From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

SINGLE-STRANDED DNA: METHODS AND APPLICATION IN NANOTECHNOLOGY

Giulio Bernardinelli

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB

ISBN 978-91-7831-531-4

Karolinska Institutet

Institutionen för Medicinsk biokemi och biofysik

Single-stranded DNA: methods and application in nanotechnology

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Biomedicum D1012, Solnavägen 9

Måndag den 7 Oktober, 2019, kl 09.30

av
Giulio Bernardinelli

Huvudhandledare:
Professor Björn Högberg
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Bihandledare:
Ana Teixeira
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Maria Lindskog
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Neurogeriatrics

Andreas Nyström
Karolinska Institutet
Institute of Environmental Medicine

Fakultetsopponent:
Professor Mats Nilsson
Stockholm University
Department of Biochemistry and
Biophysics

Betygsnämnd:
Professor C. I. Edvard Smith
Karolinska Institutet
Department of Laboratory Medicine
Division of Clinical Research Center

Professor Marcus Wilhelmsson
Chalmers University of Technology
Department of Chemistry and
Chemical Engineering
Division of Chemistry and
Biochemistry

Professor Lennart Nilsson
Karolinska Institutet
Department of Biosciences and
Nutrition

Stockholm 2019

Alla mia famiglia

ABSTRACT

Basic molecular and cell research, production of recombinant proteins, diagnostic detection of genetic mutations, construction of nanostructures and high-throughput DNA sequencing are only a few examples of the diverse set of applications that are amenable thanks to the availability of synthetic DNA polymers in biomedicine.

Strategic investments and technical progress together with the introduction of automation in the synthesis of DNA oligomers enabled to transform, in just a few decades, a process mastered only by a niche of biochemists into an affordable and available custom-made product to every scientific field.

Despite such progress, innovation soon reached a *plateau* due to intrinsic limitations of the synthesis process, putting a barrier at two hundred nucleotides as the maximum length of synthetic DNA molecules. In the meanwhile, molecular biologists closed the gap thanks to a better understanding of polymerases and the mastering of directed evolution protocols making it possible to redesign processes that are more similar to what happens in nature, taking advantage of existing and improved enzymes for the generation of long and high-quality DNA molecules. This enabled to find novel applications for DNA such as gene editing or information storage.

In this thesis I focused on the enzymatic production and functionalization of single stranded DNA. More specifically, in paper I we directed our attention to optimize the protocol for the templated enzymatic synthesis of oligonucleotides. We highlighted possible limitations of the technique and proposed a solution in employing a single stranded binding protein greatly decreasing double stranded DNA contaminants.

In paper II we further extended the workflow. In here, we focused on continuing the previous protocol to accommodate the production of chimeric DNA-protein molecular tools needed in nanotechnology where DNA is considered more a building material rather than an information rich polymer while the actuation of a particular function is operated by proteins. We worked on a minimal bacteria-derived self-tagging domain that has the capacity to establish a covalent bond with a specific DNA sequence and some applications are suggested. Paper III represents the natural extension of this work even if, in this specific case, the earlier presented rational is reverted. More specifically, a biosensor for the detection of aquatic microorganisms was produced with the characterized bioconjugation technique where the chimeric protein was used as recognition moiety and the oligonucleotide as signal amplification device through its intrinsic DNAzyme activity.

Finally, in Paper IV, we decided to use all the previously gathered knowledge – enzymatic DNA production and bioconjugation techniques – to conceive a novel basic biology investigation tool for the study of spatial organization of proteins. Here we took advantage of the possibility to grow a localized and unique DNA polymer with the ability to target proteins with DNA-protein chimeras. The resulting product is then recovered and decoded by next generation sequencing.

LIST OF SCIENTIFIC PAPERS

I. Ducani C., **Bernardinelli G**., & Högberg B. (2014). Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Research, 42(16), 10596–10604.*

II. **Bernardinelli G.**, & Högberg B. (2017). Entirely enzymatic nanofabrication of DNA-protein conjugates. *Nucleic Acids Research, 45(18), e160. https://doi.org/10.1093/nar/gkx707*

III. **Bernardinelli G.**, Oloketuyi S., Werner S. W., Mazzega E., Högberg B., de Marco A. (2019) A compact nanobody-DNAzyme conjugate enables antigen detection and signal amplification. *Manuscript*

IV. **Bernardinelli G.**, Werner S. W., Rocamonde-Lago I., Hoffecker I. T., Högberg B. (2019). RollingTag-seq: a microscopy-free DNA sequencing based approach for protein spatial organization probing. *Manuscript*

Additional papers not included in the thesis

I. Ducani C., **Bernardinelli G.**, Högberg B., Keppler B. K., & Terenzi A. (2019). Interplay of three G-quadruplex units in the KIT promoter. *Journal of the American Chemical Society.*

II. Hoffecker I. T., Yang Y. T., **Bernardinelli G.**, Orponen P., & Högberg B. (2019). A Computational Framework for DNA Sequencing Microscopy. *PNAS.*

CONTENTS

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
dNTPs	Deoxynucleotide triphosphates
ssDNA	Single stranded DNA
RNA	Ribonucleic acid
MOSIC	Monoclonal stoichiometric
TdT	Terminal deoxynucleotidyl transferase
G4	Guanine quadruplex
PAGE	Polyacrylamide gel electrophoresis
EMSA	Electro mobility shift assay
PCR	Polymerase chain reaction
RCA	Rolling circle amplification
HCR	Hybridization chain reaction
SSBs	Single-stranded DNA binding proteins
PS	Phosphorothioate
PM	Plasma membrane
PMPs	Plasma membrane proteins
POX	Peroxidase
BRET	Bioluminescence resonance energy transfer
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein

1 INTRODUCTION

DNA is an information-rich polymer[1] common to living organisms and made upon the repeating monomers, the nucleotides: made of a 2-deoxyribose sugar, a purine base - adenine (A) or guanine (G) - or a pyrimidine base - thymine (T), cytosine (C) - and a phosphate group[2] (Figure 1) As suggested by Phoebus Levene in 1919[3], a polynucleotide is a directional polymer made of those nucleotides that are connected between the fifth carbon atom of a nucleotide with the third carbon atom of the following nucleotide through phosphodiester bonds by the phosphate group. This nitrogen and phosphorous rich biomolecule was isolated for the first time at the end of the nineteenth century by the Swiss physician Frederich Miescher[4] and twenty years later Albrecht Kossel identified the five nucleobases[5]. It is only by the middle of the twentieth century that enough knowledge[6,7] was accumulated to obtain a more complete picture about its function and structure[8]. Understanding how DNA is organized also promptly suggested how molecular mechanisms work enabling a secure propagation of the genetic information.

> *"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."*
>
> *Watson, J. D. & Crick, F. H. C., 1953*

In nature, DNA is mostly found in its double stranded forms where two complementary filaments interacting through base pairing and stacking are directed toward opposite directions forming an antiparallel polymer[8]. Nevertheless, it is also possible to find organisms that make use of single stranded DNA (ssDNA)[9,10] and even in higher eukaryotes, the double stranded genomic DNA can take advantage of ssDNA structures as, for example, regulatory elements[11].

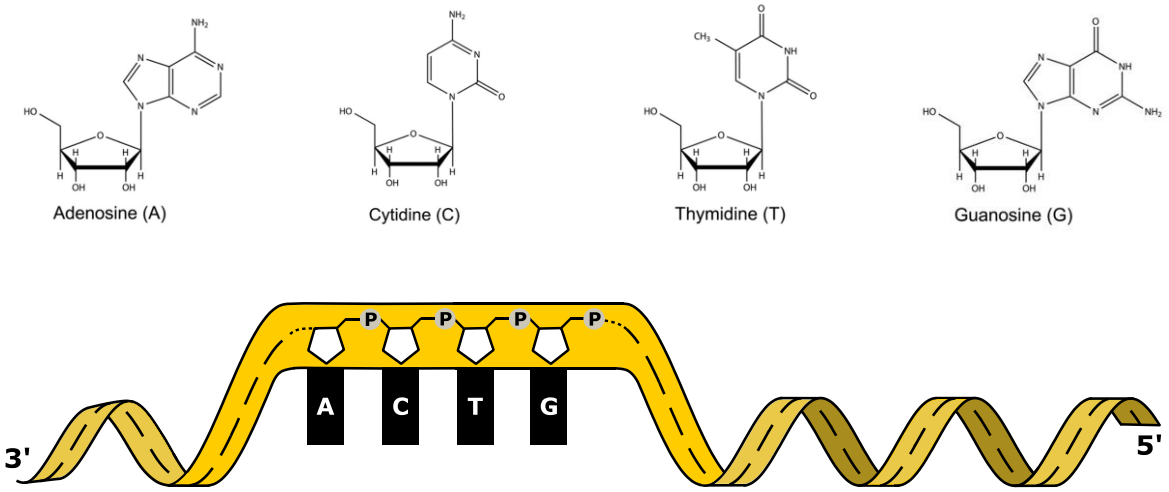

Figure 1. Structure of nucleosides and features of a single-stranded polynucleotide. DNA in its single stranded form exhibits a peculiar elastic behavior that is of major relevance for biological and non-biological processes. Such molecule forms secondary structures as a consequence of canonical Watson-Crick hydrogen bonding and other non-ordinary base paring.

Synthetic ssDNA, *in vitro*, is an extremely versatile molecule that can be exploited for a wide range of applications. Despite being rather fragile, it enables an unprecedented control over the possibility to position molecules in space with a nanometer precision when used as a construction material in nanotechnology[12]. Moreover, chemicals moieties and proteins connected to ssDNA can be organized with great precision[13]. Taking advantage of strand complementarity, ssDNA can also be designed to interact with genomic sequences or messenger RNA[14]. Furthermore, the selection of specific primary structures of a polynucleotide that could be generated rationally or combinatorically, is also responsible for the formation of intramolecular secondary structures, known as aptamers[15], able to interact with more complex players, such as proteins[16]. Selection and evolution methods[17] of specific structures enable the possibility to discover nucleic acids able to execute enzymatic-like processes[18] or behave like fluorescent molecules[19].

Another important aspect that makes ssDNA an attractive molecular tool, is the possibility to produce it in large amounts with different technologies and platforms at competitive prices. Herein the methods for production and functionalization of ssDNA will be briefly described while the stress will be put on selected applications in the enclosed papers of this thesis.

1.1 METHODS FOR THE PRODUCTION OF ssDNA

There are essentially three main strategies for the production of synthetic DNA. The method that dominates the market is the chemical synthesis on solid-phase but its intrinsic limitations drove the development of alternative strategies using enzymes *in vitro* or even taking advantage

of the molecular machinery of living organisms. The following table summarizes the main methods discussed in this thesis (Table 1).

Method	Ref.	Strategy	Fidelity or error rate	Max. Len.	Price per base	Turnover time	Template
Column based	27	Chemical	99 % /cycle	200 nt	$0.05 to $0.10	++++	-
Microarray based	28,31	Chemical	≥99,5% / cycle	350 nt	$0.0001 to $0.00001	++++	-
MOSIC	46, 47	Enzymatic	10^{-5} 10^{-6}	378 nt	$0.8 / base / µg	+++	+
TdT	51,52	Enzymatic	97,7 % / cycle	10 nt	n.d.	+	-
Phage virus	58,59	*In vivo*	n.d.	2520 nt	n.d.	-	+
In cell	60-62	*In vivo*	n.d.	20	n.d.	-	+

Table 1. Direct comparison of some of the most representative ssDNA production methods discussed in this thesis.

