>
<td>Slow off-rate modified aptamers</td>
</tr>
<tr>
<td>SP-PLA</td>
<td>Solid phase proximity ligation assayw</td>
</tr>
<tr>
<td>SulfoSMCC</td>
<td>sulfosuccinimidyl-4-([N-maleimidomethyl] cyclohexane-1-carboxylate</td>
</tr>
<tr>
<td>UDG</td>
<td>Uracil-DNA glycosylase</td>
</tr>
</tbody>
</table>
Introduction

Detecting and profiling biomolecules – terminology

There is an ever increasing number of methods available for protein analysis that all come with their own cons and pros. What methods are most suitable to use depends on the specific question one wishes to study, and what methods and samples that are available. So before digging deeper into some of all methods that are available I would like to define some of the terms that are used to describe the performance of these methods.

**Specificity** is the capability of an assay to only report something that truly is there. The term can be narrowed down into *clinical specificity*, meaning the ability to report all negative cases as such, *antibody specificity*, meaning the ability of an antibody to only bind to its intended target and nothing else, etc.

**Sensitivity** is the capability of an assay to report every case that is positive. Here *clinical sensitivity* is the ability to identify all positive cases, while *molecular sensitivity* refers to the ability to detect minute amounts of an analyte. Sensitivity can also be defined as the ability to detect small differences, or the ability to generate a lot of signal for each present target molecule. Because of its many definitions the usefulness of the term sensitivity has been debated¹.

**Limit of detection (LOD)** refers to the minimum amount of target that can be detected over background signal – lower LOD, or the maximal amount of target that can be detected before the assay saturates – upper LOD. In my thesis the term LOD will only refer to the lower LOD. This can be calculated as the concentration of an analyte that generates a signal that is 2 or 3 standard deviations above background signal.

**Dynamic range** is the span of concentrations that can be accurately measured by the assay, i.e. the range between the upper and the lower LODs. The *linear dynamic range* is the range over which various amounts of targets generate proportionate amounts of signal, important when amounts of target in a sample is determined by reference to a standard curve.

**Coefficient of variation (CV)** is the level of variation between replicate measurements related to the absolute amount that is measured. It is calculated as the standard deviation divided by the mean.
DNA – a well-defined carrier of information

This year it is 60 years since Watson and Crick published the structure of deoxyribonucleic acid (DNA). They could in detail describe the structure of the two chains of nucleotides running in opposite direction and coiling around the same axis, always with the same distance and angle between each base. There are no rules for how nucleotides can be ordered on a single chain, only that adenine pairs with thymine, and guanine pairs with cytosine on the opposite strand, so if the sequence of one strand is known the sequence of the other strand is given. Watson and Crick immediately saw how these features could build up a system where the genetic information is encoded by the exact sequence of the bases, and how that information could easily be copied as the information always occurred as a “pair of templates”. There are very accurate models for the stability of two DNA strands hybridizing to each other that can be used to accurately predict at what temperature the strands will dissociate from one another.

As Watson and Crick concluded 60 years ago, the structure of DNA makes it an excellent carrier of information. With four different bases that can be ordered without any restrictions, a chain of only 20 bases can generate more than $10^{12}$ possible combinations. Because of the capacity of coding information it has been proposed to use DNA for storage of other information than genetic information. One example is that synthetically designed DNA sequences are used as molecular barcodes in methods studying DNA, RNA and proteins. Other examples that have been proposed is to use DNA to send hidden encoded messages in microdots, or even to use it as the next-generation digital information storage, exemplified by storing journal articles, Shakespeare’s sonnets, photographs and entire books as DNA.

Tools for studying DNA

A major breakthrough when it comes to methods for studying DNA was when Kary Mullis developed the polymerase chain reaction (PCR). By using two primers hybridizing to two separate yet relatively closely located parts of the target DNA molecule to initiate polymerization in a cyclical fashion, he developed a method that could exponentially amplify the target sequence. The human (haploid) genome is built up by approximately $3 \times 10^9$ nucleotides, and therefore a primer with a length of 20 nucleotides will most likely be specific for its target. With the dual recognition by two primers the likelihood of targeting and amplifying an unintended sequence is reduced even further.

Even before the invention of PCR, Fred Sanger and co-workers had developed a strategy to detect the exact sequence of the nucleotides. Over the decades several new methods have been developed for sensitive and
specific detection and characterization of DNA. When the Human Genome Project (HGP) started to sequence the entire human genome\textsuperscript{11,12}, technologies for DNA sequencing were developed rapidly, capacities increased and cost for every base sequenced decreased with a factor of two per two years. But with the introduction of next-generation sequencing (NGS) technologies\textsuperscript{13,14} the sequencing capacities are since a couple of years back increasing way more rapidly, in fact faster than the canonical Moore’s law that reflects the evolution of computers.

However, what is still common for most methods for DNA characterization and detection is that they are based on affinity reactions via hybridization of probes that can be easily designed \textit{in silico} using the base pairing rules, perhaps assisted by DNA polymerization in order to generate new identical copies of the original template, sometimes also in combination with other enzymatic reactions such as ligation and restriction digestion.

Proteins – life’s effector molecules

DNA encodes most of the information required for building up a living cell, but it is proteins that make up most of the dry mass of the cells, and it is the proteins that carry out most of the biological functions in the cells. They build up the cells, and through their interactions they achieve the cell’s various functions, including cell cycle control, signal transduction and gene transcription and translation. Genetic information provides invaluable insights in normal as well as pathologic biological functions, and much ongoing effort is devoted to identifying genetic markers that determine the likelihood for a person to contract a particular disease. For monogenic diseases being a carrier of a certain genotype typically leads to a pathogenic phenotype with a very high probability. But many diseases are more complex than that, and may have several genetic factors associated with risks of developing disease. These genetic factors are often inherited and not acquired by somatic mutations, so they are present from birth, but it is nonetheless usually impossible to predict whether an individual will come down with the disease and if so when that will happen. What happens when the cells start to malfunction?

When the sequence of the human genome was published as a result of the Human Genome Project, it was concluded that the human genome contains somewhere between 20000 and 25000 protein coding genes\textsuperscript{15}, and today the number is estimated to be somewhere at the lower end of this range\textsuperscript{16}. However, genes often have several different transcription start sites, they give rise to several different splice variants, and translation products can subsequently be subjected to both processing steps and to a myriad of phosphorylations, glycosylations, acetylations and other post-translational modifications (PTMs). Including all these variants the number of different
human protein species may well exceed a million\textsuperscript{17}. If we include individual variation due to single amino acid polymorphisms in the human population, the number might be even larger.

It is hard to predict the number of these proteins that could be of interest to detect and characterize. Proteins that are present in blood, either as a result of active secretion or leakage from damaged cells and tissues, are highly interesting to study in order to find biomarkers for prediction, diagnosis, and treatment follow-up in diseases. In 2004 Anderson and colleagues presented a list of 1174 different gene products that are expected to be present as proteins in human serum, an estimate based on literature searches and mass spectrometry (MS) analyses\textsuperscript{18}. The human secretome has been estimated to consist of 3400 proteins\textsuperscript{19}, based on the 3400 cDNAs that Lin et al.\textsuperscript{20} identified based on 2200 genes found in algorithmic searches to have elements that made them highly suitable for secretion or that had previously been found secreted in serum.

