Proximity Ligation and Barcoding Assays

Tools for analysis of proteins and protein complexes

DI WU
Proteins are fundamental structural, enzymatic and regulatory components of cells. Analysis of proteins, such as by measuring their concentrations, characterizing their modifications, and detecting their interactions, provides insights in how biological systems work physiologically or pathologically at the molecular level. To perform such analysis, molecular tools with good sensitivity, specificity, high multiplexing and throughput capacity are needed.

In this thesis, four different assays were developed and applied to detect and profile proteins and protein complexes in human body fluids, and in cells or tissues. These assays are based on targeting proteins or protein complexes by oligonucleotide-conjugated antibodies, and subsequent proximity dependent enzymatic reactions involving the attached DNA reporter sequences.

In paper I, a solid-phase proximity ligation assay (SP-PLA) was applied to detect synthetic and endogenous amyloid beta protofibrils. The SP-PLA provided better sensitivity and increased dynamic range than a traditional enzyme-linked immunosorbent assay (ELISA).

In paper II, in situ PLA was applied to investigate the correlation between MARK2-dependent phosphorylation of tau and Alzheimer’s disease. Greater numbers of MARK2-tau interactions and of phosphorylated tau proteins were observed in brain tissues from Alzheimer’s patients than in healthy controls.

In paper III, a multiplex SP-PLA was applied to identify protein biomarker candidates in amyotrophic lateral sclerosis (ALS) disease and in the analgesic mechanism of spinal cord stimulation (SCS). Among 47 proteins in human cerebrospinal fluid (CSF) samples, four were found at significantly lower concentrations (p-values < 0.001) in the samples from ALS patients compared to those from healthy controls (follistatin, IL-1α, IL-1β, and KLK5). No significant changes of the analyzed proteins were found in the CSF samples of neuropathic pain patients in the stimulated vs. non-stimulated condition using SCS.

In paper IV, a new technology termed the proximity barcoding assay (PBA) was developed to profile individual protein complexes. The performance of PBA was demonstrated on artificially assembled streptavidin-biotin oligonucleotide complexes. PBA was also proven to be capable of profiling transcriptional pre-initiation complexes from nuclear extract of a hepatic cell line.

Keywords: proximity ligation assay, proximity barcoding assay, sensitive, in situ, multiplex, profiling, amyloid-beta, MARK, tau, CSF, biomarker, protein complexes, Alzheimer’s disease, amyotrophic lateral sclerosis

Di Wu, Department of Immunology, Genetics and Pathology, Molecular tools, Rudbecklaboratori, Uppsala University, SE-751 85 Uppsala, Sweden. Science for Life Laboratory, SciLifeLab, Box 256, Uppsala University, SE-75105 Uppsala, Sweden.

© Di Wu 2014

ISSN 1651-6206
ISBN 978-91-554-8901-4
urn:nbn:se:uu:diva-220070 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-220070)
The more precise the measurement of position, the more imprecise the measurement of momentum, and vice versa.

*Heisenberg's uncertainty principle*

*To my family and friends*
List of Papers

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


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


Contents

Introduction .......................................................................................................................... 11

DNA and proteins: polymers as targets and tools .......................................................... 12
Multiplex: when analyzing one is not enough .............................................................. 13
  Label free ....................................................................................................................... 13
  Labeled probes ............................................................................................................. 14

Specificity: a very important factor for detection of rare events .............................. 15
  Natural fidelity ............................................................................................................. 16
  Seeing more is believing ............................................................................................ 16

Quantification: about sensitivity, dynamic range and precision ............................. 17
  Exponential amplification ......................................................................................... 17
  Localized amplification ............................................................................................ 18

Single cell analysis: for heterogeneous samples ....................................................... 19
  In situ ......................................................................................................................... 19
  Sorting or diluting ..................................................................................................... 20

Detecting and profiling interacting molecules ......................................................... 20
  Affinity based ............................................................................................................ 20
  Proximity dependent ................................................................................................. 21

Neurodegenerative diseases ................................................................................ ...... 23
  Alzheimer’s disease .................................................................................................. 23
  Amyotrophic lateral sclerosis .................................................................................... 24

Present Investigations .................................................................................................. 25

Paper I: Sensitive detection of Aβ protofibrils by proximity ligation - relevance for Alzheimer’s disease .......................................................... 25
  Introduction .............................................................................................................. 25
  Summary of findings ................................................................................................. 26
  Future perspective .................................................................................................... 26

Paper II: Elevated MARK2-dependent phosphorylation of Tau in Alzheimer’s disease .......................................................... 26
  Introduction .............................................................................................................. 26
  Summary of findings ................................................................................................. 27
  Future perspective .................................................................................................... 28
Paper III: Analyses of CSF reveal decreased levels of four proteins in ALS patients, but no changes upon analgesia in patients with neuropathic pain

Background ......................................................... 28
Summary of findings ........................................... 29
Future perspective .............................................. 29

Paper IV: Profiling individual protein complexes by proximity-dependent barcoding ...................................................... 30
Background .......................................................... 30
Summary of findings .............................................. 30
Future perspective .............................................. 31

Acknowledgments .................................................. 32
References .......................................................... 35
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RCA</td>
<td>Rolling circle amplification</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy number variation</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>PLA</td>
<td>Proximity ligation assay</td>
</tr>
<tr>
<td>SMRT</td>
<td>Single molecule real time sequencing</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>Ct</td>
<td>Threshold cycle</td>
</tr>
<tr>
<td>3C</td>
<td>Chromatin conformation capture</td>
</tr>
<tr>
<td>4C</td>
<td>Circular chromatin conformation capture</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>Y2H</td>
<td>Yeast two-hybrid system</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>Aβ</td>
<td>Amyloid beta</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid precursor protein</td>
</tr>
<tr>
<td>ALS</td>
<td>Amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>SOD1</td>
<td>Superoxide dismutase 1</td>
</tr>
<tr>
<td>TDP-43</td>
<td>Transactivation response DNA binding protein 43 kDa</td>
</tr>
<tr>
<td>SP-PLA</td>
<td>Solid phase PLA</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>MAP</td>
<td>Microtubule-associated proteins</td>
</tr>
<tr>
<td>MARK</td>
<td>Microtubule affinity regulating kinases</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Co-IP</td>
<td>Co-immunoprecipitation</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>NFTs</td>
<td>Neurofibrillary tangles</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variance</td>
</tr>
<tr>
<td>KLK5</td>
<td>Kallikrein-related peptidase 5</td>
</tr>
<tr>
<td>HepG2</td>
<td>Human hepatocellular carcinoma cell line</td>
</tr>
</tbody>
</table>
Introduction

Life is a complex, yet organized process, in which enormous molecules with the size from a few ångströms to several microns work cooperatively through basic physical and chemical reactions to fulfill a variety of biological functions. Among these molecules, nucleic acids (DNA and RNA) and proteins are particularly important players. It’s amazing to see that different organisms from bacteria to humans utilize many proteins in common, and it’s equally impressive to know that all somatic cells of a human body regardless of their types and functions have essentially the same DNA content. When mistakes are introduced to these important players, such as through DNA mutations or protein misfolding, diseases may arise, causing inconvenience, suffering, and even death.