1.1.1 Chemical synthesis

Oligonucleotides have been synthetized in solution from the 1950s[20]. The synthesis method rapidly evolved, and a more efficient reaction, based on the phosphoramidite chemistry[21], was implemented on solid-phase in columns.

The opportunity to easily synthesize nucleic acid polymers made it possible to understand many different molecular processes that occur in a cell, further testifying the complementarity need and importance of chemistry and biology. Among the most important milestones achieved at the early stage of such technology, we have the deciphering of the genetic code[22] and the synthesis of the first synthetic gene encoding for the somatostatin hormone in *Escherichia coli*[23]. Moreover, the introduction of automation in the synthesis process was fundamental to reach a cost-effective scalable production of oligonucleotides, up to 96 or even 384 at the same time, in the synthesis scale ranging from 10 to 100 nmol[24].

Nevertheless, solid-phase column based synthesis of oligonucleotides is affected by limitations with respect to lengths and fidelity[25,26]. In fact even if the reaction yield of each cycle is rather high, more than 99% per coupling cycle, this will result in a final 13% yield for a desired 200 nucleotide oligo[27]. Length of the desired oligonucleotide is not the only problem, as also sudden cleavage of the polymer could happen during the deprotection step. Still connected with this step, another common problem is single base deletion. For all these reasons, a purification step of the aimed product is always advisable when long oligonucleotides are desired. Recent progress in the production steps will soon be translated in better commercially available oligonucleotides[28] and some companies have already extended the maximum length of oligonucleotides to 300-400 bases[29,30].

A further improvement in the industrial process is exemplified by the development of array-based oligonucleotides synthesis. DNA printing was first achieved with the use of photoactivated deprotection of the phosphoroamidite chemistry[31]. The invention of microarray enabled to reduce the synthesis scale, thus reducing the synthesis costs down to $0.0001 - $0.00001 per nucleotide depending on the platform used and the oligonucleotide characteristics, whereas the price for column synthesized oligos still ranges from $0.05 to $0.10 per base[32]. For these reasons, oligonucleotide pools are now the preferred starting material for gene synthesis even if the concentration of every oligo is really low.

A major difference between chemical synthesis of ssDNA and the other proposed methods is the possibility to grow a polymer in an untemplated fashion while, most of the enzymatic protocols, such as the asymmetric PCR method, rely on a sequence-verified DNA template to be further propagated. The exception to this rule is represented by those methods based on the activity of the enzymes like terminal deoxynucleotidyl transferase (TdT) or similar[52].

1.1.2 Enzymatic synthesis

Synthetic biochemistry is not the only discipline that enabled to produce high-quality oligonucleotides. In fact, molecular biologists soon understood that polymerases provide a much higher fidelity in DNA synthesis. The error rate of chemical synthesis is between 1×10^{-2} and 1×10^{-4} according to the method used[27], while enzymatic synthesis *in vitro* has reported an error rate of 1×10^{-5} to 1×10^{-6} [33,34] and this without cells proofreading machineries.

Since the development of the polymerase chain reaction (PCR) with thermostable enzymes[35] it was immediately clear that such method is much more practical, if a template sequence is available. Taking advantage of those findings a series of enzymatic strategies have been developed and optimized. The first reported approach for the production of a ssDNA molecule of interest, is an asymmetric PCR[36–38]. In this protocol, the proportion of primers used in the reaction is strongly unbalanced enabling a favorited synthesis of one strand over the other. Even if effective, this approach remains imprecise and requires a purification step as point mutations are amplified during each round. Similar approaches take advantage of 3' protected oligonucleotides and enzymatic clean-up through exonuclease treatment of the unprotected strand[39]. In alternative to exponential amplification, other protocols are based on a linear amplification taking advantage of naturally occurring ssDNA producing polymerases such as Φ29 during the rolling circle amplification (RCA)[33,40,41]. Such strategies have been extensively used to amplify genomic DNA for sequencing purposes[42], DNA and RNA detection[43–45] but, most importantly, it has been used for high quality ssDNA production within the monoclonal stoichiometric method (MOSIC)[46,47]. Notably, the production of ssDNA for the MOSIC protocol can also take place in biological systems such as bacteriophages. Other variants of the technique[48] aimed to remove the enzymatic restriction step and rather relied on autocatalytic deoxyribozymes in presence of metal ions thus reducing the freedom in the choice of the

sequence to produce[49]. Other approaches seem achievable with new molecular tools such as modified versions of the CRISPR-Cas9 system but have never been exploited for this purpose[50]. An alternative and more recent approach is based on the usage of the enzyme terminal transferase in combination with adapted nucleobases[51,52].

The unmet need for the market of long ssDNA with high degree of purity motivated the investments[53] in this field leading to the foundation of several startups such as DNAscript[54], Nuclera[55], Evonetix[56] and Moligo Technologies[57].

1.1.3 *In vivo* synthesis

A more recent approach is based on the exploitation of the life cycle of ssDNA phages. In this case, the sequence of interest, embedded in a plasmid characterized also by the presence of a bacteriophage specific packaging consensus sequence, is propagated in phage-competent bacteria in a bioreactor[58,59]. The method aims primarily to permit an agile scale-up enabling gram-scale production of ssDNA and has been developed mainly in the field of DNA nanotechnology but greatly limits the sequence freedom of choice. A few attempts of producing ssDNA[60–62] in eukaryotic cells for therapeutic purposes have been documented in the past but little or no follow-up studies have been published.

1.2 OLIGONUCLEOTIDES-PROTEINS CONJUGATION METHODS

Even if DNA offers the unprecedented ability to program the assembly of nanostructures at the nanoscale or enables precision molecular detection operations with a large degree of control that is unreachable by any other molecule, the amount of operations carried by naked DNA are limited. For this reason, most of the applications of ssDNA in nanotechnologies rely on a multitude of biomolecule conjugates to obtain the desired outcome[13,63,64].

Conjugation represents the best possible way to increase the functionality of nanostructures. Many molecules have been coupled to DNA oligonucleotides such as lipids[65], drugs[66,67] and even chemical libraries[68]. But hereon we will mainly focus on protein conjugates[69].

Despite the availability of several different approaches[70–73] (Figure 2) achieving a successful DNA-protein conjugate remains a pitfall-prone process that is highly protein dependent and that requires optimization for each new conjugate. Over-conjugation, target protein loss of function or low reaction yield represent common problems. Two main conjugation approaches intended to overcome the previously highlighted limitations will be introduced here.

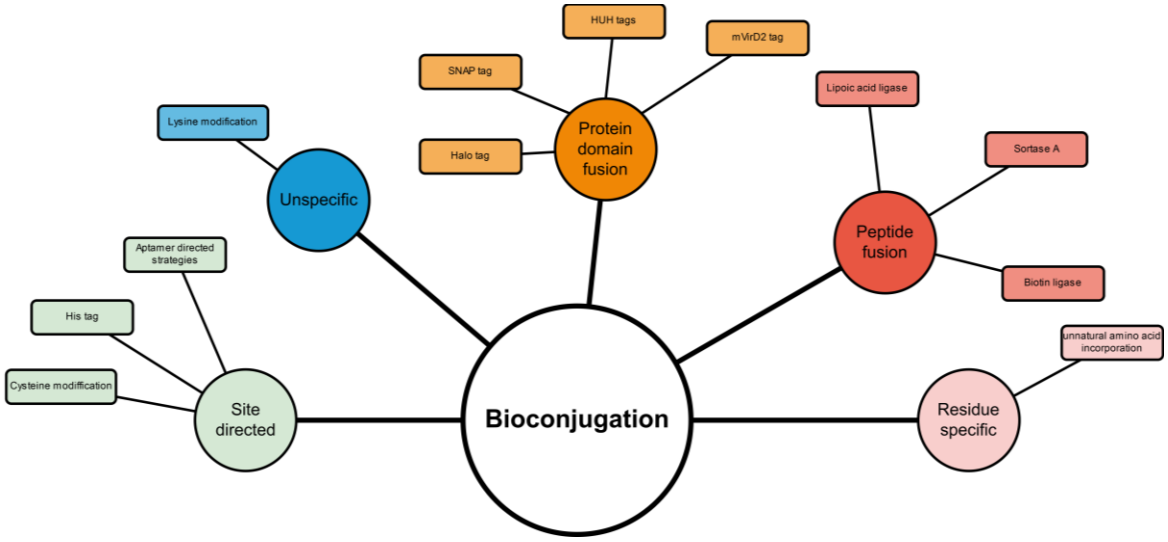

Figure 2. Main protein bioconjugation strategies discussed in this work.

1.2.1 Self-labeling tags

To avoid the above mentioned issues, conjugation strategies based on genetically encoded self-labeling tags, have been recently developed to enable a reproducible site specific labeling of the target proteins[74,75]. Nevertheless, for any of these common strategies (SNAP tag, HALO tag), as it happens to the protein as well, the DNA molecule of interest needs to be chemically modified – a process that is only feasible when starting out with synthetic DNA oligonucleotides *in vitro*.

Aiming for an automated system that enables self-assembly and functionalization through protein conjugation in physiological conditions is a challenging task for most of the listed strategies while, rather than DNA, the most successful orthogonal approaches are directed to functionalize proteins[76,77]. As previously mentioned, it has been reported that there is the possibility to produce ssDNA in eukaryotic cells[60–62]. Therefore, the possibility to use genetically encoded self-tagged domain taking advantage of viral and bacterial proteins[69,78] to covalently bind unmodified DNA becomes an attractive approach for the autonomous assembly of macromolecular complexes in living organisms.

1.2.2 Copper-free click chemistry

This relatively new approach is compliant with most of the earlier stated requirements for an efficient and biologically compatible reaction. The term was first introduced in 2001 by Barry Sharpless[79] and became a very successful strategy to efficiently link biomolecules modules with high specificity and in aqueous environment. The set of reactions was expanded over time and the copper catalyst, that could be damaging for nucleic acids[80], is not necessary to activate the alkyne if introduced in a strain-promoted destabilized form[81].

Even though such class of reactions are fast, specific and efficient for bioconjugation, both *in vitro* and *in vivo*[82,83], modifications of the biomolecules of interest are needed to introduce the moieties for the reaction to happen. For what regards nucleic acids, most suppliers provide in their catalogs[84] at an extra cost the introduction of the desired modification while proteins require a supplementary effort to become compatible for such approach. The modifications could be made by introduction of unnatural amino acids[85] or by a first, unspecific[86] or site directed[87], conjugation step with a divalent chemical handle.

1.3 ssDNA AND ssDNA-PROTEIN CONJUGATE APPLICATIONS

Since their introduction on the market, commercially available oligonucleotides have soon become affordable and indispensable instruments in the toolbox of molecular biologists. Nowadays DNA oligonucleotides are used in many other disciplines such as material sciences and data sciences[88–91].

Oligonucleotides have found applications in very diverse situations and for this reason it would not be possible to exhaustively list all of them but, for the purpose of clarity, it is possible to group the main uses of ssDNA into the following categories.

