UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXVIII CICLO DEL DOTTORATO DI RICERCA IN NANOTECNOLOGIE

HIGH-RESOLUTION NUCLEIC ACID ANALYSIS WITH A DNA NANOTECHNOLOGY APPROACH

Settore scientifico-disciplinare: BIO/14

DOTTORANDA
LUCIA CORAL Lucia Coral

COORDINATORE
PROF. LUCIA PASQUATO

SUPERVISORE DI TESI
PROF. MATTEO CASTRONOVO

ANNO ACCADEMICO 2015/2016

Abstract (English)

The goal of my research program is to develop a DNA-based nanosensor for nucleic acids analysis. I plan to use DNA Origami nanostructures that are formed by a-few-thousand-nucleotides-long, circular, single stranded (ss)DNA "scaffold" folded to form a specific shape by the action of a few hundreds of short (approx. 30 nucleotides) ssDNA "staples", which hybridize over non-consecutive regions of the scaffold. Staples can be incorporated within the structure with well-defined stoichiometry and some of them can be designed to serve as highly-specific receptor for short nucleic acids sequences. I plan to introduce a restriction site within staples adjacent to such probes to permit their steric protection from enzymatic degradation as a consequence of a probe-target recognition event in their vicinity. The restriction reaction, therefore, "writes" the amount of target molecules captured on the nanosensors by permanently modifying certain target-specific staples within the DNA nanostructure. In turn, the amount of such modified staples is associated with the amount of target molecules captured on the nanosensor surface from the solution, and can be subsequently analyzed with standard DNA quantification techniques such as quantitative PCR (qPCR), or high-throughput DNA sequencing. This research plan is based on recent results obtained in our laboratory showing that, in self-assembled DNA structures, restriction enzymatic reactions are steric-regulated in a step-wise fashion. Therefore, one goal of this PhD thesis is the development of a quantitative assay to evaluate the efficiency of enzymatic reactions within such nanostructures, as well as staples incorporation and stability for aiding fundamental studies of enzymatic reactions in DNA nanostructures. Specifically, the restriction quantification method proposed is based on a linear PCR (L-PCR) amplification reaction that involves staple-specific carriers (70-nucleotide-long ssDNA) that can fully hybridize to staple fragments produced by the enzymatic cutting. The polymerase action leads to the formation of a duplex DNA fragment 70 base pairs (bp) long for each cleaved staple, whereas the hybridization of un-cleaved staples on the carrier prevents such polymerase reaction. To study staples incorporation efficiency, the same protocol can be used, but DNA carriers are designed to hybridize the full length of the DNA staple sequence. I prepared L-PCR samples to

evaluate next-generation sequencing (NGS) quantification accuracy and L-PCR efficiency. As a proof-of-concept, I analyzed 5 staples of a triangular DNA nanostructure and obtained information on their incorporation efficiency or cleavage.

This thesis also describes the design and development of a DNA based nanosensor for improving the accuracy of short nucleic acid quantification with qPCR. The work aims at coupling qPCR with a self-assembled nanosensor, which can help overcome amplification and retro-transcription reaction bias, and circumvent the detection threshold of 2-fold concentration variation, without requiring updates to traditional qPCR instrumentation. Components of such sensor are three consecutive "foot-loop" DNA probes each carrying a target-complementary sequence in the loop. Probes are assembled over a common scaffold that joins their "feet". Each "foot" carries a restriction site and upon hybridization of three copies of the same target molecule on the respective loops, the site of each foot is destabilized (termed "bingo" configuration). Only in this case, the whole scaffold is protected from enzymatic cleavage and can be amplified with PCR. Target and bingo-scaffold concentrations are correlated by power function of 3. The results obtained demonstrate the increased accuracy of the Bingo-qPCR assay with respect to standard qPCR in evaluating small variation of enzymatic activity, and prove the feasibility of the target detection switch-based reaction.

Abstract (Italiano)

L'obiettivo finale del mio programma di ricerca è di analizzare acidi nucleici sfruttando nanostrutture di DNA come sensori. Tali strutture sono solitamente formate da una componente a singolo filamento di DNA circolare lungo migliaia di basi (scaffold) e da centinaia di sequenze di DNA corte (circa 30 nucleotidi) a singolo filamento (staples) che si ibridizzano in zone non consecutive dello scaffold e lo assemblano nella struttura prefissata. Gli staples possono agire da sonde per biomarcatori, incorporate nella nanostruttura con una stechiometria precisa. In prossimità di tali sonde può essere introdotto nella nanostruttura un sito di restrizione, in modo che l'ingombro sterico della molecola target legata alla sonda ne impedisca il taglio. Modificando in modo permanente gli staples specifici per quel target l'enzima di restrizione "scrive" nella nanostruttura il numero di molecole catturate. Quantificando con PCR quantitativa (qPCR) o Next Generation Sequencing gli staples modificati si può ottenere una misura indiretta della concentrazione del target. Questo programma di ricerca è basato su risultati ottenuti recentemente nel nostro laboratorio che mostrano come nelle nanostrutture di DNA l'ingombro sterico attorno al sito di restrizione ne regoli la degradazione in modo digitale. In questa tesi ho sviluppato un metodo per quantificare l'efficienza di enzimi di restrizione in una nanostruttura, l'efficienza di incorporazione degli staples e la loro stabilità. Il protocollo si basa su una reazione di PCR lineare (L-PCR) che coinvolge carriers lunghi 70 nucleotidi con una porzione complementare al frammento di staple derivante dalla degradazione enzimatica. La polimerasi forma un doppio filamento di DNA lungo 70 coppie di basi (bp) a partire da ogni copia di staple tagliato mentre, nel caso in cui lo staple non venga tagliato la reazione di L-PCR è bloccata. La metodica può essere utilizzata per valutare l'incorporazione degli staples, ma in questo caso il carrier ha una porzione complementare all'intera lunghezza dello staple. Ho preparato campioni di L-PCR per stimarne l'efficienza e per valutare l'accuratezza di NGS dal punto di vista quantitativo. Inoltre abbiamo analizzato l'efficienza di incorporazione o di taglio di 5 staples in una nanostruttura triangolare di DNA.

Questa tesi descrive anche la progettazione e lo sviluppo di un nanosensore volto a migliorare l'accuratezza di strumentazione biomedicale comunemente utilizzata per la quantificazione di acidi nucleici. L'innovativa metodica "Bingo-qPCR" ideata dal nostro gruppo di ricerca è volta a sfruttare la qPCR per rilevare piccole variazioni di concentrazione di acidi nucleici. La precisione della quantificazione di acidi nucleici con qPCR è negativamente influenzata da reazioni di amplificazione e retrotrascrizione, e può discriminare solo rapporti di concentrazione maggiori di 2. Il nostro lavoro, mira a sviluppare un nanosensore da accoppiare alla qPCR per migliorarne le prestazioni senza la necessità di aggiornare la strumentazione tradizionale. Il nanosensore è costituito da tre filamenti di DNA (sonde) le cui estremità (piedi) sono ibridizzate su un altro filamento di DNA (scaffold). La parte centrale di ciascuna sonda crea un loop complementare al target mentre uno dei due piedi contiene un sito di restrizione che viene destabilizzato in seguito all'ibridizzazione del target al loop. Solo se tutte e tre le sonde sono legate al target lo scaffold è completamente protetto dalla digestione enzimatica (configurazione "bingo") e può essere amplificato con PCR. La concentrazione del target e del bingo-scaffold sono correlate da una funzione potenza con esponente 3. I risultati ottenuti dimostrano una migliore accuratezza della Bingo-qPCR rispetto alla qPCR tradizionale nel valutare piccole variazioni di attività enzimatica e la capacità del target di legarsi al sensore.

Table of content

Chapter 1 – Introduction

1.1. Genetics and Genomics

1.1.1. Genetics

The study of genes sequence, their function and distribution among populations is referred to as genetics (1). Genetics techniques are used to study small groups of genes or the whole genotype of an organism, which is its gene set. Human organism inherits half of the genotype from the mother and the other half from the father. Some inherited traits are visible and determine the phenotype of an individual that is the expression of the genotype (2). Genes are DNA sequences that can be converted in messenger RNA (mRNA) sequence which acts as a template for protein production (transcription and translation processes). In fact, the gene sequence is formed by segments of 3 nucleotides (codons) that code for different amino acids, the backbone of proteins. The study of proteins can be performed with several approaches such as electrophoresis based analysis, mass spectrometer or crystallographic techniques, but none of them evaluates the activity of the protein in a cell process. To this aim genetics techniques can be used to selectively inhibit production of a protein or its activity in the cell to evaluate direct consequences of its misregulation (2).

1.1.2. Genomics

The DNA present in the cell is composed only in a small percentage by genes coding for proteins, the rest of the DNA has regulatory function on the genes, has unclear function or has no function at all. The genome is composed by all these DNA species and the discipline that focuses on genome study is called genomics. While genetics mainly analyze gene sequence in relation to the activity of the protein encoded, genomics has a structural approach and analyzes the effect of genome rearrangement on its expression regulation (2). Only 1.5 % of all of the DNA present in the cell encodes for proteins. On top of that, introns are non-coding regions localized within the genes and they are about 26 % of the genome. Pseudogenes (inactive copies of genes), and transposable elements (fragments of

DNA with the ability to change position or to create copies along the genome that account for more than 40 % of the whole DNA) complete the picture. It's worth mentioning that, after the discovery of transposable elements in the 1940s, only a few tens of years later researchers considered the hypothesis that TE might have regulatory properties. (3, 4)

1.1.3. Transcriptome

Genetic and genomic studies can be combined with the study of the transcriptome that is the set of messenger RNA (mRNA) sequences produced by a cell. While genome is almost the same for every cell in the organism, the transcriptome changes among tissues or cells with different functions or differentiation stage, and it also changes in relation to external stimuli or environmental factors (food, sedentary life style, stress) (5, 6). A subpopulation of RNA molecules is represented by microRNA (miRNA) that are non-coding RNAs, 18-25 nucleotide in length, with regulation activity on mRNA: it is believed that more than half of the transcriptome is regulated by miRNA, and it has been demonstrated that the alteration of their activity has a role in cancer and several human diseases.

miRNAs are produced in the nucleus as longer double stranded sequences that are cleaved and transported in the cytoplasm. The mature miRNA forms when the duplex dehybridizes and one strand (miRNA*) is degraded while the other strand (miRNA) is bound by Argonaute, a protein that protects miRNA from degradation and is part of the RNA-Induced Silencing Complex (RISC). RISC uses miRNA hybridization to identify the target mRNA at least by partial hybridization of the mRNA with the nucleotides from 2 to 7 at the 5' end of the miRNA. Nevertheless miRNA that hybridize to their target in their central region, or that have partial hybridization at their 3' end, have been encountered. (7, 8) In humans, miRNAs regulate the target mainly by degrading it: miRISC draws on the target enzymes that remove the protective poly(A) tales from mRNA, activating in this way its degradation. A different regulation path can be followed by miRNA if they are almost fully complementary to the target sequence. In this case the formation of a double stranded RNA fragment enables the cleavage of the mRNA by the Argonaute. This process is more frequent in plants where miRNA are more often highly complementary to the target mRNA strand. In some other cases, mRNA level is not altered, but the production of the protein encoded is prevented by translation inhibition or by the degradation of peptide. All these events lead to gene expression down-regulation. In addition target-miRNA interaction can have a feedback on miRNA concentration: miRNA can be destabilized by hybridization with the target through degradation of its 3' end. The process is not well

defined, but there is evidence that viruses like *herpes simplex* can down-regulate miRNA expression through hybridization. Considering that miRNA specificity for the target, at least in human, is low (miRNA families encoded by the same chromosome have similar sequence), it can be easily accepted that their activity can be influenced also by the presence of competing targets, and this gives rise to the possible regulatory function of the non-coding RNA present in the cell. (8)

1.2. Genetic diseases

Genetics and genomics have great impact on global health as they can for example identify genes alterations in genetic disorders, define inheritance of diseases and foresee the response to treatment. In fact the genome of each person in unique, and it can change during life: at every cell division it can undergo small alterations that can affect protein production and functionality with consequences on cell activity. In point mutations only one base pair is mutated to a different nucleobase. If this occurs in a sequence coding for a protein, the changed codon (three DNA base pairs corresponding to a specific amino acid) can be synonymous to the native with no protein alteration, or can be non-synonymous with altered protein amino acidic sequence. The non synonymous mutation can inhibit protein production or generate proteins with altered functionality. Insertions and deletions are mutations that involve DNA fragments of different lengths. One or two base pairs deletion or insertion shifts the read of the whole gene with a harmful effect on protein production. Three base pairs deletion/insertion corresponding to one codon deletion/insertion can cause the elimination or the addition of one amino acid. The relevance of this event depends also on protein structure and amino acid localization. Gene fragments, sequences containing several genes or pieces of chromosomes can be inserted or deleted. Massive insertions/deletions can involve whole chromosomes. To complete the landscape of mutations transposable elements (TE) should be cited. Insertion of TE in a gene can switch on or off its activity, and in some cases TE can produce new genes (9).
Errors in DNA replication can be due to enzymatic inefficiency. DNA-polymerase activity is not 100% accurate and even if human cells have protection methods to correct genetic alterations, sometimes mutations are spontaneous. There are several factors that can increase mutation rate such as cigarette smoke or other mutagenic molecules, but also physical agents like UV irradiation (2). Moreover errors in DNA replication can occur

during cell division. In 2003 the human genome project was completed and the whole genome of different organisms including human were sequenced. This accomplishment led to the creation of database of human mutations that cause disease (Human Gene Mutation Database). 150000 mutations are reported, involving 6137 different genes. Single nucleotide mutations and small insertion or deletions accounted for about 90%, while the remaining 10% was represented by longer deletion or insertion and other alterations (reporting year 2014) (10).

A few examples of genetic diseases are reported below:

- Sickle cell disease is due to a single nucleotide mutation in hemoglobin gene. The protein produced in patients with this mutation has reduced efficiency, they develop anemia, suffer of pain episodes, are more susceptible to infections and stroke (2, 13).
- Mutations in membrane protein that regulates sodium ions composition in the cell (Na^+-channels) can determine ventricular fibrillation and heart attack or epilepsy depending on their tissue location (2).

Depending on where the mutation occurs in terms of cell type, the alteration can be spread to progeny. Humans have germ-line cells (eggs and sperm) and somatic cells. In hereditary diseases pathological mutations in germ-line cells end in their precursors can be transferred to the embryo. Examples of hereditary diseases are:

- Emphysema, in which the activity of an enzyme is blocked by a point mutation in the corresponding gene, and causes the degradation of lung tissue dedicated to oxygen absorption.
- Familial hypercholesterolemia that is caused by mutations in the gene coding for low density lipoprotein (LDL) receptor and is characterized by high level of LDL in plasma (2).

1.2.1. Cancer

A complex disease that is characterized by the accumulation of genetic mutations of different etiologies is cancer. Typically, a series of mutations in multiple genes that increase the ability of cells to proliferate are necessary for cancer development. Pathogenic mutations mainly occur in proto-oncogenes and tumor-suppressor genes, which are involved in proliferation equilibrium control. A mutation in a proto-oncogene can turn it to oncogene, conferring to the cell the uncontrolled ability to increase growth: the mutation

can increment gene expression or gene product activity. Tumor-suppressor genes are protective against excessive cell growth and their inactivation due to genetic mutations let cells proliferate without control. Proteins that are mainly encoded by oncogenes are involved in DNA repair process, apoptosis (cell death) and cell cycle regulation (cell proliferation). Oncogenes are mainly present in somatic cells, with no risk of being transferred to the progeny, but some germ-line mutations can favor cancer development (2).

1.2.2. Copy number variation pathogenicity

A direct effect of mutations on the genome activity is the change in number of gene copies coding for a protein, referred to as copy number variation (CNV). It can concern small exons (coding sequences of the gene) that can be short DNA fragment only 50 nucleotide long, or DNA pieces up to 5 million base pairs (that can be detected also with cytogenetic microscopy). CNVs take place if there is

a. A deletion within the gene
b. A deletion of the whole gene
c. The gene contains a point mutation.

In all these cases the activity of the gene is suppressed (gene disruption) and if it can't be compensated by the second copy of the gene (e.g. one allele is disrupted and the other carries a point mutation) there is a pathogenic outcome

d. Disruption of genes with regulatory effect on the copy number variant
e. Different genes are joined in a new one.

The heterogeneity of CNV causes can lead to data misinterpretation in assessing CNV pathogenicity. CNV can cause cancer and other genetic disease like autism and mental retardation or hemophilia (10, 14).