The detailed understanding of DNA, RNA and proteins greatly relies upon the tools that are applied. The advent of large-scale technologies has provided extraordinary opportunities to study many DNA sequences and proteins in parallel. The terms, such as genome, transcriptome and proteome accurately reflect the way we study biology with those technologies. In addition to technologies, joint efforts across many research organizations all over the world have been applied in large-scale studies such as the Human Genome Project from 1990 to 2003 [1], the Encyclopedia of DNA Elements (ENCODE) Project [2], and the Human Protein Atlas project [3] from 2003. These projects have generated tremendous amounts of data and images, from which a comprehensive biological understanding is expected.

The work of this thesis was to develop and apply molecular tools to detect and profile proteins and protein complexes. Three different variants of proximity ligation assays (PLA) were applied to detect proteins, protein aggregates, protein modifications and interactions with relevance for neurodegenerative diseases. A new method, the proximity dependent barcoding assay (PBA), was developed to profile individual protein complexes. The application and development of molecular tools for analyzing proteins and protein complexes help understanding of biological processes and may be applied in clinical practice.
DNA and proteins: polymers as targets and tools

A double stranded DNA molecule consists of two antiparallel long chains of phosphodiester connected nucleotides. Each nucleotide contains one of the four nitrogenous bases: adenine (A), guanine (G), cytosine (C) or thymine (T). The sequence of the four bases along the sugar-phosphate backbone of DNA can store tremendous amount of information [4]. The human genome is estimated to be around 2 meters in length, however highly compacted [5] into the cell nucleus with a diameter usually less than 10 microns. The combined properties of double helix architecture, large capacity of information storage together with its stability and flexibility make DNA a fantastic material to fulfill important processes, such as faithful replication and phenotypic expression via transcription. Since the first genome sequencing of bacteriophage φX174 in 1977 [6], the genomes of many organisms have been sequenced including that of Homo sapiens [1].

The base pairing mechanism ensures that A bases along one DNA strand pair with Ts along its complement strand (or with U in RNA), and G residues pair with Cs. This effect makes it possible to design oligonucleotides as affinity probes [7] to target DNA or RNA molecules of interest through hybridization, as demonstrated e.g. in the Southern blot [8] and fluorescence in situ hybridization (FISH) [9]. This hybridization mechanism is also applied in various assays involving enzymatic reactions, which favor a perfect pairing between two DNA strands such as in polymerization and ligation.

The ability to be replicated, another important feature of DNA, is not only the basis of biological inheritance but also utilized in many in vitro DNA amplification techniques. In the polymerase chain reaction (PCR) [10], a DNA template can be exponentially amplified by thermal cycling controlling steps of primer annealing, amplification and denaturation. By rolling circle amplification (RCA), a circular DNA template can be isothermally amplified to form DNA concatemers.

Proteins are also polymers, consisting of single or multiple chains of amino acids. By using the combination of 20 naturally occurring amino acids with distinct chemical properties, proteins can fold by themselves or with the aid of other proteins [11] into many unique structures, and also interact or form complexes with each other to regulate essential biological processes [12,13]. The genome in each nucleated somatic cell in our body is the same; the protein content however can be very different, even in the same cell at different time points [14]. Therefore proteins are the main players in life processes, reflecting the status of a cell.

Designing probes against proteins is not straightforward compared to designing DNA or RNA probes. However, inspired by our immune system using antibodies to recognize ‘intruders’ in the form of infective organisms, antibodies labeled with reporter molecules, can be used as probes in immu-
no assays to detect proteins of interest, as demonstrated by Rosalyn Yalow and Solomon A. Berson to detect human insulin in plasma [15]. By using different labeling, e.g. fluorescence, enzymes and applying the labeled probes on different sample formats, a variety of conventional immunoassays have been developed, for instance immunohistochemistry [16], enzyme linked immunosorbent assays (ELISA) [17], western blot [18], etc. The antibody production in our body occurs through steps of random recombination, followed by selection and introduction of mutations to enhance affinity. Such process has been adopted in vitro, in systems like phage [19] and yeast display [20] to select affinity binders using a variety of protein scaffolds.

Another application of proteins as molecular tools is for enzymatic reactions. Enzymes are very specific catalysts, accelerating many metabolic reactions in our body. Many enzymes used in vivo have been adapted to work in vitro, which allows us to analyze and modify DNA and proteins in a tube just by mixing the proper enzyme and the substrates.

Identification of primary sequences, modifications and interactions of DNA and proteins can be challenging, but it can provide very useful information in research and clinically. In the following sections I will discuss challenges of bio-analytical analysis of DNA and protein, and some current methodologies and technologies to overcome these hurdles.

Multiplex: when analyzing one is not enough

Considering the complexity of life, it intuitively sounds quite unrealistic to expect that a phenotype or disease could be accurately determined by measuring only one single marker. Accumulating evidence show that in complex diseases as cancers and neurodegenerative disorders, far more than single genes or proteins are involved in the pathogenesis. Even for occasions when a single biomarker is enough; to find such a marker still requires extensive screening. In the screening, testing each single marker one by one would be very unpractical considering the sample consumption, experimental handling and expense, thus making the multiplexed analysis favorable.

Label free

In a label free manner, the analyte is measured by the properties of the molecule by itself, such as size, charge, etc. The advantage of label free methods is its potential to be highly multiplexed, if the properties of different analytes are distinguishable in the measurement. Using electrophoresis, DNA molecules can be separated according to their sizes and visualized by using DNA binding dyes.
Due to the different properties of the side chains of amino acids, proteins differ not only in sizes, but also by the charge at a certain pH. Separation of proteins based on size and isoelectric point, two-dimensional gel electrophoresis, provides improved opportunities to separate proteins [24]. The most powerful method to perform protein analysis is mass spectrometry (MS). In non-targeted MS, all proteins in a sample are normally digested into fragments, which are then ionized, separated and detected according to their mass/charge ratio in a mass spectrometer. By comparing the generated signatures with a reference, the protein content in a sample can be identified. In complex biological samples, the signals can however be dominated by abundant proteins, negatively affecting the sensitivity for the low abundant proteins. In the targeted MS the fragmented peptides are first selected, to reduce complexity of the sample and hence increasing sensitivity. Targeted MS approaches like multiple reaction monitoring (MRM) have been demonstrated for sensitive detection of large scale of proteins [25]. Since modifications of amino acids affect the mass to charge ratio, MS is also very suitable for analyses of post-translational modifications (PTMs) of proteins [26], such as phosphorylation, glycosylation, etc. One limitation of MS is that it analyzes molecules in bulk, and the spatial arrangement of a sample is lost. By using matrix-assisted laser desorption/ionization (MALDI) mass spectrometric imaging, areas in a tissue section may be sequentially selected and analyzed by MS to track the spatial arrangement of proteins in a sample [27].

Labeled probes

In an ideal affinity assay, a signal is generated only when an affinity probe binds to its target. To make such an assay multiplexed, the affinity probes should either be labeled with different reporter molecules or the different analytes should be captured on different spots that serve to identify the targets. In either case, the binding of specific probes to their targets can be distinguished.