1.3.1 Detection and / or amplification

Historically, this is the first and main application of short oligonucleotides. This could be achieved using both enzymatic and non-enzymatic methods. Among the enzymatic methods we encounter the most popular protocol in molecular biology where a couple of oligonucleotides, in this case called primers, promote a targeted enzymatic amplification of a DNA sequence via polymerase chain reaction (PCR)[92] by thermal cycling that could be used for both preparative and analytical purposes. Since then, new enzymes have been produced and protocols optimized according to specific needs. A second big family of enzymatic reactions is represented by isothermal amplification reactions. This last set of reactions is performed by a specific set of polymerases characterized by a strong strand displacement activity hence the reaction does not need a thermal induced denaturation between every new cycle. This specific activity connected with a circular template enables the rolling circle amplification (RCA). This very processive reaction has been extensively used for the purpose of amplifying the signal and being detected by hybridization but also for the generation of genomic DNA libraries for sequencing and for *in situ* sequencing applications when combined with the specific design of padlock probes that provide the amplification template. There are also other chimeric (enzymatic-non-enzymatic) reactions such as the loop mediated isothermal amplification (LAMP)[99,100].

Non-enzymatic methods have also been extensively used for the detection of nucleic acids. Those approaches are primarily based on classical Watson-Crick base pairing[93,94]. Oligonucleotides could be intrinsically fluorescent, radio-labeled or modified with small

molecules such as biotin or digoxigenin for immuno-based assays. These techniques are widely applied when it comes to decipher gene expression profiles, when labeled molecules are hybridized to synthetic oligonucleotides printed on microarrays but also to detect messenger RNA *in situ*. Lately such techniques have been further implemented and the detection of even weak signals is guaranteed by nucleic acids self-assembly such as the hybridization chain reaction (HCR) [101] or RNAScope® commercial system. The first method is based on oligonucleotides forming metastable hairpin structures that, once triggered with an initiator strand, are able to grow linear polymers made of fluorescently labeled oligonucleotides. The second method instead is based on a less dynamic assembly of labeled oligonucleotides following a "Christmas tree"-like scheme. Recently, labeled oligonucleotides have also been used as super resolution microscopy probes.[103,104].

A less canonical use of nucleic acids is as a detection tool of proteins or small molecules [15,95]. This could be achieved by a direct interaction of the oligonucleotide secondary structure, when a so called aptamer[96] is used, or by an indirect interaction when the nucleic acid handle is conjugated to a protein such as an antibody (discussed in further details in section 1.3.4).

1.3.2 Nano-construction material

ssDNA is extensively used for the production of nanostructures[105]. Such devices can be assembled from a rationally designed pool of oligonucleotides to form bidimensional[106] or tridimensional structures[107]. The easy access to short oligonucleotides enables the design of nanostructures upon a given sequence of the ssDNA from different sources such as the genome of bacteriophages[108]. To enable the compaction of DNA strands ("scaffold" and "staples") a saline buffer containing magnesium is used. DNA nanostructures can be folded thanks to direct complementarity of the sequences in a temperature ramp protocol. New generations of structures are less packed thus require less salts in the buffer for a successful folding[107]. Nanostructures can be folded naked or functionalized with chemicals or proteins to execute specific tasks[109]. Other than rationally designed nanostructures, DNA can be used to form hydrogels and matrices by RCA reaction[110].

1.3.3 Therapeutics, genome editing and gene assembly

A growing application field of ssDNA is the development of therapeutic protocols thanks to the ability to design with precision macromolecules able to target cells at different levels such as transcription, RNA splicing and gene expression. For example, DNA oligonucleotides and later chemically modified oligonucleotides have been used to disrupt putative transcription sites through the formation of triplexes[111–113]. Such macromolecules interact with the target sequence via either Watson-Crick base paring or form triple helices on polypurine stretches of DNA double helices via Hoogsteen or reverse Hoogsteen interactions. Despite the precision of such tools, synthetic oligonucleotides suffer from two main drawbacks: limited extracellular stability and difficult delivery. For those reasons, chemists developed several generations of modifications to improve oligonucleotides half-life, biodistribution and increase efficient

delivery to target cells[114], without impairing the specificity toward the target but aiming to increase the desired disruptive activity[115]. More recently, ssDNA has also been used as donor, to template the desired mutations or *de novo* insertions of DNA in genome editing protocols using the CRISPR-Cas9 system[116]. Furthermore, what fits this section is the possibility to assemble genes from oligonucleotides pools for synthetic biology applications[117].

1.3.4 DNA assisted proteomics

Many different methods have been developed over time for the detection of proteins and to highlight hypothesized protein-protein interaction. A classical thorough analysis workflow would require a multi-step approach. Following *in vitro* characterization, interactions would of course need to be validated *in vivo* as well. Additional needed steps for a full characterization require the study of the domains involved in the interaction. Moreover, functional assays need to be designed to understand the consequences of the interactions.

The choice of the optimal technique to apply depends on the question that needs to be answered but also on the specific nature of the interaction. For example, a stable interaction can be best investigated by co-immunoprecipitation[118] or far western methods[119] while a transient interaction is more difficult to capture and could require a cross-linking step or label transfer method. Briefly, techniques could be organized into four different groups:

- The main category is represented by the so-called "classical" biochemical approaches. Protein array[120], co-immunoprecipitation, affinity chromatography[121] and far western.
- The other main category of "classical" approaches involves genetic manipulations such as the yeast-two hybrid system[122], phage display[123] and the expression of protein libraries.
- Modern co-localization approaches are based on FRET[124] and BRET[125] methods and co-immunolocalization by microscopy.
- Other main techniques that cannot find a place in these classes are: surface plasmon resonance[126], X-ray crystallography[127] and cross-linking[128].

Plasma membrane proteins (PMPs) have peculiar chemical and physical properties that make their study difficult, therefore a new set of methods, mainly low-throughput optical methods, have been developed. Each method guarantees a different spatial and temporal resolution power that has to be taken into account accordingly. For example, single-molecule fluorescence imaging[129] and fluorescence correlation spectroscopy (FCS)[130] have more or less the same spatial resolution (250 nm) but FCS has a superior temporal resolution and should be preferred for highly dynamic proteins. FRET instead, enables resolution down to 5-10 nm with diffracted limited acquisition systems. Other PMPs study methods aim to reduce membrane complexity and use synthetic and cell derived membranes to be probed with atomic force microscopy[131] or other diffraction or non-diffraction limited microscopy methods such as DNA-PAINT[132] that is based on antibody-oligonucleotide conjugates.

In this context DNA became more and more popular as an additional tool for protein detection and identification[133,134]. The introduction of DNA to decode the proteome became attractive thanks to the technical improvement in the ability to sequence DNA with next generation

sequencing[135]. Among the implemented strategies involving DNA-antibody conjugates, a prominent microscopy-based technique is the so-called proximity ligation assay (PLA)[43,45,136]. Here the proximity of proteins of interest determine the assembly of a circular template for a localized RCA reaction. The RCA product is detected by hybridization with a fluorescently labeled complementary probe.

To address the low throughput achieved so far, new proteomic methods have been recently published to study cell-type specific proteomes *in vivo*[137] and proteins surrounding a protein of interest by biotin labeling [138].

All the aforementioned techniques provide a lot of information about targets vicinity, but it is not yet possible to have an amplification template guaranteed by a greater number than two analytes. One such technique that enables an open detection of an undefined number of proteins locally constrained has not yet been described.

Nevertheless, there is a growing interest in the establishment of a technique able to decipher the relations of proteins in constrained spaces as it is possible to have a bulk measurement on the protein presence in the cell and a fairly good idea of how the protein is organized but not totally conclusive yet[139,140]. Each and every method suffers from strong limitations and therefore makes it less appealing to a wide application.

1.4 DECIPHERING MEMBRANE PROTEIN ORGANIZATION

In all biological systems it is common to encounter the compartmentalization phenomenon[141]. This ability to develop specialized units with all the necessary components to accomplish specific tasks is oriented toward the overall increased efficiency of the cell's biochemical machinery and represents a clear proof of the endosymbiotic theory[142]. Nature-inspired compartments re-created *in vitro* have also been used for the optimization of synthetic biochemical processes[143].

Cell-cell communication, signal transduction and integration are some of the functions that are made possible thanks to the organization of proteins in the plasma membrane (PM). PM is one of the major cell compartments and it is now clear that proteins are not simply distributed uniformly but they are highly organized according to the cell's needs[144].

Since the 1970s a lot of theories about the organization of plasma membrane proteins (PMPs) have been developed and revised[145]. For the aim of this introduction it is possible to focus on different types of PM organization units according to their size extension. When a cell is polarized, the membrane is organized in obvious macrodomains that exhibit peculiar structures such as microvilli in small intestine epithelial cells. Other macrodomains are represented by podosomes and focal adhesions[146] or immunological synapse in leukocytes cells[147]. Even if a sharp distinction is always difficult to make, other PM functional units are microdomains that range from a few micrometers to 100 nm usually induced by ligands[13]; such domains are

thought to be organized thanks to the PM lipid composition[148] and proteins such as the arginine transporter can be found[149]. Another smaller PM unit is represented by nanodomains that have a size between 100 and 10 nm in which RAS proteins can be found[150]. Finally, nanoclusters have a size of a few nanometers such as GPI-AP clusters[151].

The clustering of PMPs has been described earlier[145,152] and the reasons that could explain this phenomenon could be stretched beyond the fact of having specialized areas to execute specific tasks. As a matter of fact, clustering enables different proteins to interact and expose hidden domains that could trigger signaling pathways. On top of that, increasing the critical mass of receptors interacting with ligands could provoke a stronger signal cascade that would affect the cell behavior[153,154].

1.5 TOWARDS AN OPTICS FREE DNA-MICROSCOPY

DNA-protein conjugation techniques enabled the generation of new powerful biological investigation tools that, among the other things, permitted a better understanding of protein-protein interactions. However, the usage of such tools is framed in rigid and binary systems able to achieve only an all-or-none output and, most importantly, where the exact number of proteins within a cluster remains unclear. To gather such information, using a high throughput technique, biological detection devices should be re-designed taking advantage of an existing approach: DNA-encoded barcodes.

Such method was already applied in pharmaceutical industry. In 1992, Sydney Brenner and Richard Lerner published a thought experiment that preconized DNA-encoded chemical libraries[155] for combinatorial chemistry. Since then, some companies such as GSK[156] and Vipergen[157] have decided to embrace such technology bringing to light some interesting molecules. The concept has since been expanded and it is also possible to profile cells individual proteomes thanks to DNA-encoded antibody libraries[68]. DNA-encoded barcodes had become even more attractive thanks to the parallel development of cost-efficient high throughput sequencing methods[158].

Inspired by this approach, a breakthrough study attempted to reconstruct the connectome labeling each neuron with a different RNA barcode and collecting the fusion products of two barcodes in synapses[159,160]. In the specific case, Zador's lab engineered a population of viruses to produce a specific protein with, on one hand, the ability to bind a randomized - barcoded - mRNAs intracellularly and, on the other hand, to be crosslinked to the same protein expressed from a neighboring cell. In such a way the crosslinked protein product brings to close proximity two unique short mRNAs that can be fused in a single molecule. In this case cell unique barcodes enable to reconstruct the connectome in a matrix.

The ability to reconstruct the spatial organization of proteins or any other biomolecule of interest thanks to localized and unique DNA sequence barcodes would enable one to investigate even protein interactions and distribution on PM[161–163]. Toward this direction a new DNA

origami-made platform, called “nanoscope” by the authors, has been recently published[164] but has not yet been used on biological samples.