1.3. miRNAs as potential biomarkers

As explained miRNA levels are under a precise regulation, and alteration of the equilibrium might be clues for pathogenic state. Considering that:

- More than 60% of the genes that code for protein in human cells are miRNA's targets

- miRNA's expression has been found to be altered in several organs dysfunctions, metabolic disorders and cancer
- miRNAs can be easily found in stable form in blood, but also at lower concentrations in saliva, plasma, and urine. In blood, they are protected from degradation by Argonauts or by lipid vesicles.
- miRNAs are stable in fixed tissues

they have been proposed as biomarkers in several diseases, cancer among others (fig. 1) (15, 16). Their stability combined with the non-invasive nature of blood tests increases the interest for these RNA small molecules. The main goal in validating miRNAs as biomarkers is to define a signature that is specific for the pathology, as many of them are found to be altered in different diseases, with miR21 being the one with expression alteration in the highest number of diseases (17). Focusing on cancer, it has been demonstrated that the analysis of expression level of different miRNA sets (from 2 to 47 in ref 7, 17) can be used for diagnosis, prognosis and prediction of treatment response (they have been found to be involved in chemoresistance). They are able not only to discriminate between normal and disease samples, but they can identify tumor subtype and developmental stage (e.g. study performed on lung cancer samples). Moreover, some miRNAs (e.g. found in adenocarcinoma) are over-expressed before tissue damage, and they might allow early diagnosis, and circulating miRNAs might be used to determine the presence of metastasis (7, 17).

Several databases are available, and effort is made to produce computational tools for comprehensive classification (18, 19). Their classification is challenging because they can originate from different genome's locations, they have differentiation stages and they can have common targets. "miRBase" database in June 2014 reported 28645 miRNA precursors in 223 species, related to 35828 mature miRNA products (15).

Gene expression regulation can be altered by genetic mutations or by copy number variation at the chromosomic level, therefore the study of the transcriptome can suggest the presence of genetic disorders. The eventual pathological condition must be confirmed by genetic studies because protein activity is influenced not only by gene expression and mRNA production, but also by the independent protein degradation, interaction with other proteins within complexes, and protein localization. Moreover, some pathological mutations that alter protein activity do not influence mRNA levels, and it is necessary to

look for alterations in the gene sequence. The advantage of transcriptome analysis is that it gives a global vision of the expression level of several proteins in a specific cellular condition (2, 6).

Breast cancer	Serum	miR-373, -155	Up
		miR-34a, -17	Down
		miR-222, -103, -23a, -29a, -23b, -24, -25	Up
	Plasma	miR-148b, -376c, -409-3p, -801	Up
Lung cancer	Serum	miR-21-3p, -205-5p, -205-3p, -141, -200c	Up
		miR-21, -24, -205, -30d	Up
	Plasma	miR-21, -155	Up
		miR-145	Down
	Exosome	miR-17-3p, -21, -106a, -146, -155, -191,-192, -203, -205, -210, -212, -214	Up
	Sputum	miR-205, -210, -708	Up
Liver cancer	Serum	miR-122	Up
		miR-500	Up
		miR-21, -122, -223	Up
		miR-21, -1, -25, -92a, -206, -375, let-7f	Up
		miR-16	Down
Gastric cancer	Serum	miR-221, -376c, -744	Up
		miR-1, -20a, -27a, -34, -423-5p	Up
	Plasma	miR-106b, -20a, -221	Up
		miR-21, -223, -218	Up
		miR-451, -486	Down
Pancreatic cancer	Serum	miR-21, -100, -155	Up
		miR-203, -369-5p, -376a, -375	Up
	Plasma	miR-21, -210, -155, -196a	Up
	Exosome	miR-21, -17-5p	Up
Colorectal cancer	Serum	miR-29a	Up
	Plasma	miR-17-3p, -92	Up
		miR-92a, -29a	Up
		miR-27b, -148a, -326	Up
		miR-221	Up
		miR-15b, -17	Up

Figure 1.1. The expression of several microRNAs in body fluids has been found altered in different cancer types. Adapted from 15. Copyright © 2015 Hindawi Publishing Corporation.

1.4. Methods in Genetics and Genomics

1.4.1. Quantitative polymerase chain reaction (qPCR)

Quantitative polymerase chain reaction (qPCR) is the most practical and cost-effective technique to quantify target DNA molecules, and therefore the most frequently utilized approach from research settings to diagnostics. All started with the invention of PCR around 1985 (20), while the implementation by Higuchi, et al. (21) opened a whole new set of possible applications.

PCR consists in the production of copies of a double stranded (ds)DNA target sequence using a cyclic reaction. The first step of denaturation at high temperature separates the strands forming the target duplex. In the second step, temperature is lowered to anneal the primers that are short single stranded (ss)DNA (about 20 nucleotides long) being

complementary to the molecule to amplify. Specifically, the reverse primer hybridizes near the 3' end of one ssDNA, while the forward primer hybridizes to the 3' end of the complementary strand. Both primers have the 3' end free for enabling the elongation process, the polymerase in fact works only in 5'-3' direction, adding nucleotides in 3' following the template sequence. Primers are added in solution in large excess as well as nucleotides. After DNA elongation, the temperature increases again for allowing denaturation and a new cycle begins. At each cycle the amplification produces two strands from one template in an exponential reaction until saturation (the solution runs out of primers and dNTPs). The annealing temperature depends on primers length and sequence, it should be the same for both of them and a few degrees lower than their melting temperature (that can be calculated by online software) otherwise primers might unfold before polymerase can work. Imperfect primer design might be compensated by the ability of the polymerase to maintain primers on the target. Elongation temperature depends on the properties of the enzyme used that might change depending on the vendor, but typically it's about 72 °C. (22) In some protocol annealing and elongation are performed at the same temperature that usually is 60°C. The advantage of using low extension temperature is to favor primer hybridization, but higher temperature might be useful for templates forming secondary structures such as hairpin or guanine tetraplex.

qPCR procedure introduced a dye in the reaction mixture of standard PCR. The dye emits fluorescence when it is intercalated in double strand DNA product. At each PCR cycle the fluorescence is registered and correlated with cycle number, producing a curve that shows logarithmic correlation between the two parameters. The Ct value (cycle threshold) is the cycle number at which the fluorescence becomes detectable (there is enough PCR product) and, in a logarithmic plot of fluorescence against cycle, a linear phase begins. Ct value in fact is correlated to the initial concentration of target in the sample: the lower the initial concentration the higher is the number of cycles to reach the threshold. The fluorescence threshold is automatically set by the instrument used, typically 10 times higher than the average background noise. Considering 100% efficiency of the PCR at every cycle the concentration of the DNA should be doubled therefore a difference in Ct = 1 between two samples means that one of them has a concentration doubled than the other one.

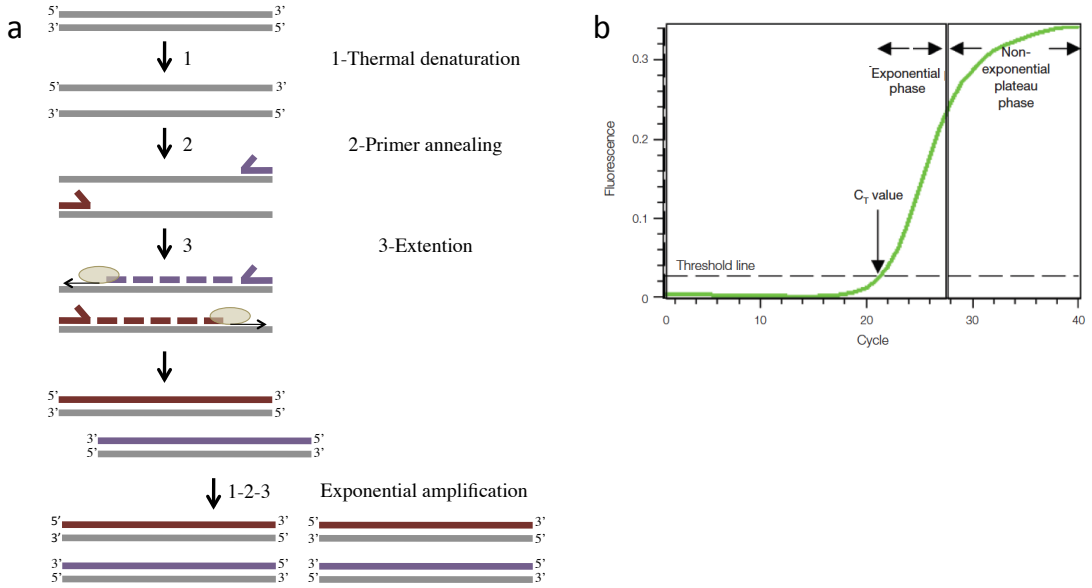


Figure 1.2. ***a.*** Polymerase Chain Reaction (PCR) schematic. In step 1 target DNA is denatured at 95°C to separate template strands. In step 2 the temperature decreases (typically 60°C) and primers hybridize to the template. In step 3 the polymerase adds nucleotides to form the missing DNA strand. Enzyme optimal temperature is typically 72°C. ***b.*** qPCR amplification curve. Ct value is the number of cycles required to reach threshold value. Threshold value indicates when the exponential phase begins. Adapted from 23. © 2006 Bio-Rad Laboratories.

At the end of the reaction, when reagents are not available or the products are limiting polymerase efficiency, there is a plateau phase in which there is no more exponential amplification. At this point, most of the instruments available on the market perform the analysis of the melting temperature by heating the sample to unfold the PCR product. Along with the thermal ramp the fluorescence decreases because the duplex DNA product is not stable, and when the melting temperature is reached the dsDNA suddenly dehybridizes thus leading to a dramatic decreases of fluorescence as the dye has no longer a dsDNA conformation to intercalate. Melting temperature is the peak point of the negative derivative of the fluorescence as a function of temperature (melting curve). This analysis helps the operator to evaluate assay specificity: the presence of more than one peak indicates the presence of secondary product that might be due to primer hybridization to non-target sequences, or due to primer-dimer formation. During primer design, one should always check for both specificity of the primer and lacking self-complementarity, although this task can require highly complex bioinformatics analysis. Secondary products should be avoided because they subtract reagents from the solution reducing reaction efficiency,

and because the fluorescence of the intercalating dye can't distinguish between product and unspecific, altering the quantification. A way to directly quantify only the desired product is to use specific labeled probes. Among tens of probes available on the market, the kind that is mainly used is the TaqMan probe: it is a DNA strand labeled at one end with a fluorescent dye, and at the other end with a quencher. The light is emitted when the Taq-polymerase with exonuclease activity used in this assay, removes and degrade the probe during elongation, separating the quencher from the dye.

Together with specificity, the efficiency of the qPCR is an important parameter. Samples very often do not contain only the dsDNA target, but might contain salt, proteins and other nucleic acids that can interfere with the chain reaction. To estimate the amplification efficiency, serial dilutions of the sample can be prepared and analyzed with the qPCR protocol to examine, towards obtaining a calibration curve of the concentration against the Ct value. In log-scale of concentration values a linear correlation is obtained, and the slope (S) is used to calculate the efficiency (E) as follows:

$E = 10^{-1/s} - 1$

E value is used to calculate the relative concentration of the sample with respect to a reference with the following equation

$[Sample] = [Reference] \times (1 + E)^{(CtReference - CtSample)}$

If E is 1, at each cycle the DNA amount is doubled (ideal condition) and the concentration of the sample corresponds is proportional to $2^{\Delta Ct}$. (22, 24, 25)

Efficiency calculation is important because it gives a global view on limiting factors such as secondary structures of primers and target DNA, as well as possible inhibitors or competitors present in solution (25). To evaluate the efficiency and perform experiments it is crucial to *i)* use the same instrument because efficiency is influenced by the instrument used mainly relating to the thermal performances, *ii)* choose and maintain the number of replicates used per sample (the higher number of replicates, the lower standard deviation) and *iii)* use constant pipetting volumes preferably ≥ 2 uL (26).

qPCR is a powerful tool and can detect very few copies (down to 3-10) of DNA in solution (27, 28, 29). Nevertheless the limit of detection should be determined for each assay as it depends on the enzyme used, on target characteristics and in general on assay efficiency (25). I.e. a sample with very low concentration of DNA can be detected at very high Ct values. Exceeding 40 cycles, however, is usually not reliable for quantification because it is affected by non-linearity associated with the reaction approaching saturation (25, 27).

Also the dynamic range should be determined specifically for the assay and it consists in the range of concentration in which the reaction is kept linear. Well-calibrated reactions can reach a dynamic range of 6 orders of magnitude (28, 29), while the accuracy limit of qPCR is about a 2-fold variation (26, 29, 30). Reducing standard deviation (by increasing replicates number for example) can bring the accuracy to 1,3-1,5 fold but this can sometimes be not sufficient, as happens when cells of interest (tumor cells) are diluted in a sample containing also healthy cells (31, 32).

qPCR is used to quantify DNA, RNA and other small nucleic acids molecules such as miRNA. In the case of DNA, one of the application is the Copy Number Variation analysis in which the assay is used to quantify the copy of the gene present in the sample. One or more amplicons (75-200 bp) are chosen within the gene of interest and they are quantified. Results are always compared with controls that usually are reference ("housekeeping") genes present in the sample and other samples with known copy number variation or with no alteration. Optimized software is available for data analysis. (33-35)

If the interest is for gene expression RT-qPCR (retro-transcription-qPCR) can be used to analyze the mRNA (36). The RNA extraction process is crucial because not only RNA needs to be separated from DNA and proteins, but also a selection of mRNA against rRNA (ribosomal RNA) needs to be performed. In fact, rRNA typically represents 90% of the whole RNA present in a cell, while mRNA constitutes only 1-2 %. Mainly, two enrichment protocols can be used for human cells depending on the % of mRNA present in the sample: selection of mRNA poly(A) sequences (during maturation a tail of adenine is added to the mRNA molecule to stabilize it, and transport it out of the nucleus (37)) if mRNA quality and amount is high enough and fits requirement of the kit used, or, alternatively, rRNA degradation. RNA in general needs to be transcribed in cDNA (complementary DNA) introducing a first error-prone biochemical step (37-40). To convert mRNA in cDNA several kits are produced containing reverse transcriptase and they can use random primers or adapters that are attached to mRNA molecules and work as primer templates. Besides retro-transcription bias another limitation in RT-qPCR as well as in qPCR is that target sequence and control sequence are not analyzed with the same assay therefore there is the need to follow defined guidelines to obtain reproducible results that can be confirmed in different laboratories. Moreover RNA quality (amount, presence of contaminants) can alter quantification even if the procedure is standardized (40). A survey on reliability of qPCR experiments has been proposed in 2009 in MIQUE Guidelines (27) and revised in 2013 in Nature Methods Commentary (40).

The analysis of miRNA is limited by the short miRNA's sequence that requires highly peculiar primers (hairpin) to be amplified, and by intrinsic structural obstacles in purification procedures (lack of poly-A tail). Nevertheless, RT-qPCR is one of the main techniques for miRNA quantification thanks to its high sensitivity, wide dynamic range, availability of the required instrumentation in almost any biological laboratory, and low costs. There have been also attempts at multiplexing the analysis, e.g. to determine miRNA feature for prostate cancer with qPCR (41).

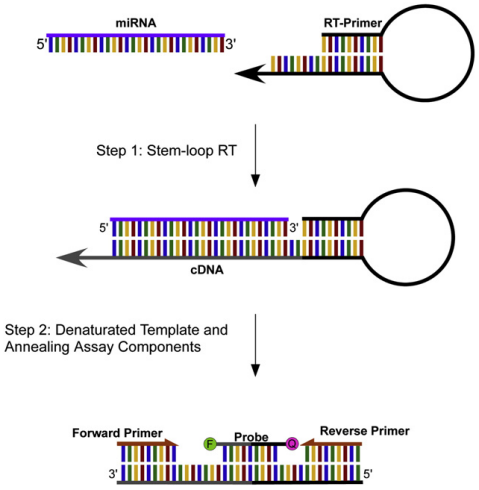


Figure 1.3. Nucleic acid amplification is limited by short template like miRNAs. Here is represented an innovative approach that exploits hairpin primers to amplify miRNAs. In step 1 (miRNA retrotranscription) the primer hybridizes to the template and is elongated using the miRNA as a template. The product of step 1 is quantified with qPCR (TaqMan assay) using a miRNA specific forward primer and a universal reverse primer (step 2). Adapted from 16. © 2013 The Canadian Society of Clinical Chemists.

1.4.2. Microarrays

To evaluate the genetic information present in tissues and cells, microarrays can also be an effective tool. They are composed by nucleic acid probes complementary to the sequence of interest printed in specific areas of the array's surface. DNA samples are fragmented, fluorescently labeled, and hybridized to probes before hybridization, while RNA samples may sometimes need an additional retro-transcription step to convert RNA in cDNA. The fluorescence signal of the sample is compared to the control, for a quantitative evaluation of the DNA/RNA present in the sample.