If the probes are labeled with distinct reporter molecules prior to analysis, they can be pooled and then allowed to simultaneously bind to their targets. The reporter molecules on the probes are then decoded and quantified. In immunofluorescence, using antibodies labeled with different fluorophores, several proteins can be visualized simultaneously in cells and tissues by fluorescence microscopy. In flow cytometry, simultaneous detection of more than 10 different proteins has been demonstrated [28]. However, one limitation of fluorescence is overbleeding between different dyes, due to the spectral overlap [29]. In mass cytometry, antibodies are labeled with distinct mass tags of rare elements, which can be identified and quantified by MS. As there is no overlap between mass tags the level of multiplexing can be far higher than with fluorophores [30,31]. DNA oligonucleotide provides another-
er means for distinct labeling. Bound affinity probes can be identified via techniques for sequencing of attached DNA strands, and the level of multiplexing is exponential to the number of bases read. In addition, the DNA oligonucleotides can be amplified, which can be used to improve the sensitivity of an assay. The decoding of the oligonucleotides needs additional analysis such as real time PCR, microarrays or sequencing [32,33,34].

Thanks to the invention of microarrays [35] and high throughput spotting technology, a large set of affinity probes can be physically immobilized as unique spots on a solid surface for capture of multiple targets in the same reaction. By scanning the intensity of each capturing spot, many targets can be simultaneously identified and quantified. Thousands of different oligonucleotides can be immobilizing on a single solid support, e.g. DNA microarrays have been used to measure the copy number variation (CNV) [36] and single nucleotide polymorphism (SNP) [37] across the whole human genome. In analogy, antibody microarrays were developed for highly multiplexed protein measurement [38]. In a different format, the suspension array, also termed xMAP technology commercialized by Luminex Corp, antibodies are not immobilized on a planar surface but on fluorescence-coded microspheres to capture specific proteins. The captured proteins are then stained with antibodies, which are labeled with one fluorophore. By analyzing the combinatorial fluorescence signals from each microsphere, the captured proteins are identified and quantified. This format allows multiplex protein detection in a single tube, thus suitable for assaying high number of analytes. One potential disadvantage of such a technology is unspecific signals. If a molecule is captured at a wrong spot, then it will generate a false positive signal as another analyte. In order to minimize cross-reactivity, it is thus necessary to optimize the capturing probes.

Specificity: a very important factor for detection of rare events

There is no doubt that sensitivity is very important for detection of a rare event. However, in other scenarios, the sensitivity might not be an issue, the main concern is instead the specificity, i.e. that we detect only what is intended. For instance, an assay can detect DNA mutation with 100% sensitivity, however with 99% specificity. If we apply it on a sample and detect 1% mutations, then we are not sure if the 1% detected mutations are real mutations or just from the false identification. In the immunoassays for protein detection, if the probes bind unspecifically, rather than the target protein, it will still generate a signal. Hence, it is crucial that unspecific binding is minimized.
Natural fidelity

The living entities of nature have evolved during a long time to create dedicated biological machineries to minimize mistakes during life. In DNA replication, a polymerase synthesizes a strand complementary to that of a DNA template with high fidelity based on the Watson-Crick base pairing mechanism. Similarly, a DNA ligase preferentially forms a phosphodiester bond between two DNA strands if nucleotides are correctly base-paired at the nick in the DNA helix. These properties of DNA enzymes have been exploited as tools for molecular biology to enhance discrimination among similar sequence variants. In PCR, by carefully designing a pair of primers, a fragment of DNA with a sequence of interest can be specifically exponentially amplified and detected [10]. By using oligonucleotides as probes to perform DNA ligation reaction using the targeted oligonucleotide as template, the DNA ligation-based assays [39,40] have been used to detect DNA and RNA molecules, with single nucleotide specificity [41,42]. As previously discussed our protein assays use antibodies to specifically recognize an antigen without targeting other proteins. The antibodies thus serve as probes for specific protein detection. However, due to the instability of tertiary structures of proteins, antibodies carrying detectable labels can bind unspecifically to a solid support or other proteins, generating false positive results.

Seeing more is believing

It might be unrealistic to require an affinity probe to bind only the target that it is supposed to. However, if an analyte is bound by many such imperfect probes, it is very likely that the identification is correct. By using many fluorescent labeled oligonucleotides, single RNA molecules can be visualized using in situ hybridization assay (FISH). Similarly, in the proximity ligation assay (PLA), only when two or even more affinity probes bind to the same protein molecule, can the oligonucleotides conjugated on each probes be brought in proximity and be ligated, forming a new DNA template, which can be amplified and detected [43]. PLA has been demonstrated to be suitable for highly multiplexed protein detection with reduced cross-activity [44]. In next generation sequencing (NGS), single DNA molecules are locally amplified to form clusters and then sequentially analyzed using polymerases or ligases [45,46,47]. In such case, although polymerases or ligases may make mistake, the chance that the mistake happens in many molecules in the same cluster is sufficiently low that the sequence can usually still be correctly identified. In third generation sequencing, DNA molecules are not amplified to form clusters. In the single molecule real time (SMRT) sequencing, developed by Pacific Bioscience, the DNA molecule is first circularized and
then sequenced during RCA. Each base in the original circular DNA is sequenced several times during RCA to improve the read accuracy [48].

Quantification: about sensitivity, dynamic range and precision

The molecular processes of our bodies are carefully regulated via the concentrations of a multitude of molecules. The sensitive and precise assessment of the concentrations at different magnitudes is thus very important, and represents an important challenge for detection methods. In human blood, albumin accounts for nearly 60% of the total protein content [49], which is around 10 orders of magnitude higher concentration than that of some low abundance proteins such as Interleukin 6 (IL 6). This broad dynamic range renders the measurement of all proteins in blood at once an impossible task with the current technologies. Sometimes the difference can be as small as only one molecule, for instance one extra copy of a gene in a cell, which nevertheless might have a great impact on a cell. In such case, a method is required to be precise enough to distinguish such subtle difference. One important factor affecting the quantification capacity of a method is the amplification method used in the assay. The signal amplification is normally necessary for a good sensitivity of an assay. But if the amplification introduces bias, it will compromise the precision. So the amplification method in an assay needs to be carefully chosen, for a specific analytical purpose.

Exponential amplification

In an exponential amplification, the products are generated from both the new products and the original templates, whereas in a linear amplification, the signals and products are only formed from the original templates. The advantage of exponential amplification is that it consumes less time to obtain large amounts of products. In a PCR, as low as only a few DNA molecules can be amplified and detected within two hours or less. In real-time PCR, the amount of amplified DNA can be measured for each PCR cycle using a non-specific DNA binding dye or via specific TaqMan probes or similar, recognizing the amplified DNA products [50]. The thermal cycles required to reach a certain fluorescent threshold is logarithmically and reversely correlated with the number of original DNA templates. Given that each PCR cycle in theory doubles the amount of an amplicon, so 20-cycle threshold cycle (Ct) difference is equivalent to an approximately 1,000,000-fold difference in terms of concentration. Therefore real-time PCR provide a very sensitive DNA detection over a large dynamic range. By combining the PCR with a microfluidic system, 96 different targets in 96 different samples can be sim-
ultaneously analyzed in one Fluidigm chip [51]. However, the amplification efficiency might not be 100% for every amplicon in each PCR cycle. This variation can be accumulated along with the PCR cycles, thus compromising the precision, especially for low copy numbers of templates requiring more thermal cycles.