Genomic studies are more common than proteomic studies according to the Pubmed entries (\textit{Figure 1}). Studies in both fields are steadily increasing but genomic analyses are increasing at a higher rate than those in proteomics.

\textit{Figure 1.} Number of publications in the fields of proteomics and genomics 1997-2012. Pubmed was searched for the terms proteomics and genomics (all fields) respectively. Review articles are included in respective searches but also plotted separately.

Tools for studying proteins

Methods for protein detection can be divided in two general groups – direct and indirect methods: In direct methods the protein(s) of interest are directly identified by determining the amino acid sequence amino acid by amino acid. In indirect methods various molecular affinity probes are used to bind to the target of interest, and then a (hopefully) specific interaction between probe and target is detected rather than the target itself.
Another way to categorize methods for detecting proteins is by purpose – into discovery methods or assays\textsuperscript{21}. Direct methods are better suited for discovery proteomics as they do not necessarily require any directed targeting. When it comes to more narrowed assays, for instance for validating candidate biomarkers, it is wise to restrict the number of proteins being detected – here affinity probes are very useful to select the target of interest. The readout can then be performed by indirect methods detecting the bound probe, or by direct methods detecting the target that was captured by the probes.

**Direct methods to detect proteins**

The protein sequenator, an instrument that automated the Edman degradation chemistry used to read the amino acid sequence of a protein or peptide, could detect the primary structure of up to 60 amino acids of a protein, with a requirement 0.25 µmoles of protein to start with\textsuperscript{22}. Further developments of the technique increased the sensitivity so that only 5 pmoles of starting material was required\textsuperscript{23}. These methods were not suited for detection of proteins in complex mixtures, so when two dimensional gel electrophoresis was developed that could resolve thousands of proteins this was a great breakthrough\textsuperscript{24}. Mass spectrometry (MS) is a technique to study the mass-to-charge ratios of ionized biomolecules. With recent advances in the techniques by introducing multiple reaction monitoring (MRM)\textsuperscript{25} and immunoaffinity enrichment strategies the limit of quantification for MS assays can be as little as picograms per milliliter (low picomolar) of the target molecule in 1 ml of sample\textsuperscript{26}. MS should have the capacity to identify more than 200 different PTMs as they affect the mass of the protein in a predictable way, but protocols have only been developed for very few of them\textsuperscript{21}.

**Immunoassays**

In the 1930s researchers gained greatly improved insights in the specificity of antibodies – how immunization could generate antibodies to basically any possible antigen, and most of all how these affinity reagents could be used to differentiate proteins\textsuperscript{27}. The same decade experiments were reported where antibodies were successfully conjugated to labels for visualization purposes, without significant loss of affinity\textsuperscript{28–30}. Such chemically modified antibodies proved useful for detecting their target molecules in tissues, as first reported by Coons and colleagues in 1942\textsuperscript{31}. Today, labels such as radioisotopes, enzymes, fluorophores and oligonucleotides are commonly added to affinity binders to detect their binding to a targeted protein. There are also methods that detects the binding without the need for certain reporter molecules, such as surface based sensors, where for instance optical (surface plasmon
resonance\textsuperscript{32} is one example of this) or electrochemical properties can be recorded to detect an interaction.

Immunoassays remain the most commonly used methods for detection of proteins\textsuperscript{33}. Yalow and Berson are often considered as the inventors of the first immunoassay for proteins in solutions – the radioimmunoassay (RIA)\textsuperscript{34}. The antigen is labeled with a gamma-emitting isotope such as \textsuperscript{125}I, or beta-emitters such as \textsuperscript{3}H. The assay is then performed as a competitive assay, where the labeled antigen is mixed in excess with the antibody along with the sample to compete with the labeled antigen for the binding sites on the antibodies. The molecules not bound by the antibodies are washed away before the radioactive signals are detect. RIAs of today have reported limits of detection similar to MRM-MS\textsuperscript{35}. Another important discovery in the 1960’s was how to immobilize antibodies on a sepharose matrix. This enabled development of automated solid phase assays\textsuperscript{36}, and the sandwich immunoassay\textsuperscript{37}.

In the next decade the enzyme-linked immunosorbent assay (ELISA) was developed by Engvall and Perlmann\textsuperscript{38}. The detection is based upon enzymatic activity from an enzyme, such as horseradish peroxidase, alkaline phosphatase or β-galactosidase, conjugated to the affinity binder. The enzyme catalyzes the conversion of a colorless substrate into a colored product that can be detected visually or more sensitively and specifically by measuring the absorbance. ELISA is today the standard assay to detect proteins with high sensitivity, both in clinical and in research laboratories. The preferred way of detecting the proteins is a sandwich assay where one antibody is used as capture and another antibody, targeting a different epitope, is used for detection.

In the beginning of the 1990s Sano and colleagues developed the immuno-polymerase chain reaction (IPCR) as a more sensitive immunoassay than ELISA\textsuperscript{39}. Instead of labeling the affinity binders with an enzyme as in ELISAs they labeled the antibodies with DNA molecules that could be amplified by PCR. Provided that nonspecific binding can be minimized IPCR is a very sensitive method, as theoretically a single DNA template is sufficient to start the exponential amplification and generate detectable levels of reporter molecules. IPCR has been reported to have very broad dynamic range and to be 100-100 000-folds more sensitive compared to ELISA\textsuperscript{40}. The amplified DNA can be detected either by gel electrophoresis after the PCR is completed, or by real-time quantitative PCR (qPCR). Since the sequence of the DNA reporter can be freely chosen, it is possible to design it to have very little similarity to sequences usually present in the lab, like genomic DNA. This makes the assay less sensitive to contamination and interference with the detection wells. In ELISA, light absorbance by the plastic material of the assay wells can generate background to the assay no matter whether the labeled antibodies are present or not. In IPCR the oligonucleotide labeled assay probes must be present in order to generate a
signal – background or not. The use of synthetic DNA sequences as reporter molecules also opens up much larger possibilities for creating multiplex assays as compared to ELISA where only a handful of enzymes and substrate are available.

The proximity ligation assay

The proximity ligation assay (PLA) requires two or more affinity binders equipped with synthetic DNA sequences to bind in close proximity in order to allow for enzymatic ligation of the DNA strands in order to generate one reporter molecule, including sequence information from two or more probes (Figure 2).

Figure 2. Schematic illustration of the proximity ligation assay. a) PLA probes are allowed to bind their target. b) When two PLA probes have bound in close proximity a connector oligonucleotide can guide enzymatic ligation. c) The ligated reporter molecule can be amplified and detected by qPCR.

