Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1192

Proximity Ligation

Transforming protein analysis into nucleic acid detection through proximity-dependent ligation of DNA sequence tagged protein-binders

BY

SIMON FREDRIKSSON

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2002

Dissertation for the Degree of Doctor of Philosophy (Faculty of Medicine) in Molecular Medicine presented at Uppsala University in 2002

ABSTRACT

Fredriksson, S. 2002. Proximity Ligation. Transforming protein analysis into nucleic acid detection through proximity-dependent ligation of DNA sequence tagged protein-binders. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1192. 34 pp. Uppsala. ISBN 91-554-5426-7

A novel technology for protein detection, proximity ligation, has been developed along with improved methods for *in situ* synthesis of DNA microarrays. Proximity ligation enables a specific and quantitative transformation of proteins present in a sample into nucleic acid sequences. As pairs of so-called proximity probes bind the individual target protein molecules at distinct sites, these reagents are brought in close proximity. The probes consist of a protein specific binding part coupled to an oligonucleotide with either a free 3'- or 5'-end capable of hybridizing to a common connector oligonucleotide. When the probes are in proximity, promoted by target binding, then the DNA strands can be joined by enzymatic ligation. The nucleic acid sequence that is formed can then be amplified and quantitatively detected in a real-time monitored polymerase chain reaction. This convenient assay is simple to perform and allows highly sensitive protein detection. Parallel analysis of multiple proteins by DNA microarray technology is anticipated for proximity ligation and enabled by the information carrying ability of nucleic acids to define the individual proteins. Assays detecting cytokines using SELEX aptamers or antibodies, monoclonal and polyclonal, are presented in the thesis.

Microarrays synthesized *in situ* using photolithographic methods generate impure products due to damaged molecules and interrupted synthesis. Through a molecular inversion mechanism presented here, these impurities may be removed. At the end of synthesis, full-length oligonucleotides receive a functional group that can then be made to react with the solid support forming an arched structure. The 3'-ends of the oligonucleotides are then cleaved, removing the impurities from the support and allowing the liberated 3'-hydroxyl to prime polymerase extension reactions from the inverted oligonucleotides. The effect of having pure oligonucleotides probes compared to ones contaminated with shorter variants was investigated in allele specific hybridization reactions. Pure probes were shown to have greater ability to discriminate between matched and singly mismatched targets at optimal hybridization temperatures.

Keywords: Proximity ligation, proteomics, SELEX, antibody, DNA microarray

Simon Fredriksson, Department of Genetics and Pathology, Section of Molecular Medicine, Rudbecklaboratory, Uppsala University, 75185, Uppsala, Sweden

ISSN 0282-7476
ISBN 91-554-5426-7
Printed by in Sweden by Universitetstryckeriet, Uppsala, 2002

Till,
Far, Mor och Bror

PUBLICATIONS

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

I. Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gústafsdóttir, S. M., Ostman, A., Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol 20(5): 473-477.

II. Gullberg, M.[1], Fredriksson, S. [1], Gústafsdóttir, S., Jarvius, J., Landegren, U. (2002). Cytokine detection by antibody-based proximity ligation. Manuscript.

III. Kwiatkowski, M., Fredriksson, S., Isaksson, A., Nilsson, M., Landegren, U. (1999). Inversion of *in situ* synthesized oligonucleotides: improved reagents for hybridization and primer extension in DNA microarrays. Nucleic Acids Res 27(24): 4710-4714.

IV. Jobs, M. [1], Fredriksson, S. [1], Brookes, A. J., Landegren, U. (2002). Effect of oligonucleotide truncation on single-nucleotide distinction by solid-phase hybridization. Anal Chem 74(1): 199-202.

[1] The first two authors of papers II and IV contributed equally to the work.

Reprints were made with the permission of the publishers.

TABLE OF CONTENTS

ABBREVIATIONS

2-DE, two dimensional electrophoresis
ASR, analyte specific reagent
CDNA, complementary DNA
CV, coefficient of variation
DIGE, difference gel electrophoresis
DNA, deoxyribonucleic acid
ELISA, enzyme linked immuno adsorbent assay
Fc-portion, fragment that crystallizes
FRET, fluorescence resonance energy transfer
HPLC, high performance liquid chromatography
HUGO, Human Genome Organization
HUPO, Human Proteome Organization
ICAT, isotope -coded affinity tags
IgE, immunoglobulin E
IL-2, interleukin-2
IL-4, interleukin-4
LOCI, luminescent oxygen channelling immunoassay
mRNA, messenger RNA
PCR, polymerase chain reaction
PDGF-BB, platelet derived growth factor BB
RCA, rolling circle amplification
RIA, radioimmunoassay
RNA, ribonucleic acid
SELEX, systematic evolution of ligands by exponential enrichment
SPA, scintillation proximity assay
Th, T-helper
TSH, tyrotropin, thyroid stimulating hormone
VEGF, vascular endothelial growth factor

1. INTRODUCTION

Nucleic acids encode the information needed for an organism to produce the protein complements of the genome. The DNA is transcribed into messenger RNA which is then translated into proteins. Sequencing of the 3.2 billion base pairs of the human genome and its ~30000 genes has now been completed[1, 2] and there is a pressing need for technological advances in protein analyses in order to clarify the functions of the encoded proteins on a global scale, and to define their roles in disease. This research has been termed proteomics, being the counterpart of genomics.

Important aspects of proteome analyses will be to measure the amounts, possible modifications, and location and co-location of specific proteins in large numbers of samples. These analyses must be sensitive, specific, applicable in high throughput and suitable for analysis of multiple proteins in parallel from individual samples. The protein detection strategy proposed in this thesis involves the simultaneous analysis of sets of proteins through a process which can be viewed as reverse translation of proteins into nucleic acid. A first step in this direction has been taken through the development of a novel method termed proximity ligation, described in papers I and II. Herein, detection of specific proteins results in the formation of a nucleic acid encoded signal which can be amplified by standard methods of molecular genetics. The sequences obtained through this amplification can potentially be analysed in a highly parallel fashion as nucleic acid products representing the individual proteins. Thereby some of the technological advances in nucleic acid analyses that propelled the sequencing of the human genome can thereby be applied to proteomics.

Papers III and IV are developments within the field of DNA microarray technology for highly parallel nucleic acids sequence analysis. Paper III describes an improved method for *in situ* synthesis of oligonucleotides where truncated products are removed by a molecular inversion procedure, leaving only full-length material with free 3'-ends useful for primer extension reactions. Paper IV investigates the effect of truncated probes in hybridization based sequence distinction on solid phases. By way of introduction, this thesis will also discuss some of the available technologies for protein analyses and different classes of protein specific binding reagents.

2. AVAILABLE TECHNOLOGIES FOR PROTEIN DETECTION

The detection and quantitation of proteins has a long and productive history[3]. The performance of various detection technologies can be evaluated according to many aspects such as sensitivity, limit of quantitation, specificity, precision, and linear range (or dynamic range). Some of these important terms are describe here.

Sensitivity, or limit of detection, is usually determined as the concentration or amount which results in a signal two or three standard deviations greater than the background signal (noise) of several replicates of a sample with no target protein present. Sensitivity can be described in amount (moles) or concentration (molar). Often sensitivities of protein detection assays are described in pg/mL but in this thesis such values have been recalculated to molar concentrations for ease of comparison. The limit of detection may not fall within the linear range of the assay where precise quantitation can be made. This is defined as the limit of quantitation. Signal to noise ratios represent the signal for a given target amount divided by the background signal.

Specificity is the capability of the assay to detect only the presence of the target protein and not to give false positive measurements by cross-reacting with other related or unrelated proteins. The precision of an assay is its ability to distinguish between closely similar concentrations of the target protein and it is often described by the coefficient of variation (%CV, also relative standard deviation) which is a relative measurement of the error in analysis. Other important features of an assay are the time it takes, the cost, suitability for automation, etc.

Many excellent techniques have been developed for detection and quantitation of nucleic acids. Most of them rely on various amplification methods exploiting the capabilities of nucleic acids to encode and allow replication of information. Current, protein detection technologies lack these features. Analyses of gene expression levels are predominantly performed at the mRNA level using techniques such as DNA microarrays or real-time quantitative PCR. These methods can only give an indication of the corresponding protein levels but do not always correlate with mRNA levels[4, 5]. Also, aspects such as secondary modifications and distribution of protein are not represented at the level of RNA. Better tools for proteomic research are needed in order to directly analyse proteins.

Most protein detection techniques involve the detection of binding events between the target protein contained in the sample and specific binding reagents, called analyte specific reagents (ASR) which is discussed in the following section.

2.1. Analyte specific reagents (ASRs)

The ASRs most commonly used in detection assays are mono- or polyclonal antibodies, but other classes of specific binders are being developed such as combinatorially generated proteins selected by phage or ribosome display[6] and nucleic acid based SELEX-aptamers[7, 8]. This section discusses some of the pros and cons of the various classes of ASRs and their uses in protein detection assays.

Two crucial, interrelated features of a binding reagent are its affinity and specificity. The affinity, expressed by the dissociation constant (K_d, unit:M^{-1}), is defined as the concentration of free ASR at half maximal binding of a limited amount of the target. The K_d-value represents the ratio between two rate constants, the dissociation rate constant (k_{diss}, or k_{off}) and the association rate

constant (k_{ass}, or k_{on}). The term association constant (K_a) is also frequently used and is simply the inverse of the dissociation constant. The dissociation rate constant can also be expressed as the half-life of the complex. The affinity of the reagent is very important since the performance of all immunoassays are affinity dependent[9]. The specificity of an ASR refers to its ability to bind only the target protein and not to cross-react with other similar or dissimilar proteins, which can give rise to false positive measurements.