A new approach to minimize the amplification bias is by introducing molecular barcodes. The DNA or RNA molecules are first attached with a unique identifier tag, and then each tagged molecule is amplified and sequenced. The molecules sharing the same unique identifier tag would indicate that they are amplified from the same original DNA template and are counted as one. By combining this approach with NGS, it is possible to reach an absolute quantification for genome sequences or transcripts [52,53,54].

Localized amplification

Another solution to solve the amplification variation is by localized amplification to form clones. By this approach, the signals are enriched locally, which can easily be distinguished from the background. Since each clone is derived from a single molecule, such a method allows absolute quantification of the original molecules. In digital PCR, the DNA templates are initially distributed randomly into many compartments in a way that each compartment contains either zero or one template. The DNA templates are amplified separately in each compartment by PCR. Signals are obtained only from the compartments containing the amplified DNA templates. By counting the total number of positive (and negative) compartments, the original number of DNA templates can be precisely accessed [55].

In a homogenous format instead of performing in separated reactions, circular DNA templates can be amplified by RCA to form a DNA concatemer locally, which can be visualized after hybridization to a fluorescent- or horseradish peroxidase (HRP) conjugated oligonucleotide. The RCA can be performed in solution and counted by flowing through a detector [56]. The RCA can also be performed on a solid support and quantified by microscopy [57], which makes it suitable for in situ analysis of biological samples, for instance, to count the DNA copy number of single cells [58].

For protein analysis, in analogy to the idea of digital PCR, digital ELISA was developed to count individual proteins [59]. If the antibodies are equipped with DNA oligonucleotides, then the measurement of proteins is converted to that of DNA. Then the previous methods to locally amplify DNA can be applied for precise protein measurement. Although the readout can be improved by localized amplification, affinity based protein assays can also be severely affected by inefficient or unspecific binding of the probes, whose contribution to the variation cannot be overcome by such approach.
Single cell analysis: for heterogeneous samples

An adult human has around $4 \times 10^{13}$ cells [60], representing a large number of lineages. Although they share the same or similar DNA content, the cells may have very different morphologies and functions. Even cells of the same lineage, each cell can be very different in terms of its RNA and protein contents. By bulk measurements, information about differences among cells is lost. It is becoming more and more important to perform RNA and protein measurements at a single cell level. However, there are several challenges for single cell analysis. First, the material from a single cell is very limited, which can be below the limit of detection of most analytical methods. Second, to fully access the heterogeneity of biological samples, a large number of individual cells must be investigated, thus requiring high throughput analysis. The first problem relies more on the sensitivity of an assay, which was discussed in the previous section. Here, I will focus on the second challenge in regard to high throughput single cell analysis.

**In situ**

In a tissue section on slide, cells are physically fixed within their natural histological context. Applying an assay directly on such fixed cells allows accessing the signals of single cells by their distinguishable geographic arrangement. By microscopy, several hundred cells can be visualized in just one image. Even when the cells are not fixed on a solid support the signals can be preserved on the level of individual cells. Flow cytometry provides a valuable opportunity to analyze individual cells in suspension [61], allowing single cells to be measured one by one by rapidly passing through a detector. Applying such an assay directly on many cells simultaneously greatly reduces the experimental handling. One of the challenges of *in situ* analysis is multiplexing, because it usually uses fluorescence as readout, with limited multiplexing capacity. By using combinations of different colors, the multiplexing capacity may be improved [62]. A further improvement is afforded by *in situ* sequencing, in which the barcoded RCA products are decoded sequentially using the sequencing by ligation protocol [63]. *In situ* sequencing has been applied to detect 40 different RNA molecules on cells or in tissue sections. Another challenge of *in situ* analysis is that the signal might become saturated in a limited surface area (the size of a cell), especially assays requiring signal amplification, e.g. RCA. This problem was improved by generating RCA products with different tags, at a set ratio. These tags can be visualized by fluorophores. When signals from one tag is saturated a more diluted one, reporting in another wavelength can be used. This will greatly improve the dynamic range [64].
Sorting or diluting

For deep analysis, such as in genome or transcriptome sequencing, cells are usually required to be lysed to release their intracellular contents. To perform single cell analysis, cells need to be isolated prior to the lysis, in order to keep the released cellular content separate. Isolating single cells can be done manually with the aid of a microscope, or using flow cytometry in a high throughput and automatic manner [65]. The recent development of micro-fluidic technology, such as Single-Cell Auto Prep System commercialized by Fluidigm integrated isolating of single cells and library preparation for sequencing in one chip.

One way to avoid laborious or instrumentation-dependent sorting is to dilute cells randomly into excessive numbers of separate water-in-oil droplets so that each droplet contains either zero or one cell, according to Poisson distribution. The development of droplet manipulation technology permits adding and removing material from the droplet, allowing large sets of separate experiments to be performed in rapid sequential order. By using droplets, although the cells are in separated compartments, they can all be handled as a one-tube reaction. This assay layout greatly reduces the reagent consumption and experimental manipulation.

Detecting and profiling interacting molecules

Progress in genome, transcriptome and proteome analyses have greatly expanded our understanding of biology. However, to further investigate how the biological regulation is carried out requires also studying the interactions among these molecules. DNA stores the genetic information in a linear format, but distinct regions along a chromosome can also bend to form loop structure in order to enhance the transcription of a gene [66]. Proteins with their unique three-dimensional structure can form important complexes, some of which can also interact with DNA or RNA to carry out transcription and translation [67,68]. The detection and characterization of interacting molecules have additional challenges as what has been discussed earlier for DNA or protein analysis alone. Current methods for studying molecular interactions mainly rely on the affinity or proximity between the molecules.

Affinity based

A functional molecular interaction usually lasts longer than a random collision to ensure an expected reaction to occur, such as enzymatic reaction. Based on this assumption, physiologically interacting molecules should have a higher affinity to keep them together. In co-immunoprecipitation (Co-IP),
the protein complex is captured by using the antibody against a known protein. The unknown interacting partners will also be pull down, depending upon their affinity to the known protein. Then the captured proteins can be further validated by Western blot or analyzed by MS to identify the proteins in the complex. This approach provides the golden standard to study in vivo protein interaction. The challenge of this approach is to maintain the protein complexes, especially for those transiently interacting proteins.

In a different assay format under the affinity assumption, the molecule can be used as a bait to allow accessing to a pool of molecules, and pulling down its interacting partners. This is useful to analyze the interactions between biological materials and synthetic products, such as protein-ligand screening.

Proximity dependent

Another evidence showing two molecules are interacting is that they are close in the native biological environment. The closeness can provide them a higher chance to be joined together than the molecules stay apart. By analyzing the joined two molecules or parts from both of them, the original interacting molecules can be identified. For protein complex analysis by using bivalent cross-linkers, covalent bonds can be formed between the proximal amine groups on proteins. Then the cross-linked peptides can be further analyzed by MS to identify the interacting proteins [69]. By using this approach large sets of protein complexes can be profiled without prior knowledge. Similar as cross-linked proteins, two contacting DNA fragments can be ligated into one DNA fragment by using DNA ligase. In the chromatin conformation capture (3C) [70], and related technologies termed as 4C [71] and Hi-C [72], interactions between DNA fragments in cells can be detected.