PLA has been used in various liquid samples such as plasma, serum, cerebrospinal fluid and cell lysates to detect endogenous cytokines and other potential biomarkers, both single proteins, protein aggregates (e.g. Aβ protofibrils), and larger membrane coated vesicles (e.g. prostasomes), infectious agents, protein-protein interactions and protein-DNA interactions. With the in situ PLA, endogenous levels of proteins, protein-protein interactions, phosphorylations and glycosylations have been successfully detected in fixed cells, tissues and on membranes.

PLA was first developed as a sensitive assay for detection of proteins without the need for a solid support that may cause background due to unspecific adsorption of reporter molecules, but still requiring dual recognition to ensure specificity. In the first PLA, aptamers with attached reporter DNA sequences were used. Upon proximal binding of two aptamers to the same target molecule, the two reporter DNA sequences could be ligated into one reporter molecule that could be amplified and detected by
The assay has been further developed to use antibodies as the affinity binders in the PLA probes. In the solid phase PLA (SP-PLA), a solid support has been introduced to further increase the specificity of the assay and to allow for more complex samples to be analyzed. Apart from detecting proteins in solutions such as plasma, serum, cell lysates and cerebrospinal fluid, an in situ variant of PLA has been developed for detection of proteins and protein complexes in fixed cells and tissue sections, where the two PLA probes guide the ligation of connector oligonucleotides into a circle. This circle can then work as a template in a rolling circle amplification (RCA) reaction that generates one big concatemer of DNA, still attached to the target protein via the affinity binder.

PLA is very suitable for studying a large number of proteins and protein-protein interactions in a highly multiplex fashion as DNA is used as a reporter molecule and thus all methods that are available for detecting DNA can be used as readout. The DNA reporter molecules generated by PLA in solution was originally detected by qPCR, also when scaled up into multiplex assays. The qPCR readout is a limitation to the assay since even if the protein recognition can be made in multiplex, the readout is limited to singleplex (or at least very few simultaneous reactions) with qPCR. To overcome the problem with a singleplex readout for a multiplex protein detection assay, Darmanis et al. have developed ProteinSeq, a multiplex (36-plex) SP-PLA with next generation sequencing as readout.

Functionalizing affinity binders with reporter molecules

An important consideration for constructing immunoassays is how to accurately attach a reporter molecule to an affinity reagent. Early immunoassays employed covalent coupling of enzyme reporters to antibodies by means of homobifunctional crosslinkers such as glutardialdehyde. Direct chemical coupling of reporter molecules to affinity binders has been simplified by the introduction of heterobifunctional crosslinkers. SulfoSMCC (sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate) is one such linker that is commercially available and has been used to couple thiol-containing reporter molecules to primary amines of protein binding reagents. The reaction can also be reversed to couple amine-containing reporter molecules to site-specific cysteins on the binder. Direct chemical coupling has the advantage that the reactions are irreversible so there is no chance that the reporter molecule will detach from its binder during the reaction. However, optimizing protocols for conjugation and purification can be tedious. Excessive activation at random positions of the binders may destroy the binding sites. On the other hand, if many unlabeled binders are present during the assay these will compete for binding with the labeled binders and thus reduce signal and thereby also the sensitivity of the assay.
Secondary reagents, targeting the constant parts of antibodies, have been developed based on antibodies from a different species than that of the primary antibody targeting the protein of interest\textsuperscript{60}, or bacterial proteins (such as protein A and protein G) binding to Fc-parts of antibodies. The very strong interaction between biotin and streptavidin is also commonly used as a mean to indirectly label biotinylated protein binders with reporter molecules via this linker protein. These indirect methods have the drawback that while the coupling may be very strong it is still reversible, so there is a chance that the reporter molecule can fall off and attach to a different primary binder, which is unacceptable in multiplex assays. It is also important that the reporter molecule can only attach to one type of primary binder, which may limit sandwich immunoassays or other dual recognition assays to using different types of binders for the different binding events.

For recombinant binders several methods have been developed to allow site-specific covalent labeling. One approach is to use “self-labeling proteins” such as O\textsuperscript{6}-alkylguanine-DNA-alkyl-transferase (often referred to as “SNAP-tag”)\textsuperscript{61}, or Haloalkane-dehalogenase (“Halo-tag”)\textsuperscript{62} fused to the binder of interest and then letting the fusion protein react with its substrate benzylguanine (BG) and chlorohexane (CH) respectively, attached to a reporter molecule. By using these fusion tags highly specific and efficient coupling reactions can be achieved. A drawback of this approach is that these tags are large, often the same size as the recombinant binder itself, and the properties of the tags can therefore affect both the functional properties of the binders and the efficiencies by which the protein can be expressed in \textit{E. coli}\textsuperscript{63}.

Lately efforts have been made to modify the genetic code so that unnatural amino acids with desirable functions in their side chains can be incorporated in recombinant proteins. tRNA/synthetase pairs have been evolved so that codons that are normally recognized as stop codons instead code for an unnatural amino acid\textsuperscript{64}. This approach has been used to site-specifically incorporate alkyne or azide derivatives into recombinant proteins to allow for easy conjugation of reporter molecules via click chemistry\textsuperscript{65}. Click chemistry is a concept coined by Nobel laureate Barry Sharpless and includes chemical reactions that are characterized by “simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation”,\textsuperscript{66} exactly the things that are of importance for successful conjugation of reporter molecules to affinity binders.

Large scale initiatives to produce binders to all human proteins
Several initiatives are underway to generate affinity binders towards all human proteins. The Human Protein Atlas, led by Mathias Uhlén\textsuperscript{67} has in the
latest version of the atlas (11.0) published data on 18,000 polyclonal antibodies raised against protein epitope signature tags (PrESTs), protein fragments between 50 and 150 amino acids in size, to protein products of 15,000 human genes. Their goal is to generate a collection of at least two polyclonal antibodies per human protein, and also constructing an atlas where all antibodies have been used to investigate the expression of the proteins in various tissues, malignant and healthy, along with immunofluorescence images of cell lines to visualize the cellular localization of all proteins.

Several other initiatives to generate and validate affinity binders have been started including the Antibody Characterization Program by the United States National Cancer Institute (NCI; http://antibodies.cancer.gov/), the European consortia Affinomics (http://www.affinomics.org/) and AffinityProteome (http://www.proteomebinders.org/), and the German Antibody Factory. The different projects produce different types of binders – antibodies, various recombinant antibody fragments, other recombinant protein scaffolds, DNA aptamers, and other small molecules. The company SomaLogic has set up the SOMAscan assay with over 1000 SOMAmers (Slow Off-rate Modified Aptamers), aptamers with T residues replaced by dUTP with side chains mimicking the side chains of amino acids at the 5-position. The aptamers are selected by SELEX (Systematic Evolution of Ligand by Exponential enrichment) to have very low off-rates\textsuperscript{68–70}.

The different inherent features of the binders as well as the methods by which they are selected and matured can affect what targets and applications they are suitable for. It will therefore be very interesting to follow comparisons between different classes of binders when the initiatives are mature enough to allow such comparisons.