Today the generation of new tools to answer biological questions at the nanoscale represents an unmet need that could be achieved in the near future combining previously developed techniques. This would enable us to broaden our knowledge of basic molecular mechanisms of the living cell with an unprecedented throughput and resolution.

2 AIMS

The works collected in this thesis aim to both explore the understanding of basic molecular mechanisms underlying single-stranded DNA amplification and enzymatic conjugation as well as applying such knowledge to the development of tools and methods for biological investigation at the nanoscale.

Paper I: To investigate the causes and possible countermeasures of the generation of an unwanted – double stranded DNA – product in the generation of single stranded DNA via rolling circle amplification.

Paper II: To take advantage of a naturally occurring enzymatic process in pathogenic bacteria to develop and optimize a novel DNA-protein conjugation method.

Paper III: To implement a novel bio-sensing strategy with the use of a chimeric DNA-protein conjugate.

Paper IV: To deliver a set of molecular tools and protocols for the study of protein clustering organization at the nanoscale enabling topological reconstruction via next generation sequencing without the need of microscopy.

3 MATERIALS AND METHODS

Here some of the most relevant methods used in the appended works are introduced. A detailed explanation of relevant protocols and techniques are included in the manuscript section.

3.1 ENZYMATIC PRODUCTION OF SSDNA (MOSIC)

The production of unmodified oligonucleotides of different lengths was carried out with the method developed in our lab[46], the so called MOSIC protocol. Briefly, the generation of an oligonucleotide of interest is ensured by the following steps:

1) The generation of a clonal template;

2) The linear RCA amplification reaction *in vitro*;

3) The enzymatic digestion and purification of the oligonucleotide of interest.

3.1.1 Generation of RCA template

A MOSIC template is generated either by a sequence verified clonal plasmid or oligonucleotides assembly. In the first case the plasmid containing the pseudogene template (BioCat) is harvested from a bacterial culture, digested to remove the unwanted features (i.e. antibiotic resistance, origin of replication) and the fragment of interest is re-ligated in a mini-circle like template form using the T4 ligase (0.25 U/ µl, Thermo Scientific) in 1x rapid ligation buffer at 22 ºC for 2 hours. In the latter case, synthetic oligonucleotides (IDT) are circularized by hybridization to a bridging oligonucleotide (100 mM potassium acetate; 30 mM HEPES, pH 7.5) in a 1 to 1 ratio (1 µM for each oligonucleotide). Fifteen µL of such reaction are combined with 5 units of T4 ligase (Thermo Scientific) in a diluted ligation buffer (1:20) supplemented with 1 mM ATP. The twenty µL reaction is incubated at 22 ºC for 2 hours and the product of interest is purified first by enzymatic removal of unligated oligonucleotides with the addition of 1 µL of ExoSAP-IT (Thermo Scientific) and 1 µL of Exonuclease III (Thermo Scientific) for 30 minutes at 30 ºC and later by extraction from a 10% denaturing (8 M urea) polyacrylamide gel by diffusion in a 10 mM Tris-HCl pH 7.5 solution at 40˚ C for 16 hours and concentrated using the QuiaEX II kit (Quiagen) according to the manufacturer instructions. The MOSIC template sequence encloses the following features: one or more oligonucleotide template sequences interspaced by short self-folding sequences with the aim to generate a restriction site for enzymatic cleavage of the polymer. Moreover, the sequence hosts also a nicking site or hybridization target site for an oligonucleotide to prime the enzymatic amplification reaction.

3.1.2 Linear RCA amplification

The combination of the so-prepared template with a RCA amplification reaction mixture: Φ29 polymerase (0.5 U/µl, Thermo Scientific), dNTPs (1 mM each, Thermo scientific) in a 1x Φ29 reaction buffer 33 mM Tris acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 % (v/v) Tween 20 and 1 mM Dithiothreitol (DTT), Thermo Scientific) is then incubated at 30

ºC. Such reaction is known to generate a great amount of product in short time. When a high yield and purity of the product is paramount, a long reaction (12-16 hours) is supplemented with T4 gene 32 ssDNA binding proteins (100 ng/µl, NEB) to inhibit the formation of dsDNA products. The reaction is stopped with a denaturation step (20 minutes, 75 ºC).

3.1.3 Enzymatic digestion and purification of the oligonucleotide of interest

The so-generated ssDNA polymer is enzymatically digested at the "hairpins" site, with the use of a type IIs restriction enzyme BseGI - GGATG(2/0)^ (Thermo Scientific) or BtsCI[165]. The choice of this enzyme guarantees maximum freedom in the design of oligonucleotide sequences. The digestion reaction is carried out with a 1:100 dilution of the enzyme for 12 hours at 50 ºC in 1x CutSmart (NEB) buffer. The degree of digestion and the purity of the desired oligonucleotide are assessed by a denaturing polyacrylamide gel (PAGE), 20% polyacrylamide, 20% formamide and 8 M urea dissolved in 1× TBE buffer. Nucleic acids are stained with SYBR® Gold (Thermo Scientific) and images are acquired using a Las 4000 imager (GE). The oligonucleotide can be extracted directly from the gel by diffusion or used as is after a desalting/buffer exchange step in an Amicon spin filter (Millipore).

3.1.4 Electrophoretic characterization of ssDNA

Despite denaturing (8M Urea) PAGE gel represents the best option to characterize oligonucleotides, native agarose gel (2% or above), ethidium bromide stained, are used to highlight the formation of undesired dsDNA[47].

3.2 RECOMBINANT PROTEINS DESIGN, EXPRESSION AND CHARACTERIZATION

The open reading frame encoding proteins expressed in this thesis were derived from NCBI Genes database or from the literature[166–168]. The coding sequence was optimized for the bacterial expression[169] and synthetized as dsDNA fragment (IDT) or fusion products that were obtained by PCR. Such fragments were cloned in expression vectors, such as pET-16b, under the control of an IPTG inducible T7 promoter. For affinity purification purpose a 6 or 10 His tag was also added in frame at the C or N terminus of the recombinant protein. A small-scale production of the recombinant protein was expressed in E. coli BL21 pLyss and IPTG induced overexpression of the protein of interest was validated by SDS-PAGE. Large scale production and purification was carried out by the protein science facility (Karolinska Institutet/SciLifeLab).

3.3 DESIGN OF MOSIC PSEUDOGENES AND OLIGONUCLEOTIDES FOR PROTEIN CONJUGATION

In silico analysis of the sequence enabled to reduce secondary structures that might hinder or inhibit the conjugation yield. For this purpose, the prediction tool Nupack[170] (http://www.nupack.org) was used to evaluate the complexity of both synthetic oligonucleotides and MOSIC-produced oligonucleotides. The template design of MOSIC oligonucleotides was generated via the online tool available on our website

(http://www.hogberglab.net/software/). Oligonucleotides to be conjugated with the mVirD2 tag were fused at the 5' with the T1 target sequence[69] while those to be used with click chemistry were ordered (IDT) with a 5' azido moiety.

3.4 DNA-PROTEIN CONJUGATION

3.4.1 mVirD2 tag mediated conjugation of oligonucleotides

The self-tagging domain mVirD2 was used alone or in fusion at the N terminus with other proteins for the sequence specific conjugation of an oligonucleotide. The conjugation reaction yields a site-specific conjugation product in biological conditions. Generally, protein and oligonucleotide are mixed at a 1:1 ratio in TKM buffer (50mMTris–HCl pH 8, 150 mM KCl, 1 mM $MgCl_2$, 10% glycerol) and incubated at 37 ºC for 1 to 2 hours in a PCR thermocycler. When using synthetic oligonucleotides, phosphorothioate bonds are introduced in the conjugation site to further increment the reaction yield[69].

3.4.2 Sortase A mediate conjugation of oligonucleotides

This two-step protocol takes advantage of the Sortase A enzyme (SrtA)[171]. Such enzyme recognizes a specific peptide tag (LEPTGG) and has the ability to cut the polypeptide at the glycine position and fuse it to another protein substrate with a N terminus glycine. The introduction of a DBCO-amine compound (Sigma Aldrich) in the reaction enables the site-specific modification of proteins for strain-promoted azide-alkyne cicloaddition[81] with an azido modified oligonucleotide. Briefly, a reaction containing 50 µM of the recombinant affibody, 150 µM of Sortase A and 10 mM of DBCO-amine (Sigma Aldrich) in Sortase A Buffer (HEPES 20 mM, NaCl 150 mM, $CaCl_2$ 10 mM) was incubated for 2.5 hours at 25˚ C. The conjugation product (affibody-DBCO) was purified flowing the crude reaction through a HisPur cobalt superflow agarose resin (Thermo Scientific) for 1.5 min at 1000 g. The coeluted DBCO-amine compound was removed using an Amicon® 3 kDa spin filter (Millipore). The conjugate was quantified with Bradford assay. The conjugation reaction was assessed in a Tris-Tricine 16% gel (Novex).

3.5 ELECTRO MOBILITY SHIFT ASSAY

The degree of conjugation with the mVirD2 tag was assessed with native PAGE gels (a fresh solution comprising polyacrylamide (19:1) final concentration 10% (Biorad), 0.5x TBE buffer and glycerol (2.5%, v/v) was mixed and polymerized by the addition of fresh 10% APS solution, final concentration 10%, and TEMED, final concentration 1%, in a mini-protean gel system (Biorad). TBE 0.5x was used as running buffer and the gels were run at room temperature for 30 min at 40 V/cm. Such gels enable to detect the intrinsic fluorescence of modified oligonucleotides or GFP protein, to stain all nucleic acids with SYBR® Gold (Thermo Scientific) and are compatible with total protein GelCode blue stain reagent (ThermoFischer).

3.6 ONE-POT SYNTHESIS OF OLIGONUCLEOTIDES AND CONJUGATION WITH MVIRD2-TAGGED PROTEINS

The sequential oligonucleotides synthesis and conjugation, with no need of buffer exchange, was made possible by the compatibility among amplification, digestion and conjugation. Typically, a MOSIC digestion reaction with BseGI is heat inactivated (75 ºC for 20 minutes) and slowly cooled down to 37 ºC. The mVirD2 chimeric protein is directly added to the reaction and incubated for 1 to 2 hours previous assessment on a 12 % Bis-Tris PAGE gel.

3.7 PREPARATION OF G-QUADRUPLEX OLIGONUCLEOTIDES

The stabilization of G-quadruplex structures in the oligonucleotide (IDT) used for signal amplification in conditions suitable for the conjugation with the mVirD2-C1 protein was optimized. The conjugation buffer was modified by increasing its potassium chloride concentration (modified TKM buffer: mTKM: containing 150 mM KCl, 3.5 mM $MgCl_2$, 10% glycerol and 50 mM Tris HCl, pH 8.0). Proper folding of secondary structures was enabled by heating the sample in mTKM buffer at 95 ºC for 5 minutes in a PCR thermocycler and gradually cooled down to 4 ºC at a rate of 0.1 ºC/s. Evaluation of the folding was made by gel electrophoresis and peroxidation assay.