In gene expression analysis, DNA microarray can be used for CNV studies or for the identification of deletions and insertion in array-Comparative Genomic Hybridization (a-CGH). In this technique the region of interest of the genome is amplified by PCR and than hybridized to the array. Sample and control are labeled with different fluorescent dyes,

they are mixed together and hybridized to the array which contains specific probes at specific locations on the surface. The fluorescence signal correlates with sequence concentration. Probe length ranges from 60 to 170000 bp depending on the instrument used (42). RNA microarrays have been used to analyze mRNA in transcriptome studies. As mentioned, the protocol for RNA detection is almost the same as for DNA, with the addition of an optional retro-transcription step. RNA microarrays were used for the first classification in subtypes of breast cancer, by analyzing the expression pattern of a set of characteristic genes. (6)

In genotyping, microarrays have been used to identify single nucleotide polymorphisms (SNP) mutations. One of the approaches is the Affimetrix assay in which the array contains probes for both the native and mutated sequences. The DNA sample is labeled with a fluorescent dye and the signal is geographically linked to the different positions occupied by native and mutated probes in the array. When using genomic DNA, a selective amplification is performed before hybridization to favor the binding of the fragments of interest. (43)

For miRNA quantification microarrays are the most diffuse technique. Typically miRNAs of the sample are purified and labeled with a fluorescent dye. The hybridization to a specific area of the array identifies the miRNA and fluorescence intensity correlates with the concentration. The limiting step is miRNA purification and labeling; therefore, label-free protocols have been developed to reduce bias. Duan D. et al. (fig. 1.4) hybridize the miRNAs on the microarrays without a labeling step, but the probe contains a portion that is complementary to a reporter sequence (with a fluorescent dye) that is short, and its hybridization is favored when the probe is already hybridized to the target (21, 44).

Probe length to use depends on the task of the assay. Microarrays used for SNP detection will have shorter probes than microarrays used for CNV studies. Melting temperature is a fundamental parameter for hybridization efficiency and is mainly influenced by strand length and GC%. In short probes a minimal variation in the sequence like a point mutation has high impact on the melting temperature, also depending on the position of the mismatch (SNP in the middle of the sequence are more relevant). Therefore 20-50 nucleotides long probes are mainly used for SNP detection. Probes used for aCGH ranges from 60 to 170000 bp depending on the instrument used, and allow higher hybridization specificity reducing the background noise. (6, 42, 45) Microarrays are able to analyze several sequences from different samples at the same time with SNP identification and reproducibility at about 99,5 % (16, 42). The main drawback in microarrays is that the

analysis is focused on already know sequences, therefore the discovery of new mutation is basically impossible. Moreover hybridization rates are not linearly proportional to concentration: at high concentration the array is saturated and at low concentration the hybridization is penalized. In turn, heterogeneous samples presenting differentially represented sequences might be associated to underestimation of the less concentrated components due to a dominating background. These issues are solved in NGS, that can sequence and quantify at once, not influenced by sequences relative concentration. Considering the ever decreasing costs of NGS (associated with increasing throughput), its revolutionary approach to nucleic acid analysis can replace microarrays. (43)

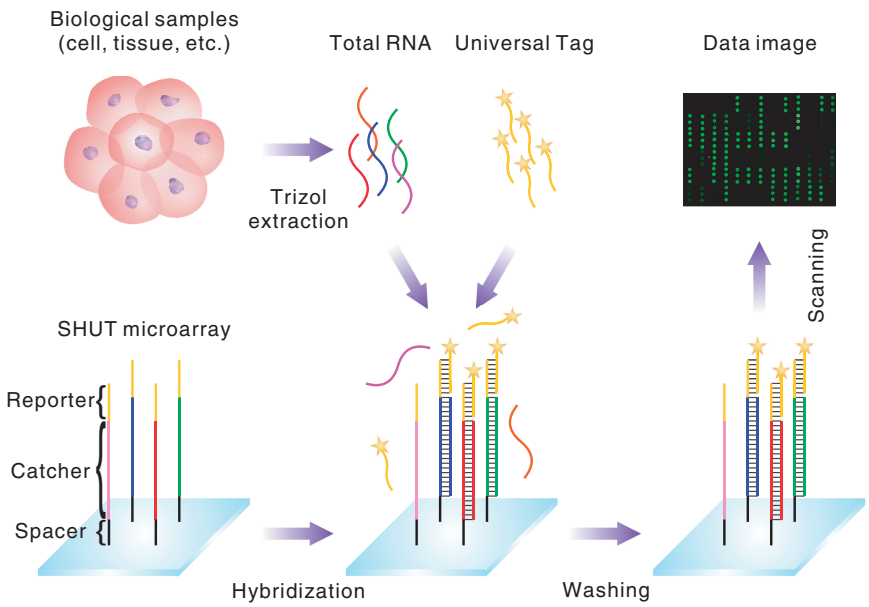


Figure 1.4. To avoid miRNA labeling in microarray analysis Duan D. and coworkers designed a probe with a portion complementary to the target (catcher) and a portion complementary to a universal tag (reporter). MiRNA extracted from the cells are mixed to the universal tag and hybridized on the microarray: the short tag hybridizes to the reporter sequence only if the catcher is already hybridized to the target. (44)

1.4.3. Next Generation Sequencing

The term Next Generation Sequencing (NGS) indicates a class of techniques that allow high-throughput and cost-effective analysis in comparison to other technologies based on the first-generation sequencing approach introduced by Sanger et al. in 1977. The key idea of DNA sequencing is based on the combination of knowledge in polymerase chain reaction (developed by Kary Mullis in the early eighties (20)) with the idea of

incorporating nucleotides lacking the hydroxyl group in 3' to block the polymerization. Performing PCR reaction in four different tubes containing all the necessary reagents (including all four of the standard deoxynucleotides dATP, dGTP, dCTP and dTTP) plus one of the four deoxynucleotides with a polymerization-terminating modification in 3' produced a set of DNA fragments of different lengths with the last nucleotide being A, T, C, or G in the different tubes. By running each sample on gel, the lengths of fragments could be determined by accurate fragment separation according to the molecular weight. By combining the produced bands from the four samples, the researchers were able to obtain the correct input sequence.

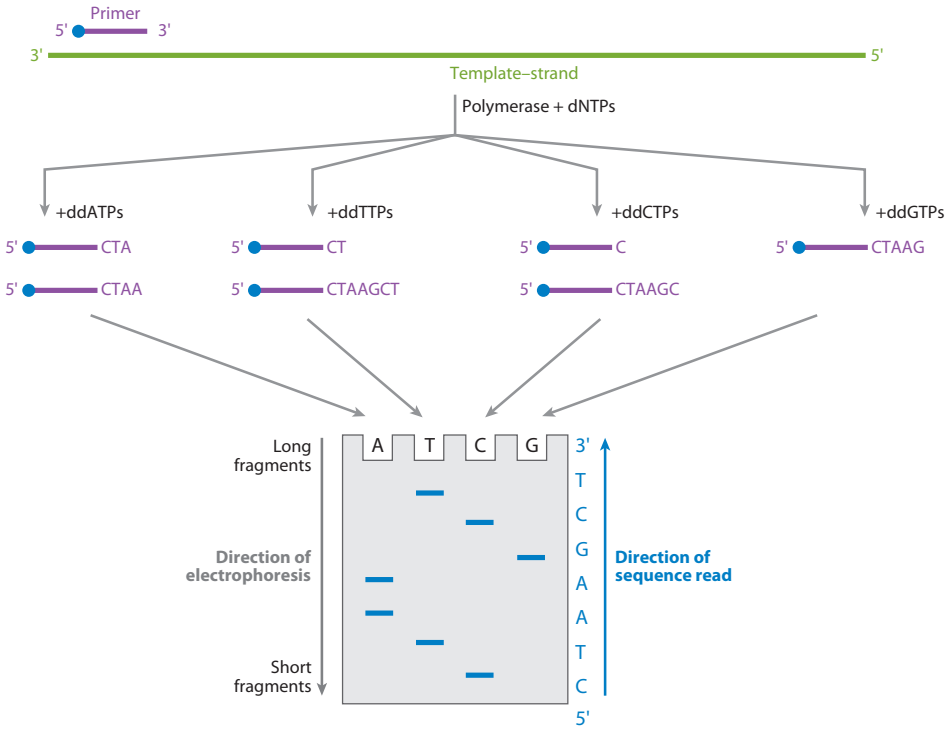


Figure 1.5. Sanger sequencing schematics. In the first step 4 different PCR are performed. Each reaction contains the template and all the reagents, but one of the nucleotides lacks the 3'-OH to block the polymerization. In this way all the DNA fragments produced in one tube will end with adenine, in another tube with cytosine and so on producing a set of DNA fragments of different lengths with the last nucleotide being A, T, C, or G in the different tubes. Each sample is then analyzed with gel electrophoresis and fragments length is determined from the shortest to the longest. By combining length from the four samples researchers were able to obtain the final sequence. (46). © 2013 by Annual Reviews.

Setting up the technique was however not so simple, as it was prone to error mainly due to gel electrophoresis (gel preparation, gel resolution, band detection with X-rays and performed by eyes). In the following ten years there have been improvements in many

aspect of the technique: *i)* the commercialization of fluorescently labeled nucleotides terminator allow to dismiss the X-ray band identification for a fastest process that in addition allowed to mix the different nucleotides in the same sample, *ii)* the engineered enzymes with higher processivity and accuracy; *iii)* the introduction of capillary gel electrophoresis that increased the automation and reduced gel and sample manipulation by the operator. The latest advancement in Sanger sequencing instrumentation can perform the analysis on 384 samples at once in about 2 hours (47). In the last 10 years, novel sequencing technologies have been developed based on the real-time monitoring of the DNA synthesis process. The approach termed sequencing-by-synthesis is the current state-of-the-art.

1.4.3.1. Illumina

Sample preparation for sequencing is called library preparation (see fig. 1.6-a). The procedure starts with sample fragmentation (enzymatically or with sonication) followed by end-repair and DNA is ligation to a dsDNA sequence called adapter at both ends. Adapters serve as template for primers during sequencing and in some kits it is possible to perform enzymatic fragmentation and adaptor ligation in the same reaction (38). Library preparation can introduce bias due to polymerase amplification: e.g. polymerase activity is negatively influenced by the GC reach segments and by repetitive sequence composition. The starting amount of DNA required depends on the kit used (in general from micrograms to picograms), but low DNA mass has to be compensated by amplification, and the more amplification cycle are performed, the highest bias is produced. When the sample is ready, it is loaded in the flow-cell, a chip of glass with micro-channels functionalized to anchor the ssDNA fragments. The flow-cell surface hosts single stranded sequences of optimized adapters that can hybridize to the ssDNA fragment during a peculiar bridge amplification reaction to create colonies of identical ssDNAs starting from a single copy of ssDNA fragment (see fig. 1.6-b). In this phase the amplification is performed with normal nucleotides only to amplify the DNA molecules for subsequent detection. After such a "solid-state" amplification, the sequencing-by-synthesis process can begin (see fig. 1.6-c): using the ssDNA colony as template, the polymerase adds fluorescently labeled nucleotides (four different dyes, one for each base) to polymerize the missing strand of each fragment to form a complete duplex. A deoxynucleotide-dependent fluorescence signal is detected and registered at each nucleotide addition. The fluorescent labels linked to nucleotides stop the polymerase reaction and are enzymatically removed after detection

of the fluorescence signal. After a washing step a new cycle begins with the addition of a new labeled nucleotide, and so on. In this way, millions of colonies are sequenced at the same time. Illumina allows the so called paired-end sequencing to increase accuracy. Specifically, after the whole DNA fragment has been sequenced, new colonies are generated and the strands are sequenced using the other primer. In term of performance, The Illumina Technology allows reads of 150-300 pb and the error is about 0,5%. (46, 48, 49).

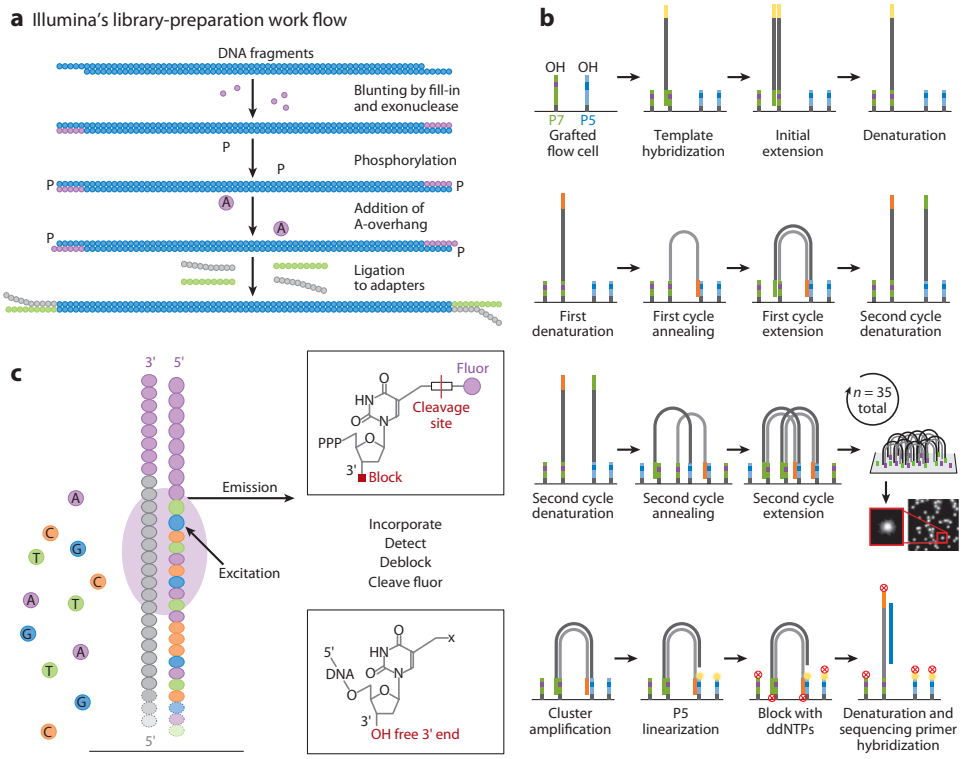


Figure 1.6. Illumina sequencing schematic. a. Template DNA is repaired and ligated to the adapters. b. The DNA is denatured and the strands hybridize to the complementary sequence of the adapter on the flow cell surface. A polymerase converts the ss/ds hybrid DNA in a duplex fragment with one of the two strands linked to the flow cell. After a denaturation step the strand that is not covalently bound to the surface is washed away. Bridge amplification starts after "first denaturation" step: in this reaction the adapter sequence not bound to the surface hybridizes to the complementary sequence on the flow cell and the polymerase creates the complementary strand. After several cycles of bridge amplification and denaturation colonies of ssDNA fragments are obtained and the amplification is stopped using ddNTPs (dideoxynucleotides lacking the 3'-OH). c. the sequencing by synthesis concept is depicted. At each cycle one fluorescently labeled nucleotide is added to the sequence and the signal of the whole colony is detected (signal amplification). At the end of the cycle the fluorophore is enzymatically removed and another labeled nucleotide is added (46).

1.4.3.2. IonTorrent

Also here library preparation is required to ligate adapters, which are necessary for sample amplification in Ion Torrent chip. The samples are then emulsified with beads in 1:1 ratio to create micelles containing on average a single bead, a single DNA strand and the necessary PCR reagents. The bead surface is functionalized with strands complementary to the adaptor sequence that are used as primers for the sample DNA. After completion of DNA amplification, the bead's surface is covered with identical dsDNA copies. The strand that is not covalently attached to the bead is denatured, while the surface-bound strand serves as template for the sequencing-by-synthesis reactions. The Ion Torrent chip is comprised by millions of microwells, each one containing a single bead, connected to a semiconductor that measures the pH at every nucleotide addition and converts it into a nucleotide-dependent voltage output. At every step of nucleotide addition to the sequence under polymerization, one hydrogen proton is released, thus leading to a pH change in the microwell. At every step only one type of deoxynucleotide flows in the chip and a change in the voltage of each well is separately recorded only if that nucleotide is added to the sequence. Ion Torrent avoids error due to deoxynucleotides labeling and, compared to Illumina, also reduces errors associated to the enzymatic cleavage of the deoxynucleotide-linked dye. The bias in this technology is mainly due to the impaired synthesis and pH meter saturation in presence of sequence repeats in the chip. In term of performance, the error of the technology is around 1%, for read lengths of 200-400 bp. (46, 50, 51)

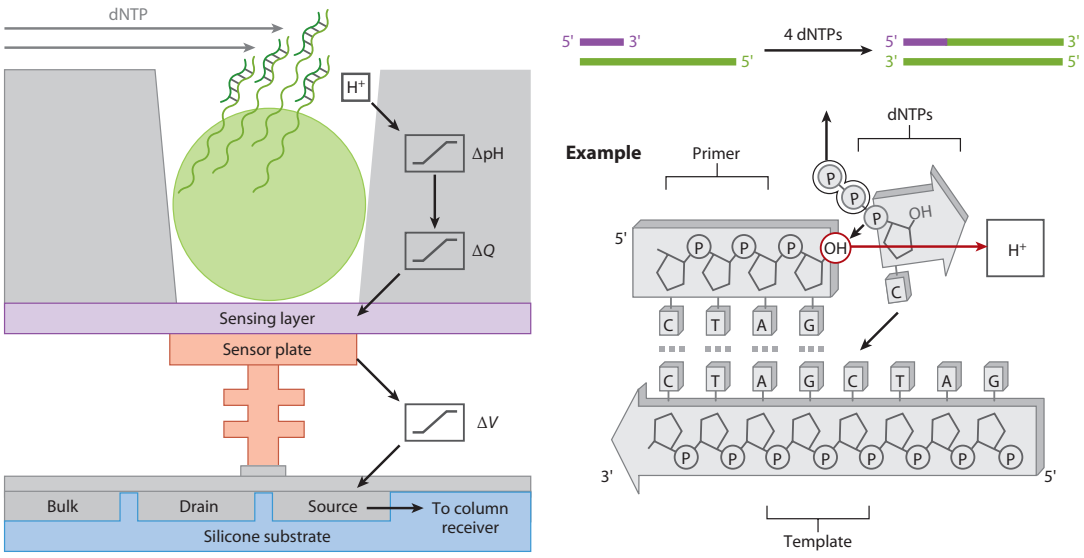


Figure 1.7. Target DNA is linked to the surface of a bead in 1:1 ratio and then amplified. Later the DNA is denatured and a solution containing only A, T, G or C flows in the chip, the primers anneal and the DNA

polymerase adds the nucleotide following template sequence. At every addition the pH changes and is recorded for each well. After a washing step a second cycle starts with a different nucleotide. At every cycle the instrument correlates pH changes with addiction of a specific nucleotide in a certain well (46). © 2013 by Annual Reviews.