2.1.1. Antibodies

Antibodies are raised against a specific target protein by immunization of a host animal. The antibodies derived from immunization are polyclonal, meaning they represent antibodies produced by many lymphocyte clones, and are specific for different binding sites on the target protein. The immunoglobulin fraction can be purified from serum or antibodies may be affinity purified by selection for binding to immobilized target protein in order to remove most non-specific antibodies. Affinity purified polyclonal antibodies are readily and commercially available. The performance of polyclonal antibodies is batch dependent and such preparations can not be renewed with exact reproducibility. In comparison to nucleic acid based binders (see below) antibodies are not as stable in storage and may not be reversibly denatured.

Monoclonal antibody preparations with specificity for single epitopes can be produced by hybridoma cell culture and they can be distributed as a regeneratable and consistent source[10]. The affinity of antibodies are said to reach an apparent affinity ceiling during in vivo maturation by hypermutation[11] to a $K_d > 0.1$ nM. This may reflect that the B cell receptor internalization rate is too rapid to permit discrimination of antibodies with a further reduction in dissociation kinetics[12, 13]. This does not mean that antibodies can not have higher affinity than 0.1 nM but that in vivo selection usually is not sufficient. Antibody libraries can also be displayed on the surface of phages[14], yeast[15], or through ribosome display[16] and then selected in vitro to yield affinities in the low picomolar range, K_d ~20 pM. Libraries may also be mutagenized prior to the selection process[17] and single-chain antibody fragments have reached K_d's of 48 fM and dissociation kinetics with a half-life greater than 5 days[15]. The increase in affinity usually derives from decreased dissociation rate constant while on rates are fairly constant.

2.1.2. Random libraries

Protein based binders can be selected from random libraries for affinity to a given target. These naïve libraries can be screened using various types of display technologies which link phenotype with genotype[6]. The success of such selections is often dependent on the size of the library. Naïve phage display libraries are limited in size and complexity by the bacterial transfection step ($<10^9$) while *in vitro* ribosome display and mRNA display technologies are capable of handling greater complexity. These display technologies are faster in generating analyte specific reagents compared to monoclonal antibody production by mouse immunization. In the proteomics area the ease of obtaining ASRs of high quality will be essential.

As an example of the use of random libraries, Wilson *et al* used an mRNA display technique based on the covalent fusion between *in vitro* translated peptids and their cognate mRNAs via a puromycin residue linked to the 3´end of the randomized mRNAs[18, 19]. The starting library consisted of about 10^{13} random peptides, each 88 amino acids in length, which where selected for binding to streptavidin. Peptide sequences of 5 nM K_d affinity were found; compared to

a previous experiment using phage display with a library of lower complexity these mRNA display selected peptides bound with three to four orders of magnitude higher affinity. The authors also anticipate that their selection technique will be able to find sub-nanomolar affinity reagents.

Another selection strategy is the use a starting material composed of a naturally occurring polypeptide scaffold with a structured fold, reviewed in[20]. Many different types of protein scaffolds have been mutagenenized and selected for binding, including ones based on lipocalin[21], the Z-domain of protein A (affibodies)[22], and others.

2.1.3. SELEX-aptamers

Generally, short single stranded nucleic acids are thought of as linear molecules with no specific secondary structure. However, certain sequences can fold into rigid structures through interchain hybrid segments, resulting in an overall stable tertiary structure. Large libraries of randomized nucleic acid sequences of RNA or DNA can be screened for affinity towards a specific target molecule and used in assays. The process has been termed SELEX (Systematic evolution of ligands by exponential enrichment) and the functional sequences found are called aptamers, from the Latin *aptus*, to fit[7, 8]. These aptamers can then bind to a given target protein or a small molecule with specificity and high affinity [23, 24]. Aptamers can also be selected to catalyse enzymatic reactions.

Typically, the library is synthesized using nucleic acid chemistry with a randomized sequence flanked by fixed sequences used for amplification by PCR. The library is iteratively selected for binding to the immobilised target protein and non-bound sequences are removed by washing or e.g. nitrocellulose filtration. Bound sequences are then amplified by PCR and the product is made single stranded and reselected for binding to the target in a process repeated about 8 to 15 times. The final selections are cloned and sequenced and a family of similar nucleic acid structures are usually found, figure A. These are then synthesised and performance tested for affinity and specificity. The sequences are also optimized by truncating the molecules into minimal high-affinity ligands. Aptamers are often modified for increased resistance to nucleases[25], a process which can also be incorporated into the selection process by using nuclease resistant nucleotides such as 2'-modified RNA-triphophates[26, 27].

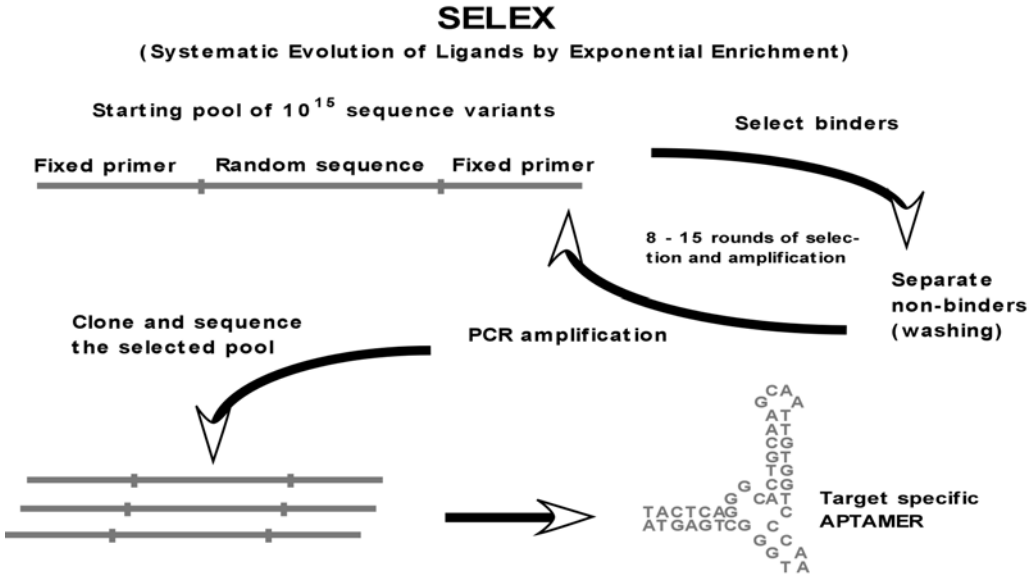

Figure A. General outline of a SELEX protocol.

These reagents hold great promise for analytical[28, 29] as well as therapeutic applications[30]. Since they consist of nucleic acid of between 25 and 40 nucleotides of single stranded DNA or RNA, they posses a number of important features. They can be readily produced through synthesis chemistry, making them a consitent and renewable source, and the conjugation of reporter groups such as fluorophores is trivial. They are also very stable in storage and can be repeatedly denatured and renatured, a feature not common for proteins including antibodies. The selection process is done entirely *in vitro* with no need for animal or cellular hosts. Immunization in vivo can be problematic for toxic proteins and for highly conserved proteins. As an aptamer can be readily synthesized they are electronically transferable reagents well suited as a general and easily accessible resource stored in as information in a database, which eliminates the need to share material between labs.

Aptamers have been used in a number of standard assay formats in place of antibodies. For example in sandwich type detection assays[31], histochemistry[32], flow cytometry[33], fluorescence polarization[34], and biosensors[35]. Aptamers coupled with fluorophores can monitor the presence of the target analyte in homogenous assays by changes in fluorescence intensity[36, 37]. Aptamers have also be constructed as molecular switches, also called allosteric ribozymes. These “aptazymes” are composed of a catalytic aptamer component fused to an activator domain which upon binding a cofactor enhances the catalytic property of the aptazyme[38, 39]. Selex aptamers are further discussed in section 2.3.1.3 regarding protein arrays (photo-aptamers) and in paper I.

2.2. Detection of single proteins

2.2.1. Solid phase heterogeneous assays

One of the first assays using a specific binding reagent was a radioimmunoassay (RIA) for the detection of insulin[40]. The competitive RIA technology is based on the competition of binding between variable amounts of the target protein in the sample and a fixed amount of a radiolabeled competitor to the antibody. The amount of bound labeled protein, after appropriate separation of non-bound sample and competitor, correlates inversely to the target amount present in the sample. Detection limits in this type of assay depend on the specific activity of the label and the affinity of the immobilised antibody used[3] and sensitivities in the pico molar range have been reported[41]. This assay is particularly useful for targets of low molecular weight that cannot bind two antibodies, see below, and uses the fact that radiolabelling of low molecular weight competitors does not alter the specific binding interaction between target and antibody.

Rather than using a labeled competitor, Wide *et al* [42] and Miles and Hales[43] initiated the use of labeled detection antibodies in the late 1960's, and they later also introduced the use of dual specific sandwich type assays[44, 45]. This two-site sandwich assay is the most commonly used detection method today. A first antibody is immobilized on a solid support which captures the target protein and a second antibody is subsequently added containing a detection label measuring the amount of protein. The requirement for two specific binding events to the target protein at separate epitopes in order to generate a signal increases the specificity of the assay. The sensitivity and linear range is increased compared to the single antibody-based competitive immunoassay. The sandwich assay requires a number of washing steps in order to remove the nonbound secondary antibody. The background signal generated in this assay along with the strength of the signal itself sets the limit of detection. Much effort has been devoted to the development of stronger signals to permit sensitive detection. Examples of reporters include enzyme labels (ELISA, enzyme linked immuno adsorbent assay[46]) and fluorescent labels[47, 48]. Frequently used enzyme labels are horseradish peroxidase or alkaline phosphatase, which can catalyze reactions that convert substrates into products detectable with spectroscopy. These enzymatic detection systems amplify the signal through continuous catalytic conversion over a period of time.

However, with very strong signals from the label, the nonspecific binding of labeled secondary antibody to the solid support will give background problems that must be minimized by extensive washing[3]. Probably the strongest signal available is achieved through the immuno-PCR technique which involves a PCR template coupled to the secondary antibody[49]. After extensive washing this template is amplified by PCR and detected and quantified in gels or using real-time PCR[50]. Remarkable sensitivities have been reported using this technique with detection of around 1000 molecules of the antigen[49, 51, 52]. Others report sensitivities of 150,000 [50]and $12x10^6$ molecules[53].