3.8 CONJUGATION SCALE UP AND PURIFICATION OF THE OLIGONUCLEOTIDE-PROTEIN BIOSENSOR

Small scale conjugation reactions were performed similarly to what previously indicated while preparative conjugation reactions were performed in a 45 µl volume. The typical reaction comprises 22.5 µl of mTKM buffer, 22.5 µg of the fusion protein mVirD2-C1 and folded G4 oligonucleotides in molar ratios of 1:1 or 2:1 and incubated at 37 ºC for 2 hours. The crude reaction was speared in a 10 % native PAGE stained with SYBR® Gold (Thermo Scientific) and bands of interest were excised. Gel slices were soaked in TBS 1x buffer supplemented with 150 mM KCl for 4.5 h at 35°C under constant agitation (300 rpm). Next, the extracts were run through a 0.45 µm nylon filter device (VWR) in a bench-top centrifuge (90 sec at 16,800 rcf) and assessed for purity by gel electrophoresis.

3.9 COLORIMETRIC DNAZYME POX ASSAY

Prior to the experiment, a 4 µM hemin (Sigma-Aldrich) working solution was prepared by diluting a freshly thawed stock solution (1 mM in DMSO, stored at -20° C, protected from light) with mTKM buffer. Ten µL of this hemin stock solution were mixed with an equal volume of folded G-quadruplex oligonucleotides (1.6 µM, if not indicated differently) or gel-purified mVirD2-C1 DNAzyme conjugates (0.16 µM), respectively, and incubated in the dark for 30 min at 25 °C in a PCR thermocycler. At time point 0 of the reaction, 10 µL of freshly prepared 20 mM ABTS (Roche) and 3 µL of 3.75 mM H_2O_2 (Sigma-Aldrich) were added to the hemin-DNA mixture and the samples were mixed thoroughly for a few seconds (Figure S3A). The final working concentrations were 1.2 µM hemin, 0.5 µM DNA, 6 mM ABTS and 350 µM H_2O_2 in either mTKM buffer or, for experiments on gel-purified conjugates, in 1x TBS buffer containing 150 mM KCl. Two µL of the reaction mixture were then loaded onto a

μDrop™ Plate (Thermo Scientific) and the change of absorbance at 420 nm and 25°C was recorded every minute on a Varioskan Lux microplate reader (Thermo Scientific) with a measurement time of 250 milliseconds per read. Absorbance data were plotted with GraphPad Prism 8 after subtraction of background signal (absorbance in the absence of G-quadruplex DNA).

3.10 BRANCHED RCA IN SOLUTION AND ON DNA NANOSTRUCTURES

In this specific case, a circular ssDNA (1 ng/μL) serves as template in an RCA reaction primed by an oligonucleotide - P1- (100 nM) in a typical reaction mixture (0.5 U/μL Φ29 (Thermo Scientific), dNTPs 1 mM each in 1x Φ29 buffer). The branched amplification is triggered by a second oligonucleotide species - P2 - (100 nM) complementary to the previously generated DNA polymer. Such reaction is generally incubated for 30 minutes at 30 ºC and stopped by heat inactivation.

3.11 PREPARATION OF "READY TO SEQUENCE" LIBRARY AND NEXT GENERATION SEQUENCING

The information-dense DNA fragment generated during the RCA reaction is collected with the use of streptavidin functionalized magnetic beads (Thermo Scientific) and used as template for a two-step PCR.

4 RESULTS AND DISCUSSIONS

4.1 PAPER I

4.1.1 Φ29 polymerase driven RCA produces mainly dsDNA

In paper one, we studied the RCA with the aim to optimize the amplification reaction for the production of ssDNA oligonucleotides. In such process, we discovered that most of the product resulting from long incubations (> 10h) is in the form of dsDNA. This has been proved using two different templates, pUC19 and pBluscript II SK(+), on which an RCA reaction is initiated from a 3' end generated by a nicking endonuclease. In such experiment, the product of a 24-hour amplification – in absence or presence of single-stranded DNA binding proteins (SSBs) - is assessed in an agarose gel as is or after treatment with restriction endonucleases. It is pretty noticeable how the presence in the reaction mix of SSBs, in this case the T4 gene 32 protein, affects the total amount of DNA produced. However, a high molecular weight polymer is still noticeable in the gel well. Moreover, the treatment of the RCA product with restriction endonuclease in absence of SSBs leads to the predicted digestion pattern of regular dsDNA while, in the presence of SSBs, this pattern is absent. Such findings have also been confirmed by an additional experiment in which the RCA product generated from the pUC19 template in presence of a gradient of SSBs, as for the restriction assay, was combined with the dsDNA specific fluorescent dye PicoGreen®. Here it is noticeable how including the T4 gene 32 protein at a concentration of 100 ng/μL almost abolished the recorded fluorescence.

4.1.2 T4 gene 32 SSB prevents the formation of dsDNA in RCA

To investigate if SSBs simply inhibits RCA or rather specifically prevents the formation of dsDNA we assessed the product quality of an enzymatically produced 378 bases long oligonucleotide with the MOSIC protocol. Seven independent RCA reactions – in presence or absence of SSBs - using a nicked minicircle enclosing the oligonucleotide template were stopped after 30 minutes, 1, 3, 6, 9, 12, 24 hours. The obtained products were first digested with the restriction endonuclease BseGI and size separated on an agarose gel. It was then possible to appreciate that the digestion product of reactions incubated in absence of SSBs is running, according to the ladder, as a dsDNA of almost 400 bp while in presence of SSBs in the reaction most of the digested product is running faster (at the height corresponding to the 300 bp band of the ladder) as it is ssDNA. Nevertheless, the ethidium bromide is not staining with the same efficiency ssDNA and dsDNA therefore it is not possible to compare the reactions yield. For this reason, an additional experiment was designed. The same template (p378), was used for new RCA reactions in presence or absence of SSBs for up to 72 hours. After digestion, a part of the product was assessed on a native agarose gel to determine the ssDNA/dsDNA ratio while the second part of the sample was used for total DNA quantification in a denaturing PAGE gel stained with SYBR® Gold. The relative band intensity was used to determine the amount of total, single-stranded and dsDNA produced at each time point. It is

interesting to notice that in absence of SSBs, ssDNA is produced first and later (> 12 hours) converted to dsDNA while this is inhibited with SSBs without affecting the reaction yield.

4.2 PAPER II

4.2.1 Characterization of the self-tagging domain mVirD2

The Protein VirD2 from *Agrobacterium tumefaciens* is known to drive the genetic material transfer of pathogenic genes in the host cell. In this work we exploited this ability to generate a compact self-tagging domain for site specific conjugation of ssDNA to a protein of choice. The native protein exhibits two main functions, the N-terminus domain is involved with the recognition of the DNA sequence consensus that triggers the nicking coupled with the bond formation and a C-terminus helicase capable of unwinding dsDNA. It has also been demonstrated that a conserved tyrosine at position 29 is involved in the bond formation with the DNA molecule. For those reasons a minimal self-tagging domain was established by truncating the protein at the end of the N-terminus domain (amino acid 204) producing a 24.5 kDa. The tag was produced alone or in fusion with the green fluorescent protein (GFP) and its ability to recognize and establish a covalent bond with oligonucleotides was assessed with electro mobility shift assays where both nucleic acids and proteins were stained. This first assay confirmed that the purified minimal recombinant protein is active in binding the oligonucleotide in a dose dependent manner. Moreover, the presence of the GFP did not hinder the conjugation. A limitation of the technique itself resides in the reversibility of the conjugation reaction that imitates the effective conjugation to env. 40% of the proteins while the DNA is entirely bound to proteins if the latter are present in excess. We faced this limitation by developing some strategies to increase the conjugation yield by using chemically modified oligonucleotides. We observed that 3' end locked oligonucleotides could safely be incubated in a conjugation reaction in presence of exonucleases (3' to 5' activity) while the nicked byproduct fragment produced in the reaction is degraded. The degradation of this substrate reduces the chances for the reverse reaction to happen. A second strategy regards the use of phosphorothioate (PS) modified oligonucleotides. In this case, the reverse reaction is unfavorable for the charges distribution on the residues involved in the bond.

4.2.2 Sequential synthesis and conjugation of oligonucleotides

The main motivation behind the development of this conjugation strategy is to extend the capabilities of the previous implemented method for the enzymatic synthesis of oligonucleotides combining a conjugation step in a biologically neutral condition. To validate this combined protocol, the conjugation was first tested in the RCA restriction enzyme buffer. Later, two different pseudogenes were produced enclosing respectively a 103-nucleotide long oligonucleotide and two oligonucleotides of 87 and 57 bases. All oligos produced enzymatically for conjugation have been elongated at the 5' sequence with the T1 consensus (17 nucleotides long). Different PAGE gels enabled to highlight the different steps of the sequential method in respect of the DNA or the protein.

4.2.3 Applications of the new conjugation technique

As a general proof of relevance of the technique a few applications have been proposed. In the first case a DNA aptamer generated enzymatically was coupled with a HER2 aptamer. This biosensor was proved to associate with the recombinant protein *in vitro* (in solution or blotted) or to find application in sating biological samples. As a continuation of the project the protein was also transiently expressed extracellularly as chimera of the PDGF receptor's transmembrane domain. The incubation of a labeled oligonucleotide with the cells in PBS buffer supplemented with magnesium enabled to label the cells membranes. Another proof of concept was to explore the possibility to use mVirD2 DNA binding tag to direct transcription activator with the use of oligonucleotides pre-complexed to a luciferase plasmid.

4.3 PAPER III

4.3.1 Generation of a DNA-protein conjugate for antigen detection

In this work a specific application suggested in the previous paper was further explored: the construction of a biosensor for the detection of the toxic algae *Alexandrium minutum.* In this case the oligonucleotide is used as signal propagation tool by mean of the intrinsic capabilities of G rich oligonucleotides to operate as peroxidase-like DNAzymes while the mVirD2 domain was designed, produced and purified in fusion with a nanobody with a yield of 30-50 mg/L. The ability of the fusion nanobody to bind its antigen was assessed by immunostaining against *A. minutum* cells and detected with an anti His-tag secondary antibody. While the ssDNA binding activity was validated before with the use of electromobility shift assays, the biosensor was produced substantially as before, exception made for the oligonucleotide's secondary structure that needed a folding step and a buffer supplementation with further KCl (150 mM) to stabilize the G-quadruplex structure. Such deviations did not impact the conjugation ability furthermore the presence of the bound protein did not affect the selected DNAzyme peroxidation activity. To obtain a maximal response a set of previously validated DNAzymes, with the wanted function, were tested as for their ability to retain the secondary structure after conjugation (EMSA) and to induce a peroxidation (ABTS-H_2O_2 system) once complexed with hemin. The selected sequence T2-CatG4 is a fusion of the mVirD2 PS modified binding domain (T2) connected through a 3 or 12 nucleotide linker with the DNAzyme sequence (CatG4). The linker purpose is to reduce interactions between the protein and the DNAzyme even though a direct comparison of the different lengths did not influence the activity in ABTS oxidation or conjugation therefore we chose to conduct further experiments with T2-CatG4a (12 nucleotides spacer).

4.3.2 Production scale up and validation

The main challenge of this project was the production of enough conjugate for the validation. For this reason, a novel strategy is here proposed. Large scale reactions were incubated as before in a modified buffer (150 mM KCl) and verified by EMSA. To validate the activity of

the biosensor it was necessary to remove the unreacted oligonucleotides and proteins. For this purpose, the most effective strategy was to select the right conjugation product using a preparative PAGE gel. The so obtained biosensor was used in immune staining to check if the oligonucleotide hindered the antigen recognition or impaired the specificity for the target. Finally, the purified biosensor was used in an ELISA assay and absorbance of the oxidized ABTS substrate was measured a t 420nm. The recorded signal revealed a dependency on the cell number used on the assay that is clearly above the background signal.