Next Generation Sequencing methods are based on the counting of single molecules by deriving the signal from clusters of clones, and therefore it does not need sample dilutions or a standard to determine the amount of DNA molecules present in the sample. However biases due to fluorescence labeling, incomplete chemical reaction, amplification in library preparation, or polymerase inaccuracy have to be taken into account. In general, the noise accumulation increases with the reads length. Short reads are difficult to use for *de novo* sequencing (sequencing of an unknown genome) because the overlapping fragments of DNA are too short to determine how the obtained sequences should be reassembled. There are however different approaches to help arrange sequence in the correct order, and one is the mate-pair sequencing, in which long (2-5 Kbp) fragments are functionalized with adapters and circularized to bring together the ends. After fragmentation there is a selection for the strands containing the end of the original sequence that can now be sequenced. Mate-pair can be combined with single reads or pair-end reads to reconstitute the genome. This kind of analysis is often unnecessary because human genome sequence and sequences from other organisms are now available as reference (46). Targeted sequencing can be performed using libraries from the whole genome and enriching the interesting portions. Alternatively, the selection can be obtained by using PCR products to create libraries. In the first case, probes are used to fish the sequences of interest. This approach, however, can exclude mutated sequences or sequences difficult to be chemically obtained (i.e. repetitive sequences). In the second case, the amplicons generated by PCR can be sequenced as they are, or they may need fragmentation. In general, PCR amplification bias can alter the original sequence and the quantification. (38)

1.4.3.3. Third Generation Sequencing

To overcome amplification biases and reads short length, a new generation (Third Generation Sequencing) is now emerging with single molecule detection. Pacific Bioscience has developed and commercialized (2010) a technology that can sequence up to 10000 nucleotides with an error of 15%. Sequencing is performed by a DNA-polymerase with reduced processivity, fixed on the surface of an extremely sensitive fluorescence

detector (zero-mode waveguide). The nucleotides added are fluorescently labeled with a different dye and the instrument records the light signal at every incorporation event. Here the error is mainly due to the sensitivity of the camera, to the polymerase activity and to the amplification steps that are still required prior to sequencing. (46)

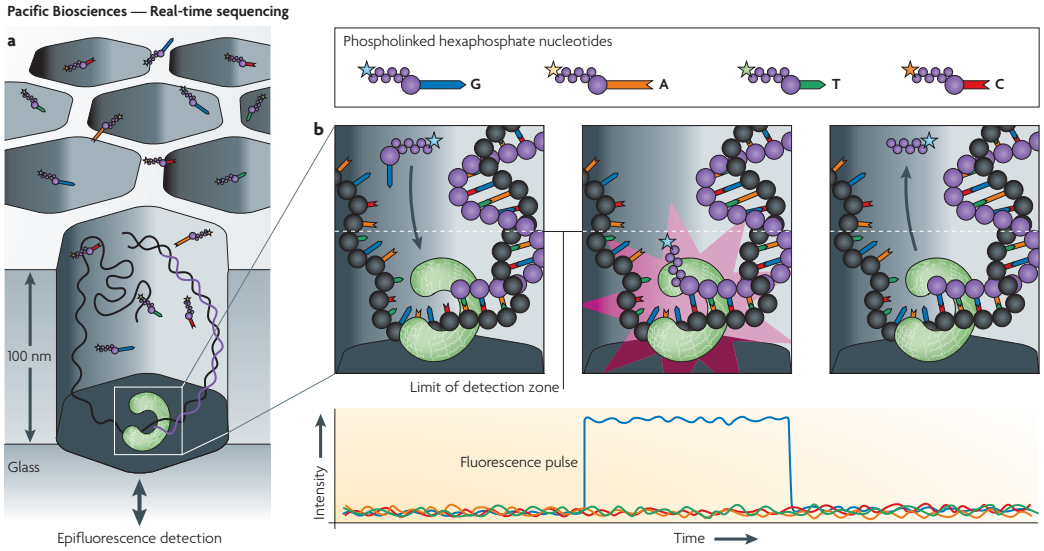


Figure 1.8. In Pacific Bioscience technology polymerase is fixed on the surface of the fluorescence detector (zero-mode waveguide) and each nucleotide is labeled with a different fluorophore (39). © 2010 Macmillan Publishers Limited.

NGS can be used to analyze not only DNA but also, mRNA and miRNA among other RNA molecules. As explained in qPCR section bias is introduced by retro-transcription reaction. Moreover, in some cases it is important to maintain the strand orientation (antisense, long, non-coding RNA can have regulatory effect on gene expression) that is not possible with standard cDNA production. The most adopted solution is to synthesize the strand that is complementary to the one of interest with uracil, that prevents amplification of the latter (38). miRNA analysis with NGS can be challenging for the same reasons explained for mRNA, and, in addition, be due to the complexity of miRNA sample preparation (purification of miRNA from other RNA), and/or their short length. miRNAs are usually 18-25 nucleotides long and their signal can be lost in the adapter-dimer noise.

1.5. Single-cell genomics

So far genomic and transcriptomic studies have been performed on cell populations to obtain a sufficient DNA/RNA amount. Despite the important information gained in the past, the scientific community is now moving toward single-cell analysis for more detailed data. In fact population analysis averages results and differences between cells and details in cellular processes such as cancer development are lost. (12)

The main challenge in single-cell genomic analysis is that only 7 pg of DNA are present in one cell. The required amplification can introduce errors in the sequence or be more efficient with some sequences compared to others altering quantitative analysis. It is important to select the protocol that suits better the experimental conditions. An example is the random primers amplification (MDA, multiple displacement amplification uses random primers and a highly efficient strand displacing DNA polymerase (phi29), (6, 38)) preceded by a linear pre amplification to reduce bias. For transcriptome single-cell analysis only 1 to 0,1 pg of mRNA is estimated to be available and more than 85% of the transcripts have only 100 copies per cell. (12, 38)
Both microarray and NGS have been successfully used for copy number variation analysis. NGS has some advantages: i) it works sequence-untargeted and can perform wider analysis; ii) it can detect unknown mutations; iii) it can identify structural modifications. Nevertheless, NGS data can be biased by amplification artifacts or by imbalanced amplification. NGS and microarrays together with qPCR have been used also for mRNA analysis and quantification. The main problem in this kind of experiment is the retro-transcription of RNA in cDNA because only 5-25% of mRNA might be converted in cDNA. This enormous lack of efficiency has more consequences when the starting nucleic acid amount is extremely low in abundance like in single-cells. Nevertheless, reaction efficiencies typically increases in small volumes. (12)

Studies have been performed on gametes single cells were higher mutation amount has been found compared to population cell analysis. Other studies on single cells from embryos demonstrated that at the beginning of life the probability of mutation is extremely high and that the genetic makeup for diseases forms with the first cell cycles. The study of cancer with single-cell resolution helped in discover new mutations and genomic structural

alterations, along with the fact that higher mutation rate was detected. The importance of single-cell analysis emerged also from transcriptome sequencing as for example it can help in identifying different subsets in a cell population. Single-cell analysis of the transcriptome can be used for circulating tumor cells as well, which are difficult to analyze due to their low number. (12)
Single-cell analysis can be applied to the clinic in prenatal tests of fetal cells from maternal blood, in assisted reproduction before implantation and in cancer diagnosis from circulating tumor single cells. (12)

The emerging field of single-cell analysis is generating an answer from the technological aspect to help scientists in handle tiny samples. Major accomplishments have been reached by Fluidigm corporation that produces a microfluidic based instrument to isolate single cells, extract nucleic acid and perform amplification with up to 96 single cell per run. (38, 52)

1.6. DNA Nanotechnology

1.6.1. DNA hybridization

Sequence-specific DNA hybridization folds two single-stranded DNA molecules in a double-helix structure by pairing complementary bases with non-covalent, Watson-Crick interactions. Non-conventional DNA secondary structures are present in nature and are functional in biological systems. Stem-loop motifs (intra molecular base pairing of one ssDNA complementary regions to form a double helix that ends in a single stranded loop) and G-quadruplex (four stranded structure formed by Hoogsteen hydrogen bonding of tandem-repetitive guanine-rich DNA) are mainly relevant examples (2, 53). The work by John Santa Lucia (54) is a comprehensive evaluation of previous studies to create a model for DNA thermodynamic prediction and which is now the reference for open source, nucleic acids design, online tool Nupack. As reported in this article prediction of the thermodynamic of DNA strands has to take into account several parameters:

- Oligonucleotide concentration: the affinity between two molecules in solution increases at higher concentration.

- Interaction between neighboring base pairs. The “nearest neighbor method” approximates the binding energy of single nucleic acid base pairs taking into account the stabilization power of other base pairs in proximity.
- Base pair interaction strength (GC > AT: A-T pairing consists in the formation of two hydrogen bonds, whereas G-C pairing occurs with the formation of three hydrogen bonds).
- Strand terminal base pair (GC or AT). The terminal bases are typically the most prone to dissociate because are not stabilized by other base pairs. This spontaneous dissociation, in hybridization unfavorable conditions, can proceed along the sequence.
- Length of nucleic acid strand is an important parameter as each base pair stabilizes the neighboring ones, and it should also be taken into account to estimate salt concentration best conditions.
- Buffer composition in terms of pH, ion quality (monovalent cations Na^+, K^+ or divalent cation Mg^{++}) and concentration (55).

In addition to their essential biological functions, nucleic acids are promising building blocks for artificial nanostructured systems that have a potentially broad range of applications, from biology and medicine to information technology. Nucleic acid-based molecular nanoengineering has experienced a rapid development, as it takes advantage of the inherent capacity of nucleic acids to self-assemble by Watson-Crick base pairing. Moreover, progress has been ably supported by the commercial availability of custom DNA and RNA molecules with a broad range of modifications, as well as diverse enzymes that can recognize, process or otherwise modify nucleic acids with high efficiency and selectivity. In living organisms, DNA is usually not found as a single stranded molecule (ssDNA), but in the form of a duplex helical structure (dsDNA). During DNA hybridization, two strands running in the opposite direction (anti-parallel) are coiled around the same helical axis and bound together by hydrogen-bond nucleobases pairings. Each type of nucleobase on one strand typically forms a purine-pyrimidine pairing with just one type of nucleobase on the other strand. Adenine pairs with thymine to form two hydrogen bonds, while guanine pairs with cytosine to form three hydrogen bonds: this is coupling rules are called Watson-Crick base-pairing.
Sequences complementarity is the founding principle of DNA duplication and underlies its application to nanotechnology. (1, 2, 56)

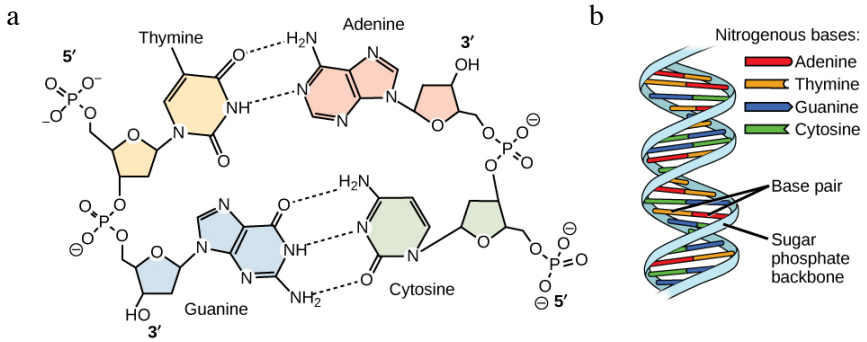


Figure 1.9. Deoxyribonucleic acid (DNA) is an organic polymer composed by repeating units called nucleotides. Each nucleotide is formed by *i)* a nucleobase, which is a nitrogen-containing organic compound; *ii)* a deoxyribose, which is a pentose sugar; *iii)* a phosphate group. DNA backbone is formed by phosphodiester bonds between phosphate groups of adjacent sugar rings. Nucleobases are cytosine (C), guanine (G), adenine (A) and thymine (T), and are covalently bound to sugar. The sequence direction of DNA strand is identified by the presence of phosphate group (5' end) or the hydroxyl group (3' end). The association of complementary ss-DNA strands is called DNA hybridization: two strands running in the opposite direction (anti-parallel) are coiled around the same helical axis and bound together by hydrogen-bond nucleobases pairings (Watson-Crick interaction). Adenine pairs with thymine to form two hydrogen bonds, while guanine pairs with cytosine to form three hydrogen bonds. Adapted from 56.

1.6.2. Toehold exchange

Strand displacement can take place spontaneously because dsDNA is usually partially dehybridized at both ends due to thermal fluctuations. Toehold mediated strand displacement reactions involve one single stranded oligonucleotide z, and a duplex DNA fragment formed by strand x and strand y (fig. 1.10-a). One of the two strands of the duplex (x) has a single stranded non-complementary region (toehold) that is hybridized by strand z acting as a seed of hybridization z-x. Once the toehold is hybridized, the reaction proceeds along the whole x strand with the formation of an intermediate xyz-strand structure that evolves in the xz-strand structure. Typically, the toehold is 5-8 nucleotides long. If x and z are fully complementary the reaction ends with the formation of a stable, fully duplex DNA strand, but if the invading strand z is shorter than x, a new toehold will remain active on strand x for a secondary displacement. This second type of reaction is called toehold exchange. (57) To design strands x, y and z researchers can consider parameter such as strand GC %, length and melting temperature, and the reaction can be tested is silico using several programs (Nupack, mfold, Vienna package, HYTHER) that differ mainly for strand input number and can predict the behavior of the strands in

solutions considering their concentrations, buffer salt concentration (Mg^{2+}, Na^{+}) and the temperature at which the reaction is performed. The accuracy of the thermodynamic predictions depends on the complexity of the reaction, still requiring experimental optimization reaction condition. An example of nucleic acid sensor which combines DNA strand displacement and DNA nanostructures is the one developed by Dongfang and co-workers, in which a strand displacement reaction cascade takes place on the surface of a DNA nanostructure molecule, which is then analyzed with atomic force microscopy (AFM). The hybridization of a target miRNA on a 2-strand probe on the nanostructure displaces a signal DNA molecule (ssDNA oligo labelled with biotin) that activates a second strand displacement reaction to generate the output dsDNA. The presence of biotin in the output area of the DNA nanostructure is detected with AFM. (58)

1.6.3. Molecular beacon

This class of molecules is one of the first, simplest and most widely used among the hybridization based molecular devices. They are stem-loop structures labeled at both ends of the strand with a dye and a quencher respectively (59). The loop part acts as a sensor for target molecule detection (fig. 1.10-b). Upon hybridization of the target the linearization of the structure leads to the separation of the dye from the quencher resulting in fluorescence emission. There are some structural constraints about loop and stem length therefore the beacon should be specifically optimized in relation with the target (e.g. pathogens nucleic acid or PCR products). To successfully monitor PCR reactions with molecular beacon, they should hybridize to their targets at the annealing temperature of the PCR, but keep the closed conformation at this temperature if they do not hybridize to the target. In other words the melting temperature (Tm) of the stem should be higher than annealing temperature and lower than the beacon-target complex Tm. Typically a probe length of 22–30 nucleotides can be used, and the Tm of the probe-target hybrid should be 7–10°C above the annealing temperature of the PCR. Loop length affects beacon specificity because longer loop sequences stabilize probe-target hybrids even in presence of mismatches. The design of beacon should be varied to fit specific requirements and this can be done with useful, open-source, online software, or by the probe suppliers. (60, 61)

1.6.4. DNA tweezers

DNA tweezers are simple DNA complexes that can form an open or closed state. In the open state, a scaffold strand is partially hybridized to 2 secondary strands (probes). In the

middle of the scaffold there is a short single stranded region (a few nucleotides) that allows scaffold bending upon hybridization of a fourth strand (fuel strand) to the single stranded ends of the probes (closed state). The tweezer can switch from such a closed conformation back to the open conformation by hybridization of a reset strand that displaces the fuel strand. Fuel and reset strands can act subsequently, thus originating a cyclic DNA tweezer reaction. By labelling the scaffold with a dye and a quencher at 3' and 5' ends, respectively (FRET couple), the motion of such a nanodevice can be monitored in real time with a fluorescence microscope or with gel electrophoresis. Minghui Liu and coworkers demonstrated how DNA tweezers can be utilized to control biological reactions. They hybridized to one arm of the tweezer an enzyme (G6pDH), and its cofactor (NAD^+) to the other. In turn, the enzyme is active only in the closed state in which it interacts with the cofactor (fig 1.10-c). The group was able to cyclically control the enzymatic activity using fuel and reset strands. The ability to accurately control DNA tweezers reactions opens to a wide range of applications such as drug delivery and biological reaction control (62, 63).