2.2.2. Liquid phase homogenous assays

Assays that do not require a solid phase or washing steps are said to be homogenous. These assays are simple and quick to perform and they are more easily automated, which is especially advantageous for high throughput applications. There are several strategies for homogenous assays for single targets but most of them exhibit low sensitivity. One assay based on fluorescence resonance energy transfer (FRET) exhibited detection levels around 100 pM of the antigen[54]. The

scintillation proximity assay (SPA) is widely used for high throughput screening and has been reported to detect around 20 pM of human serum albumin[55]. A fluorescence polarisation assay specific for PDGF-BB using aptamers as affinity reagent can detect 220 pM of the protein[34]. The luminescent oxygen channelling immunoassay (LOCI) is also used for high throughput screens and is more sensitive than FRET and SPA. Two different latex particles are used with binding ligands at their surface which aggregate the particles in the presence of the target. The first particle has a photosensitizer which generates a singlet oxygen molecule upon illumination. This singlet oxygen only travels a short distance in the solution and if the second latex particle is within close proximity its chemiluminescent dye generates light. A LOCI based thyrotropin (TSH) assay has been capable of detecting down to 4 amoles or 90 fM of the target in serum[56]. LOCI has also been miniaturized for use with small volume samples with high sensitivity, 2 pM of TSH in 10 nL[57].

2.3. Detection of multiple proteins

In the age of proteomics, simultaneous analyses of several target proteins will be essential. This is of vital importance since proteins do not act alone but in concert. It is crucial to capture as much information as possible especially when using precious non-renewable samples from biobanks. Two general strategies can be envisioned for acquiring large amounts of information from a sample; detecting multiple proteins in a single assay (parallel multiplexing), or splitting the sample into minute aliquots and analysing them one by one for single proteins using sensitive technology (serial multiplexing).

2.3.1. Protein microarrays

The concept of protein microarrays is a promising tool for analyses of several proteins simultaneously in individual assays[44]. The technology is similar to DNA microarrays where many spatially separated features represent distinct analyses at each position over a small surface. The individual targets are captured onto their specific feature and the amounts are most commonly recorded as fluorescence, by scanning with a laser to illuminate a reporter dye. Other detection methods include chemiluminescence, mass spectrometry, radioactivity, or electrochemistry, reviewed by Templin *et al* [58].

2.3.1.1. *Fluorescently labeled sample*

RNA expression measurements are frequently conducted by labeling the cDNA complement and analyzing these on a DNA microarray. In a similar detection strategy for proteins is to directly label the proteins in the samples with a fluorophore and use an immobilized capture antibody[59]. In this manner, two samples can be differentially labeled and analyzed on the same chip to determine relative protein expression levels. This only requires one specific antibody deposited on the array. Drawbacks with such a scheme are evident including the limited specificity through the use of single antibody binding, and the covalent modification of the target proteins that can influence their binding interaction with the antibody. In one study, only 20% of 115 arrayed antibodies showed specific and accurate performance for concentrations down to nM[59].

2.3.1.2. *Sandwich type assays*

By using two specific antibodies for each target, individual features may be considered miniature sandwich immunoassays where the samples need not be labeled. The signals in protein microarrays must remain localized in order to maintain the spatial separation. The secondary antibodies can for example be fluorescence labeled[60, 61], or oligonucleotide labeled and amplified through a localized rolling circle amplification[62], or enzymatically labeled where the product is locally deposited such as tyramide amplification. An antibody array using tyramide amplification has been marketed by Perkin-Elmer Life Sciences.

Robinson *et al* spotted 196 distinct biomolecules for the detection of autoantigens with a sensitivity of about 6 pM [61]. The sandwich type array was capable of distinguishing autoimmune diseases when analysing patient sera. However, some of the biomolecules they spotted, for example histones, lost their three dimensional conformation and could not be used in the assay. Compared to DNA microarrays, the attachment of proteins to surfaces is often difficult and will at times lead to inactivating conformational changes, a major challenge for protein microarrays[63].

Pawlak *et al*. immobilised antibodies for a sandwich assay of 8 cytokines with a sensitivity of about 0.3 pM and without any noticeable cross-reactivity[60]. The antibody pairs were chosen with great care to avoid cross-reactivity and the authors anticipate that an array of 10-20 specificities would be feasible. Cross-reactivity is further discussed below.

Schweitzer *et al.* designed an amplifiable and localised signal for protein microarrays using a universal detection antibody specific for the Fc-portion of the secondary antibody of the sandwich[62]. This detection antibody was coupled to an oligonucleotide with free 3'-end capable of initiating a localized rolling circle amplification (RCA)[64]. The amplification produces a concatamer of single stranded DNA, detected by hybridization of fluorescence labeled oligonucleotides. They compared the RCA amplification with direct fluorescence detection and reported a 1000-fold improvement. Their 75-plex cytokine assay displayed variable sensitivities for individual targets ranging from about 0.03 pM to 70 pM with a median sensitivity of 0.7 pM, roughly equal to the standard ELISAs using the same antibodies. If instead individual ELISAs had been performed for all these targets, 1000-fold more sample would have been consumed, illustrating the power of protein microarrays.

The issue of cross-reactivity is of vital importance as the number of analytes increases. The dual specific sandwich immunoassay of single proteins is very selective since it requires two binding events. But in an array format using many secondary antibodies the likelihood of cross-reactivity and unspecific signals will increase with the complexity of the array and the problem will be even greater since analytes typically vary over large concentration spans. In a multiplexed sandwich array, any single nonspecific binding event an antibody and a target will amount to a nonspecific signal.

2.3.1.3. *Photo-aptamers*

Another approach to multiplex protein detection on arrays involves the use of photo-SELEX aptamers[65, 66]. These analyte-specific reagents are composed of nucleic acids which makes their spotting and stability similar to ordinary DNA microarrays i.e. with better preservation of function than protein based reagents. The photo-aptamers are selected for specific binding and also for uv-crosslinking to the target protein in an automatable selection process producing reagents with high affinity and specificity. The uv-crosslinking step adds an additional degree of specificity since it is dependent on the exact orientation of the aptamer-target binding for efficient crosslinking. Non-specifically bound proteins are less likely to have the required proximity between the bromodeoxyuridine nucleotide of the aptamer and a crosslinkable amino acid of the protein. After covalent crosslinking the array is extensively washed removing all non-specifically and non-covalently bound proteins. Since the capture reagents are composed of DNA the array can be stained with a protein-specific reagent for quantitative fluorescence detection. The performance of photo-aptamer arrays has not yet been reported in the literature but the feasibility was demonstrated at a recent seminar[67].

2.3.2. Fluorescent microbead and cytometric assays

Fluorescent microbeads and flow cytometry has also been used for multiplex protein detection, reviewed by Vignali[68]. Microbead particles can be distinguished according to their distinct proportions of red and orange dyes, enabling the separate detection of at least 64 different particles (Luminex). Each particle type is then converted into a protein specific miniature sandwich immunoassay with a common green detection fluorescence. The detection is usually carried out by fluorescence labeled streptavidin binding to the biotinylated secondary antibody in the sandwich. A 15-plex cytokine assay with high sensitivity has been reported capable of detecting down to ~0.5 pM protein in 100 μL[68]. Another 6-plex assay for the cytokines representing Th1/Th2 response is commercially available with equal sensitivity(BD-Biosciences).

2.3.3. 2-D gels and mass specrometry

Two dimensional gel electrophoresis and mass spectrometry has become a standard technology of proteomics endeavours and is currently run in several labs. Protein samples are run in an isoelectric focusing gel in the first dimension, then size separated on a SDS-PAGE gel in the second dimension. The gel is then stained for protein spots which are picked out by robotics and digested into peptide fragments by proteases such as trypsin. These peptide mixtures are analysed by mass spectrometry and the fragment sizes of the peptides are profiled against databases for identification of specific proteins. The mass spectrometer analyses the peptide fragments on a mass/charge basis.

This proteomics strategy has advantages over protein detection using ASRs. First no ASRs are needed and “all” proteins in the sample can be analsed without any predetermined bias for a certain subset of proteins. This means that one can find variation in expression levels of specific protein, and only later indentify them via mass spectrometry. Using mass spectrometry, small changes in molecular mass of the peptide fragments can be detected and identified as post-translational modifications such as phosphorylation etc.

The drawbacks of 2-DE gels are many. They are very laborious to run and require a high degree of expertise to obtain consistent results and they severely lack sensitivity. Gygi *et al* showed that low- and mid-abundance proteins are not readily detected by this technology, rendering about

half of the yeast proteome invisible[69]. These low abundance proteins are often thought to be the most interesting and those responsive to physiological stimuli such as transcription factors and protein kinases[70, 71]. Mass spectrometry can detect zeptomole amounts of sample but the sensitvity of the combined technology including the gel separation is limited by spot identification in 2-DE with sensitivity limits in the low femtomole level[72, 73]. Also, hydrophobic and large proteins such as transmembrane receptors do not enter the second dimension of the gel[74]. When analysing serum and cerebrospinal fluid samples, these contain 99 % albumin and globulins which mask the more interesting low abundance proteins[74]. Purification strategies are used to remove these high abundance proteins prior to sample application on the gel.

Quantitation is also difficult with these techniques. Quantitative comparisons can be made of two samples run on two separate gels but is difficult due to the inconsistency of the gel runs and the intensity of a peptide peak in a mass spectrum is not quantitative[74]. However, by differential labeling fo the two samples and running them on the same gel the problem can be reduced. This type of differential labeleing can be done with isotopic methods which give small shifts in mass/charge ratios, ICAT[70]. For differential protein expression without mass spectrometry, samples can be fluorescence labeled with different dyes (Cy3, Cy5) and run on the same gel, so called difference gel electrophoresis (DIGE)[75].