4.4 PAPER IV

4.4.1 Design of a DNA-based approach for the detection of the organization of proteins of interest

In the present work we designed a strategy to collect information from the reciprocal localization of proteins of interest (POI). The amplification reaction relies on a rolling circle amplification of a circular ssDNA template embedding a unique molecular identifier (ID) that was produced by annealing to a padlock oligonucleotide and enzymatic ligation. Such a template is able to produce a long homo-polymer propagating locally the distinctive ID. The polymer is designed in such a way to enable hybridizations with a second species of uniquely identified oligonucleotides linked to protein binders. Those oligonucleotides can be extended, in the same reaction, on the polymer template thus providing a fusion product bearing the two IDs sequences (ID1 and ID2). The ability of producing such information-rich fragments, retrieving the embedded information and giving an interpretation to the collected data was assessed *in vitro*. In first place an RCA reaction is triggered with the use of the oligonucleotide P1. The presence of biotinylated P2 oligos enables the generation of the fusion strand and its selection from the mass of amplification product. A first step of enzymatic restriction enables to reduce the complexity of the sample and select with more precision the fusion product with streptavidin coated magnetic beads. The collected product is amplified by PCR for increasing the number of molecules and including the necessary adapters for the sequencing part.

4.4.2 Proof of concept and validation of the proposed workflow

Such described design was first tested *in vitro* on streptavidin-coated magnetic beads. This experiment proved that, once correctly designed, the oligonucleotides involved in the reaction are able to 1) generate an RCA product and 2) promote the fusion of DNA molecules - including unique molecular identifiers – during the same reaction. The information-rich fragment was at first PCR amplified and sequenced. Bulk sequencing (Sanger) of the information-rich pool of molecules highlighted the structure of the desired barcode. Similar experiments were reproduced in solution or on DNA nanostructures and modified in such a way to produce an NGS compatible library via a 2-step PCR protocol. DNA was sequenced with an Illumina Nextseq instrument. Analysis showed the possibility to recover single ID1/ID2 combinations and to reconstruct a connection network.

Next steps to prove the validity of the method will involve the use of fixed cell lines with known expression levels for the two antigens of choice (HER2 and EGFR). Different cell lines have been screened with the PLA method as a comparison method.

5 CONCLUSIONS

Historically, the miniaturization of devices in biomedicine aimed to reduce the costs by lowering the consumption of limiting reagents. A macroscopic example of this could be the invention of hand micropipettes that enabled the development of molecular biology. A similar motivation also underlined the development of DNA nanotechnology now more than 20 years ago even though researchers are still looking for a "killing" application of the method for which the metaphor of the colossus with clay feet would give a good description. The capability of investigating biological reactions at the nanoscale would provide us precious information to decipher physiological and pathological phenomena. However, too many intrinsic limitations of nanostructures - i.e. construction, stability, functionalization - often lead to an increasing level of complexity to solve problems resulting in an overcomplication of the strategy. More fundamental work is indeed needed to extend the applicability of such tools.

The overall aim of the papers presented here is to take a step back and work on the improvement of the basic building materials – DNA oligonucleotides (Paper I), DNA-protein conjugates (Paper II). As earlier discussed, there is a growing need of high-quality DNA for precision application not only connected to the construction of DNA nanostructures but also for example in the production of nucleic acids-based therapeutics or genome editing applications. In this thesis we proved that limitations of traditional solid-state chemical synthesis can be overcome with the use of enzymatic strategies. Such strategies have room for improvement and will most probably become a competitive alternative in the next decade. The ability of exploiting enzymatic systems to build biomolecules enables the integration of several modules that are traditionally separated as, for instance, proved with the combination of synthesis and targeted conjugation to proteins.

As a secondary aim of this thesis, the outcomes of the new tools provided were leveraged to be applied for the solutions of problems. The *trait d'union* with this section is represented by the possibility of using the minimal unit of a DNA-protein conjugate as a biosensor (Paper III). In this case, we wanted to prove how environmental point of analysis devices could greatly benefit from the modular assembly, a key concept in nanotechnology, to produce consistent, cost-effective and time saving biosensors.

The same technology also enabled the production of monomeric binders conjugated with DNA in a 1:1 ratio. They have been used in the last presented manuscript (Paper IV). Starting from this point we wanted to prove that it is indeed possible to (a) locally grow a DNA amplification product able to (b) create a bipartite network among biomolecules of interest encoding (c) a reciprocal spatial organization in a couple of DNA barcodes. Even though a minimal proof of concept of the approach has been provided, the so far obtained data look encouraging. Further work is indeed necessary to fully validate the methodological approach and its capabilities.

6 ACKNOWLEDGEMENTS

Here it is, after ~50.000 hours, including ~50 lab meeting presentations, 6 EScoDNA ITN meetings, 2 lab relocations and a few more wrinkles and gray hairs I reached the conclusion of my journey as a PhD student. What an experience! All this would have not been possible without the help, support, company and encouragement of many people.

First, I would like to express my deepest appreciation to my Thesis committee; my supervisor **Björn** and all my co-supervisors **Ana**, **Mia**, **Andreas** and my mentor **Patrizia**. I am extremely grateful to my past, present and - who knows - future *colleagues* **Mino**, **Ferenc**, **Iris**, **Marco**, **Yang**, **Yunshi**, **Ioanna**, **Ian**, **Igor**, **João**, **Esther**, **Anurupa**, **Erik**, **Alan**, **Yin** **and Masa**, and our students **Sally**, **Larsen**, **Sólrun**, **Victoria**: you all made me learn a lot and I very much appreciated all the time we spent together inside and outside the lab. Thank you as well to all my friends, especially to **Kati**, **Lisa**, **Gabriele**, **Iñes** and **Teresa**. I would also like to thank all the peers and supervisors of the **EScoDNA ITN network** and I would like to recognize the help that I received from **Lise**. In particular, I would like to extend my sincere thanks to **Ebbe** and his lab crew **Guido**, **Denis**, **Steffen**, **Ane**, **Cody** and likewise to **Kurt's** lab members, especially to **Ilenia** and **Mattia** who kindly hosted me in Aarhus. I am also very pleased for all the time spent with the 2017 Stockholm iGEM team and especially with the "biocontainment" team **Hrista** and **Achilleas** and the communication team **Jutta** and **Rasmus**. I would also like to extend my sincere thanks to all the people in the biomaterials division / B9 quarter for making the working environment a better place and especially my gratitude goes to **Björn R.**, **Christos**, **Elena**, **Toon**, **Alessandro**, **Christina**, **Trixie**, **Hannah** but also to **Katia**, **Hannes**, **Lara**, **Lorenzo**, **Luca**, **Sarah**, **Hampus** **and Ileana** and all the RNA team as well as to **Chris**, **Maggie**, **Allie**, **Hélène** and all the **Stevens group**. Additionally, I would like to thank our former neighbors at neuroscience **Domenico**, **Ben**, **Anil**, and **François**. I also need to mention my collaborator from Slovenia, **Ario** and the people at the **PSF facility** for making my projects possible. Before thanking my biggest fan club at home, a special mention is reserved to "**Marie Curie**" for the funding of the first half of this adventure. Per finire desidero ringraziare tutta la mia famiglia in particolare mio fratello **Francesco** ed i miei genitori **Stefano** e **Rossana**. Giunto a questo punto non posso che concludere ringraziando la persona che ogni giorno, indipendentemente dal sole o dalla pioggia, mi ha dato il suo sostegno incondizionato e trasmesso la forza e motivazione che ha reso possibile tutto ciò; grazie alla mia piccola, dolce, nuova famiglia &.

7 REFERENCES

1. Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of the Pneumococcal types : induction of transformation by a desoxyribonucleic acid isolated from Pneumococcus type III. *J. Exp. Med.* (1944).

2. Neidle, S. *Principles of Nucleic Acid Structure. Principles of Nucleic Acid Structure* (2008). doi:10.1016/B978-0-12-369507-9.X5001-8

3. BASS, L. W. PHOEBUS AARON THEODOR LEVENE 1869-1940. *Science (80-.).* (2006). doi:10.1126/science.92.2392.392

4. Miescher, F. Ueber die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen* (1871). doi:10.1002/andp.18491530607

5. JONES, M. E. Albrecht Kossel, a biographical sketch. *Yale J. Biol. Med.* (1953).

6. Franklin, R. E. & Gosling, R. G. Evidence for 2-chain Helix in crystalline structure of sodium deoxyribonucleate. *Nature* (1953). doi:10.1038/172156a0

7. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. Molecular structure of deoxypentose nucleic acids. in *50 Years of DNA* (2016). doi:10.1038/nature01396

8. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171**, 737–738 (1953).

9. Tischer, I., Gelderblom, H., Vettermann, W. & Koch, M. A. A very small porcine virus with circular single-stranded DNA. *Nature* (1982). doi:10.1038/295064a0

10. Krupovic, M. & Forterre, P. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann. N. Y. Acad. Sci.* **1341**, 41–53 (2015).

11. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).

12. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).

13. Shaw, A. *et al.* Spatial control of membrane receptor function using ligand nanocalipers. *Nat. Methods* **11**, 841–6 (2014).

14. Ke, R., Mignardi, M., Botling, J., Wählby, C. & Nilsson, M. In situ sequencing for RNA analysis in preserved tissue and cells In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

15. Gijs, M. *et al.* Improved aptamers for the diagnosis and potential treatment of HER2-positive cancer. *Pharmaceuticals* **9**, (2016).

16. Kankia, B. I. & Marky, L. A. Folding of the thrombin aptamer into a G-quadruplex with Sr2+: Stability, heat, and hydration. *J. Am. Chem. Soc.* (2001). doi:10.1021/ja010008o

17. Chai, C., Xie, Z. & Grotewold, E. SELEX (systematic evolution of ligands by exponential enrichment), as a powerful tool for deciphering the protein-DNA

interaction space. *Methods Mol. Biol.* **754**, 249–258 (2011).

18. Zhou, W., Ding, J. & Liu, J. Theranostic dnazymes. *Theranostics* **7**, 1010–1025 (2017).

19. Huang, H. *et al.* A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **10**, 686–691 (2014).

20. M Michelson, B. A. & Alexander Todd, S. R. Synthesis of a Dithymidine Dinucleotide Containing a 3',5'-Internucleotide linkage. *Biochim. Biophys. Acta J . Biot. Chern* **8350208**, (1952).

21. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites-A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).

22. H. G. Khorana, H. Büchi, M. H. Caruthers, S. H. Chang, N. K. Gupta, A. Kumar, E. Ohtsuka, V. Sgaramella, and H. W. Progress in the Total Synthesis of the Gene for ala-tRNA. *Cold Spring Harb Symp Quant Biol* **33**, 35–44 (1968).

23. Itakura, K. *et al.* Expression in Escherichia coli of a Chemically Synthesized Gene for the Hormone Somatostatin. *Science (80-.).* **198**, 1056–1063 (1977).

24. Rayner, S. *et al.* MerMade: An oligodeoxyribonucleotide synthesizer for high throughput oligonucleotide production in dual 96-well plates. *Genome Res.* **8**, 741–747 (1998).

25. Semenyuk, A. *et al.* Cartridge-based high-throughput purification of oligonucleotides for reliable oligonucleotide arrays. *Anal. Biochem.* **356**, 132–141 (2006).