In an indirect manner, the interacting molecules can be labeled with reporter molecules that upon proximity can generate an overlapping signal. One example is using antibodies labeled with different fluorescence to target the proteins of interest. If the two fluorescent signals are always overlapping under microscope, it indicates that the two proteins may be interacting. The development of super-resolution microscopy has made it possible to detect single proteins in a protein complex [73].

It is also possible to label the probes with reporter molecules that upon in proximity can generate a new detectable signal. Förster resonance energy transfer (FRET) was discovered by the German scientist Theodor Förster in 1946 [74]. It describes the phenomenon that when a donor fluorophore is exited, the energy is transferred from the donor fluorophore to a nearby (within several nanometers) acceptor fluorophore that will emit light at a longer wavelength. The transfer efficiency is inversely dependent on the sixth power of the distance between the donor and acceptor fluorophores, thus it can be used as a molecular ruler to measure distance between mole-
cules at nanometer scale. By labeling two proteins with fluorescent donor and acceptor tags, FRET can be applied to study *in vivo* protein interaction to visualize two proteins interaction in a living cell [74]. The *in vivo* system provides real time protein interaction measurement in biological environment. However the proteins of interest need to be transfected and co-expressed with fluorescent reporter molecules, and thus is not suitable for analyzing endogenous pathological samples.

Another useful labeling moiety is oligonucleotides, which can be manipulated to trigger a variety of enzymatic reactions once in proximity. In PLA, a pair of antibodies conjugated with different oligonucleotides are allowed to bind their targets. If the targets are interacting, then the oligonucleotides are secondarily brought in proximity and may be ligated into one new DNA template carrying the identities of the two interacting molecules. The new DNA template can be amplified by PCR and analyzed by dual tag microarray or sequencing. By designing a set such pairs of oligonucleotides and conjugating them to different antibodies, a set of protein interactions can be analyzed [75,76]. In a different format, the oligonucleotides brought in proximity can help additional oligonucleotides to be ligated into circular DNA template, which can be amplified locally to form a DNA concatemer at the site where an interaction occurred. This allows detection of protein interaction *in situ* at subcellular resolution. By incorporating different oligonucleotide motifs in the RCA product, *in situ* PLA can be used to simultaneously visualize several protein complexes in cells or tissues [77].

PLA with its variants provides useful tools to analyze pair-wise molecular interactions. However, due to the topological nature of ligation, joining two per reaction, this approach is not suitable for profiling complexes consisting of more than two molecules. To overcome this limitation, I have developed a new concept, proximity dependent barcoding, to identify molecules participating in the same molecular complex by representing these with identical, unique sequence tags. The process creates pools of DNA strands where each newly joint sequence contains both the identity of the molecule participating in a complex and also a unique identifier sequence characterizing each molecular complex. The combinations of tags generated in the procedure are determined via NGS. By sorting all the recorded sequences according to the identity tags for each complex, the molecular components from individual complexes can be revealed.

One means to instantiate this general concept is to separate individual complexes into isolated compartments, each of which also contains many copies of a unique tag sequence, which can be incorporated in DNA strands that also encode the identity of affinity reagents that have bound members of the complex. However such a procedure would be very laborious when analyzing many complexes, and it would be favorable to perform the assay in a homogenous manner. We amplified circular oligonucleotides containing random sequences to make RCA products, each carrying copies of identical
of unique tags. Then we used these RCA products to barcode oligonucleotides attached to affinity reagents brought in proximity by recognizing molecules participating in individual complexes. Monomers of the RCA products serve as templates for the molecules to carry on DNA polymerization, through which the molecules from the same complexes obtain the same unique tag. We expect that this approach will prove useful for profiling any molecular complex at sub-micron scale, as long as the components in the molecular complexes themselves are distinct DNA strands or can be labeled with oligonucleotides.

One limitation of proximity-based interaction detection is the false positive signal. In a dense biological environment, molecules can be physically close but not functionally interacting. It is important to estimate what is the minimal distance to generate a proximity signal. If the molecules are too dense, then the proximity signals might just come from random colocalization instead of specific interactions. Therefore, control experiments are required to rule out that signals are a consequence of chance co-localization.

Neurodegenerative diseases

Neurodegenerative diseases are characterized by progressive neurological dysfunctions, and neuron death. Although they have distinct symptoms, such as dementia for Alzheimer’s disease, and difficulties with voluntary movements for amyotrophic lateral sclerosis, they share several molecular similarities, for instance the ‘prion-like’ misfolded proteins are found in these and also many other neurodegenerative diseases [78].

Alzheimer’s disease

Alzheimer’s disease (AD) was discovered by and named after Alois Alzheimer in 1906. It mainly affects the elderly population and is the most common type of dementia. There are around 26.6 million people globally (2007) suffering from AD or similar dementias, and the number of patients is expected to double by 2050 [79]. Despite great efforts, AD remains incurable and brings enormous suffering to the patients, and it is a great burden on the families and society.

The exact cause of AD is still unknown, but there are mainly two pathological pathways believed to be relevant for the disease. In the amyloid cascade hypothesis, proposed by John Hardy and David Allsop in 1991, amyloid precursor protein (APP) mismetabolism and beta-amyloid deposition are the initiating events in the disease process [80]. APP is a transmembrane protein, expressed in many tissues with yet unclear functions. In the AD pathological pathway, APP is sequentially cleaved by β-secretase and γ-
secretase to release amyloid beta peptide (A\(\beta\)), which is shown to be neurotoxic via various pathways, for instance by being involved in impairing mitochondrial function [81]. This hypothesis was also supported by genetic studies showing that several mutations of APP gene were related to the familial AD with an early onset [82,83].

In the tau pathology, the tau protein is phosphorylated and dissociates from the microtubules. The phosphorylated tau proteins can be hyper phosphorylated and eventually aggregate into neurofibrillary tangles (NFTs) [84]. Because the normal function of tau is to stabilize the microtubules [85], the dissociation of tau would potentially impair the integrity of microtubules and induce neuron death.

Considering the lack of detailed understanding about the mechanisms of AD, molecular biomarkers indicating the state of the disease would be clinically very useful for providing good diagnosis and prognosis. Currently elevated levels of total tau and phosphorylated tau, and a decreased level of A\(\beta\)\(_{42}\) in the cerebrospinal fluid (CSF) are found to be important diagnostic and prognostic markers of AD [86,87,88]. An elevated level of soluble A\(\beta\) oligomers in CSF from AD patients has been reported [89], suggesting that it could be also an important biomarker candidate for the disease.

**Amyotrophic lateral sclerosis**

Amyotrophic lateral sclerosis (ALS) is characterized by the progressive degeneration of motor neurons. The patients’ voluntary movements are affected, with initial symptoms as involuntary limb movements and speech awkwardness, followed by swallowing and breathing difficulties as the disease progresses to the late phase. ALS is currently incurable, and most of patients die within 5 years from the onset.