3. PRESENT INVESTIGATION

The aim of the present investigation was to develop a methodology which would enable highly sensitive and parallel analysis of protein concentrations using nucleic acid based technologies. Through quantitative conversion of the presence of a protein into an amplifiable nucleic acid, using a pair of ASRs, a sensitive and highly specific assay termed proximity ligation was developed. Conversion of specific proteins into DNA can be viewed as an artificial reverse translation where individual proteins can be identification encoded by taking advantage of the information carrying abilities of nucleic acids.

3.1. PAPER I, Protein detection using proximity-dependent DNA ligation assays.

Paper I describes the basic properties of proximity ligation. The target protein is analysed using two so called proximity probes each with a target specific binding moiety (ASR) and a reactive DNA oligonucleotide component. Binding by two proximity probes to a target molecule at different sites brings the reactive oligonucleotides into close proximity in the sample solution. After the incubation has reached binding equilibrium, a mixture containing all components for a proximity dependent reaction between the oligonucleotides and detection of this reaction is added. This mixture contains a large excess of a connector oligonucleotide, or splint, capable of hybridizing to the 3´-end of probe #1 and the 5´-end of probe #2 thus forming a substrate for a ligation reaction uniting the two probes and thereby forming a new nucleic acid sequence. Two proximity probes which are close to each other will hybridize to one and the same connector forming the ligation substrate and the assay signal. Probes which are not in proximity will hybridize to one connector each rendering them unable to undergo ligation, figure B. A high degree of specificity is ensured by the requirement of dual and proximate recognition of the target protein.

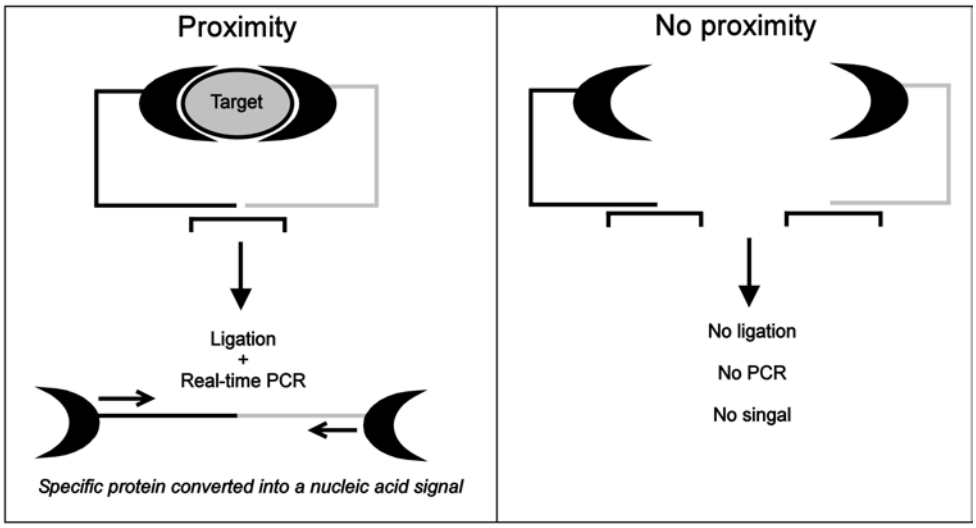

Figure B. The principle of proximity ligation. Probes which have bound their target protein gain a large relative increase in concentration which enables one connector oligonucleotide, added in great molar excess, to hybridize to two probes generating a ligation product (Proximity panel). This united DNA sequence is specifically amplified and detected by real-time PCR. Non-bound probes which are not in proximity will receive one connector each rendering them unable to undergo ligation. (No proximity panel).

The ligation reaction is detected by real-time PCR[76, 77] where the primers are placed on one probe each making the amplification specific for the ligation product. Real time PCR uses the 5´exonuclease activity of the DNA polymerase to degrade a dual labeled probe which hybridizes to the amplification product. Upon degradation, the reporter fluorophore is cleaved from the quencher fluorophore which decreases the fluorescence resonance energy transfer between the two labels. This increases the emission of the reporter which is monitored over the entire PCR in a fluorometric PCR instrument. Real-time PCR technology is very sensitive due to the exponential amplification of PCR capable of detecting fewer than 10 amplicons and it has a wide (10^7–fold) linear range.

The binding moieties of the proximity probes used in paper I were SELEX aptamers. Two homogenous assays were designed for the detection of the homodimeric platelet derived growth factor (PDGF-BB) and the monomeric alpha-thrombin. The PDGF-BB assay displayed a very high sensitivity capable of detecting as few as 24 000 molecules of the protein. This is a 1000-fold greater sensitivity compared to a commercially available ELISA for the same protein, figure 3. The assay performed equally well in various biological fluids such as fetal calf serum, cerebrospinal fluid, and Eagle's minimal growth medium for cell culture, figure 4.

There are some important parameters for the assay which were optimized for high signal to noise ratio at a fixed amount of target. First, in figure 2b various incubation volumes and proximity probe concentrations were tested. All reactions were ligated with a mix resulting in a 50 μL PCR. With lower volumes, the sensitivity increases due to fewer background ligation events and the signal does not increase with higher volumes since the probe concentration determines the degree of target binding by the proximity probes. Optimal proximity probe concentrations were also tested and is discussed in section 3.2, paper II.

The connector oligonucleotide was designed not to give rise to any amplification products in the absence of the ligase enzyme. This was accomplished by using only ten hybridizing nucleotides at each end of the proximity probes. With more complementarity, the DNA polymerase can extend across the ligation junction giving rise to a ligation independent product (data not shown). In figure 2a, the connector concentration was varied along with the ligation time. At high connector concentration the ligation is completed quickly both for signal and background and does not increase over time since nearly all probe ends rapidly boind one connector each. High connector concentration is beneficial since probes not in proximity will bind one connector each rendering them unable to undergo ligation. We refer to this as the splint-effect, illustrated in figure B. At lower connector concentrations the reaction increases over time in both signal and background. This is probably due to the inefficient and slow hybridisation to proximity probe ends at lower concentrations.

An experiment was conducted to demonstrate the advantages of dual and proximate recognition when applied to solid phase assays, figure 6 paper I. As described above in section 2.2.1, the background of sandwich type solid phase assays is often limited by the non-specific binding of the secondary reporter reagent. A first polyclonal antibody specific for PDGF-BB was immobilised in a reaction vessel. The sample was then applied and detected either by one PDGF-BB specific aptamer carrying a directly PCR amplifiable tag (immuno-PCR approach), or by using two ligatable proximity probes. Non-specific binding to the reaction vessel is as likely in both cases but the probability of two proximity probes binding the surface sufficiently close to one another to generate a ligation product is very low reducing the assay background significantly and increasing

sensitivity. This version of proximity ligation also adds an extra element of specificity by requiring three binding events. Larger sample volumes can be analysed this way than in the homogenous assay making the assay more sensitive with respect to the concentration of target analysed, compared to the homogenous format.

3.1.1. Perspectives on paper I

The SELEX DNA-aptamers are ideally suited as analyte-specific binding moieties of the proximity probes. The attachment of the reactive nucleic acid component is trivial as the entire probe is composed of single stranded DNA. Figure C illustrates PDGF-BB bound by two aptamer-based proximity probes. These reagents are also very stable in storage and they can be shared between labs, allowing standardization of protein detection assays. In comparison, the performance of polyclonal antibodies depends on batches used. The activity of antibodies can frequently deteriorate upon storage or reuse whereas aptamers can be frequently denatured and renatured. However, only a small number of aptamers are currently available. About 100 aptamers with good specificity and affinity have been reported in the literature[65]. To my knowledge PDGF-BB and alpha-thrombin are currently the only proteins to which two aptamers can bind simultaneously with high affinity at two separate sites.

To increase throughput, the selection of aptamers has been automated by several groups using robotics[78-82]. These automated selection strategies have been devised to select only one aptamer per target and the family of aptamer sequences found after a SELEX usually all bind the same epitope on the target. There would be a need to design SELEX protocols for selecting pairs of aptamers for each target for conversion into proximity probes, a strategy which should be feasible (L. Gold personal communication).

Since the sequence generated after ligation of the two proximity probes can be chosen at will, it is possible to encode sequence motifs representing specific proteins for detecting many proteins in parallel. Quantitative and qualitative analysis of such sequences can be performed using several available techniques. For example, during the amplification variously labeled molecular beacons can monitor the amplifications in real-time. Four amplicons have been simultaneously analysed using molecular beacons and possibly six- or seven-plex analyses are anticipated[83].

End-point measurements of multiple amplification products by using DNA microarrays could greatly increase the levels of multiplexing. This identification-tagging strategy, also called molecular bar-coding[84, 85] or zip codes, has been frequently employed using many molecular strategies and arrays of sequence tags. For example; minisequencing using arrays[86-88] or fluorescent microbeads[89] ; RNA templated oligonucleotide ligations analysed on fiberoptic DNA microarrays[90] ; and multiplexed padlock probe ligations (J. Banér et al, submitted). Over 1000 parallel padlock probe ligations have been performed in parallel and analysed using tag-microarrays (T. Willis et al, CSHL Genome Sequencing and Biology meeting 2002). Recently, 5916 bar-coded yeast deletion strains were simultaneously analysed for growth fitness using a generic DNA microarray[91]. Generic tag arrays can thus create a universal and highly multiplexed analysis platform for DNA polymorphisms, RNA quantitation, and proteins using bar-coded proximity ligations.

Multiplexed proximity ligations offer some advantages over protein microarrays in regard to the issue of cross-reactivity, since each proximity probe can be designed to ligate only to its partner through unique sequence compositions in the ligation site for each proximity probe pair. Single

cross-reactive binding events between target and ASR will thereby not generate nonspecific signals as in protein microarrays, discussed in section 2.3.1.2.

Using sandwich type protein microarrays, the relative quantitation of two samples must be done by comparing two array readouts. However, with proximity ligation, two samples can be amplified and applied to the same array and read relative to one another through the use of two distinct fluorophores, a technique often used for RNA expression analyses.