Finding the monospecific antibody – an impossible task?

Several researchers have expressed that the lack of specific antibodies, or pairs of specific antibodies, is the greatest problem for developing new immunoassays\textsuperscript{19,71}. It is tempting to dream about the monospecific affinity binder that recognizes its intended target and only that target. But it seems likely that there is no such thing as a monospecific affinity binder, but that every single antibody will be likely to show some affinity for a large number of different antigens\textsuperscript{72–75}.

The term “monospecific antibody” has recently been used to describe an immune serum, i.e. a polyclonal antibody mixture, that has been subjected to affinity purification against the antigen or parts of it as synthetic peptides or recombinant protein fragments\textsuperscript{76,77}. This terminology appears improper since only antibodies with very low or no affinity for the intended target have been removed from the polyspecific polyclonal serum. Any antibodies with specificity for both the targeted antigen and something else are still there.
Proteins may interact with each other but there are no rules as is the case for DNA that allows one to predict complementarity between two proteins. Proteins that are similar to each other of course have a higher chance to have cross-reactive antibodies, but some antibodies can distinguish differences of as little as a single amino acid or even a difference of two or more atoms, making it very difficult to predict what cross-reactivity can be expected and selected against for an antibody.

Kramer et al. extensively analyzed the binding promiscuity of an HIV-1 specific anti-p24 monoclonal antibody against a peptide library. They revealed five non-homologous peptides that bound to the same antibody region and were able to compete with recombinant p24 for binding. Based on these five peptides they derived more than 1000 single substitution analogs that were still able to bind the antibody. Analyzing these analogs they could identify so-called supertopologies, i.e. the actual structural elements that are required for binding. By bioinformatic analyses the authors found almost 6000 proteins from various species containing any of these supertopologies. Analyzing a few of these proteins for binding to the anti-p24 monoclonal showed that the antibody could actually detect these proteins in denatured and/or native form. So there are means to predict tendencies to cross-reactivity, but those methods are far too tedious to be deployed for very large sets of binders.

Developing highly specific assays with promiscuous binders

The method of choice for developing a diagnostic assay to detect single proteins in blood with high sensitivity has for a long time been the sandwich-ELISA. By using one antibody for capture and another antibody targeting a different epitope for detection the risk for cross-reactivity is dramatically reduced as it is unlikely that two different antibodies by chance would cross-react with the same target. This strategy is not suitable for multiplex assays, however. When the number of antibody pairs is increased the likelihood of antibodies cross-reacting with each other or each other’s targets also increase.

The research community lacks sufficient information about validation of commercially available affinity binders. Antibodies are only specific under a defined range of conditions and must therefore be evaluated in an application-specific manner. This type of validation is of course costly and labor-intensive. It is therefore of great value if all validation information about affinity binders becomes publicly available. Kinetic parameters (association/dissociation constants), epitope specificity, selectivity and cross-reactivity are among the things that are valuable to know about the binders. Screening of antibodies on microarrays, both planar and bead based has been proposed as a means to cost efficiently validate antibodies for cross-reactivity against a large number of proteins. This would enable
rapid screening for cross-reactivity against hundreds or thousands of proteins or peptides, but it would still provide a biased screening against the proteins that are chosen and suitable to be immobilized on such arrays.

Characterizing affinity binders in this manner requires considerable efforts, resulting in a very high cost for each binder. Therefore methods have been developed to try to circumvent the need for exclusively monospecific binders, such as sandwich immunoassays, combining immunoprecipitation with size separation to improve the specificity of the results, or using the SP-PLA where not only two but three antibodies need to bind different epitopes on the same target protein in order to generate a signal. In SP-PLA the risk for cross-reactivity is also decreased due to the fact that the antibody pairs that generate the reporter molecules have to be exact combinations – any two PLA probes could not generate an amplifiable reporter DNA strand, but only pairs intended to target the same molecule.

Larry Gold and colleagues report that their SOMAmers have very low cross-reactivity by themselves due to the chemical nature (a negatively charged phosphodiester backbone of DNA). Using binders that have minimal chances of cross-reactivity is of course a key to developing very specific assays, but it is not the only reason for the success of the SOMAscan assay. A major part is probably the fact that the assay has been developed in parallel with the methods for selecting the binders. By doing so they could select binders with slow dissociation rates and use them in an assay that requires a slow dissociation rate for detecting a target. As explained by the authors, the SOMAscan does not require a dual recognition but instead a dual dependence on one binding event: the recognition between the target and the SOMAmer has to first withstand a capture of the target via the binder immobilized to a solid support and thereafter detection of the binder captured to the target immobilized to a solid support.

Larry Gold and colleagues report that their SOMAmers have very low cross-reactivity by themselves due to the chemical nature (a negatively charged phosphodiester backbone of DNA). Using binders that have minimal chances of cross-reactivity is of course a key to developing very specific assays, but it is not the only reason for the success of the SOMAscan assay. A major part is probably the fact that the assay has been developed in parallel with the methods for selecting the binders. By doing so they could select binders with slow dissociation rates and use them in an assay that requires a slow dissociation rate for detecting a target. As explained by the authors, the SOMAscan does not require a dual recognition but instead a dual dependence on one binding event: the recognition between the target and the SOMAmer has to first withstand a capture of the target via the binder immobilized to a solid support and thereafter detection of the binder captured to the target immobilized to a solid support.

Complexes of proteins generate biological function in health and disease

In the search for new biomarkers a big hope has been set towards proteins that may have a role in the pathological process of the disease. Major processes in the cells are typically executed by sets of greater than 10 proteins working together. Studies of how proteins interact and not only their mere presence are keys to gaining functional insights, and complexes of proteins may prove important biomarkers. When Leigh Anderson reviewed all protein-based assays approved by the US Food and Drug Administration (FDA) he found that only 3 of 109 unique protein targets in plasma or serum were protein complexes.
A much debated biomarker is the prostate specific antigen (PSA). The reason for debate is the fact that neither the sensitivity nor the specificity is 100%, i.e. some men tested and found to have normal levels of marker turn out to have prostate cancer, and some men tested positive turn out not to have the disease. Several studies have shown that by discriminating between free PSA and PSA bound by serum proteins the specificity of the tests can be increased, i.e. the number of false positive patients, that have to undergo unnecessary biopsies and worry about having a severe disease, can be decreased\textsuperscript{90,91}.

Detecting complexes of biomolecules

Several methods have been developed to study protein-protein interactions \textit{in situ}. A large proportion of the methods require genetic modification of the proteins of interest to generate a signal. These methods include Förster resonance energy transfer (FRET)\textsuperscript{92,93}, bioluminescence resonance energy transfer (BRET)\textsuperscript{94} and various protein complementation assays\textsuperscript{95} such as bimolecular fluorescence complementation assays\textsuperscript{96}, all have their own advantages and disadvantages. One advantage is that upon interaction the proteins generate the signal themselves so there is no need for disrupting or fixing the cells prior to detection. What is common for all these methods though is that they require genetic labeling of the proteins of interest, which means that the interactions have to be studied in model systems. It is thus impossible to use these methods on clinical samples from human patients.