There is no clear clue about the cause of the disease. Molecular investigation of ALS in 1993 lead to the identification of a mutation in the superoxide dismutase 1 (SOD1) gene [90]. Since then, more than 160 mutations of SOD1 with relevance for ALS have been identified [91]. SOD1 is responsible for clearing superoxide radicals in the body. Both mutant and wildtype forms of SOD1 have been found to be able to misfold into aberrant conformation, which is speculated to be neurotoxic for motor neurons [92]. Mutations in proteins for RNA processing such as TDP-43 have also been demonstrated to be of importance for progress of ALS are [93]. The efforts to identify protein biomarkers for ALS in human CSF or blood have yet not resulted in a reliable diagnostic biomarker for ALS.
Present Investigations

Molecular tools can be used for detection and characterization of proteins and their interactions and posttranslational modifications; thereby they can offer detailed insights in the pathology of diseases. In this regard, the PLA has been shown to permit high-performance protein analyses in a wide variety of contexts [43]. In this thesis, three variants of PLA, solid phase PLA (SP-PLA), in situ PLA, and multiplex SP-PLA were applied to detect proteins and protein complexes from human body fluid, cell lines and tissue sections. In addition, a new method, the proximity dependent barcoding assay (PBA), was developed to profile individual protein complexes.

Paper I: Sensitive detection of Aβ protofibrils by proximity ligation - relevance for Alzheimer's disease

Introduction

Amyloid-beta (Aβ) peptides, processed from amyloid beta precursor protein (APP) [94], can aggregate into Aβ fibrils, and deposit as plaques outside neurons in the brain, which is one of the hallmarks of Alzheimer’s disease (AD). The histochemical and in vivo detection of Aβ fibrils in the brain has great value to understand the progress of AD [95]. However, recent studies have shown that soluble Aβ oligomers/protofibrils are more toxic than the insoluble fibrils, therefore holding potential as an important blood or CSF biomarker of AD [96].

By using the same monoclonal antibody as both capture reagent and enzyme-linked probe, sandwich ELISA has been able to sensitively detect both synthetic and endogenous Aβ protofibrils [97]. The dual binding requirement in sandwich ELISA renders it capable of detecting Aβ protofibrils with a minimum of two monomer Aβ peptides. In a SP-PLA, the target protein is first captured by an antibody on a solid support, and then probed by two antibodies each conjugated with an oligonucleotide that can be ligated to form a new reporter oligonucleotide upon binding to the same target [98]. This assay arrangement assures that SP-PLA has triple binding requirement, and it is thus suitable to detect protein aggregates with a minimum of three monomer proteins. Based on the triple binding requirement of SP-PLA, we
aimed to set up an assay for sensitive and specific detection of larger Aβ protofibrils by using one monoclonal antibody against the identical epitope of Aβ peptide.

Summary of findings

We achieved detection of synthetic Aβ protofibrils at a concentration of 0.1 pg/ml in buffer and 0.27 pg/ml in 10% human cerebrospinal fluid (CSF). These results were up to 10-folds better than a sandwich ELISA using the same antibody as capture and probe. Moreover, the SP-PLA had a linear dynamic range of five orders of magnitude, whereas the linear range for ELISA reached only to two orders of magnitude.

We applied our assay to detect endogenous Aβ protofibrils in brain homogenates from mice. Elevated concentrations of Aβ protofibrils were observed in samples from transgenic mice (APP<sub>Arc-Swe</sub>), harboring a mutation in the human associated with AD, compared to those from non-transgenic mice. We also found that the protofibril concentrations revealed by SP-PLA was lower than that revealed by sandwich ELISA and a homogenous phase PLA, both requiring dual binding to produce a detection signal. We interpret this to reflect the ability of SP-PLA for specific detection of Aβ oligomers/protofibrils composed of three or greater numbers of monomers.

We did not observe significantly higher concentration of Aβ protofibrils in the CSF samples from AD patients (n=4) compared to samples from age matched controls (n=4) using our assay, but the small numbers of samples investigated limits the conclusions that can be drawn from this experiment.

Future perspective

The requirement for triple binding makes SP-PLA a general tool to detect protein- and peptide aggregates. It is feasible to design the assays so that four or more target recognitions are required for detection [99]. However, the requirement for multiple binding events may also compromise the sensitivity of the assay, necessitating careful binder validation and assay optimization.

Paper II: Elevated MARK2-dependent phosphorylation of Tau in Alzheimer's disease.

Introduction

Tau belongs to the family of microtubule-associated proteins (MAP), and mainly functions to stabilize the axonal microtubules. In the pathological
pathway of AD, tau may be phosphorylated and dissociate from the microtubules, compromising the axonal integrity. The dissociated tau may further be phosphorylated and aggregate into tau tangles, serving as an important hallmark of AD. The microtubule affinity regulating kinases (MARKs) have been suggested to play an important role in the early phosphorylation of tau leading to aggregation [100]. Therefore, the investigation of MARK-dependent phosphorylation of tau may increase our understanding of the disease, and provide hints for clinical treatment.

Co-immunoprecipitation (Co-IP) is a conventional method to detect protein interactions. Since this assay is performed in bulk, histological information is lost, and it thus fails to reveal where the proteins interact. In situ PLA uses a pair of antibodies conjugated with different DNA oligonucleotides as PLA probes to target pairs of proteins of interest. If the proteins interact, the two bound antibodies are brought in proximity, and consequently also the conjugated oligonucleotides. The DNA oligonucleotides are used to template formation of a circular DNA oligonucleotide through enzymatic ligation. This circular oligonucleotide is locally amplified via RCA to generate a detectable DNA cluster, easily visible under the microscope. Since the signal is directly generated at the location of the interacting proteins, their histological context is preserved [101]. In this study, we aimed to apply in situ PLA to investigate the role of one of the MARK family, MARK2, in its interaction with tau protein in cells and AD brain tissue sections.

Summary of findings

We first validated that the MARK2 antibody only recognizes the MARK2 protein, as opposed to other MARK isoforms. We co-transfected rhTau 3T3 cells, stably expressing human tau proteins, with each of the MARK isoform plasmids (MARK1-4) together with green fluorescent protein (GFP) plasmids. After applying rabbit MARK2 antibody, and then secondary PLA probes against rabbit IgG, we observed increased in situ PLA signals only in the MARK2 transfected cells, indicating that the MARK2 antibody is directed specifically against the MARK2 protein and does not recognizes the other MARK isoforms.

By using primary mouse and rabbit antibodies respectively against tau and MARK2 protein, followed by secondary in situ PLA probes against mouse and rabbit IgG, we observed elevated MARK2-tau interactions in MARK2 transfected cells. We also found that pSer262 tau was elevated in the transfected cells. After pre-treatment of the cells with staurosporine, a protein kinase inhibitor, the MARK2-tau interaction and pSer262 were found at decreased levels, however without a loss of MARK2 or tau expression. This confirmed that the MARK2-tau interaction is not due to random colocalization between MARK2 and tau, and the interaction is relevant to the phosphorylation of Ser262 of tau protein.
In post-mortem brain sections from AD patients and healthy controls we observed an overlapping pattern of MARK2 protein and MARK2 mRNA signals in neurons in the DG and CA regions of the hippocampus using immunohistochemistry and in situ hybridization. We observed significantly elevated levels of MARK2-tau interactions in the hippocampus region of AD patients compared to those in non-AD controls. Analyses of the adjacent tissue sections stained with the pSer262 antibody revealed the correlation of the formation of NFT and MARK2-tau interaction in AD.