As noticed by Yeakley *et al*, a certain problem becomes pertinent when using multiplexed PCR with universal primers and tags[90]. If the mixture of tag-sequences contains both high and low abundance amplicons then these will reach the product inhibited plateau phase of PCR at the same time, due to the fact that the various amplicons can re-anneal to each other and stop amplifying since they contain the same primer sequences. As a result, low copy number sequences may not have reached a level of amplification where they are detectable. Yeakley *et al.* suggest that generally low abundance genes and high abundance genes should be separately amplified and then pooled before application to the microarray. This poses a potential problem since exact expression levels are unknown prior to the experiment.

An alternative means of amplifying DNA with high quantitative precision which is not product inhibited is through rolling circle amplification[92, 93]. Proximity ligation reactions can be designed to generate circular amplicons in proportions to the target protein (S. Gustafsdottir *et al.* unpublished and M. Gullberg *et al.*, unpublished). Rolling circle replication can also be performed in consecutive reactions in order to give the amplification power required by proximity ligation and also in multiplex using tags (F. Dahl *et al* manuscript in preparation). Rolling circle amplification combined with proximity ligation can also provide local information on where a certain protein, or protein complex, is residing in a cell (M. Gullberg *et al*, unpublished).

Proximity ligation offers sensitive detection of proteins in regards to small sample amounts. This is of importance when samples are limited such as biobank samples, small samples from laboratory animals, and for serial multiplexing. The technology does not only enable the detection of proteins but also of protein complexes where the proximity probe pair has specificity for two individual proteins that may or may not be in complex. Which also enables the screening of substances capable of disrupting specific protein-protein interactions.

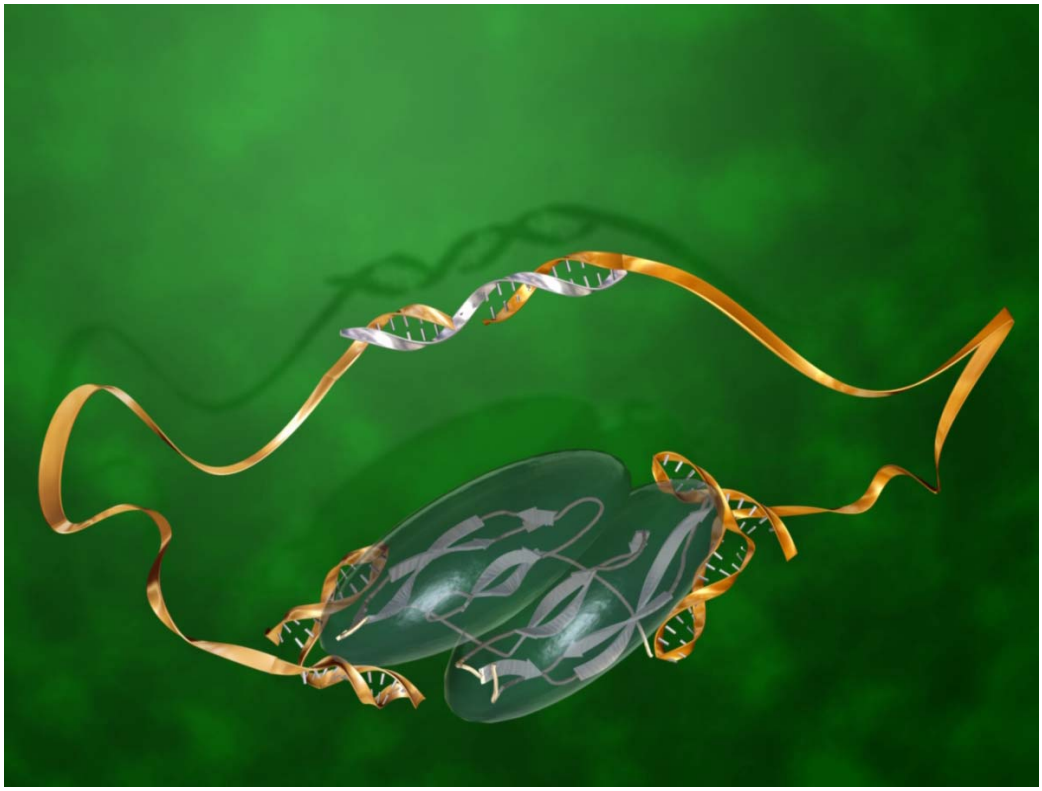

Figure C. PDGF-BB bound by two aptamer-based proximity probes hybridized to a connector oligonucleotide, illustrated by Jonas Jarvius.

3.2. PAPER II, Cytokine detection by antibody-based proximity ligation.

Paper II discusses in greater depth the theoretical foundations for proximity ligation and reports data on various types of binders that can be converted into proximity probes by coupling of ligatable oligonucleotides. Antibodies are shown to be suitable reagents for proximity ligation. The paper also attempts to define what concentrations of the proximity probes which should be used in order to keep the background ligations to a minimum. With increasing probe concentrations the potential for target-independent ligation of probes increases as the square of the concentration increase, table 1. At higher probe concentrations the probability of two proximity probes being in proximity as a random event in the solution increases. The optimal concentrations of proximity probes was first investigated in paper I, figure 2b. We found that the concentration should be kept at around 20 pM in order to give a stable but low background signal. At 20 pM about 1000 background ligations arise which are quantified with high precision using real-time PCR. This concentration is well below the dissociation constant (K_d) of typical proximity probe/target interactions (~150 pM for PDGF-BB detection paper I). As a consequence only a few targets will have bound one proximity probe and the square of this probability of binding is the probability of a target being bound by two proximity probes, given that they have the same affinity for the target. As a result, only 1 out of 25 PDGF molecules were detected in experiments reported in paper I. In paper II we also describe a model for the performance of proximity ligation as a function of the affinity of the binders. Using standard affinity calculations we found that a concentration of 20 pM proximity probes is optimal for most reagents regardless of their affinity, and that the sensitivity of the assay is directly related to the affinity of the reagents used. This is also true for the case when individual probes in a pair have different affinities for the target, table 2 paper II.

In figure 2, we have made theoretical standard curves for proximity ligation reactions using various dissociation constants for the interactions (dashed lines), and compared those with experimental results for PDGF-BB and thrombin from paper I, as well as a monoclonal antibody-based insulin assay (see below). The performance of these three proximity probe pairs correlated well with the expected results in regards to the previously reported affinities.

In this paper we also investigate the performance of various antibodies as affinity reagents for proximity ligation and describe two strategies for coupling the antibody with the oligonucleotide. A matched monoclonal antibody pair with distinct binding sites on insulin were converted into a proximity probe pair using SMPB (succinimidyl 4-[p-maleimidophenyl]butyrate) chemistry. The antibody was derivatized with a maleimide function and then coupled to either a 3'- or 5'-thiol-modified oligonucleotide, resulting in a covalent thioether linkage. These conjugations were then purified and used as proximity probes for insulin detection displaying a sensitivity about 30 pM in a 1 μL sample.

Purification of the proximity probes is important for the performance of proximity ligation since free antibodies can occupy target binding sites without contributing to the signal, and free oligonucleotides contribute to the background. The negative effect of binding reagent impurities is greater when high affinity reagents are used compared to low or moderate affinity binders, as shown in table 3. Purification of the aptamer-based proximity probes in paper I was simple since they consist entirely of DNA. Two oligonucleotide parts were preparatively ligated together to ensure a long high quality nucleic acid of about 80 nucleotides. The ligation products were purified by a preparative denaturing PAGE, yielding sufficient product for about one million homogenous

assays. If the proximity probe is synthesized in one piece, a typical yield of a 0.2 μmol scale synthesis would give product sufficient for 200 million assays.

Another strategy for coupling oligonucleotides to antibodies presented in the paper is through a streptavidin-biotin linkage. Biotinylated antibodies were incubated with a covalent conjugate between streptavidin and oligonucleotide. This creates a very stable, yet non-covalent, linkage between the antibody and the oligonucleotide. At certain molar ratios of antibody and streptavidin-oligonucleotide conjugate supramolecular complexes can form[51]. We took advantage of this and purified the complexes using a simple procedure of passing them through a gel filtration spin column with a high molecular weight cut-off, thereby removing both free antibodies and free oligonucleotides.

The streptavidin conjugated oligonucleotides can be used as universal reagents capable of converting any matched pair of biotinylated monoclonal antibodies or a biotinylated polyclonal antibodies into a proximity probe pairs by simply incubating the antibodies with the streptavidin-oligonucleotide conjugates, followed by a simple separation step.

Three polyclonal antibodies were converted into proximity probes, anti-IL-2, anti-IL-4 and anti-VEGF. Polyclonal antibodies are readily available affinity reagents and they are typically specific for several epitopes on the target protein, enabling detection through proximity ligation. In a homogenous format, the VEGF assay displayed sufficient sensitivity to detect ~0.5 pM targetin 1 μL sample, while the IL-2 and IL-4 assays detected at least 3 pM target, figure 3. The VEGF assay performed equally well in the presence of 100% fetal calf serum and levels of VEGF could be measured in serum and plasma by reference to a standard curve.

3.2.1. Perspectives on paper II

Paper II describes the affinity-dependence on the sensitivity of homogenous proximity ligation assays. This renders the assay also capable of estimating the affinity between two proteins or other molecules. It should also be feasible to measure association rate constants by varying incubation times between probes and sample. As the assay is affinity dependent, the performance of assays can be estimated on the basis of the properties of new reagents during assay setup. However, the attachment of the oligonucleotides will probably at times affect the affinity of the binding reagents.

If there is an affinity ceiling for *in vivo* matured antibodies, as discussed in section 2.1.1, then no standard homogenous proximity ligation assay using a polyclonal antibody will reach a sensitivity higher than that achieved with 0.1 nM K_d reagents, around 20 000 molecules or 0.03 pM in 1 μL. This degree of sensitivity should be sufficient for many protein detection applications.