When it comes to studying protein-protein interactions in unmodified biological samples the options are scarce. Methods such as immunofluorescence with two dyes co-localizing and FRET\textsuperscript{97} can be used to study protein-protein interactions \textit{in situ}. In the \textit{in situ} PLA\textsuperscript{50} the signal from each protein or protein complex can be amplified by RCA, thus improving sensitivity so that single proteins or interactions in protein complexes can be visualized.

One commercially available test to study protein-protein interactions in patient tissue samples is the VeraTag test, that is used to investigate homo- and heterodimers of epidermal growth factor receptors and other members of this receptor protein family. The technique utilizes a pair of antibodies each linked to a fluorescent reporter and a photosensitizer molecule, respectively\textsuperscript{98}. When the photosensitizer molecules are illuminated with 670 nm light, they cleave any reporters in close proximity via generated free radical oxygen. The liberated reporters are then recorded using conventional capillary electrophoresis.

However none of these described methods are really suitable for multiplex detection of protein-protein interactions as the number of interactions that can be studied simultaneously in the same sample is restricted by the capacity of fluorescence channels. For multiplex studies of
protein-protein interactions among endogenous proteins the golden standard has so far been combining co-immunoprecipitation (Co-IP) with MS, to look for all interaction partners of one targeted protein. In theory, the number of interaction partners to the bait that can be detected this way is unlimited, but only one bait can be investigated at the time.
Investigations

Paper I – Sensitive plasma protein analysis by microparticle-based proximity ligation assays

Introduction
The concentrations of proteins present in blood plasma vary over a very broad range: albumin is present at a concentration around 35-50 mg/ml, while cytokines such as IL6 are present at 0-5 pg/ml. Both these proteins are measured and used as clinical markers today. The abundance of these two markers thus differs by a factor of 10^{10}. As Anderson and Anderson illustrated, finding one IL6 molecule among all these albumin molecules is a task similar to that of finding one individual person in the entire human population of the world^{99}. In the search for new biomarkers, it is not unlikely that proteins leaking out from injured tissue and ending up in concentrations even lower than that of IL6 will be of clinical interest. Assuming that a protein could be present in concentrations of one molecule per microliter plasma then that protein would be present in a 10^{14}-fold lower amount than albumin, but it could conceivably still be measured with good precision in realistic one ml blood samples, given sufficiently sensitive methods^{100}. This illustrates the vast need for specificity of protein detection in clinical contexts.

PLA has been described as a very sensitive method for protein detection^{41,54}, however with the homogenous assay format the assay is quite sensitive to potential inhibitors of DNA ligases or polymerases that may be present in samples to be analyzed. Introducing a solid support offers a means to remove potential inhibitors by washing, and it would also impose a requirement for triple recognition of the antigen, serving to further increase the specificity of the assay.

Aim of the study
The study aimed at establishing a robust protocol for solid phase PLA on magnetic microparticles, allowing the development of assays against various target proteins with high specificity and sensitivity. We also wanted to demonstrate how the method may be used for clinical applications.
Summary of findings

We established a protocol for SP-PLA on magnetic bead microparticles (*Figure 3*), and applied it for detection of nine different proteins. All assays showed LODs in the low femtomolar range, broad dynamic ranges (5-6 orders of magnitude), and intra-assay CV% between 7 and 25%. In comparisons to state-of-the-art sandwich ELISAs all assays had equal or better limits of detection. The ELISA tests required 100 µl sample input while only 5 µl sample is required for SP-PLA, so in terms of actual numbers (not concentrations) of molecules detected SP-PLA performed better than ELISA for all nine analytes.

*Figure 3*. Schematic illustration of SP-PLA. a) Capture antibodies are immobilized on paramagnetic beads. b) Sample is added to allow for capture of the target. c) After washing away unbound sample matrix PLA probes are added to the captured target. d) When two PLA probes have bound in close proximity a connector oligonucleotide can guide enzymatic ligation. e) The ligated reporter molecule can be amplified and detected by qPCR.

All assays were set up in 10% chicken plasma and 10% chicken serum, to validate the performance in a complex sample matrix and not just in buffer. For one analyte, vascular endothelial growth factor (VEGF), we tested the assay performance in 100% chicken plasma, 100% chicken serum, or even whole blood, and found that it was as good as in buffer. We detected GDF-15, proposed as a biomarker for recurrence of myocardial infection, in 40 patients divided in two groups and 20 controls. For one group we determined the assay characteristics and found intra-assay CV% of 15-16%, and inter-assay CV% around 35%. The correlations between the concentrations of GDF-15 measured by SP-PLA versus ELISA were between 0.89 and 0.95.

In comparison to already established PLA protocols we found that SP-PLA performed equally well for detecting VEGF, IL8 and IL6 in buffer as sandwich PLA in polycarbonate tubes described by Ericsson et al. The performance was also very similar to that reported for homogenous-phase PLA with dual or triple recognition, and IPCR. 


Perspectives on paper I

Since the establishment of the SP-PLA protocol this has been used for sensitive detection of single analytes, such as Aβ-protofibrils\textsuperscript{43}, and spores of \textit{Bacillus globigii}\textsuperscript{104}. The PLA on microparticles has been further developed to allow for multiplex protein detection, detecting 36 proteins in the same sample\textsuperscript{58}. Here the authors could show how the PLA improves the possibilities of multiplexing compared to standard sandwich immunoassays.

Paper II – Sensitive detection of aggregated prion proteins by proximity ligation assay

Introduction

Prion diseases are characterized by the accumulation of a protease resistant isoform (PrP\textsupersc{Sc}) of the normally occurring cellular prion protein (PrP\textsupersc{C}). The word prion comes from the fact that the diseases are transmitted via proteinaceous infectious particles\textsuperscript{105}, a discovery that gave Stanley Prusiner the Nobel Prize in Physiology and Medicine in 1997. There is a critical need for early diagnostics since the diseases caused by prions, such as Creutzfeldt-Jakob disease (CJD), scrapie, bovine spongiform encephalopathy (BSE), and chronic wasting disease, are transmissible even long before the onset of the clinical disease.

Classical methods to detect aggregated PrP\textsupersc{Sc} have lacked sufficient sensitivity to detect the low levels of infectious proteins that are present in blood from infected individuals. Recently, several methods have been developed that allow detection of PrP\textsupersc{Sc} in blood or cerebrospinal fluid (CSF) from patients with CJD. One strategy used for developing sensitive assays for PrP\textsupersc{Sc} detection makes use of the aggregation inducing properties of PrP\textsupersc{Sc} by allowing infected samples to seed the conversion of PrP\textsupersc{C} into PrP\textsupersc{Sc}. This is done in the protein misfolding cyclic amplification (PMCA) assay\textsuperscript{106} and the more recently developed quaking-induced conversion (QuIC) assay\textsuperscript{107}. QuIC has been used to analyze CSF samples from sporadic CJD patients\textsuperscript{108}.