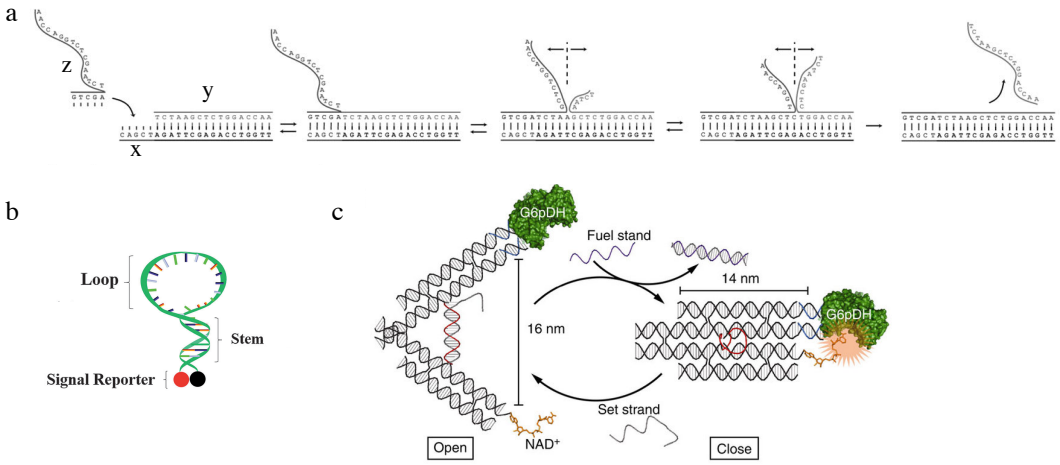


Figure 1.10. ***a.*** Toehold mediated strand displacement schematic: single stranded oligonucleotide z hybridizes to the complementary single strand region (toehold) of strand x. While the hybridization proceeds along strand x, strand y is displaced. ***b.*** Molecular beacon schematic: stem-loop structures labeled at both ends of the strand with a dye and a quencher respectively. ***c.*** DNA tweezer schematic: a long ssDNA is folded in a two-arm configuration by secondary short strands. Free nucleotides in the middle of each arm allow the hybridization of a short single strand DNA (in red) that in the open configuration is hybridized to the complementary strand to form the duplex, while in the closed configuration is free. Two secondary strands were used to link the enzyme (G6pDH) and its cofactor (NAD+) to the tweezer. The enzyme is active only in the closed state in which it interacts with the cofactor. Adapted from 61, 62, 63.

1.7. Ordered DNA nanostructures [1]

The hybridization of short DNA sequences in solution allows formation of diverse structural motifs of several nanometers in width, such as double-helical and single-helical regions, sticky ends, hairpin loops, bulge loops, junctions, and crossovers (64). With the aid of computer design, the assembly can be scaled up to form structures from a few to several hundreds of nanometers in width (64-67). A novel approach to the design of water-soluble, dense DNA nanostructures (also termed "DNA origami") was described by Rothemund in 2006, who created planar structures that are formed via Watson-Crick hybridization of several, short ssDNA sequences (also termed "staples") with a long circular ssDNA molecule (the "scaffold") (65). The ssDNA scaffold, thermally treated in the presence of a large excess of staples, provides reproducible formation of different shapes ("DNA origami")(fig. 1.11). Since then, several studies demonstrated the functionalization of DNA origami with different probes, including aptamers, DNA, proteins, gold nanoparticles (68-71), as well as antibodies for immuno-detection capability (72). In most of these cases, the probes are linked to selected staples after their incorporation within the scaffold-staples-based nanostructure, formed during a single-step self-assembly process (73). Initial studies demonstrated that DNA self-assembly can generate materials exhibiting precise spatial positioning of fluorescent labels, and possessing exquisite optical properties (74, 75). For instance, Niemeyer and coworkers realized a supramolecular Forster Resonance Energy Transfer (FRET) system, based on a linear structure, stabilized by several ssDNA staples, that provided a defined distance between a fluorescent protein and a synthetic chromophore (76). Similarly, Tinnefeld and coworkers exploited DNA Origami technology to generate a two-dimensional "molecular breadboard" allowing a programmed spatial arrangement of fluorophores in a grid-like fashion. This was used to demonstrate alternative energy-transfer pathways dependent upon incorporation of a "jumper" dye at specific positions (77).

[1] This paragraph derives from my publications Castronovo M. *et al.* Effects of Nanoscale Confinement on the Functionality of Nucleic Acids: Implications for Nanomedicine, *Current Medicinal Chemistry*, 20, 38 (2013); Nicholson A. W. *et al*. Emergent Properties and Functions of Nanoconfined Nucleic Acid Architectures, *RNA and DNA Diagnostics*, 183-204 (2015).

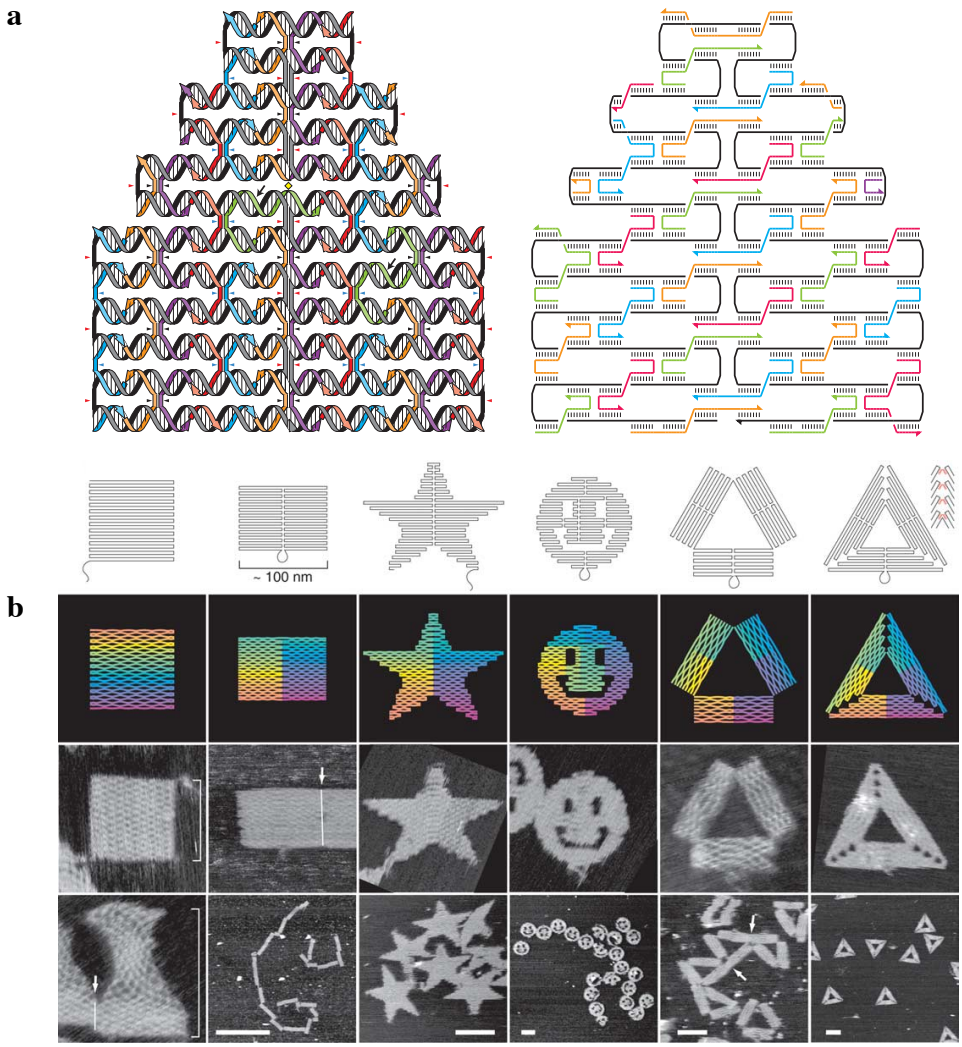


Figure 1.11. ***a.*** Schematic of DNA nanostructure. The grey line is the long ssDNA scaffold and the colored lines are the staples. Arrows indicate where staples switch from one helix to another generating a crossover. ***b.*** With the same scaffold (M13mp18 phage) is possible to obtain different shape of DNA nanostructures using different staples arrays. Imaging performed with atomic force microscopy. Adapted from 65. © 2006 Nature Publishing Group.

1.7.1. Staple hybridization efficiency

Pinheiro et al. (78) designed a water-soluble DNA nanostructure to determine how steric crowding may influence DNA hybridization. The nanostructure was a planar, rectangular tile comprised of six parallel double helices, joined by helix-spanning, single-stranded oligonucleotides. The tile possessed adjacent, 20-nt overhangs, with one of the overhangs serving as the binding site for the target probe, while the other overhangs representing "off-target" probe binding sites. The steric factors evaluated included (i) the site of binding

of the target probe in the tile; (ii) the presence of DNA flanking the hybridization site; and (iii) formation of a double-helical structure between the target probe and other components of the tile. Using real-time fluorescence spectroscopy the authors found that duplex formation involving the target probe overhang caused enhanced fluorescence, reflecting the ability of the newly-formed double helix to displace the dye from its intercalation site within the core of the tile structure. More efficient hybridization was obtained when the target probe was flanked by blunt-ended duplexes, or was located at an outermost (edge) position. In the latter case, the target probe would be flanked by only a single off-target probe. When the target probe is located at inner positions, the kinetics of hybridization are largely insensitive to whether the flanking off-target probes are single- or double-stranded. Additional experiments showed that the hybridization rate dramatically decreases when the target probe is bent (kinked), rendering the hybridization site less accessible from solution. The authors argue that observed differences in kinetics primarily reflect differences in the frequency of productive collisions, and that the slower hybridization kinetics reflects an impaired nucleation step. In summary, this study has demonstrated the significant impact of accessibility within the context of highly dense DNA tiles.

Xixi Wei and coworkers functionalized two staples within a 2D rectangular DNA nanostructure with a dye and a quencher (FRET couple) to monitor the assembly process. They created several versions of the rectangle removing different pattern of strands surrounding the labeled staples, and they analyzed the melting profile of the couple measuring fluorescence. They observed a cooperative hybridization in which staples stabilize each other, with main contribution by staples on the same helix. In another experiment, they placed the FRET couple in 3 different positions and evaluated the melting profile. They observed lower stability of staples in proximity of helix end (79).

Yonggang Ke and coworkers analyzed the incorporation efficiency of 12 staples in a 60-helix bundle 3D structure, 6 staples were chosen in external helices and 6 in internal helices. They used a hybridization-based assay to extract staples of interest from the structure followed by gel electrophoresis analysis for staple detection (the accuracy of gel quantification was estimated to be about 10%). They demonstrated that internal staples (mean length of 57 bp) have 20% lower incorporation efficiency than external staples (mean length of 54 bp). They also found higher incorporation efficiency for longer compared to shorter staples (80).

It is generally accepted that DNA hybridization is hindered in highly dense monolayers (81). Similarly, the formation of highly dense, self-assembled 3D DNA origami

nanostructures is impeded as the number of inter-helix connections ("crossovers") increase. These structures are formed by the hybridization of short ssDNA staples with multiple segments of the long ssDNA scaffold. The amount of inter-helix connections inevitably increases with nanostructure size and complexity, and the concomitant hindrance of self-assembly reflects the effects of molecular crowding. The study of Yan, Shi and coworkers suggest that crossovers introduce repulsive forces and steric hindrance that impede the action of ssDNA staples on the ssDNA scaffold during self-assembly (82). The authors demonstrated that the yield of self-assembly of a 3D DNA cuboid structure (with DNA duplexes arranged in a square lattice) increases by diminishing the number of cross-overs only within the core of the cuboid, thus preserving the desired shape.

1.7.2. DNA nanostructures applications

Intensive parallel development efforts in fluorescence optical imaging has led to super-resolution techniques that validate the proposal that biomarker detection in tissues can be achieved through single-molecule counting (83). Besides the diagnostic importance of single-molecule sensitivity, there has been a pressing need to circumvent the current limits to the number of simultaneously localizable biomarkers in cells and tissues. This has spurred the development of super-resolution approaches based on combinations of dyes as "barcodes," allowing multiplexed biomarker imaging (84).

DNA Origami provided a revolutionary approach to this problem: Yin, Shih and coworkers created up to 216 different biomarker barcodes by the site-specific fluorescent labelling of different self-assembled DNA nanostructures (shaped as rods, several hundred nm in length) linked to antibodies. This was accomplished using four fluorescent dyes corresponding to four different colours. Barcode species were unambiguously decoded using diffraction-limited, total internal reflection fluorescence (TIRF) microscopy (85). Up to five barcodes with higher spatial information density were demonstrated via construction of super-resolution barcodes, with features spaced by ~40 nm (85) (fig. 1.12). This imaging approach, termed DNA-PAINT, relies on the continuous association/dissociation reactions involving the Origami-affixed ssDNA molecules, and the complementary, fluorescently-labelled ssDNA (imager) molecules in solution. The latter molecules are ubiquitously present during imaging, but are preferentially localized to the corresponding hybridization sites on the Origami surface. Transient binding between imager and docking strands produces fluorescence "blinking," allowing stochastic super-

resolution imaging (86). Using the same approach, Yin, Shih and coworkers demonstrated that super-resolution fluorescence imaging can provide multiplexed detection of multiple proteins in microtubules and mitochondria in a single fixed cell sample, using a single dye. Here, antibodies linked to specific short ssDNA molecules were used as docking strands, while the complementary imager strands (labelled with the same dye) were exchanged, step-by-step, thus allowing the multiplexed detection of up to four proteins by super-resolution imaging, in a layer-by-layer fashion (86).

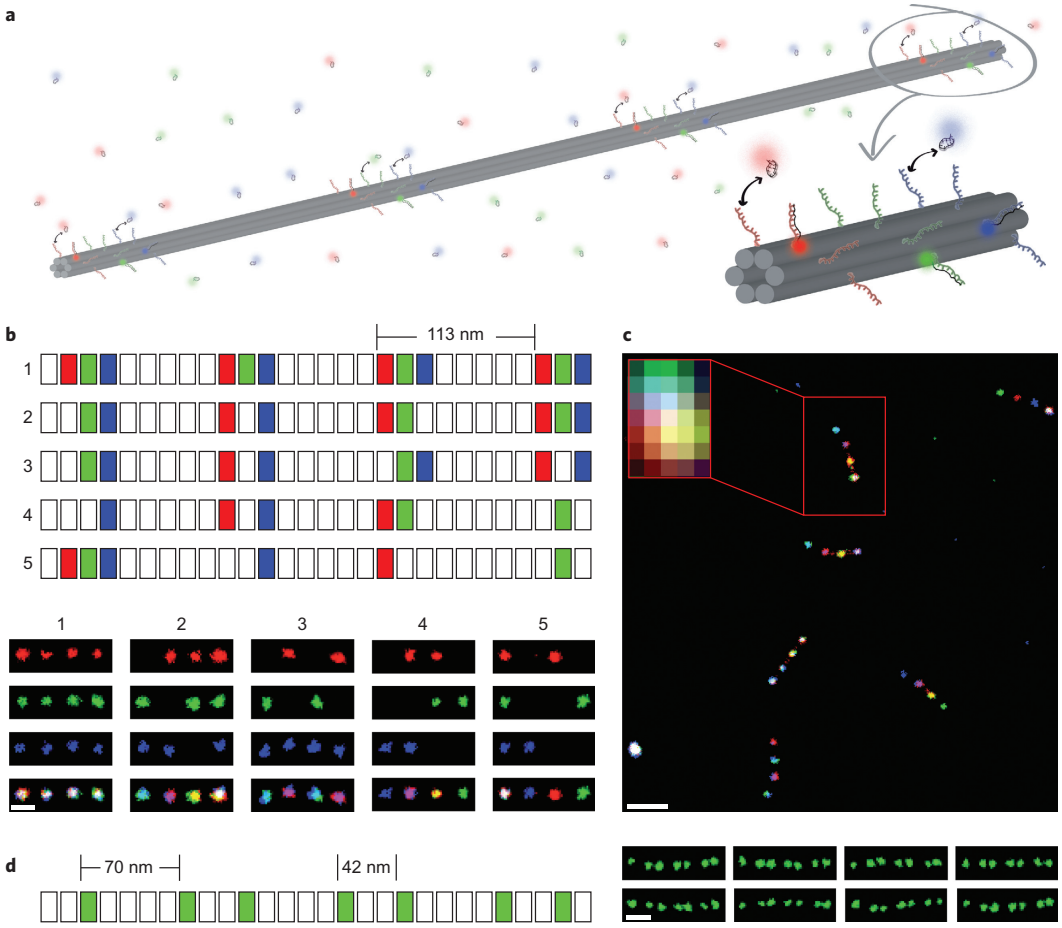


Figure 1.12. ***a.*** The imaging approach, termed DNA-PAINT, relies on the continuous association/dissociation reactions involving the Origami-affixed ssDNA molecules, and the complementary, fluorescently-labelled ssDNA (imager) molecules in solution. The latter molecules are ubiquitously present during imaging, but are preferentially localized to the corresponding hybridization sites on the Origami surface. Transient binding between imager and docking strands produces fluorescence "blinking," allowing stochastic super-resolution imaging (86). The depicted nanostructure is functionalized with docking strands in four different areas to create a barcode. ***b.*** Schematics (top) and super-resolution images (bottom) of five barcodes obtained with the same stick-shaped DNA nanostructure. ***c.*** Mixture of barcodes presented in b. ***d.*** Using the same fluorescent label an asymmetric barcode was designed separating docking strand regions by 70 or 42 nm (85). © 2012 Macmillan Publishers Limited.