Future perspective

In this study, MARK2, a member of the MARK family of protein kinases, was demonstrated to interact with tau protein, suggesting its role in the development of AD. This raised questions about the activities of other isoforms of MARK, and specifically about their roles in the phosphorylation of tau protein in AD. A follow up study was performed to investigate all the MARK isoforms, showing that among the four MARK proteins MARK4 has the strongest interaction with tau, and correlated best with the disease development [102].

Paper III: Analyses of CSF reveal decreased levels of four proteins in ALS patients, but no changes upon analgesia in patients with neuropathic pain

Background

Neurological disorders cause great suffering and a considerable burden to the family and society. Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease annually affecting around 4-6 per million people around the world [103]. Despite great efforts and investments, there is still a lack of accurate diagnostic assays and means for proper treatment, partially due to the lack of knowledge about the molecular pathways of the disease, in part related to limit access to pathological materials. CSF circulating in brain and spinal cord provides an important and promising matrix for neurological research and clinical investigation. However, unlike other body fluid samples that can easily be acquired, CSF sampling is more complex, and can cause side effect such as severe headaches. Therefore, control samples from neurologically healthy donors are in short supply, rendering clinical studies difficult. In addition, the amount of CSF obtained from individuals is also limited, making assays that require substantial amounts of material unpractical.
Multiplexed immunoassays provide tools to analyze many targets simultaneously, not only saving on samples, but also on experimental handling. One concern of multiplex immunoassays is the commonly observed increased cross-reactivity between the affinity binders compared to assays for single proteins, potentially leading to false negative or positive findings. However, the multiplex SP-PLA, by requiring triple antibody-antigen recognition, has proven to be a very specific assay, suitable for analyzing many proteins in parallel, and without increased background due to cross-reactivity upon multiplexing compared to simplex assays [104].

In this study, we aimed to expand a previously described 35-plex SP-PLA to 47-plex to analyze a panel of proteins including inflammatory and neurotrophic markers in human CSF samples from neurologically healthy control patients, from ALS patients, and from neuropathic pain patients in the stimulated vs. non-stimulated condition using spinal core stimulation (SCS), with the intention both to obtain reference protein concentration values from normal individuals, and also to identify potential biomarkers for ALS disease and analgesia.

Summary of findings

By using recombinant protein dilutions, we found that 46 out of the 47 proteins in our panel gave a dose-dependent response. After applying our assay on CSF samples from 72 urology patients with no neurological diseases, we found that 19 proteins were detectable in at least 95% of all patients.

By comparing all the 47 proteins from ALS patients (n=20) and controls (n=20), we found that four proteins (follistatin, IL-1α, IL-1β, and KLK5) were significantly lower (p-values < 0.001) in the ALS patients compared to the controls. But we found no significant differences among the analyzed proteins in the neuropathic pain patients in the stimulated vs. non-stimulated condition using SCS.

Future perspective

In this study, we demonstrated that multiplex SP-PLA is a useful tool for analyzing proteins in human CSF samples. We measured several protein concentrations that have not been previously reported for non-neurological diseases. Our hope is that these values can serve as reference values in other clinical studies. In the ALS biomarker screening, we compared 20 ALS patients with 20 controls, as an early phase of biomarker discovery. Larger ALS patient and control cohorts will be required to validate the four biomarker candidates identified in this study.
Paper IV: Profiling individual protein complexes by proximity-dependent barcoding

Background

Proteins interact with each other to regulate a variety of important biological pathways inside cells [12,13], and the detection and characterization of these interacting proteins or protein complexes can thus provide valuable information about these cellular events at the molecular level, holding promising clues for biomarker finding and therapeutic intervention.

The identification of proteins in complexes requires a multiplexed assay. Mass spectrometry provides the most unbiased and highly multiplexed assays, capable of de novo identifying proteins [105]. However, it measures the protein composition with relatively poor sensitivity, and it thus has limited suitability for heterogeneous or diluted samples. Affinity based methods, by labeling the binders with distinguishable reporters, can generate a signal to indicate where proteins are interacting [75]. However, in such methods, the generation of signals normally reflects pair-wise interactions, while they are unsuitable as tools to examine higher order protein complexes.

In this study, we aimed to profile individual protein complexes using a novel proximity barcoding assay (PBA). In this assay, every protein in each complex is barcoded with a DNA tag, which after sequencing can reveal both the identity of the protein, and which individual complex it originated from.

Summary of findings

We first validated our methods on artificially assembled protein complexes. We incubated streptavidin tetramers with four different biotin-modified oligonucleotides (bio-oligos) separately or jointly, to make each streptavidin tetramer bind only one type of bio-oligos or a random combination of four different bio-oligos. We also incubated streptavidin tetramers at a constant concentration with increasing amounts of bio-oligos, to construct streptavidin tetramers that on average bound different numbers of bio-oligos. By applying our method on these differently prepared streptavidin-bio-oligo complexes, we found that the PBA-based profiling of the complexes were consistent with the expected profiles.

We applied our method to profile transcription pre-initiation complexes. A TFIID antibody was immobilized on a solid support, and a pool of PBA antibody-oligonucleotide conjugates directed against TFIIA, TFIIIB, TFIID, TFIIF, TFIIH, RNA polymerase II and CTCF were applied together with a negative control probe. We observed that in the nuclear extracts from the
human hepatocellular carcinoma (HepG2) cell line, these proteins form complexes. In the absence of nuclear extracts, we only observed complexes containing one or two PBA probes, which likely originated from nonspecific binding of PBA probes on solid support.

Future perspective

We demonstrated in this study that PBA could be used to profile individual artificially assembled protein complexes. As a proof of concept, we demonstrated that the method could be applied to profile protein complexes from biological materials. It is necessary to apply the procedure in several well-characterized and known biological complexes in order to investigate the accuracy of profiling revealed by PBA. Thereafter, the technology can be applied on unknown biological complexes that are not individually profiled before, such as exosomes.
Finally it comes to the part I’m most enthusiastic to write, in which I want to express my sincere gratitude to the people for their contribution, support and friendship during these years, especially:

**Masood Kamali-Moghaddam**, my main supervisor, thanks for choosing me as your PhD student, and believing in me these years. You were always available to solve any problem I had, and taught me not only how to do an excellent work in science but also important characteristics that a scientist should have.  
**Ulf Landegren**, my co-supervisor, I feel very lucky to have the opportunity to work in such a wonderful lab you built. Thanks for inspiring me with your great wisdom, broad knowledge and the elegant way you design molecular tools.  
**Ola Söderberg**, another co-supervisor of mine, thanks for sharing your expertise in biology, pathology, molecular tools, and the combination of the three. You explain things clearly and with a lot of patience no matter in the lab or a swimming pool.