Methods based on affinity reagents have also been developed for detection of PrP\textsupersc{Sc}. Edgeworth and colleagues\textsuperscript{109} combined a novel solid-state capture matrix with an immunoassay to develop a promising assay for distinguishing non-symptomatic variant CJD patients from healthy controls.

Kamali-Moghaddam et al.\textsuperscript{43} have used a monoclonal antibody selective for Aβ protofibrils in a SP-PLA to develop a sensitive assay for detection of Aβ aggregates. The published assay has a limit of detection of 0.1 pg Aβ protofibrils per ml, which makes it – with further improvements – a promising tool for less invasive diagnostics in Alzheimer’s disease. The rational for this assay is that only when Aβ is aggregated in protofibrils can
the same monoclonal antibody bind to three identical epitopes in proximity, as is required to generate a signal in SP-PLA.

Another approach where several copies of a monoclonal antibody must bind in proximity has been developed for detection of both Aβ protofibrils and PrPSc using variants of fluorescence correlation spectroscopy (FCS), where fluctuations in fluorescence intensities are studied. In this approach, fluorescence labeled monoclonal antibodies are added to a sample to bind PrP. Antibodies binding PrPSc will co-localize as there are several of the same epitopes available in the aggregated PrPSc, while the antibodies are distributed randomly in solution when no aggregated protein is present. A method based on this approach has proven sensitive enough to detect PrPSc in CSF from BSE infected cattle and in blood plasma from scrapie infected sheep.

**Aim of the study**

In paper II our aim was to develop a sensitive assay based on SP-PLA for detection of aggregated prion proteins without the need to amplify the infectious agent. Unlike the situation for the assays we developed in paper I, we did not have access to a source of aggregated protein that could be diluted in standard curves with given concentrations. Our aim was therefore to demonstrate the potential of PLA as a sensitive tool for PrPSc-detection by assaying dilutions of brain homogenate from hamsters infected with the prion strain 263K under conditions where no detection signals were recorded from negative control.

**Summary of findings**

We established SP-PLAs with two different monoclonal antibodies, 3F4 and 6H4 (Prionics) (Figure 4), and we showed that SP-PLA has a potential for detection of PrPSc. Our assays showed high sensitivity and specificity as the brain homogenates could be diluted as much as 10 million fold and still be detected over background. The performance of the assay is similar independently of whether the brain homogenate was diluted in buffer or in 10% hamster plasma (data not shown).

**Perspectives on paper II**

So far we have only worked with brain homogenates from a hamster model for the PrPSc detection. It will next be important to test the assay in real biological samples to see if the performance is sufficient to develop an *ante mortem* test for prion diseases. In comparison to the methods based on FCS, the SP-PLA does not require any complicated optical instrumentation, but only a qPCR instrument that can be regarded as standard lab equipment. In the results presented in paper II we only used 5 µl of a highly diluted sample. Since the assay starts with capture to a solid support the sample input can probably be increased to further improve the sensitivity of the assay.
Figure 4. Schematic illustration of SP-PLA for detection of aggregated PrPSc. a) Monoclonal capture antibodies are immobilized on paramagnetic beads. b) A dilution of guanidine-HCl-denatured brain homogenate sample is added to allow PrP to be captured by the capture antibody. c) After washing away unbound sample matrix PLA probes - consisting of the same monoclonal antibody coupled to streptavidin-oligonucleotide conjugates with free 3’-ends and free 5’-ends respectively - are added to the captured target. d) When two PLA probes have bound in close proximity a connector oligonucleotide can guide enzymatic ligation. e) The ligated reporter molecule can be amplified and detected by qPCR.

We have developed assays with two different monoclonal antibodies, indicating that it is the requirement for binding by three copies of the same monoclonal antibody rather than the specific properties of a certain clone of antibody that ensures the assay performance. Thus it should also be possible to develop assays with similar performance for other aggregated proteins, for instance to detect PrPSc from a different species.

Paper III– Profiling cellular protein complexes by proximity ligation with dual tag microarray readout

Introduction

There are several methods available to study protein-protein interactions, but if the aim is to study interactions of endogenous proteins the procedures typically depend on target recognition by an affinity binder. A very common method to study protein-protein interactions is to perform CoIP to capture one target protein of interest such that any proteins bound to that target will be captured together with it. These other proteins can subsequently be identified by Western blot or MS. MS permits detection of large sets of interactants, but for each experiment only one bait is used.

PLA has a great potential for characterization of protein-protein interactions in a highly multiplex fashion, as synthetically designed oligonucleotide tags can be used to barcode the antibody probes recognizing the proteins participating in interactions. If two oligonucleotide-modified antibodies are brought in proximity by binding an interacting pair of proteins, then the antibodies can be joined by ligation, giving rise to an
amplifiable reporter molecule containing tag sequences that identify both antibodies.

The dual tag microarray (DTM) requires perfect complementarity of oligonucleotides printed on the array and the tag sequences located at each end of the reporter DNA strand in order to ligate that strand into a DNA circle that can subsequently be amplified by RCA. DTM greatly enhances specificity for microarray detection and has been proven to be a suitable readout for PLA\textsuperscript{102}.

**Aim of the study**

The aim of this study was to establish a protocol for low multiplex detection of interactions in an all-against-all fashion (i.e. not limited to one bait with many preys) in liquid samples such as cell lysates. We also aimed to establish a protocol for probing the interactions *in situ* in fixed cells or tissues, but with the readout *in vitro* via DTM.

**Summary of findings**

To my knowledge this is the first time an immunoassay is described that can detect all possible pair-wise interactions within a targeted set of proteins. The method is illustrated in Figure 5. As a proof of principle we analyzed cell lysates to detect all binary interactions between four proteins of the NFκB family, and also a house keeping protein, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) to control for total protein input.

We further proved that we could detect a protein (SMAD4) in fixed cells immobilized on a glass slide, and then cleave off the generated reporter molecules to perform the readout on the DTM. This shows that our method is not limited to protein interactions in liquid samples, but that we also should be able to use solid samples such as tissue sections.

**Perspectives on paper III**

We have proven that it is possible to detect all binary interactions among a small set of proteins. It would be interesting to scale up the approach and study a larger set of proteins and to target medically interesting protein interactions.