DNA Origami technology is not limited to the creation of smart devices for imaging, but also opens up the prospect of nanomachines that can control their shape and perform operations in a programmable manner in biological systems (62, 87). This has provided the basis for development of novel carriers for the delivery and release of drugs in a cell-specific manner (72, 88–91). For instance, Högberg, Teixeira and coworkers utilized rod-like DNA Origami to determine the spatial distribution of ligands, thus affording accurate distance manipulation at the nanoscale. Thus, ligand spacing can be controlled independently of ligand concentration (92). The device, termed ligand "nanocaliper," displays precise, programmed patterns of ligands on its surface. The tested hypothesis is that cell signaling is usually performed using a group of receptors and ligands, and that within the cell membrane, receptors are confined in spatially organized membrane domains, such as lipid-rafts. The authors applied nanocalipers to map the relationship between the nanoscale spacing of ligand and cell surface receptor activation resulting in a cellular response. It was determined that the nanoscale distribution of the ligand ephrin-A5 affects signal transduction and cell behavior. Cells stimulated with an ephrin-A5 ligand nanocaliper, with 40 nm spacing, showed higher levels of activating phosphorylation of the EphA2 receptor, and lower invasive properties of the breast cancer cells (MDA-MB-231) than with the cells stimulated with the monomeric ligand, or dimeric ligands spaced at 100 nm. In summary, this study demonstrates an unprecedented application of DNA Origami for stimulating a cellular response by controlling the "display architecture" of nanoscale-confined cell surface ligands (92).

In combination with atomic force microscopy (AFM), two-dimensional DNA Origami have been utilized as nanochips for biosensing applications (93, 71). Lindsay, Yan and coworkers created a DNA origami as an RNA detector, potentially applicable to gene expression analysis at the single-cell level (93). The method however, would be challenging to implement, as RNA detection requires that the nanochips are first immobilized on a surface, then sample-imaged by high-resolution AFM, allowing the enumeration of individual RNA molecules immobilized on their respective nanochip (93). In contrast, the combination of fluorescence imaging with self-assembled DNA nanostructures has led to novel approaches that also underscore the impact of fluorescence-based tools on tissue imaging.

Obviously, concerns about biostability and biocompatibility must be addressed, as nucleic acid-based nanotechnology is still in its infancy (91, 94, 95).

1.7.3. DNA nanostructures resistance to enzymatic digestion

Given the thrust to develop applications in drug delivery or in vivo biosensing, the fate of DNA origami in cell lysates or in the presence of specific nucleases has been investigated. Castro et al. found that the incubation of different 3D DNA nanostructures with exo- or endonucleases provide disparate results (96). However, compared to a circular dsDNA plasmid, the DNA nanostructures more resistant to DNase I action. Meldrum, Yan and coworkers examined the behavior and stability of triangular, rectangular, and cuboidal DNA nanostructures in mammalian cell lysates (97). While the DNA origami structures were apparently unaffected by incubation in the cell lysates at different temperatures, with no strong interactions with cellular components as noted by nondenaturing gel electrophoresis, the plasmid DNA was shown to interact with one or more lysate proteins of unspecified identity. In addition, a rectangular DNA nanostructure, functionalized with ssDNA overhangs that can bind a short ssRNA, was shown to retain its functionality after incubation in cell lysates. Meldrum et al. argue that the enhanced resistance of the DNA nanoassemblies in the presence of the myriad cell components reflects the high charge density and molecular rigidity of the assemblies, which hinders protein engagement, perhaps due to preventing conformational changes required for recognition.

With the new knowledge of the stability of nanoconfined DNA structures, biocompatible carriers were designed that could deliver a siRNA molecule into a xenografted tumour. This approach provided a nearly four-fold increase in siRNA half-life (98). Also, DNA origami containing intercalated doxorubicin molecules allowed increased drug internalization in cultured mammalian cells, and a concomitant strong reversal of drug resistance in doxorubicin-resistant cancer cell lines (91).

Chapter 2 – Thesis outline

2.1. Project 1: DNA-nanostructures enzymatic restriction quantification

DNA nanostructures can be used as innovative, biocompatible, versatile nanosensor. They are formed by a long, circular DNA single strand (thousand nucleotides) folded to form a specific shape by hundreds of short (30 nucleotides) DNA single strands (staples) that hybridize over non-consecutive regions of the scaffold. Staples incorporation within the structure follows well-defined stoichiometry and they can be modified with organic or inorganic molecules. Staples carrying a tail of DNA, RNA or LNA can be used as highly-specific receptor for short nucleic acids sequences or proteins. In our nanosensor design receptor staples are modified with a restriction site that will be protected from enzymatic degradation by the presence of the target. The restriction reaction therefore "writes" the amount of target molecules captured on the nanosensor surface by permanently modifying certain inherent DNA staples. Such modified staples can be subsequently analyzed with quantification techniques such as quantitative (q)PCR, or Next Generation Sequencing (NGS).

For the development of such nanosensor the enzymatic efficiency and staples stability within the nanostructure need to be clarified. Recent results obtained in our laboratory showed that, in the folded DNA structure, several restriction enzymes are able to process only a few of their corresponding restriction sites because the other are sterically protected from enzyme access/binding. Specifically, I participated in a study of the effect of 14 different restriction enzymes on a rectangular and a triangular DNA Origami (structures published by Rothemund P.W.K. (65)), originated from the same scaffold (M13mp18 genomic DNA), and found enzyme-dependent and shape-dependent fragmentation patterns (fig. 2.1). Such highly distinct fragmentation patterns reflect the fact that only a few restriction sites are accessible to their respective enzymes, while the others are fully protected from enzymatic processing. Recent results in our laboratory have examined enzymatic reactions in a site-

specific manner, by processing mutated nanostructures containing a single activated site. This approach allowed to uncover the key structural determinants of restriction reactions in planar DNA Origami, but remains laborious as it requires the generation of tens of different nanostructures per enzyme (Stopar A. et al., *manuscript in preparation*). The approach, in turn, is non-practical for extending our study to more complex DNA nanostructures (e.g. 3D vs. 2D) or different type of enzymes. Therefore, one of my goals is to develop a quantitative tool for the parallel analysis of enzymatic processing within "wild-type" DNA nanostructures, as well as other biochemical reactions such as DNA staples incorporation efficiency, or the overall nanostructure stability in complex environments, with particular interest for nuclease-containing biological mixtures such as cell lysates, and blood serum. The novel methodology that I plan to develop will be also applied as readout for the nanobiosensors that I plan to develop.

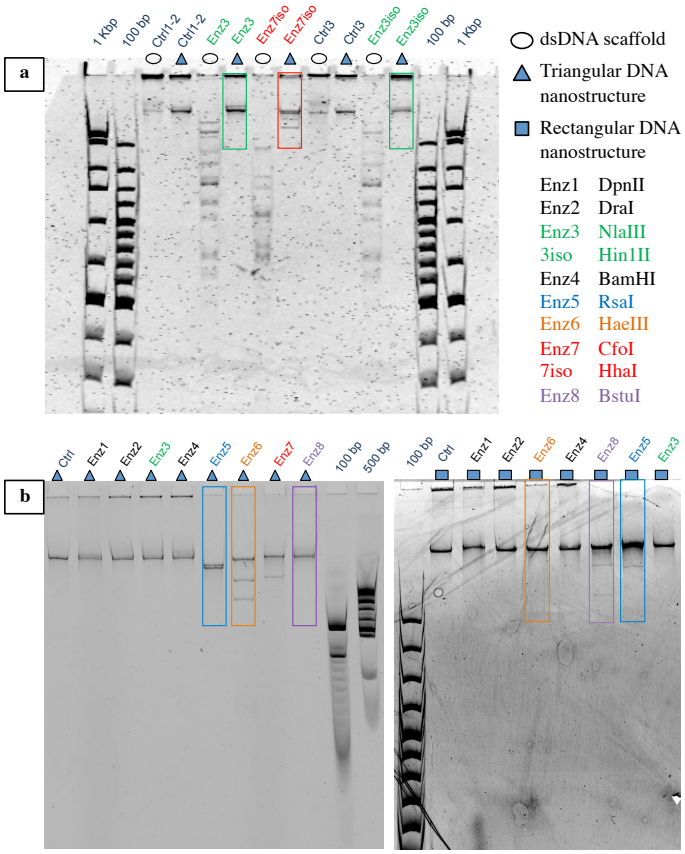


Figure 2.1. ***a.*** Enzyme 3 and the izoschizomer 3 and 7 digest the duplex DNA strand with the same sequence of the scaffold of the triangular DNA nanostructure. Nevertheless, only enzyme 7iso partially digests the triangle. ***b.*** Enzyme 5 and 6 produce an enzyme-dependent fragmentation pattern of the triangular DNA nanostructure, but do not digest the rectangle. Enzyme 8 has opposite behavior. The enzymes, therefore, display a DNA shape-dependent fragmentation capability. Controls are nanostructure or dsDNA we incubated in the enzyme buffer lacking the enzyme.

2.1.1. Experimental design

The methodology designed in this thesis for analyzing restriction reactions within DNA nanostructures focuses on the detection of enzyme-specific DNA staple fragment vs. unmodified staples as reaction products. A linear PCR reaction (L-PCR) converts ssDNA staples fragments in duplex DNA fragments that can be analyzed with NGS. For the preparation of NGS libraries (DNA amplification and adapters ligation), in fact, the target DNA has to be double-stranded. In the L-PCR step, staple-specific carriers (70-nucleotides-long ssDNA) can hybridize to their corresponding, shorter, ssDNA staples (or staple ssDNA fragments). This leads to the formation of 70 base pair (bp) long dsDNA fragments by means of an enzymatic elongation reaction. To quantify enzymatic processing in DNA nanostructures, we designed carriers containing the sequence complementary to one of the two fragment of the cleaved staple. If the staple is cleaved by restriction enzyme, it will fully hybridize to the carrier, otherwise the partial hybridization of an unmodified staple will stop polymerase activity (see the diagram at the bottom in fig. 2.2-a). To quantify staple incorporation efficiency, I designed carriers with a segment complementary to the full staple sequence. In this case, enzymatic elongation of the staple-carrier construct always produce a 70-bp-long duplex (see fig. 2.2-b). In both cases, the number of outputted copies of linear PCR products per each staple is equal to the number of copies of target staples.

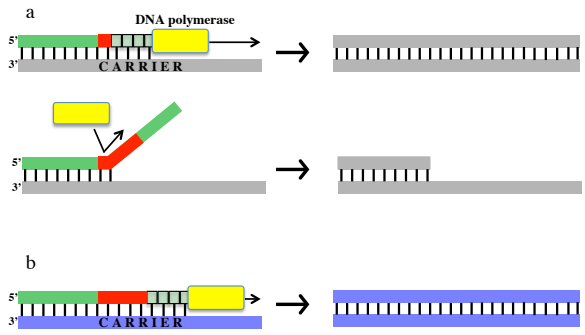


Figure 2.2. Schematics of linear PCR converting ssDNA staples (or inherent fragments) in duplex DNA ready for NGS library preparation. ***a*** The reaction is able to discriminate between cleaved and un-cleaved staple. ***b*** The reaction can be applied to the analysis of unmodified staples by using a different carrier, to which the unmodified staple fully hybridizes.

The use of carriers reduces the influence of DNA sequence and/or length on the efficiency of the amplification reactions required for NGS or qPCR analysis. In addition, several carriers can be used in parallel to evaluate site-specific enzymatic processing or staple

specific incorporation in a DNA nanostructure without requiring modifications to such nanostructures. As a proof-of-concept, I focused on the MspI restriction reaction in a two-dimensional, triangular, DNA nanostructure (designed by Rothemund P.W.K. (65)) that contains 18 restriction sites for the enzyme. Five staples lacking the restriction site were chosen as positive and negative controls (fig. 2.3):

- 3 staples carry no MspI restriction sites and are placed in peculiar region of the triangle (negative controls A, B and C);
- 1 staple forms a stem-loop structure that protrudes from the triangle surface, and carries a MspI restriction site in the stem (positive control Hairpin);
- 1 staple contains a ssDNA segment that protrudes from the planar surface of the DNA triangle and is hybridized to a complementary strand forming a short, duplex DNA segment containing a MspI restriction site (positive control dsDNA).

The described controls and one staple containing a MspI restriction site (staple X, the shortest, cleaved staple) were analyzed in the preliminary experiments described in this thesis.

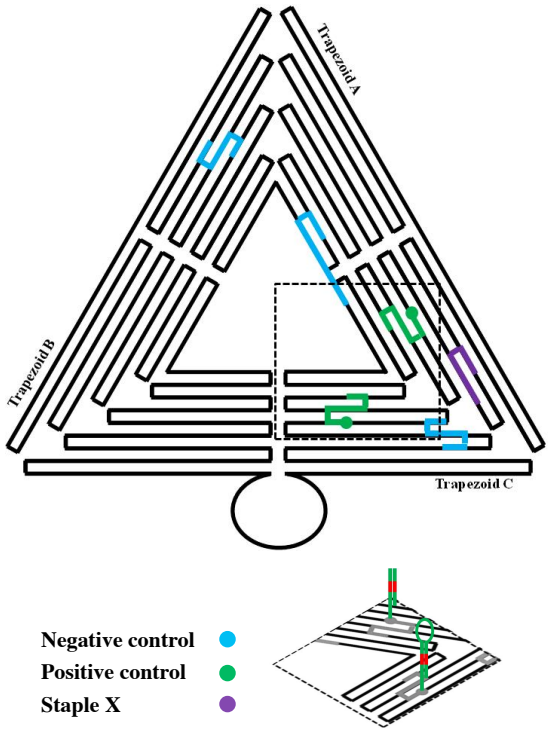


Figure 2.3. Diagram of the triangular DNA nanostructure indicating the location of staples selected for the proof-pf-concept analysis, and the inherent motifs.

2.2. Project 2: DNA based nanosensor for accurate nucleic acid quantification (Bingo-qPCR assay)

Nucleic acid quantification with qPCR can be biased by amplification reactions and, in the case of RNA, by retro-transcription reaction. Moreover, the detection threshold of 2-fold variation in DNA concentration is typically circumvented with more complex approaches such as digital-PCR or NGS. Our work aims at coupling qPCR with a self-assembled nanosensor that can help overcome the aforementioned difficulties without requiring updates to traditional qPCR instrumentation.