Antibodies, antibody fragments, and also aptamers, can have higher affinity than 0.1 nM K_d through *in vitro* selection and maturation processes (se section 2.1). Theoretically, a proximity probe pair with an affinity of 0.01 nM K_d each would be expected to result in a 100-fold increase in sensitivity, compared to 0.1 nM K_d, thereby as little as 200 molecules could be detected, if less than 20 pM of the proximity probes are used in order to limit the number of background ligations.

Using the same basic principles of blocking the target sites by excess unconjugated antibodies described in paper II, proximity ligation can also be used for qualitative screening of molecules capable of disrupting molecular interactions (Fredriksson *et al.* unpublished).

3.3. PAPER III, Inversion of *in situ* synthesized oligonucleotides: improved reagents for hybridization and primer extension in DNA microarrays.

Paper III involves a novel chemical strategy for *in situ* synthesis of high quality and high complexity DNA microarrays.

3.3.1. Background on DNA microarrays

DNA microarrays are composed of many distinct and specific DNA sequences (features) spatially separated on a small surface and used for many types of genetic analyses. There are two general ways of making DNA microarrays. Oligonucleotides are either presynthesized and then deposited by spotting on a planar glass surface and covalently coupled, or oligonucleotides may be constructed by direct synthesis on the surface (*in situ* synthesis), both reviewed by Pirrung[94]. The quality of the oligonucleotides can be well controlled using the spotting technique, since they can be purified after synthesis, and high yield chemistries can be used during synthesis. The actual spotting also gives an element of purification since only oligonucleotides with complete 5'-ends will attach to the solid support through a 5'-coupling group (for example a 5'-amine, reactive with an aldehyde derivatised surface). However, in very high density arrays the task of presynthesis and purification of many oligonucleotides becomes laborious and difficult and typically such arrays are of larger dimensions compared to ones obtained through *in situ* synthesis.

In situ synthesized microarrays of high complexity are usually manufactured using photolabile protected phosphoramidites and photolithographic masks that spatially direct the incorporation of nucleotides. For each cycle of synthesis the protecting groups are removed using light directed through the masks for all those features that are to receive a particular nucleotide at that position in the strand. One problem associated with the photochemistry used for *in situ* synthesis is the high degree of impure products that result from incomplete photo-deprotection. These impurities arise as truncated oligonucleotides.

3.3.2. Results of Paper III

Paper III describes a novel technology for removing these impurities of *in situ* synthesis by secondarily attaching the 5'-end of full-length products to the solid support after synthesis, thereby forming looped structures. Next, the 3'-end of all oligonucleotides is cleaved releasing all non full-length products while leaving the pure oligonucleotides with free 3'hydroxyls. The solid support was first simultaneously derivatized with an oligonucleotide building moiety and a capture moiety, figure 2. The building group was selectively deprotected first and after oligonucleotide synthesis and phosphorylation, the levulinyl protecting group of the capturing moiety was removed and allowed to react with the 5'-end of the oligonucleotide resulting in the formation of looped structures. The 3'-end of the oligonucleotide was then specifically cleaved by ammonia treatment. Various proportions of the building and capturing moieties were investigated for yield and inversion efficiency as analyzed by capillary electrophoresis, figure 3. The increased purity of the inverted product was also shown, figure 4.

The *in situ* inversion results in pure oligonucleotides that expose free 3'-ends. The oligonucleotides can therefore be used as primers for DNA polymerase catalysed reactions. This was demonstrated by performing minisequencing[95] and pyrosequencing[96] with the support-bound

oligonucleotides, figure 5. Figure D shows a graphical illustration of the molecular inversion reaction.

3.3.3. Perspectives on paper III

When sorting many target nucleic acids on a DNA microarray the discrimination achieved between differentially stable DNA-hybrids is not as great as when also including an enzymatic steps such as polymerisation or ligation. For example, when using tag-arrays for sorting many target sequences, false crosshybridizations are not as likely when all targets are in roughly the same concentration. This is the case in genetic analyses of single nucleotide polymorphims. However, with transcript analyses of mRNA or protein (through tagged proximity ligation) the various tags will differ greatly in concentration and the risk for mismatched crosshybridization will increase. The quantitation of rare transcripts or proteins could become overshadowed by abundant ones at high levels of multiplexing. By having a liberated 3'-hydroxyl from an *in situ* inversion, or a spotted oligonucleotide, a polymerisation step could possibly increase the discrimination and quantitation power of tag-sequences.

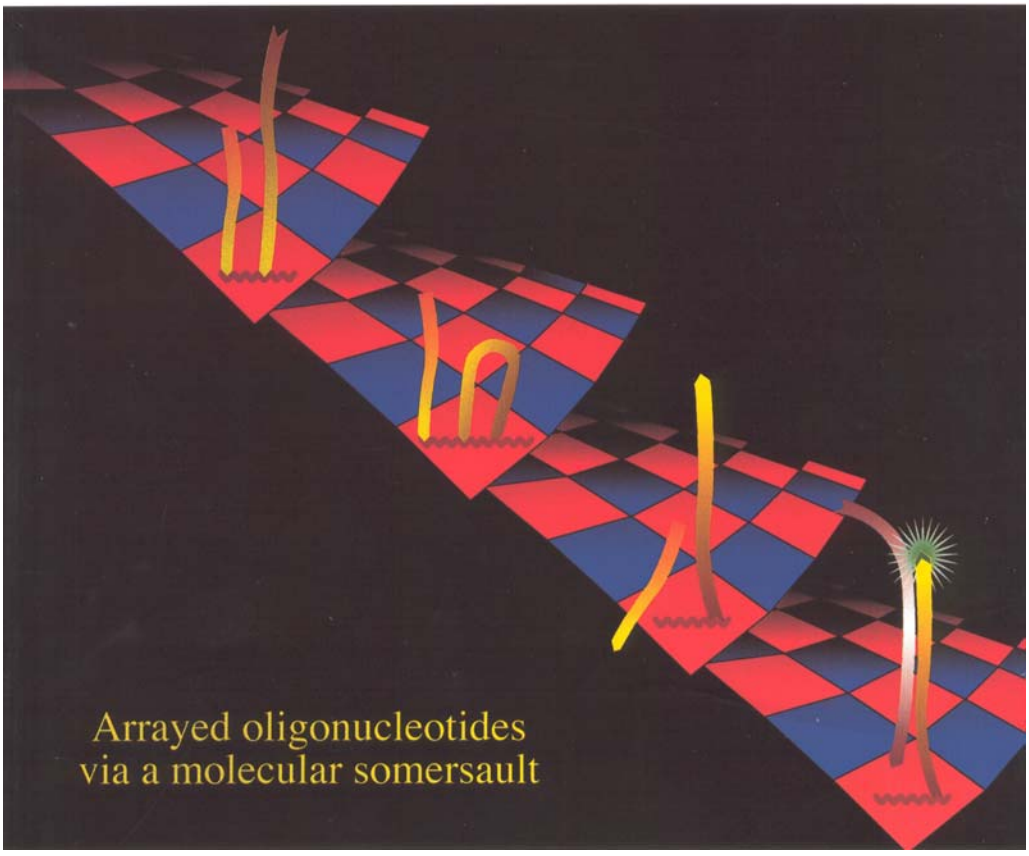

Figure D. Oligonucleotide inversion follow by a polymerization step templated by a hybridized target. As appeared on the cover of Nucleic Acids Research, 1999, 27(24), illustrated by Ulf Landegren.

3.4. PAPER IV, Effect of oligonucleotide truncation on single-nucleotide distinction by solid-phase hybridization.

In situ synthesized arrays have stepwise synthetic yields of 82-92%, which results in only 2-20% full-length probes for a 20-mer[97]. Paper IV reports an investigation on the effect of oligonucleotide impurities on the discriminatory ability in allele-specific hybridization on a solid phase. In short, we wanted to see if a purification strategy such as *in situ* inversion (Paper III) would significantly improve the performance of DNA microarrays. Considering the high degree of truncated oligonucleotides, the non-purified *in situ* synthesized arrays currently used exhibited surprisingly good results[98]. In an evaluation of the Affymetrix p53 mutation detection chip compared to Sanger sequencing, point mutations were readily detected but some frameshift mutations not coded for in the array could not be seen with the chip[99, 100].

In the model study of paper IV we immobilized biotinylated oligonucleotides of various lengths and amounts in streptavidin-coated wells. The amounts and lengths were chosen to simulate different stepwise synthetic yields and did not include the very shortest oligonucleotide impurities below a variable nucleotide position located centrally in the full-length probes, table 1. These experiments were not designed to exactly reflect the situation on an *in situ* synthesized chip but to give an indication of the influence by impurities. A target oligonucleotide was added to matched and mismatched immobilised oligonucleotide mixes and the amount of hybridization was monitored using the SYBR green intercalating dye over a wide temperature range. The highest discrimination ratio was achieved using 100% pure oligonucleotides at the optimal temperature, figure 1 and 2. Using oligonucleotides contaminated with truncated variant a multitude of hybrids of distinct hybridization stability arise between probes and target. This reduces the ability to distinguish matched and mismatched targets at only a particular temperature. Interestingly, the impurer mixes were able to discriminate at lower temperatures, probably due to the representation of shorter oligonucleotides capable of distinction at lower temperatures. This observation has recently been confirmed by others using non-purified *versus* HPLC-purified oligonucleotides, spotted and investigated at a single temperature[101]. In that study, non-purified probes showed higher match/mismatch ratios but a reduced capability to detect low target amounts. Since their investigation was only performed at one temperature, an optimal temperature could possibly be found where pure probes would perform better than impure ones. Also as mentioned, spotting of oligonucleotides involves some degree of purification since only full-length oligonucleotides will attach through the 5'-amine used for coupling.

In practical microarray experiments, large numbers of sequences with different optimal temperatures of distinction are analysed simultaneously. In order to distinguish all these sequences at a single temperature the impure probes may actually have a positive effect as they tend to increase the temperature window of distinction but at a reduced discrimination ratio. If sequences can be chosen to have the same optimal stringency temperature as can be done with tag-arrays then pure probes would be preferred.