The main limitation during the work with paper III was to find good binders – binders that recognize the target with sufficient affinity and specificity, but without competing with the studied interactions upon binding. In this paper we used an extensive binder validation pipeline, but to enable scaling up and study a larger set of proteins, the validation of the binders would have to be simplified.
**Figure 5.** Schematic illustration of PLA for detection of protein-protein interactions with DTM readout. a) PLA probes are allowed to bind their target in the assay format most suitable for the application (homogenous assay, SP-PLA, or in situ PLA). b) When two PLA probes have bound in close proximity a connector oligonucleotide can guide enzymatic ligation. The connector oligonucleotide carries a cassette barcode sequence that is incorporated into the reporter molecule as a sample barcoding sequence. c) The connector oligonucleotide is degraded by exonucleases. d) Reporter molecules from two different samples are pooled to lock their ratio during PCR-amplification to allow comparison. e) The samples are prepared for DTM readout by restriction enzyme digest to expose the protein tags and f) uracil-DNA glycosylase (UDG) treatment to ensure that all molecules are single stranded. g) The reporter molecules are then hybridized to the DTM, ligated into a circle and h) allowed to template RCA. The resulting RCA product is detected by hybridization of Cy3 or Cy5 labeled detection oligonucleotides hybridizing to the sample barcoding sequence.
When this work was initiated DNA sequencing was still an expensive luxury that was not cost-efficient to use for the read out of synthetically designed sequence tags. To decode synthetic tags methods based on hybridization was the cost-efficient way to do it. But hybridization requires synthesis of a complementary strand to allow the detection. With the sequencing capacities increasing and the prices dropping\textsuperscript{116}, sequencing has become the natural choice for reading out experiments like this.

Paper IV – A DNA-mediated search for optimal combinations of protein binders

Introduction
As previously described, binding by two or even more affinity reagents is commonly used to ensure specificity of immunoassays. It is a major hurdle in the development of new immunoassays, however, to find pairs or trios of monoclonal binders that can simultaneously bind the target, especially since antibody performance is very much dependent on the assay conditions\textsuperscript{73,82}. In one study it was found that out of 5000 commercially available antibodies from different companies, less than 50% were scored as recognizing their target\textsuperscript{84}. Methods are therefore needed where several antibodies, or other classes of affinity binders, can be validated under conditions as similar to the intended assay conditions as possible.

In paper III we showed that PLA can be used to probe protein interactions in solution or in fixed cells to generate reporter molecules that can be detected \textit{in vitro} with methods allowing high levels of multiplexing\textsuperscript{117}. In this paper we will use a similar approach to detect protein binders simultaneously interacting with their target antigen.

Aim of the study
The aim of the study was to set up a PLA-based protocol to detect pairs of monoclonal antibodies optimally suited to detect specific proteins among a given set of binders.

Summary of findings
We set up a protocol for converting 20 monoclonal antibodies targeting IL12 into PLA-probes so that they can all be used in the same experiment to detect what pair that generates the highest signal over background, and that is thus the most promising pair for a dual recognition immunoassay (Figure 6). Each antibody is separately attached to oligonucleotides with either free 3’- or 5’-ends. Apart from allowing us to study all pairwise combinations among the set this ensures that we have a control for bias from the oligonucleotide system. After ligating the reporter molecules we used
fusion-primers for the PCR-amplification to enable easy sample preparation for readout by NGS using Ion Torrent PGM technology.

When testing the 20 monoclonal antibodies against IL12 we found that one combination of tags is present in higher numbers than the others, indicating that the corresponding pair of antibodies should be optimal to use as a pair in for instance a sandwich immunoassay or in PLA. This combination generated almost half of all reporter molecules from this sample, balancing reaction products from all others of the 380 antibody combinations. The experiment was performed with the probes with free 5’-ends as capture probes, and repeated with the probes with free 3’-ends for captured. Even though the antibodies were labeled with different barcoding sequences the same antibody pair stands out as the most promising pair, and in both experiments the same antibody is suggested for capture. This antibody pair has been identified by others to be the best pair in a sandwich ELISA.
Figure 6. Schematic illustration of PLA for binder pair validation. a) Each clone of binder is split in two and conjugated to DNA oligonucleotides with unique barcode sequences. b) PLA-probes with either free 3’- or 5’- ends are immobilized on paramagnetic beads, each bead receiving monoclonal antibodies from one clone. c) The target is added to be captured by the immobilized PLA-probes, followed by washes to remove unbound target. d) The PLA-probes with the opposite polarity (free 5’- ends if probes with free 3’- ends were used for capture, and vice versa) are added and allowed to bind. e) After washes, a connector oligonucleotide hybridizes to the arms of two PLA-probes brought in proximity by binding to the same target molecule. The connector molecule is also used to introduce a sample barcoding sequence while guiding enzymatic ligation. f) The ligation products then serve as template in a PCR to amplify the amount of reporter molecules, followed by identification of binder combinations by sequencing these reporter molecules.
Perspectives on paper IV
The presented method has been used to identify the most promising pair of monoclonal binders against a cytokine. It would be very interesting to apply the method to validation of binders against other targets. As shown in paper III, probing can take place not only in liquid samples but also in fixed cells, still allowing a multiplex readout \textit{in vitro}. I therefore believe that this method can be used to validate binders to be used for probing dual recognition assays for \textit{in situ} detection of proteins in cells and tissues.

It would also be very interesting to apply the method to other types of affinity binders, and perhaps to combinations of clonal binders of different origin, targeting the same proteins. Binders that are recombinantly produced can be easily modified with chemical groups for site-specific attachment of oligonucleotides as previously discussed so the proposed method might be particularly useful for such binders as conjugation procedures could be easier to perform and avoid risks that the coupling chemistry might destroy the binding site of the affinity reagents.

In SP-PLA three epitopes on the target molecule are bound simultaneously. The search for optimal trios of affinity reagents is even more demanding than the search for pairs, and I have therefore also designed an oligonucleotide system which I will use to identify the best trios of binders (\textit{Figure 7}).
Figure 7. Schematic illustration of PLA for binder trio validation. a) Each clone of binder is split in three and conjugated to DNA oligonucleotides with unique barcode sequences. b) PLA-probes with either free 3’- or 5’- ends are immobilized on paramagnetic beads, each bead receiving monoclonal antibodies from one clone. c) The target is added to be captured by the immobilized PLA-probes, followed by washes to remove unbound target. d) The PLA-probes with the opposite polarity (free 5’- ends if probes with free 3’- ends were used for capture, and vice versa) and the probes with both free 5’- and 3’- ends are added and allowed to bind. e) After washes, connector oligonucleotides hybridize to the arms of the PLA-probes brought in proximity by binding to the same target molecule. One of the connector molecules is also used to introduce sample barcoding sequence while guiding enzymatic ligation. f) The ligation products then serve as template in a PCR to amplify the amount of reporter molecules, followed by identification of binder combinations by sequencing these reporter molecules.
Conclusion

Will the number of proteomic studies performed ever surpass the number of genomic studies? There is certainly some biological rational to aim for proteomic studies rather than genomic studies. Proteins are the effector molecules of life carrying out most biological functions. Proteins are more likely to vary in concentrations, modifications, and interaction patterns depending on the state of health of an individual, while the genome is more static (even though also DNA is subjected to changes such as epigenetic modifications and somatic mutations). Therefore proteins have been suggested to have greater potential to serve as excellent biomarkers to enable early diagnosis and prognosis of severe diseases. However, it should not be forgotten that the genome encodes the proteome via RNA so by studying the genome and the transcriptome important insights are also obtained about the state of the proteome. In my thesis I have not discussed studies at the transcriptome level but these are of course also very important to gain more insight in the proteome and biological processes in health and disease.