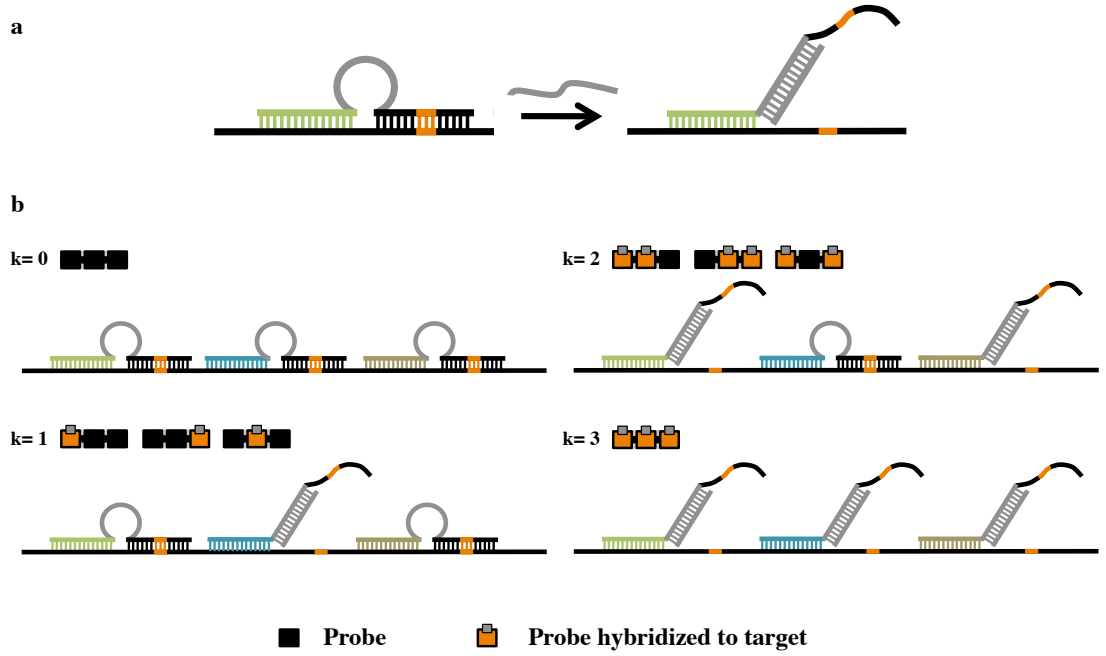


Figure 2.4. a Schematic representation of the sensing element in the Bingo-qPCR nanodevice: the probe has two feet (green and black) hybridized to the scaffold (black line), and one of the two elements carries the site (orange) of a specific restriction enzyme. The target hybridizes to the probe-loop, thus unfolding the restriction site (switch reaction). **b** Schematic representation of full Bingo-qPCR nanodevice in the presence of the target: the target (grey square) is less concentrated than the probes, therefore only a fraction of the sensing DNA loops are converted into duplex DNA, leading to disruption of the associated restriction sites. The disruption of three restriction sites over the same nanodevice copy is, therefore, significantly less likely.

2.2.1. Experimental design

Components of such sensor are three "foot-loop" DNA probes sequentially assembled over a long segment of ssDNA termed scaffold. Each probe carries a target-complementary sequence in the central loop, and scaffold complementary sequences at both termini (the

"feet"). Each probe carries a restriction site within only one of the two feet. According to our design, as the probe fully hybridizes to the scaffold, a structure that cannot bind the target is formed. As the concentration of the target in solution increases, a binding-competent form in which the site-carrying foot is dissociated from the scaffold is favored. In this form, probe-target recognition is allowed, thus leading to a stiffening of the sensing element (the loop) as the inherent single stranded sequence is converted in a stable duplex. Such stiffening prevents the complete probe-scaffold re-association, thus disrupting the restriction site (fig. 2.4-a.). In the presence of the target, it is possible that three restriction sites are disrupted over the same scaffold leading to a special configuration termed "bingo" (with a certain frequency that will be discussed in the following sections). Only in this case, all of the restriction sites are disrupted in the nanosensor, and it turn the whole scaffold is protected from enzymatic cleavage. In this way, the full sequence of the bingo-scaffold can be amplified with PCR. Bingo-scaffold can be quantified in almost any laboratory for molecular biology with standard qPCR instrumentation. As previously described this technique is based on the amplification reaction of the dsDNA of interest, monitored in real time by detecting fluorescence emission of a dye intercalated in the duplex product. The fluorescence signal is directly proportional to the amount of DNA product present in the solution and has a sigmoid behavior. Each cycle of the reaction proceeds along denaturation, primer annealing and polymerization steps and at each cycle DNA doubling should occur. The cycle number at which fluorescence raises above noise is called Ct. In the analysis of two samples, i.e. A and B, in which, for example, the target concentration in A is higher than the target concentration in B, the Ct shift between the fluorescence curves associated with the two samples will be $\Delta Ct = Ct_B - Ct_A$.

The ratio between the two sample is expressed as follows

$$\frac{[A]}{[B]} = 2^{Ct_B - Ct_A} \longrightarrow \Delta Ct = \log_2 \frac{[A]}{[B]}$$

Equation 1

2.2.2. Theoretical model

2.2.2.1. Using a Bingo-qPCR nanodevice with n probes, a two-fold variation in target concentration corresponds to 2^n-fold variation in bingo-scaffold concentration

In the Bingo-qPCR nanodevice probes are in excess with respect to the target, leading to various configurations of probe occupancy on the same nanodevice (fig. 2.4-b). The number of scaffold copies carrying none, one, two or three target molecules can be calculated

following the binomial distribution, which determines the probability of having ***k*** successful (binding) events in a sequence of ***n*** Bernoulli trials (i.e. the active probes in the same nanodevice). The number of combinations of ***k*** successful binding events over ***n*** probes is given by the Binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Equation 2

A Bernoulli trial is a random experiment with exactly two possible outcomes, "success" and "failure", and the probability of success (***p***) is the same every time the experiment is conducted. Conversely, the probability of failure is ***q*** =1-***p.*** In this way, the probability of a Bernoulli process leading to **k** successes in **n** trials is expressed by the Binomial distribution:

$$\Pr(X = k) = \binom{n}{k} p^k \cdot q^{n-k}$$

Equation 3

This formula provides a proper probability distribution according to the Newton Binomial Theorem, that proves that the sum of the probabilities over all of the combinations is one.

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = (p+1-p)^n = 1^n = 1$$

Equation 4

In our case the probability of success ***p*** is defined as the probability that the target binds to a probe in the nanosensor. In the simplified case in which the probe concentration is much higher than the target concentration, ***p*** can be approximated with the ratio between these two concentrations:

$$p = \frac{[\text{target}]}{[\text{probe}]}$$

Equation 5

The probability to obtain a bingo-scaffold (P) is the probability of having k=3 successes over n=3 recognition events.

$$P = \binom{n}{k} p^k q^{n-k} = \binom{3}{3} p^3 q^{3-3} = p^3$$

Equation 6

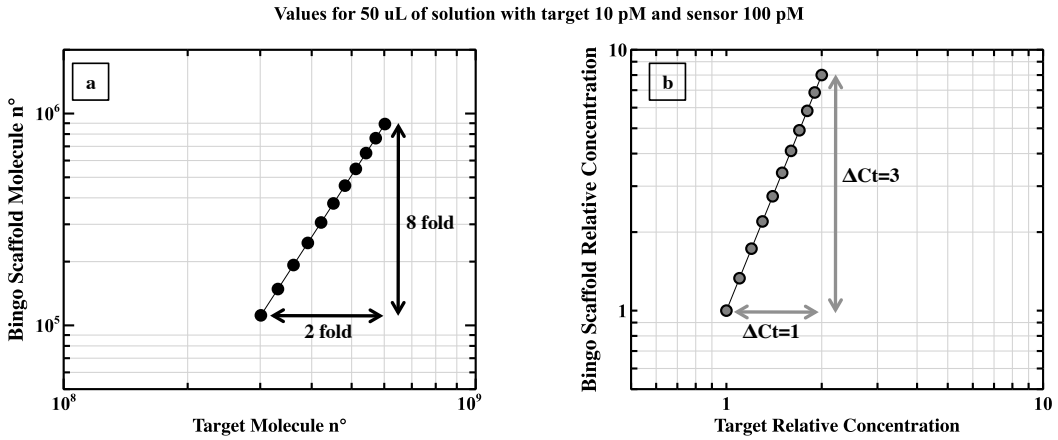
I turn, target concentration and bingo-scaffold concentration are correlated by the following power function:

$$[\text{Bingo Scaffold}] = [\text{Scaffold}] \times P = [\text{Scaffold}] \times p^3 = [\text{Scaffold}] \times \left(\frac{[\text{target}]}{[\text{probe}]} \right)^3$$

Equation 7

This allows transforming a two-fold variation of target concentration into 8-fold variation of bingo-scaffold concentration (plot 2.1. a). Having a nanosensor with more than 3 probes will further increase the signal gain, thus improving detection accuracy. The accuracy limit of qPCR is about a 2-fold target variation. Reducing standard deviation (by increasing replicates number for example) can bring the accuracy to 1,3-1,5 fold. In the standard qPCR, sample dilution is however required, therefore making the accurate analysis of micro-samples (volumes < 1μL) nearly impossible. If data replicates follow normal distribution, for a 100% accurate quantification of ΔCt=1 the standard deviation will need to be lower than 1/6. In other words, to quantify a target variation lower than 2-fold, the standard deviation has to be < 0.167, which is laborious to obtain experimentally. Therefore, in standard diagnostic assays ΔCt=1 is typically considered the minimum value measurable with qPCR. (26-29-30)

With the Bingo-qPCR assay, a 2-fold target concentration variation is converted into ΔCt=3. In fact, if target has a 2 fold increase from sample A to sample B, the bingo-scaffold has a variation of 8 fold ($2^{k=3}$) corresponding to a ΔCt=3 (plot 2.1. b).



Plot 2.1. a. The Bingo-qPCR detection mode converts a 2-fold variation in target amount into a 8-fold variation of bingo-scaffold amount. **b.** qPCR output: a ΔCt value of 1 is converted in a ΔCt of 3.

2.2.2.2. Target concentration calculation

To obtain target concentration from Bingo-qPCR assay we calculate the corresponding amount of un-cleaved restriction sites in the sample.

Let's consider having sample X and Y, in which only X has been incubated with the target (as Y is the Negative CTRL corresponding to the maximum amount of cleavable restriction sites in the sample). To obtain the restriction efficiency for sample X let's quantify the fraction of unmodified restriction sites with qPCR in X and Y, by amplifying either the ssDNA bingo-scaffold or the portion of scaffold that surrounds each restriction site (i.e. separating the analysis of each site associated with the probes A, B, and C – this procedure is termed "site-specific" qPCR assay) and calculate the ratio of the values found for X and Y.

Site specific qPCR assay calculations.

In the site-specific qPCR assay (for A, B or C) we can calculate the ratio of un-cleaved restriction sites A, B or C between sample X and Y as follows.

$$\frac{[X_A]}{[Y_A]} = 2^{Ct(Y_A)-Ct(X_A)} \qquad \frac{[X_B]}{[Y_B]} = 2^{Ct(Y_B)-Ct(X_B)} \qquad \frac{[X_C]}{[Y_C]} = 2^{Ct(Y_C)-Ct(X_C)}$$

These values should reflect the experimental quantification of the probability $\boldsymbol{p}$ to find an un-cleaved restriction site in A, B or C after enzymatic restriction.

Bingo-qPCR assay calculations.

In the Bingo-qPCR assay, we can calculate the ratio of bingo-scaffold in sample X and Y as follows.

$$\frac{[X_{Bingo}]}{[Y_{Bingo}]} = 2^{Ct(Y_{Bingo})-Ct(X_{Bingo})}$$

These values should reflect the experimental quantification of the probability $\boldsymbol{P}$ to find the integer bingo-scaffold in sample X after enzymatic restriction. The probability $\boldsymbol{p}$ to find an unmodified restriction site independently in A, B or C in sample X is such that $\boldsymbol{p}^3=\boldsymbol{P}$ (see equation 6).

$$p = 2^{\frac{Ct(Y_{Bingo})-Ct(X_{Bingo})}{3}}$$

2.2.2.3. The Bingo-qPCR assay increases the statistical accuracy of qPCR by 3-fold

The uncertainty **σ** of the exponential function $f(x)=n^{kx}$ is approximated by the equation

$$\sigma \approx n^{kx} \cdot k \cdot \ln(n) \cdot \sigma_x$$

where σ_x is the uncertainty associated with the independent variable x.

In our experiments, p is the probability to find an un-cleaved restriction site in A, B or C after enzymatic restriction and therefore is expressed by

$$p = 2^{\Delta Ct} = 2^{\frac{\Delta Ct_{(Bingo)}}{3}}$$

The uncertainty σ_p associated with p calculated with the standard assay can be approximated by

$$\sigma_p \approx 2^{\Delta Ct} \cdot \ln 2 \cdot \sigma_{\Delta Ct} \qquad \text{where} \qquad \sigma_{\Delta Ct} = \sigma_{Ct_X} + \sigma_{Ct_Y} .$$

The uncertainty σ_p^* associated with p, but now measured with a Bingo-qPCR assay can be approximated by the following formula

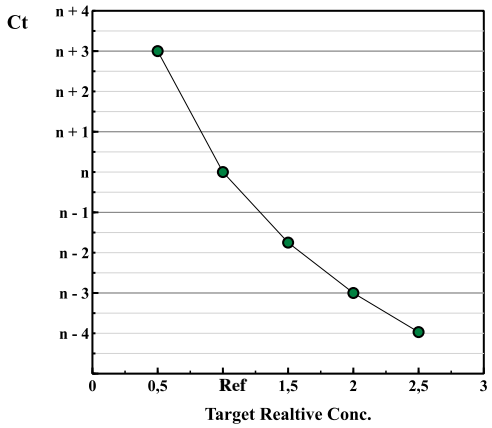
$$\sigma_p{}^* \approx \frac{1}{3} \cdot 2^{\frac{\Delta Ct_{(Bingo)}}{3}} \cdot \ln 2 \cdot \sigma_{\Delta Ct_{(Bingo)}}$$

indicating therefore that $\sigma_p^* = \sigma_p/3$.

In the general case of a Bingo-qPCR nanodevice comprised of k probes, the uncertainty associated with the measure of p using the Bingo-qPCR assay vs. a standard PCR assay becomes $\sigma_p^* = \sigma_p/k$, ideally with an arbitrarily low value of k.

2.2.2.4. Probe:Target ratio in bingo model

Equation 7 is acceptable if the target-to-probe ratios << 1, and therefore a prior semi-quantitative determination of target concentration, e.g. by using qPCR or other standard techniques should be performed (reference sample). For instance, using a 1:10 target-to-probe ratio allows the analysis of target concentration variations in a narrow range (2,5-fold). Of note, with a 1:10 target-to-probe ratio, the bingo-scaffold concentration is a 1000-folds lower than the initial scaffold concentration (Eq. 7) (plot 2.2.).



Plot 2.2. The reference sample has target-to-probe ratio 1:10. Bingo-qPCR assay allows to discriminate between concentration of 0,5-1,5-2-2,5 fold with respect to the reference with at least $\Delta Ct = 1$.

Chapter 3 – Materials and Methods Project 1

3.1. Triangular DNA nanostructure enzymatic treatment

3.1.1. Triangular DNA nanostructure folding

For accurate information about design of the nanostructure see Appendix I.

The DNA origami folding solution contains:

- single strand phage M13mp18 final concentration 10 nM (New England Biolabs (NEB), see appendix I for the sequence)
- 207 different staples final concentration 50 nM (Biomers, see appendix I for the sequences)
- folding buffer (Tris 10 mM, MgCl2 12.5 mM, Sigma Aldrich).

The solution undergoes thermal treatment that consists in a denaturation step at 95°C for 10 minutes and a ramp of decreasing temperature to 20°C with 1°C/min rate. The instrument used is the T100 Thermal Cycler (Bio-Rad).

3.1.2. Triangular DNA nanostructure purification with PEG

The solution used to purify DNA nanostructures contains 15% PEG (polyetilene glycol) and 0.5 M NaCl. 1 volume of PEG-NaCl solution is added to the sample and the mixture is homogenized by gentle pipetting. After 30 minutes of incubation in ice, the sample is spinned at 12000 g for 30 minutes (using a refrigerated centrifuge at 4°C). The supernatant is removed and the nanostructure pellet is resuspended in half volume of Tris 10 mM. The protocol is repeated a second time, and the pellet is resuspended in one volume (99).

3.1.3. Triangular DNA nanostructure enzymatic digestion

MspI enzyme was purchased from New England Biolabs (NEB). Recommended volume for enzymatic reaction is 50 µl. Typical enzymatic reaction contains 25 µl of 10 nM DNA Origami, 1-fold enzymatic buffer Cutsmart (NEB) and 1μL of the enzyme. The amount of

enzyme was calculated from the producer definition of enzyme unit: “the amount of enzyme required to digest 1 ug of λ DNA in one hour at 37°C in a total reaction volume of 50 μL”. Thermal treatment is carried out at 37 °C for 1 hour in water bath. The reaction is blocked by placing samples in ice, freezing or by incubating the sample at 80°C for 20 minutes.

3.2. DNA sequences in silico analysis

There are several programs available online to analyze DNA strands. For this project, we mainly used IDT Oligoanalyzer (100) and BLAST (101). To design carriers we evaluated the ability of each sequence to form self-dimer, hairpins, hetero-dimers with other carriers, staples (cleaved and not cleaved) or with the scaffold. All carriers were tuned by manually changing their sequence to avoid strong secondary structures.

3.2.1. IDT OligoAnalyzer

OligoAnalyzer is a web application of IDT website that allows analysis of DNA sequence. The input sequence can be analyzed in terms of secondary structure formation (hairpin, self-dimer, hetero-dimer) in customized conditions of salt (Mg^{++}, Na^{+}) and DNA concentration. The output is expressed as melting temperature of the structure formed and as Gibbs free energy (G), enthalpy (H) and entropy (S). Other data such as GC content, molecular weight and extinction coefficient can be obtained. It can be used for basic analysis involving maximum 2 strands.