26. Badi, N. & Lutz, J.-F. Sequence control in polymer synthesis. *Chem. Soc. Rev.* **38**, 3383–3390 (2009).

27. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: Technologies and applications. *Nature Methods* **11**, 499–507 (2014).

28. LeProust, E. M. *et al.* Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010).

29. Biolegio. No Title. https://www.biolegio.com/products-services/long-ol

30. Biosynthesis. No Title. https://www.biosyn.com/long-dna-oligonucleotide-sy

31. Fodor, S. P. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).

32. Hughes, R. A. & Ellington, A. D. Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).

33. Reagin, M. J. *et al.* Templiphi: A sequencing template preparation procedure that eliminates overnight cultures and DNA purification. *J. Biomol. Tech.* **14**, 143–148 (2003).

34. Esteban, J. A., Salas, M. & Blanco, L. Fidelity of φ29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* **268**, 2719–2726 (1993).

35. Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. & Khorana, H. G. Studies on polynucleotides. XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* **56**, 341–361 (1971).

36. Citartan, M. *et al.* Asymmetric PCR for good quality ssDNA generation towards DNA aptamer production. *Songklanakarin J. Sci. Technol.* **34**, 125–131 (2012).

37. Marimuthu, C., Tang, T.-H., Tominaga, J., Tan, S.-C. & Gopinath, S. C. B. Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst* **137**, 1307 (2012).

38. Poddar, S. Symmetric vs asymmetric PCR and molecular beacon probe in the detection of a target gene of adenovirus. *Mol. Cell. Probes* **14**, 25–32 (2000).

39. Dongran Han1, Xiaodong, Cameron Myhrvold, Bei Wang, Mingjie Dai, Shuoxing Jiang, Maxwell Bates, Yan Liu, Byoungkwon An, Fei Zhang, Hao Yan, P. Y. Single-stranded DNA and RNA origami. *Science (80-.).* **358**, (2017).

40. Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099 (2001).

41. Silander, K. & Saarela, J. Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol. Biol.* **439**, 1–18 (2008).

42. Johne, R., Müller, H., Rector, A., van Ranst, M. & Stevens, H. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in Microbiology* **17**, 205–211 (2009).

43. Bagchi, S., Fredriksson, R. & Wallén-Mackenzie, Å. In Situ Proximity Ligation Assay (PLA). *Methods Mol. Biol.* **1318**, 149–159 (2015).

44. Ke, R., Nong, R. Y., Fredriksson, S., Landegren, U. & Nilsson, M. Improving Precision of Proximity Ligation Assay by Amplified Single Molecule Detection. *PLoS One* **8**, (2013).

45. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **20**, 473–477 (2002).

46. Ducani, C., Kaul, C., Moche, M., Shih, W. M. & Högberg, B. Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides. *Nat. Methods* **10**, 647–52 (2013).

47. Ducani, C., Bernardinelli, G. & Högberg, B. Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Res.* **42**, 10596–10604 (2014).

48. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552**, 84 (2017).

49. Gu, H. & Breaker, R. R. Production of single-stranded DNAs by self-cleavage of rolling-circle amplification products. *Biotechniques* **54**, 337–343 (2013).

50. Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K. & Doudna, J. A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell* **60**, 398–

407 (2015).

51. Lee, H. H., Kalhor, R., Goela, N., Bolot, J. & Church, G. M. Enzymatic DNA synthesis for digital information storage. *bioRxiv* (2018). doi:10.1101/348987

52. Palluk, S. *et al.* De novo DNA synthesis using polymerasenucleotide conjugates. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4173

53. Perkel, J. M. The race for enzymatic DNA synthesis heats up. *Nature* (2019). doi:10.1038/d41586-019-00682-0

54. No Title. Available at: http://www.dnascript.co/.

55. No Title. Available at: https://www.nuclera.com/technology/science/.

56. No Title. Available at: https://www.evonetix.com/.

57. No Title. Available at: http://www.moligotechnologies.com/.

58. Shepherd, T. R., Du, R. R., Huang, H., Wamhoff, E. C. & Bathe, M. Bioproduction of pure, kilobase-scale single-stranded DNA. *Sci. Rep.* (2019). doi:10.1038/s41598-019-42665-1

59. Kick, B., Praetorius, F., Dietz, H. & Weuster-Botz, D. Efficient Production of Single-Stranded Phage DNA as Scaffolds for DNA Origami. *Nano Lett.* (2015). doi:10.1021/acs.nanolett.5b01461

60. Datta, H. J. & Glazer, P. M. Intracellular generation of single-stranded DNA for chromosomal triplex formation and induced recombination. *Nucleic Acids Res.* **29**, 5140–7 (2001).

61. Chen, Y., Ji, Y. J. & Conrad, C. Expression of ssDNA in mammalian cells. *Biotechniques* **34**, 167–171 (2003).

62. Chen, Y. & McMicken, H. W. Intracellular production of DNA enzyme by a novel single-stranded DNA expression vector. *Gene Ther.* **10**, 1776–1780 (2003).

63. Shaw, A. *et al.* Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nat. Nanotechnol.* (2019). doi:10.1038/s41565-018-0336-3

64. Grossi, G., Dalgaard Ebbesen Jepsen, M., Kjems, J. & Andersen, E. S. Control of enzyme reactions by a reconfigurable DNA nanovault. *Nat. Commun.* (2017). doi:10.1038/s41467-017-01072-8

65. Langecker, M. *et al.* Synthetic lipid membrane channels formed by designed DNA nanostructures. *Science* **338**, 932–6 (2012).

66. Thazhathveetil, A. K., Liu, S. T., Indig, F. E. & Seidman, M. M. Psoralen conjugates for visualization of genomic interstrand cross-links localized by laser photoactivation. *Bioconjug. Chem.* (2007). doi:10.1021/bc060309t

67. Liu, T., Song, P., Märcher, A., Kjems, J., Yang, C. and Gothelf, K. V. Selective Delivery of Doxorubicin to EGFR+ Cancer Cells by Cetuximab–DNA Conjugates. *ChemBioChem* **20**, 1014–1018 (2019).

68. Bailey, R. C., Kwong, G. A., Radu, C. G., Witte, O. N. & Heath, J. R. DNA-encoded antibody libraries: A unified platform for multiplexed cell sorting and detection of

genes and proteins. *J. Am. Chem. Soc.* **129**, 1959–1967 (2007).

69. Bernardinelli, G. & Högberg, B. Entirely enzymatic nanofabrication of DNA-protein conjugates. *Nucleic Acids Res.* **45**, e160 (2017).

70. Kukolka, F. & Niemeyer, C. M. Synthesis of fluorescent oligonucleotide-EYFP conjugate: Towards supramolecular construction of semisynthetic biomolecular antennae. *Org. Biomol. Chem.* (2004). doi:10.1039/b406492e

71. Debets, M. F. *et al.* Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* (2010). doi:10.1039/b917797c

72. Yang, Y. R., Liu, Y. & Yan, H. DNA Nanostructures as Programmable Biomolecular Sca ff olds. (2015). doi:10.1021/acs.bioconjchem.5b00194

73. Rosen, C. B. *et al.* Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. *Nat. Chem.* (2014). doi:10.1038/nchem.2003

74. Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

75. Cong, Y. *et al.* Site-specific PEGylation at histidine tags. *Bioconjug. Chem.* (2012). doi:10.1021/bc200530x

76. Hino, N., Hayashi, A., Sakamoto, K. & Yokoyama, S. Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code. *Nat. Protoc.* **1**, 2957–2962 (2007).

77. Hohsaka, T. & Sisido, M. Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology* **6**, 809–815 (2002).

78. Lovendahl, K. N., Hayward, A. N. & Gordon, W. R. Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. *J. Am. Chem. Soc.* **139**, 7030–7035 (2017).

79. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie - International Edition* (2001). doi:10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5

80. Schweigert, N. *et al.* DNA degradation by the mixture of copper and catechol is caused by DNA-copper-hydroperoxo complexes, probably DNA-Cu(I)OOH. *Environ. Mol. Mutagen.* (2000). doi:10.1002/1098-2280(2000)36:1<5::AID-EM2>3.0.CO;2-4

81. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* (2006).

82. Chang, P. V. *et al.* Copper-free click chemistry in living animals. *Proc. Natl. Acad. Sci.* (2010). doi:10.1073/pnas.0911116107

83. Yoon, H. I. *et al.* Bioorthogonal Copper Free Click Chemistry for Labeling and Tracking of Chondrocytes In Vivo. *Bioconjug. Chem.* (2016). doi:10.1021/acs.bioconjchem.6b00010

84. IDT. Available at: https://eu.idtdna.com/site/Catalog/Modifications/ClickChemistry/8.

85. Tyagi, S. & Lemke, E. A. Genetically Encoded Click Chemistry for Single-Molecule FRET of Proteins. in *Methods in Cell Biology* (2013). doi:10.1016/B978-0-12-407239-8.00009-4

86. Gong, H. *et al.* Simple Method to Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Protein Detection in Single Cells. *Bioconjug. Chem.* (2016). doi:10.1021/acs.bioconjchem.5b00613

87. Trads, J. B., Tørring, T. & Gothelf, K. V. Site-Selective Conjugation of Native Proteins with DNA. *Acc. Chem. Res.* (2017). doi:10.1021/acs.accounts.6b00618

88. Bornholt, J. *et al.* A DNA-Based Archival Storage System. *ASPLOS '16 - Proc. Twenty-First Int. Conf. Archit. Support Program. Lang. Oper. Syst.* 637–649 (2016). doi:10.1145/2872362

89. Bornholt, J. *et al.* Toward a DNA-Based Archival Storage System. *IEEE Micro* **37**, 98–104 (2017).

90. Willner, I. Biomaterials for Sensors, Fuel Cells, and Circuitry. *Science (80-.).* **298**, 2407–2408 (2002).

91. Van Den Beucken, J. J. J. P. *et al.* Fabrication, characterization, and biological assessment of multilayered DNA-coatings for biomaterial purposes. *Biomaterials* **27**, 691–701 (2006).

92. Mullis, K. B. The unusual origin of the polymerase chain reaction. *Sci. Am.* (1990). doi:10.1038/scientificamerican0490-56

93. Conner, B. J. *et al.* Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci.* **80**, 278–282 (1983).

94. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879 (1989).

95. Fang, X., Sen, A., Vicens, M. & Tan, W. Synthetic DNA aptamers to detect protein molecular variants in a high-throughput fluorescence quenching assay. *ChemBioChem* **4**, 829–834 (2003).

96. Lee, J. F. Aptamer Database. *Nucleic Acids Res.* (2003). doi:10.1093/nar/gkh094

97. Bülbül, G., Hayat, A. & Andreescu, S. ssDNA-Functionalized Nanoceria: A Redox-Active Aptaswitch for Biomolecular Recognition. *Adv. Healthc. Mater.* **5**, 822–828 (2016).

98. Goux, E. *et al.* A colorimetric nanosensor based on a selective target-responsive aptamer kissing complex. *Nanoscale* **9**, 4048–4052 (2017).

99. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **28**, E63 (2000).

100. Kishi, J. Y., Schaus, T. E., Gopalkrishnan, N., Xuan, F. & Yin, P. Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* (2017). doi:10.1038/nchem.2872

101. Evanko, D. Hybridization chain reaction. *Nat. Methods* **1**, 186–187 (2004).

102. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science (80-.).* **270**, 467–470 (1995).