One reason for the currently much greater number of genomic studies is probably the convenience of performing the experiments. It is easier to study DNA since it is possible to readily copy the study objects, and probes can be designed in silico following simple rules. But not only the technologies are contributing to the relatively ease of studying DNA. The nature of the targets is also of importance. DNA is present in the form of four times $3 \times 10^9$ bases per cell nucleus. Any particular nucleotide will represent one part in $12 \times 10^9$. Since proteins vary so much in their concentration one particular amino acid in a given type of protein can constitute anything from one in $5 \times 10^4$ to one in $5 \times 10^{11}$ of all the amino acids in proteins in a cell. This illustrates the need for highly precise and sensitive methods for protein detection and profiling, where probably no method suites all needs.

In this thesis I present a method that improves specificity and sensitivity of protein detection by using three recognition events per target and allowing exponential signal amplification by PCR. I have applied this method for detecting candidate biomarkers and aggregated PrP. I have also developed a method for profiling protein-protein interaction networks. Through all these projects I have encountered the difficulty of identifying suitable antibodies. Therefore I developed a method for validating a large set of binders in one experiment, where assay conditions can be set to be as similar as possible to where the real assay will be performed.
Acknowledgements

This thesis only contains a wee fraction of everything that I’ve learnt during my time as a PhD student. Many people have contributed to that. There are many things that could be said but I’ll limit myself to “TACK OCH FÖRLÅT!”

There are some people that I am especially grateful to:

**Ulf Landegren**, my main advisor, thank you for believing in me and telling me that you do so. Thanks for all the advice in my projects and on how to be a scientist!

**Masood Kamali-Moghaddam**, I’m glad that you accepted to become my co-supervisor. Thank you for bearing with me when I am exaggerating critical review of results!

I would also like to thank **Olle Ericsson** and **Katerina Pardali** for introducing me to the lab and advising me in the very beginning. Another person who has been there from the very beginning is **Rachel Nong** - thanks for everything: from encouraging me to come back to the MolTools group for my PhD, through EINAR, via NY and DC, to advice and support when it was time to write my thesis!

Thanks to all past and present group members for making the Moltools group such an exciting, interesting and enjoyable place to work at! Thank you **Christina** for all pleasant coffee breaks! Thanks also for keeping this playground somewhat organized, awesome cakes, and for trying to answer all my stupid questions! Big thanks to **Elin** and **Johanna** for keeping the lab run smoothly! **Johan O**, thanks for taking care of our computers! And thanks to **Erik** for keeping track of the money and travel company in Japan! Thanks to **Delal, Carla, Lena**, and **Anders** for organizing the lab in the past!

Thanks to **Spyros** for sarcasms as well as valuable comments on how to proceed with my projects. I would also like to thank **Gucci, Johan V, Mikaela, Di, Anne-Li, Lei, Tonge, Liza** and **Junhong** for sharing my SP-PLA endeavors. Special thanks to **Junhong** for being a great student, collaborator, and teaching colleague! Thanks to **Caroline** for bringing up the
ethical questions regarding our work as scientists, and also for chairing my defense! I would also like to thank Lotta for being a great collaborator teaching me a lot about prions, and now being a critical newcomer in the group, constantly asking me critical questions about our methods to help me prepare for my defense. Andries, thanks for early morning company on the train and nice discussions on biological questions that would be interesting to study! Thanks to Rasel for your everlasting energy and positive spirit!

Thanks to Ola Söderberg for helpful tips about cell culture and building your team of cell culturers from whom I could “borrow” stuff. Your team, Karin, Linda, Carl-Magnus, Björn, Gaëlle and Agata are not only kind providers of cells but also good friends, travel partners, teaching colleagues etc. Thank you so much!

Thanks to Mats Nilsson and the padlock people: Rongqin, Anja, David, Anna, Camilla, Marco, Elin, Malte and Annika! Lotte and Elin, thank you for being excellent office mates answering my questions about DNA sample purification, sequencing, and bioinformatics!

My office mates in the past, Sara, Ida, Irene, and Malin, thanks for nice chats in the office as well as in the lunch room, and thanks for all the yummy cakes! Big thanks also to Rebecka and Lily, for being patient students ready to learn, but most of all for reminding me how fun science is, even if that means preparing buffers for an entire day. Thanks to Mike for proving that Englishmen can have a good taste of wine and trying to keep track of all binders! Joakim, thanks to you I will make sure that the projector presents all different colours as different!

Thanks to all the people I’ve met in various PhD student councils! I’ve learnt so much from you! Thank you Jenny Ehnberg, Taha Ahmed, and the rest of Doktorandnämnden 2012-13 for a fun, exciting, and productive year! I would also like to thank Sofia Cassel and Andreas Rydberg for our time working with the brochure. It was exciting to write something together with students from different faculties. Thanks also to Johan Lindqvist, Joakim Engström and the rest of MDR 2011-12! And thanks to everyone in the IGP PhD student council over the years!

Coco Norén, Carl-Gustav Spangenberg, and Irène Johansson, I think it should be mandatory for PhD students to be part of the presidium of the Academic Senate. It taught me so much of what means to be a researcher. Merci beaucoup! Tusen tack!

I am very grateful to Marie Allen for giving me such freedom to develop the course labs I have been teaching. Thank you, you gave me an opportunity to
develop as a teacher! Thanks also to Maria and Victoria for good collaboration and teaching me more about forensics!

Christina M, Ulla, Birgitta, Pirkko, Tuulikki, and all other past and present members of the IGP administration, thank you for your kind patience and all help with rules, payments and paper work! Thanks to Avhandlingsproduktion for your patience and correcting all the little style issues that I would never have caught until it was too late!

Thanks to all collaborators! Thank you Alan Sawyer, Pascal Galea and Daniel Laune, Aleksandra Balliu and Lars Baltzer, and the high throughput team in Zurich for teaching me more about protein binders! I am especially grateful to Jonas Schaefer for giving me the opportunity to try the most complicated conjugation protocol so far! Thanks to Jean-Philippe Deslys and Tommy Linné for teaching me more about prions! And thanks to Lars Johansson and Thomas Lundbäck for increasing my knowledge about drugs and drug screening!

My family and friends, thank you for always being there for me with love, care and support! Thanks for convincing me that I am capable of doing whatever I love to do! Familj och vänner, tack för att ni alltid finns där och stöttar och tror på mig! Tack för att ni har övertygat mig om att jag klarar vad som helst så länge det är det jag vill göra! Mamma och pappa, ni är mina största inspirationskällor! Ni skulle bara veta hur många gånger under de här fyra åren som jag har frågat mig vad ni skulle ha gjort i samma situation... Tack för allt!
References


116. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [Internet]. [cited 2013 Aug 13];Available from: www.genome.gov/sequencingcosts

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 923

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.