4. DISCUSSION

This thesis has described a novel protein detection technology and an improved method for the synthesis of DNA microarrays. Proximity ligation has been has been used to detect single proteins by conversion of detection signals to into a nucleic acid signals. Many formats for the detection of these nucleic acids exist besides the real-time PCR presented here. There is a need for simultaneous analysis of many proteins in future proteomic applications. Samples will be analyzed for many qualities either by splitting them into minute aliquots for serial multiplexing, or parallel multiplexing in single assays. Proximity ligation can be applied with either of these strategies since small sample amounts can be analyzed at high throughput for serial multiplexing enabled by the high sensitivity. As only additions of mixtures to the sample are required, first the incubation mixture and then the combined ligation and PCR detection mixture, the assay is easily automated by robotics with no need for washing steps. For highly parallel multiplexing, the use of DNA microarrays to analyze large numbers of identification tagged proximity ligations reactions, one for each target protein, is anticipated. The "reverse translation" of proteins through proximity ligation into nucleic acid signals expands the capacity of genetic technologies such real-time PCR and DNA microarrays into universal platforms for the analysis of DNA, RNA, and proteins.

Not only is the detection and quantification of proteins important but also their interactions, modifications and cellular or subcellular localization. Proximity ligation can be used for assaying protein-protein interactions through the use of probes specific for each member of a protein complex. The efficiency of inhibitory ligands can also be investigated by assaying their potency in disrupting complexes, a potential application for drug candidate screening.

The two-site binding concept ensures specificity of detection in many standard solid phase assay formats. Proximity ligation requires not only dual-specific binding but has an added requirement for proximity of interaction further increasing specificity especially in parallel multiplexing. Furthermore, specificity can be increased by the use of a third analyte specific reagent which may be immobilized in a solid phase format. Also a homogenous triple-specific assay is under development where two ligation events unite three oligonucleotides into a PCR template. This can increase specificity and also sensitivity by lowering background ligations, enabling the use of greater concentrations of the proximity probes. Background ligations derive from spurious probes fortuitously located in sufficient proximity to hybridize to a common connector oligonucleotide then consequently ligated. The probability of target independent ligation of triple-spedific probes is substantially lower, compared to dual-specific probes. An obvious drawback is the need for three analyte specific binders to each analyzed protein.

A relevant question is which type of binder should be used in proximity ligation. This thesis describes the use of aptamers and monoclonal or polyclonal antibodies. We have also used a small molecule binder, biotin, for the detection of streptavidin and a protein antigen in combination with an aptamer for the detection of a specific IgE antibody (unpublished). International projects are now being proposed to generate large panels of specific binders to the human proteome, and discussions are ongoing as to how this should be conducted and what classes of binders should be aimed for. Much like the Human Genome Organization, HUGO, which sequenced the human genome, a Human Proteome Organization (www.hupo.org) has been formed to "assist in the coordination of public proteome initiatives". Proximity ligations can be used along with any binder with sufficient specificity and affinity, as long as a pair of them can be coupled to oligonucleotides.

The second part of this thesis involves an improved method for synthesis of highly pure DNA microarrays and the effect of impurities on solid phase hybridization reactions. DNA microarrays of high complexity, such as those manufactured through *in situ* synthesis by Affymetrix, are widely used today for large scale RNA expression and single nucleotide polymorphism analsis. Oligonucleotide synthesised in these arrays contain a high degree of impurities due to premature termination of synthesis. The inversion chemistry described in paper III can remove these impurities, leaving only the full-length oligonucleotides now with a free 3´-hydroxyl.

Now, one can question the need for such purifications considering the performance of *in situ* synthesized microarrays. We investigated the effect of truncated oligonucleotides in solid phase supported allele-specific hybridization reactions in paper IV. Lower match to mismatched ratios were observed with impure oligonucleotide populations compared to pure ones, but with an increased temperature window of discrimination for impure populations. Solid phase supported primer extension reactions have higher sequence resolution compared to hybridization alone. The free 3'-end generated in the purification step can be used for on-chip sequencing reactions or allele-specific primer extension. When using generic tag-arrays, cross-reactivity must be avoided and is a potential problem as the number of targets increase and especially when these are present in a range of concentrations as for expression analysis of RNA or protein through proximity ligation. Primer extension reactions from the 3'-ends of the inverted oligonucleotides can increase the specificity of the analysis of large numbers of tag-sequences in variable amounts compared to hybridization alone.

5. ACKNOWLEDGEMENTS

The work in this thesis was done at the department of Genetics and Pathology, Rudbecklaboratory. It has been supported by many people in different ways and I would like to take the opportunity to express my appreciation.

Especially to my encouraging supervisor Ulf Landegren for providing such an excellent, creative and open environment to work in. Your confidence in me has been very inspiring.

My co-supervisor Marek Kwiatkowski for spending so much time teaching me the chemistry of nucleic acids, good luck with Quiatech.

Mike Taussig, thanks for your comments on the thesis. I hope your stay in Uppsala will be fruitful.

All the administrators at Rudbeck, especially Jeanette Backman for always being helpful.

Thanks to all my molecular medicine colleagues. I chose to pursue my doctors degree here because you were such a great bunch of people. Anette, Maritha, Mats N., Charlotta, DOA, Anders I., for teaching me all the techniques in the lab. Mats N., for all our creative discussions that keep the motivation high. Johan, for the great parties at Fjället, and Fredrik, thanks for optimizing the gimlet to perfection. Sigrun, for always being so enthusiastic over our work as well as our parties. Jonas, thanks for letting me use your incredible artistic images of proximity ligation and helping out with my troubled car. Ola, hope your prox-work goes well. Maryam, thanks for keeping it all together in the lab. Mats G., its been great working together developing and discussing this technology with you. It's truly a pleasure to have ones colleges as ones friends and good luck to all the newcomers, Olle, Ara, Johan S., Jonas, Dieter, Susanna, "no retreat, no surrender!"

My former students Osama, Frida and Nicklas who did so much good work and helped me find out what you can do with proximity ligation and sambuca, and how they actually work. The splint-effect?

My collaborators Kristian Pietras and Arne Östman at the Ludwig Institute, Uppsala. Magnus Jobs and Anthony Brooks at CGB, Karolinska. Lena Marklund and Nils-Erik Heldin at our department.

Also thanks to my three supervisors during the UGSBR-year, Marek Kwiatkowski, Mostafa Ronaghi and Måns Ehrenberg, who made me realize how much fun science can be.

Eric, Peder, Staffan, John, Martin S., Martin P., Sara, Ann, and all those other people who I had the pleasure of making music with in my spare time. Ted, your clear-sightedness never ceases to amaze along with your jul-rim and fifa 97-02. Madeleine, for your support through all those good times.

The people behind Capricell, this is for you, Börje, Ann-Christine and Linus.

6. REFERENCES

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921. (2001).
2. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-1351. (2001).
3. Ronald, A. & Stimson, W.H. The evolution of immunoassay technology. *Parasitology* **117**, S13-27. (1998).
4. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730. (1999).
5. Anderson, L. & Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-537. (1997).
6. Li, M. Applications of display technology in protein analysis. *Nat Biotechnol* **18**, 1251-1256. (2000).
7. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510. (1990).
8. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822. (1990).
9. Butler, J.E. Solid supports in enzyme-linked immunosorbent assay and other solid- phase immunoassays. *Methods* **22**, 4-23. (2000).
10. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497. (1975).
11. Di Noia, J. & Neuberger, M.S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-48. (2002).
12. Batista, F.D. & Neuberger, M.S. Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. *Immunity* **8**, 751-759. (1998).
13. Foote, J. & Eisen, H.N. Kinetic and affinity limits on antibodies produced during immune responses. *Proc Natl Acad Sci U S A* **92**, 1254-1256. (1995).
14. Chen, Y. et al. Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. *J Mol Biol* **293**, 865-881. (1999).
15. Boder, E.T., Midelfort, K.S. & Wittrup, K.D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* **97**, 10701-10705. (2000).
16. Hanes, J., Schaffitzel, C., Knappik, A. & Pluckthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat Biotechnol* **18**, 1287-1292. (2000).
17. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391. (1994).
18. Wilson, D.S., Keefe, A.D. & Szostak, J.W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc Natl Acad Sci U S A* **98**, 3750-3755. (2001).
19. Roberts, R.W. & Szostak, J.W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A* **94**, 12297-12302. (1997).
20. Skerra, A. Engineered protein scaffolds for molecular recognition. *J Mol Recognit* **13**, 167-187. (2000).
21. Skerra, A. Lipocalins as a scaffold. *Biochim Biophys Acta* **1482**, 337-350. (2000).