103. Raap, A. K. *et al.* Ultra-sensitive fish using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum. Mol. Genet.* **4**, 529–534 (1995).

104. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–318 (2014).

105. Seeman, N. Nucleic acid nanostructures and topology. *Angew. Chemie Int. Ed.* **37**, 3220–3238 (1998).

106. Benson, E. *et al.* Computer-Aided Production of Scaffolded DNA Nanostructures from Flat Sheet Meshes. *Angew. Chemie - Int. Ed.* **55**, 8869–8872 (2016).

107. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).

108. Kick, B., Hensler, S., Praetorius, F., Dietz, H. & Weuster-Botz, D. Specific growth rate and multiplicity of infection affect high-cell-density fermentation with bacteriophage M13 for ssDNA production. *Biotechnol. Bioeng.* **114**, 777–784 (2017).

109. Shaw, A. *et al.* Spatial control of membrane receptor function using ligand nanocalipers. *Nat. Methods* **11**, (2014).

110. Ali, M. M. *et al.* Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.* **43**, 3324 (2014).

111. Guntaka, R. V., Varma, B. R. & Weber, K. T. Triplex-forming oligonucleotides as modulators of gene expression. *International Journal of Biochemistry and Cell Biology* **35**, 22–31 (2003).

112. Knauert, M. P. & Glazer, P. M. Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Hum. Mol. Genet.* **10**, 2243–2251 (2001).

113. Faria, M. & Giovannangeli, C. Triplex-forming molecules: From concepts to applications. *J. Gene Med.* **3**, 299–310 (2001).

114. Rogers, F. A., Manoharan, M., Rabinovitch, P., Ward, D. C. & Glazer, P. M. Peptide conjugates for chromosomal gene targeting by triplex-forming oligonucleotides. *Nucleic Acids Res.* **32**, 6595–6604 (2004).

115. Havre, P. A., Gunther, E. J., Gasparro, F. P. & Glazer, P. M. Targeted mutagenesis of DNA using triple helix-forming oligonucleotides linked to psoralen. *Proc. Natl. Acad. Sci.* **90**, 7879–7883 (1993).

116. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

117. Quan, J. *et al.* Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.* **29**, 449–452 (2011).

118. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).

119. Wu, Y., Li, Q. & Chen, X. Z. Detecting protein-protein interactions by far western blotting. *Nat. Protoc.* **2**, 3278–3284 (2007).

120. Melton, L. Protein arrays: Proteomics in multiplex. *Nature* **429**, 101–107 (2004).

121. Kasai, K. Affinity chromatography. *Tanpakushitsu Kakusan Koso Protein Nucleic Acid Enzym.* **31**, 1143–1151 (1986).

122. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences* **10**, 2763–2788 (2009).

123. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-.).* **228**, 1315–1317 (1985).

124. Clegg, R. M. Fluorescence resonance energy transfer. *Curr. Opin. Biotechnol.* **6**, 103–110 (1995).

125. Bacart, J., Corbel, C., Jockers, R., Bach, S. & Couturier, C. The BRET technology and its application to screening assays. *Biotechnology Journal* **3**, 311–324 (2008).

126. Zeng, S., Baillargeat, D., Ho, H.-P. & Yong, K.-T. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. *Chem. Soc. Rev.* **43**, 3426 (2014).

127. Gregory, N. W. Elements of X-Ray Diffraction. *J. Am. Chem. Soc.* **79**, 1773–1774 (1957).

128. Fancy, D. A. & Kodadek, T. Chemistry for the analysis of protein–protein interactions: Rapid and efficient cross-linking triggered by long wavelength light. *Chem. Biochem.* **96**, 6020–6024 (1999).

129. Lommerse, P. H. M. *et al.* Single-Molecule Imaging of the H-Ras Membrane-Anchor Reveals Domains in the Cytoplasmic Leaflet of the Cell Membrane. *Biophys. J.* **86**, 609–616 (2004).

130. Wawrezinieck, L., Rigneault, H., Marguet, D. & Lenne, P. F. Fluorescence correlation spectroscopy diffusion laws to probe the submicron cell membrane organization. *Biophys. J.* **89**, 4029–4042 (2005).

131. Whited, A. M. & Park, P. S.-H. Atomic force microscopy: a multifaceted tool to study membrane proteins and their interactions with ligands. *Biochim. Biophys. Acta* **1838**, 56–68 (2014).

132. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

133. Sano, T., Smith, C. L. & Cantor, C. R. Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science (80-.).* (1992). doi:10.1126/science.1439758

134. Cao, Y., Kopplow, K. & Liu, G. Y. In-situ immuno-PCR to detect antigens. *Lancet* (2000). doi:10.1016/S0140-6736(00)02696-9

135. Nong, R. Y., Gu, J., Darmanis, S., Kamali-Moghaddam, M. & Landegren, U. DNA-assisted protein detection technologies. *Expert Review of Proteomics* (2012). doi:10.1586/epr.11.78

136. Gullberg, M. *et al.* Cytokine detection by antibody-based proximity ligation. *Proc. Natl. Acad. Sci.* **101**, 8420–8424 (2004).

137. Alvarez-Castelao, B. *et al.* Cell-type-specific metabolic labeling of nascent proteomes in vivo. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.4016

138. Bar, D. Z. *et al.* Biotinylation by antibody recognition—a method for proximity labeling. *Nat. Methods* (2017).

139. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nature Reviews Molecular Cell Biology* (2019). doi:10.1038/s41580-018-0094-y

140. Gatto, L., Breckels, L. M. & Lilley, K. S. Assessing sub-cellular resolution in spatial proteomics experiments. *Current Opinion in Chemical Biology* (2019). doi:10.1016/j.cbpa.2018.11.015

141. Alberts, B. *et al. Molecular Biology of the Cell 6e. Garland Science* **6**, (2014).

142. Martin, W. *et al.* Modern endosymbiotic theory: Getting lateral gene transfer in- to the equation. *J. Endocytobiosis Cell Res. Int. Soc. Endocytobiol. Endocytobiosis Cell Res. J. Endocytobiosis Cell Res. J. Endocytobiosis Cell Res.* **23**, 1–5 (2012).

143. Li, X. & Liu, D. R. DNA-templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie - International Edition* **43**, 4848–4870 (2004).

144. Jacobson, K., Mouritsen, O. G. & Anderson, R. G. W. Lipid rafts: at a crossroad between cell biology and physics. *Nat. Cell Biol* **9**, 7–14 (2007).

145. Kusumi, A., Ike, H., Nakada, C., Murase, K. & Fujiwara, T. Single-molecule tracking of membrane molecules: Plasma membrane compartmentalization and dynamic assembly of raft-philic signaling molecules. *Seminars in Immunology* **17**, 3–21 (2005).

146. Wu, C. Focal adhesion: a focal point in current cell biology and molecular medicine. *Cell Adh Migr* **1**, 13–18 (2007).

147. Dustin, M. L. The immunological synapse. *Cancer Immunol. Res.* **2**, 1023–33 (2014).

148. Anderson, R. G. W. A Role for Lipid Shells in Targeting Proteins to Caveolae, Rafts, and Other Lipid Domains. *Science (80-.).* **296**, 1821–1825 (2002).

149. Malinska, K., Malinsky, J., Opekarova, M. & Tanner, W. Visualization of protein compartmentation within the plasma membrane of living yeast cells. *Mol. Biol. Cell* **14**, 4427–4436 (2004).

150. Prior, I. A., Muncke, C., Parton, R. G. & Hancock, J. F. Direct visualization of ras proteins in spatially distinct cell surface microdomains. *J. Cell Biol.* **160**, 165–170 (2003).

151. Friedrichson, T. & Kurzchalia, T. V. Microdomains of GPI-anchored proteins in living cells revealed by crosslinking. *Nature* **394**, 802–805 (1998).

152. Garcia-Parajo, M. F., Cambi, A., Torreno-Pina, J. A., Thompson, N. & Jacobson, K. Nanoclustering as a dominant feature of plasma membrane organization. *J. Cell Sci.* **127**, 4995–5005 (2014).

153. Lingwood, D. *et al.* Cholesterol modulates glycolipid conformation and receptor activity. *Nat. Chem. Biol.* **7**, 260–262 (2011).

154. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–175 (2014).

155. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci.* **89**, 5381–5383 (1992).

156. Wu, Z. *et al.* Cell-Based Selection Expands the Utility of DNA-Encoded Small-Molecule Library Technology to Cell Surface Drug Targets: Identification of Novel Antagonists of the NK3 Tachykinin Receptor. *ACS Comb. Sci.* **17**, 722–731 (2015).

157. Blakskjaer, P., Heitner, T. & Hansen, N. J. V. Fidelity by design: Yoctoreactor and binder trap enrichment for small-molecule DNA-encoded libraries and drug discovery. *Current Opinion in Chemical Biology* **26**, 62–71 (2015).

158. Shendure, J. *et al.* DNA sequencing at 40: Past, present and future. *Nature* (2017). doi:10.1038/nature24286

159. Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975–987 (2016).

160. Peikon, I. D. *et al.* Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res.* **45**, (2017).

161. Weinstein, J. A., Regev, A. & Zhang, F. DNA Microscopy: Optics-free Spatio-genetic Imaging by a Stand-Alone Chemical Reaction. *Cell* (2019). doi:10.1016/j.cell.2019.05.019

162. Boulgakov, A. A., Xiong, E., Bhadra, S., Ellington, A. D. & Marcotte, E. M. From Space to Sequence and Back Again: Iterative DNA Proximity Ligation and its Applications to DNA-Based Imaging. *bioRxiv* (2018). doi:10.1101/470211

163. Hoffecker, I. T., Yang, Y., Bernardinelli, G., Orponen, P. & Högberg, B. A Computational Framework for DNA Sequencing-Based Microscopy. *bioRxiv* 476200 (2018). doi:10.1101/476200

164. Schaus, T. E., Woo, S., Xuan, F., Chen, X. & Yin, P. A DNA nanoscope via auto-cycling proximity recording. *Nat. Commun.* **8**, 696 (2017).

165. Rosa, J., Fernandez-Gonzalez, E., Ducani, C. & Högberg, B. BtsCI and BseGI display sequence preference in the nucleotides flanking the recognition sequence. *PLoS One* (2018). doi:10.1371/journal.pone.0202057

166. van Kregten, M., Lindhout, B. I., Hooykaas, P. J. J. & van der Zaal, B. J. Agrobacterium-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a type IV secretion system translocation signal. *Mol. Plant. Microbe. Interact.* **22**, 1356–1365 (2009).

167. Orlova, A. *et al.* Synthetic affibody molecules: A novel class of affinity ligands for molecular imaging of HER2-expressing malignant tumors. *Cancer Res.* (2007). doi:10.1158/0008-5472.CAN-06-2887

168. Glasgow, J. E., Salit, M. L. & Cochran, J. R. In Vivo Site-Specific Protein Tagging with Diverse Amines Using an Engineered Sortase Variant. *J. Am. Chem. Soc.* (2016).

doi:10.1021/jacs.6b03836

169. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki376

170. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).

171. Parthasarathy, R., Subramanian, S. & Boder, E. T. Sortase A as a novel molecular 'stapler' for sequence-specific protein conjugation. *Bioconjug. Chem.* (2007). doi:10.1021/bc060339w