My colleagues and friends in the lab: **Anne-Li**, thanks for being a joyful collaborator and also a nice friend to talk to and laugh with. I don’t remember how many things we did together, but I’m pretty sure that each of them was fun! **Junhong**, thanks for working closely with me in the last two years of my PhD and always believing that our project would succeed; also thanks for your friendship inside and outside the lab. **Liza**, thanks for your smiling, which reminds me that stress shouldn’t stop me from feeling good. **Lotta W**, thanks for taking the responsibility to organize our projects, continuing our tradition to detect ‘prion-like’ proteins. **Felipe**, thanks for coming back to Sweden joining the PLA club and your advice on my thesis. **Rachel**, thanks for sharing your smart solutions for experiments, philosophical views about science and wild imagination about the life in future. **Lei**, my office mate, thanks for sharing your spirit to make things easier, faster, bigger and brighter. **Tonge**, thanks for all the occasional scientific or hilarious talks. **Caroline**, thanks for teaching me how to write a proper subtitle. **Andries**, thanks for making me be aware of how ignorant I am in cell biology by each of your presentations. **Rasel**, thanks for showing your enthusiasm for science, and warmness to people every day. **Carl-Magnus**, thanks for the long discussion in the pub about science, thesis writing and defense. **Karin**, thanks
for always being energetic in organizing something for other people. **Linda**, thanks for encouraging of workout and FIKA, I wish I had attended more. **Gaëlle**, thanks for your accompany of snowboarding in the warm winter. **Björn**, thanks for your special humor and the fun you made from various objects. **Axel**, thanks for helping with HPLC. **Agata**, thanks for bringing various top ics worth talking about. **Lotte M**, thanks for the discussion about beads and droplets. **David**, thanks for your endless interests in RCA. **Elin F**, thanks for helping me dealing with any type of sequencing results regardless of their messiness. **Camilla**, thanks for the discussion in the Berzelii PhD forum. **Johan O**, thanks for your generous help about all computer-related problems and mathematical puzzles. **Erik**, thanks for chairing various events, from Christmas dinners to PhD defenses including mine. **Joakim**, thanks for all the interesting discussions. **George**, thanks for bringing microfluidics to our lab. **Elin E, Christina C and Johanna**, thank you all for your great contribution to make everything tidy and smooth in the lab, and arranging enjoyable activities outside of the lab every year.

Thank you **Mats Nilsson** for your enthusiasm to discuss new ideas and being able to tell whether it can work or not right away. Also thanks to **Marco M, Anna E, Annika, Anja, Malte, Elin L, Tomasz, Monica, Thomas, Ivan**, for all the fun we had in Uppsala and your hospitality in Stockholm.

Special thanks to the members not working in our lab anymore: **Rongqin**, for sharing your interesting ideas in science and business. **Spyros**, for your critical and valuable comments about my work. **Johan V**, for sharing your entrepreneur adventure and always asking my progress in lab. **Maria**, for solving problems that trouble many people in the lab. **Gucci**, for being a driving force in our project, and a modest friend in daily life. **Irene**, for being my in situ PLA tutor and helping me with fly anatomy under microscope. **Ida**, for your positive attitude. **Pier**, for introducing me to the radioactive kingdom. **Reza**, for teaching me cell culture. **Yanling L**, for being my first bench neighbor. **Yuki**, for making things neat. **Henrik, Magnus** and **Olle**, for inspiring me with selectors. **Mathias**, for your clever comments. **Linnea**, for your cheerful attitude. **Shanjun**, for sharing the student office. **Kalle**, for demonstrating how to behave well in a thesis defense. **Tim**, for being the founder of ICA Conze and **Sara, Malin, Katerina, Michaela, Jenny, Carla, Lena, Delal**, thank you all for making our lab a pleasant place.

I want to give my gratitude also to the people outside our lab. **Ulla Steimer and Christina Magnusson**, thank you for answering all kinds of questions from me and helping me arrange my half time review and thesis defense. **Eva and Mats G**, thanks for the excellent biostatistical analysis. **Marco C**, thanks for making the ChiP (not from potatoes). **Sonchita**, thanks for the productive collaboration. **Maria L**, thanks for the story about forensic sci-
ence. **Sara B**, for arranging interesting PhD activities. **Sandra**, thanks for sharing the teaching experience. **Gang**, thanks for the discussion about cloning. **Cecilia** and **Ulrika L** thanks for the help with IonTorrent and Miseq sequencing. **Gerald, Ulrika G, Jan and Hampus**, thanks for the discussion and help with patent application.

还有在乌普萨拉认识的朋友们：孙峪，付曦，胡劲之，高爽，王潇宇，张晶，李黎，孙纬伦，石成硕，周研，李晓梦，李娜，张昕，焦响，吴逸飞，박성규，徐长刚，刘佳音，胡双威，孙阳，宋冬岩，骆静辉，柴谦，袁殷，王琳，李素真，刘尚锋，马海莎，许勍，崔涛，毛昉，郁笛，靳川，黄华，姜望舒，赵金，玛希，郭小虎，华恺，张蓬，于子泉，张璐，丁舟洁，徐晖，张磊，张研宇，张靖佶，张博，陆禄，赵雅妮，熊安琪，谢源，任毅，李晓文…，感谢和你们一起在瑞典的时间，你们的热情让我从来没有感觉到孤独过。

爸爸妈妈，谢谢你们为我做的一切，并且一直相信我，支持我做所有我喜欢做的事情。

Finally, thank you **Yanling C** for everything we explored together in the last 10 years… and I want to explore more with you in the rest of my life.
References

intermediate complexes in transcription initiation by RNA
polymerase II. Cell 56: 549-561.
13. Carter AP, Clemons WM, Jr., Brodersen DE, Morgan-Warren RJ,
in individual cells at the single molecule level. Nature 440:
358-362.
15. Yalow RS, Berson SA (1960) Immunoassay of Endogenous
Plasma Insulin in Man. Journal of Clinical Investigation 39:
1157-1175.
16. Coons AH, Creech HJ, Jones RN (1941) Immunological properties
of an antibody containing a fluorescent group. Proceedings of
the Society for Experimental Biology and Medicine 47: 200-
202.
17. Engvall E, Perlmann P (1971) Enzyme-linked immunosorbent
assay (ELISA). Quantitative assay of immunoglobulin G.
Immunochimistry 8: 871-874.
of Proteins from Polyacrylamide Gels to Nitrocellulose Sheets
- Procedure and Some Applications. Proceedings of the
National Academy of Sciences of the United States of America
76: 4350-4354.
Vectors That Display Cloned Antigens on the Virion Surface.
Science 228: 1315-1317.
combinatorial polypeptide libraries. Nat Biotechnol 15: 553-
557.
21. Sanger F, Nicklen S, Coulson AR (1977) DNA Sequencing with
Chain-Terminating Inhibitors. Proceedings of the National
Academy of Sciences of the United States of America 74:
5463-5467.
22. Lander ES, Consortium IHGS, Linton LM, Birren B, Nusbaum C,
et al. (2001) Initial sequencing and analysis of the human
The sequence of the human genome. Science 291: 1304-1351.
24. O'Farrell PH (1975) High resolution two-dimensional


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.