22. Nord, K. et al. Binding proteins selected from combinatorial libraries of an alpha- helical bacterial receptor domain. *Nat Biotechnol* **15**, 772-777. (1997).
23. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. Diversity of oligonucleotide functions. *Annu Rev Biochem* **64**, 763-797 (1995).
24. Hermann, T. & Patel, D.J. Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820-825. (2000).
25. Eaton, B.E. et al. Post-SELEX combinatorial optimization of aptamers. *Bioorg Med Chem* **5**, 1087-1096 (1997).
26. Green, L.S. et al. Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor. *Chem Biol* **2**, 683-695 (1995).
27. Ruckman, J. et al. 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J Biol Chem* **273**, 20556-20567 (1998).
28. Jayasena, S.D. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem* **45**, 1628-1650 (1999).
29. Clark, S.L. & Remcho, V.T. Aptamers as analytical reagents. *Electrophoresis* **23**, 1335-1340. (2002).
30. Osborne, S.E., Matsumura, I. & Ellington, A.D. Aptamers as therapeutic and diagnostic reagents: problems and prospects. *Curr Opin Chem Biol* **1**, 5-9 (1997).
31. Drolet, D.W., Moon-McDermott, L. & Romig, T.S. An enzyme-linked oligonucleotide assay [see comments]. *Nat Biotechnol* **14**, 1021-1025 (1996).
32. Charlton, J., Sennello, J. & Smith, D. In vivo imaging of inflammation using an aptamer inhibitor of human neutrophil elastase. *Chem Biol* **4**, 809-816. (1997).
33. Ringquist, S. & Parma, D. Anti-L-selectin oligonucleotide ligands recognize CD62L-positive leukocytes: binding affinity and specificity of univalent and bivalent ligands. *Cytometry* **33**, 394-405 (1998).
34. Fang, X., Cao, Z., Beck, T. & Tan, W. Molecular aptamer for real-time oncoprotein platelet-derived growth factor monitoring by fluorescence anisotropy. *Anal Chem* **73**, 5752-5757. (2001).
35. Potyrailo, R.A., Conrad, R.C., Ellington, A.D. & Hieftje, G.M. Adapting selected nucleic acid ligands (aptamers) to biosensors. *Anal Chem* **70**, 3419-3425 (1998).
36. Jhaveri, S. Designed signalling aptamers that transduce molecular recognition to changes in fluorescence intensity. *J Am Chem Soc* **2000**, 2469-2473 (2000).
37. Jhaveri, S., Rajendran, M. & Ellington, A.D. In vitro selection of signaling aptamers. *Nat Biotechnol* **18**, 1293-1297. (2000).
38. Robertson, M.P. & Ellington, A.D. Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res* **28**, 1751-1759. (2000).
39. Hartig, J.S. et al. Protein-dependent ribozymes report molecular interactions in real time. *Nat Biotechnol* **20**, 717-722. (2002).
40. Yalow, R.S. & Berson, S.A. Assay of plasma insulin in human subjects by immunological methods. *Nature* **44**, 1648-1649 (1959).
41. Goldsmith, S.J. Radioimmunoassay: review of basic principles. *Semin Nucl Med* **5**, 125-152. (1975).

42. Wide, L., Bennich, H. & Johansson, S.G. Diagnosis of allergy by an in-vitro test for allergen antibodies. *Lancet* **2**, 1105-1107. (1967).
43. Miles, L.E. & Hales, C.N. Labelled antibodies and immunological assay systems. *Nature* **219**, 186-189. (1968).
44. Ekins, R.P. Ligand assays: from electrophoresis to miniaturized microarrays. *Clin Chem* **44**, 2015-2030 (1998).
45. Self, C.H. & Cook, D.B. Advances in immunoassay technology. *Curr Opin Biotechnol* **7**, 60-65. (1996).
46. Engvall, E. & Perlman, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871-874. (1971).
47. Kricka, L.J. Selected strategies for improving sensitivity and reliability of immunoassays. *Clin Chem* **40**, 347-357. (1994).
48. Kricka, L.J. Ultrasensitive immunoassay techniques. *Clin Biochem* **26**, 325-331. (1993).
49. Sano, T., Smith, C.L. & Cantor, C.R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **258**, 120-122 (1992).
50. Sims, P.W., Vasser, M., Wong, W.L., Williams, P.M. & Meng, Y.G. Immunopolymerase chain reaction using real-time polymerase chain reaction for detection. *Anal Biochem* **281**, 230-232. (2000).
51. Niemeyer, C.M. et al. Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR. *Nucleic Acids Res* **27**, 4553-4561. (1999).
52. Saito, K. et al. Detection of human serum tumor necrosis factor-alpha in healthy donors, using a highly sensitive immuno-PCR assay. *Clin Chem* **45**, 665-669. (1999).
53. Sanna, P.P., Weiss, F., Samson, M.E., Bloom, F.E. & Pich, E.M. Rapid induction of tumor necrosis factor alpha in the cerebrospinal fluid after intracerebroventricular injection of lipopolysaccharide revealed by a sensitive capture immuno-PCR assay. *Proc Natl Acad Sci U S A* **92**, 272-275. (1995).
54. Ullman, E.F., Schwarzberg, M. & Rubenstein, K.E. Fluorescent excitation transfer immunoassay. A general method for determination of antigens. *J Biol Chem* **251**, 4172-4178. (1976).
55. Hart, H.E. & Greenwald, E.B. Scintillation proximity assay (SPA)--a new method of immunoassay. Direct and inhibition mode detection with human albumin and rabbit antihuman albumin. *Mol Immunol* **16**, 265-267. (1979).
56. Ullman, E.F. et al. Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc Natl Acad Sci U S A* **91**, 5426-5430. (1994).
57. Dafforn, A., Kirakossian, H. & Lao, K. Miniaturization of the luminescent oxygen channeling immunoassay (LOCI(TM)) for use in multiplex array formats and other biochips. *Clin Chem* **46**, 1495-1497. (2000).
58. Templin, M.F. et al. Protein microarray technology. *Trends Biotechnol* **20**, 160-166. (2002).
59. Haab, B.B., Dunham, M.J. & Brown, P.O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* **2** (2001).
60. Pawlak, M. et al. Zeptosens´protein microarrays: A novel high performance microarray platform for low abundance protein analysis. *proteomics* **2**, 383-393 (2002).

61. Robinson, W.H. et al. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* **8**, 295-301. (2002).
62. Schweitzer, B. et al. Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol* **20**, 359-365. (2002).
63. Walter, G., Bussow, K., Lueking, A. & Glokler, J. High-throughput protein arrays: prospects for molecular diagnostics. *Trends Mol Med* **8**, 250-253. (2002).
64. Lizardi, P.M. et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* **19**, 225-232. (1998).
65. Brody, E.N. et al. The use of aptamers in large arrays for molecular diagnostics. *Mol Diagn* **4**, 381-388 (1999).
66. Golden, M.C., Collins, B.D., Willis, M.C. & Koch, T.H. Diagnostic potential of PhotoSELEX-evolved ssDNA aptamers. *J Biotechnol* **81**, 167-178. (2000).
67. Carter, J.D. Developing Research Proteomics and Clinical Diagnostics Assays Using Photoaptamer Arrays. *Tides, Oligonucleotide and Peptide Technology Conferences* (2002).
68. Vignali, D.A. Multiplexed particle-based flow cytometric assays. *J Immunol Methods* **243**, 243-255. (2000).
69. Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* **97**, 9390-9395. (2000).
70. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999. (1999).
71. Kodadek, T. Protein microarrays: prospects and problems. *Chem Biol* **8**, 105-115. (2001).
72. Naaby-Hansen, S., Waterfield, M.D. & Cramer, R. Proteomics--post-genomic cartography to understand gene function. *Trends Pharmacol Sci* **22**, 376-384. (2001).
73. Wickware, P. Proteomics technology: Character references. *Nature* **413**, 869-875 (2002).
74. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837-846. (2000).
75. Unlu, M., Morgan, M.E. & Minden, J.S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-2077. (1997).
76. Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. Real time quantitative PCR. *Genome Res* **6**, 986-994. (1996).
77. Saiki, R.K. et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354. (1985).
78. Sooter, L.J. et al. Toward automated nucleic acid enzyme selection. *Biol Chem* **382**, 1327-1334. (2001).
79. Cox, J.C. & Ellington, A.D. Automated selection of anti-protein aptamers. *Bioorg Med Chem* **9**, 2525-2531. (2001).
80. Cox, J.C., Rudolph, P. & Ellington, A.D. Automated RNA selection. *Biotechnol Prog* **14**, 845-850. (1998).
81. Drolet, D.W., Jenison, R.D., Smith, D.E., Pratt, D. & Hicke, B.J. A high throughput platform for systematic evolution of ligands by exponential enrichment (SELEX). *Comb Chem High Throughput Screen* **2**, 271-278. (1999).
82. Cox, J.C. et al. Automated acquisition of aptamer sequences. *Comb Chem High Throughput Screen* **5**, 289-299. (2002).

83. Tyagi, S., Marras, S.A. & Kramer, F.R. Wavelength-shifting molecular beacons. *Nat Biotechnol* **18**, 1191-1196. (2000).
84. Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat Genet* **14**, 450-456. (1996).
85. Giaever, G. et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet* **21**, 278-283. (1999).
86. Fan, J.B. et al. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* **10**, 853-860. (2000).
87. Hirschhorn, J.N. et al. SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc Natl Acad Sci U S A* **97**, 12164-12169. (2000).
88. Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L. & Syvanen, A.C. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res* **30**, E70-70. (2002).
89. Chen, J. et al. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res* **10**, 549-557. (2000).
90. Yeakley, J.M. et al. Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol* **20**, 353-358. (2002).
91. Giaever, G. et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387-391. (2002).
92. Baner, J., Nilsson, M., Mendel-Hartvig, M. & Landegren, U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res* **26**, 5073-5078 (1998).
93. Nilsson, M., Gullberg, M., Dahl, F., Szuhai, K. & Raap, A.K. Real-time monitoring of rolling-circle amplification using a modified molecular beacon design. *Nucleic Acids Res* **30**, E66-66. (2002).
94. Pirrung, M. How to Make a DNA Chip. *Angew. Chem. Int. Ed.* **41**, 1276-1289 (2002).
95. Syvanen, A.C., Aalto-Setala, K., Harju, L., Kontula, K. & Soderlund, H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684-692. (1990).
96. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363, 365. (1998).
97. Pirrung, M., Fallon, L. & McGall, G. Proofing of Photolithographic DNA Synthesis with 3′,5′-Dimethoxybenzionyloxycarbonyl-Protected Deoxynucleoside Phosphoramidites. *J. Org. Chem.* **63**, 241-246 (1998).
98. Patil, N. et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719-1723. (2001).
99. Wikman, F.P. et al. Evaluation of the performance of a p53 sequencing microarray chip using 140 previously sequenced bladder tumor samples. *Clin Chem* **46**, 1555-1561. (2000).
100. Ahrendt, S.A. et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc Natl Acad Sci U S A* **96**, 7382-7387. (1999).
101. Relogio, A., Schwager, C., Richter, A., Ansorge, W. & Valcarcel, J. Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res* **30**, E51-51. (2002).