3.2.2. BLAST (Basic Local Alignment Search Tool)

BLAST is typically applied to comparing nucleotide or protein sequences to sequences from databases. It can find sequence similarities and complementarities with three different algorithms: Megablast, Discontinuous megablast and Blastn. We used Blastn to analyze carrier complementarities to scaffold, staples and other carriers. This algorithm finds short matches (down to 7 nucleotides) that might be relevant on a DNA sequence of 70 nucelotides.

3.3. DNA gel electrophoresis

When analyzing DNA complexes containing duplex and single stranded portions, molecular weight DNA ladder bands cannot be used as an effective molecular weight reference.

3.3.1. Polyacrylamide gel electrophoresis (PAGE)

The Mini-PROTEAN Tetra cell Bio-Rad has been used for 1-D vertical gel electrophoresis. Polyacrylamide 29:1 (40 %) was purchased by Thermo-Fisher. TEMED was purchased by Bio-Rad and stored at 4°C. APS powder was purchased by Sigma and the stock solution (10% w/v in water) was stored -20 °C.

TBE (Tris base 0.9 M, Boric Acid 0.9 M, EDTA 20 mM, single component purchased by Sigma Aldrich) running buffer was prepared in-house.

The gel is prepared by mixing polyacrylamide 29:1, TBE and water. Immediately before pouring the gel, TEMED and APS 10% were added to the solution in 1:1000 and 1:200 ratio respectively. After pouring the mix in the gel caster, the comb was positioned gently in the gel preventing air bobbles formation on the bottom of the wells. Empty gels were pre run to equilibrate ions inside of it. Sample loading was performed using glycerol enriched (20% final glycerol concentration) loading buffer (TriTrack DNA Loading Dye (6X) Thermo Fisher Scientific). Each gel was run in TBE buffer 1X at constant voltage.

Gel staining was performed by soaking the gel in ethidium bromide solution (10 mg/ml) at the end of the run. Excess of intercalating agent is removed by washing the gel with water before image acquisition (ChemiDocTM XRS, Quantity One Gel Doc software). Gel images were visualized and analyzed with ImageJ.

3.3.2. Agarose gel electrophoresis (AGE)

Agarose gel was prepared and run using Bio-Rad Mini-Sub® Cell GT System. Agarose gel is prepared by heating a solution of TAE 10 X (Tris-acetate 400 mM, EDTA 10 mM) containing the specific percentage (w/v) of agarose powder (2 %). Attention was paid in order to dissolve the powder without boiling the solution to avoid agarose degradation. Agarose solution was poured into gel caster with the comb already positioned. The comb is removed when the gel solidifies. To load samples (7 μL) the gel is immersed in cold (4° C) running buffer (TAE) and the addition of glycerol containing loading buffer (2 μL)

facilitates the deposition of samples on the bottom of the wells. Gel was run at constant voltage (80 V). The 10 base pairs molecular weight marker O'Range Ruler (Thermo Scientific) was used (3 μL). 105 ng of duplex DNA were loaded in gel.

3.3.3. ImageJ gel bands quantification

For comparing gels, these should have the same characteristics, and therefore I usually chose to invert the gel and crop it full size. To calculate background, I generated the plot profile of the whole gel area, exported values and plotted them in an excel scatter plot. For each band, I selected one value on the left and one on the right of the band and used the mean value as background.

To calculate gel band intensity, I rotated 90° the image of the gel and generated the plot profile of each sample lane using fixed dimension of the selected area. The section has to fit tightly the bands of each sample excluding the neighboring background areas. Values of the profile are exported in Excel file, and the calculated background is subtracted. These values can be plotted to obtain the normalized profile.

To obtain a comparable value, the intensity of the bands is calculated as "Full Width at Half Maximum" integral of the curve of the band. To identify the half maximum value the peak maximum is selected and divided by two. Band intensity is obtained by the sum of all the points below this value.

3.4. Linear PCR

The instrument used is the T100 thermal cycler Bio-Rad. Master Mix used is 2G Fast Hot Start Readymix by Resnova. Each reaction has final volume of 25 μL. DNA nanostructure concentration is 2 nM, other experiments might have different concentration for technical reason (appendix I).

3.5. DNA quantification with spectrophotometer

The instrument used was Varian Cary 5000 at the Department of Pharmaceutical Chemistry in Trieste University. The measurements were performed on DNA samples with expected concentration of 100 μM diluted 1:100 in water. The scanning range was set from 500 to 210 nM. To calculate the concentration, extinction coefficient values from Sigma

Aldrich were used, while the path length of the beam light through the material sample is 1 cm. The photometric accuracy of the instrument is less than 0.00025. From quantified staples stock solutions at 100 μM, dilutions at 10 μM were obtained and used in the experiments.

3.6. DNA gel extraction

3.6.1. DNA gel extraction from acrylamide gel

To excise the gel slice containing the DNA I used a scalpel and followed band shape (particularly relevant for bent bands). For gel extraction QIAEX® II Kit by Qiagen was used with few modifications to the original protocol. I added 2 volumes of diffusion buffer (v/w) to the sliced bands and crushed them using a palette knife. After 2 hours incubation at 50°C samples were centrifuged at 10000 g for 30 sec and the supernatant was removed with a pipette. Using a Whatman GF/C filter (Sigma Aldrich, pore size 1.2 μm) any residual polyacrylamide was removed. I added 6 volumes of buffer QX1 to 1 volume of sample. QIAEX II buffer was vortexed for 30 seconds and then added to the sample. During the 10 minutes incubation at room temperature, the mixture was vortexed every 2 min to keep QIAEX II in suspension. Now DNA is bound to QIAEX II beads. 2 minutes of centrifuge at 10000 g bring the beads to the bottom of the tube and the supernatant can be removed. In most cases the supernatant solution still contained beads, therefore I centrifuged it for 2 minutes at 10000 g and removed the clean supernatant. To collect together the two pellets, I added 400 μL of washing buffer PE to the first and 100 μL to the second, then I resuspedend and mixed the solutions. The washing step is repeated by centrifuging at 10000 g for 2 minutes, discarding the supernatant and adding 500 of PE buffer to resuspend the pellet for the second time. The pellet is then air-dried for 10–15 min and 20 µl of Tris HCl 10 mM pH 8.5 are added to resuspend the pellet by vortexing. After incubation for 10 min at room temperature the solution is spinned for 30 seconds and the supernatant, which now contains the purified DNA, is transferred to a clean tube. The elution step is repeated to increase yield and the final volume of purified DNA is 40 μL.

3.6.2. DNA extraction from agarose gel

The same kit used for acrylamide gel extraction was used (QIAEX® II Kitby Qiagen), but in this case no modification to the protocol was introduced.

3.7. DNA EtOH-Na Acetate precipitation

Na-Acetate 3M pH 5.2 (10X) and 2 volumes of ethanol 100% were added to Linear PCR (L-PCR) product and the solution was mixed by inversion. The tube was then placed in vertical position at -80˚C for 2h (or -20°C over-night), and subsequently centrifuged at 18000 g at +4˚C for 30 min.

The supernatant was removed with a pipet leaving 10 μL on the bottom of the tube. To wash DNA pellet 1 volume of cold ethanol 70% was added, and the solution mixed by pipetting. After Centrifuging at 18000 g at +4˚C for 30 min, the supernatant was removed with a pipette leaving 10 μL on the bottom of the tube. Ethanol excess was removed by spontaneous evaporation. 10 mM Tris pH 8.5 was used to reach final volume.

3.8. Duplex folding protocol

To form the duplexes, the oligos used in the experiment have been quantified with spectrophotometer, while the two complementary strands of each sequence were hybridized separately in 1:1 ratio.

3.8.1. Protocol 1

The thermal ramps tested are reported in figure 3.1. Folding buffer used for protocol a was NaCl 50 mM and Tris-HCl 10 mM with final concentration of duplex DNA of 100 ng/μL to easily visualize DNA on gel. To analyze DNA with NGS, the DNA needs to be salts-free and therefore samples were purified with EtOH precipitation.

3.8.2. Protocol 2

In this new experiment we used the DNA nanostructure folding buffer and the thermal protocols b-c of figure 3.1, which are consistently slower than protocol 1, to fold carrier of negative control A and the complementary sequence.

a	
95°C	2 minutes
94 - 25 °C	40 sec/°C
25°C	2 minutes

b	
95°C	10 minutes
94 - 62 °C	0.5°C/min
62°C	10 minutes

c	
95°C	10 minutes
94 - 50 °C	0.5°C/min
4°C	2 minutes

Figure 3.1. Duplex folding protocols

3.9. DNA sample analysis: quantification and molecular weight profile evaluation

3.9.1. DNA quantification: Qubit High Sensitivity DNA Kit

The protocol is suitable for sample volumes between 1-20 μL. We typically used 1 μL of sample. The assay is highly selective for double-stranded DNA (dsDNA) and is accurate for initial sample concentrations from 10 pg/µL to 100 ng/µL. For each sample, 200 μL of working solution is prepared by diluting the dsDNA HS Reagent (intercalating dye) 1:200 in dsDNA HS Buffer. 1 μL of sample is added to 199 μL of working solution, using tubes provided by the kit that are compatible with Qubit laser. After vortexing for 2-3 seconds and a 2-minutes incubation at room temperature protected from light, the sample is placed in the instrument to perform the measurement. The temperature is critical for the assay, and therefore, to repeat the measurement on the same sample, the latter should be left at room temperature for 2 minutes, and protected from light. Instrument calibration was performed every second day using standard solutions included in the kit (standard volume used is 10 μL). The instrument calculates the concentration of the sample by the equation Cf x Vf = Ci x Vi where

- Cf is final concentration measured by the the instrument
- Vf is final volume (200 μL)

- Vi is initial volume that can be 1 to 20 μL and is specified by the operator before the measurement.

3.9.2. Agilent 2100 Bionalayzer

The Agilent 2100 Bionalayzer instrument can perform quantitative and qualitative analysis of nucleic acids and proteins or flow cytometry with a "lab on a chip" setup. Electrodes, optics for detection and a chip holder form the instrument core. Four different chips are available, each one for different application. DNA chip contains an interconnected set of microchannels that is used for separation of nucleic acid fragments based on their size as they are driven through it electrophoretically. The chip is loaded with a mixture of gel and dye using the syringe of the priming station. DNA ladder and samples are added to the wells after the chip has been filled with the gel-dye mix. For our purposes we used:

- Agilent High Sensitivity DNA Kit (sizing range 50-7000 bp and input DNA amount between 5-500 pg in 1 μL)
- Agilent DNA 1000 Kit (sizing range 25-1000 bp and input DNA amount between 0.1–50 ng in 1 μL).

The output of the DNA analysis is a plot of fluorescence as a function of time or base pairs (bp). Agilent DNA 1000 Kit in the range between 25-100 bp has 5 bp resolution and ± 10% accuracy. Agilent High Sensitivity DNA Kit in the range between 50-600 bp has 10% resolution and ±10% CV (coefficient of variation) accuracy.

3.9.3. LabChip GX (Caliper Life Science)

The LabChip GX assays exploits gel electrophoresis to size and quantify DNA, RNA and proteins and allows reducing time requirements and sample handling by transferring the process on a chip. In the microfluidic system, there is the separation channel filled with gel that is connected to buffer reservoir and to the microplate in which samples are loaded. Electrodes are present in sample wells. Molecules are bound to a fluorescent dye added to samples and the software plots fluorescence intensity versus time for each sample. To calibrate sizing and quantitation, a molecular weight ladder and internal standards at known concentration are added to the sample. The instrument needs accurate calibration because sample composition influences the amount of solution sipped into the chip.

3.10. NGS sample preparation

The kit (Ovation® Ultralow Library Systems Nugen) used to prepare NGS libraries from L-PCR product required 1-100 ng of DNA.

3.10.1. NGS standard sample

Extracted DNA was quantified with Qubit, a second measurement has been performed after 2 hours, and the mean value was used to prepare the pool. Before library preparation we doubled sample concentration in centrifugal evaporator at 30°C (IGA Technologies Services). From the pool, we obtained two libraries using 10 ng of sample.

3.10.2. Linear PCR efficiency sample

To perform this analysis three pools were prepared:

- All carriers were pooled to reach a final concentration of 2.5 μM each
- Staples in cleaved configuration were pooled to reach a final concentration of 200 nM each
- Staples without restriction site were pooled to reach a final concentration of 300 nM each.

In L-PCR mix (Vf = 500 μL) 18 cleaved staples carrying the restriction site, 3 negative controls, 2 positive controls and structural staples had 2 nM concentration, while carriers were used in 10-fold excess. Folding protocol a of figure 3.1 was used to prepare External CTRL (final concentration 100 ng/μL in 50 μL), which was then added to L-PCR mix at the same concentration of staples. Thermal protocol A was the one used for L-PCR on 20 fractions and gel extraction has been performed thereafter (we loaded in each well only one L-PCR reaction (25 μL) as suggested by an intermediate sample check gel analysis). 18 gel bands were extracted, samples were purified with EtOH precipitation and resuspended in 20 μL of water. We used 40 ng as DNA input for library preparation.

3.10.3. Triangular DNA nanostructure NGS sample 1

60 L-PCR were performed on digested triangular DNA nanostructure and purified by gel extraction with an expected recovered DNA of about 150 ng, that would allow to repeat library preparation. From gel extraction about 530 μL of purified product were obtained. This volume was aliquoted and concentrated with the EtOH DNA precipitation protocol

(section 3.6). To be sure that ethanol was completely removed from the purified sample prior NGS sample preparation, I repeated the Ethanol-Na Acetate procedure omitting the final washing step with ethanol at 70%. I used 10 μL (about 50 ng of DNA) as input for library preparation.

3.10.4. Triangular DNA nanostructure NGS sample 2

A second sample was prepared with the same process of sample 1, but a lyophilization step was introduced in the protocol after enzymatic reaction. This approach aimed at concentrating the sample before L-PCR to reduce the number of reaction and therefore the number of gel bands to extract. 11 enzymatic reactions were pooled together and fractionated in aliquots of 147 μL. Each aliquot was lyophilized and then resuspended in 49 μL of 20 mM Tris HCl. The product was used to prepare 19 L-PCR reactions (protocol A) with positive control carrier, negative controls carrier and short staple A29 carrier. The product was purified with gel electrophoresis, extracted with Quiaex kit, and finally concentrated with ethanol DNA precipitation. Two libraries were prepared with 3.1 μL of sample each, and one of them was prepared starting from the ligation step (the end repair step was omitted).

3.10.5. Library preparation: Ovation® Ultralow Library Systems Nugen

NGS libraries were prepared by the IGA Technology Services staff. The protocol followed was Ovation Ultralow DR Multiplex Systems 1–8 and 9–16 which has been developed according to Illumina protocol. It supplies a set of adaptors for multiplex sequencing. For detailed description see reference 102. We started the process from the End Repair step in which the enzyme adds missing nucleotides to the sticky end in 5'. This step is fundamental for the ligation of the adaptors because the enzyme used needs blunt ends to work properly. This enzyme feature also leads to an increase of reaction specificity as it brings the non-duplex DNA fraction in the library to a negligible level (carrier-integer staple hybrids that might be present in the sample are in turn suppressed). The ligation product was purified using Agencourt RNAClean XP Beads as specified by Nugen kit and then amplified with primers complementary to the adaptor used. 15 cycles of PCR were performed (the maximum number of cycles suggested by the protocol is 18). After another step of purification with Agencourt beads, DNA purity was checked with BioAnalyzer.

3.11. Next Generation Sequencing (NGS)

3.11.1. NGS analysis

Sequencing and data analysis were performed by Iga Technology Services in Udine on HiSeq or MiSeq Illumina instrumentation. Each sample was sequenced with reads length of 100 bp.

3.11.2. NGS data analysis

The algorithm used to analyze our samples follows customized steps, which have been defined to include in the analysis the highest percentage of reads obtained with sequencing.

1. **Removal of 26 nucleotides from the end of the sequence (trimming)**
   Reducing the length of the sequence reduces mismatches frequency in fact the ends of the sequence are more prone to errors because of polymerase low accuracy, chemical and physical stress and adaptor ligation reaction. This value (26 nt) has been chosen as the minimum value that allows more than 90% alignment.
2. **Removal of short and long sequences**
   Sequences can be minimum 45 nucleotides long (one nucleotide more than trimmed sequence) and maximum 70 nucleotides long (the expected value for sequenced dsDNA fragments).
3. **Search for exact carrier sequence**
   Smith-Waterman alignment algorithm has been used with 3 mismatch allowed. Sequences aligned univocally are counted as "*perfect match*", sequences aligned to more than one reference sequence are counted in the "*multiple carrier*" list.
4. **Search for both tags in the sequence**
   Each sequence has a specific feature at both ends: the index (5' end) and the staple sequence (3' end). These are the tags that the algorithm looks for at this step and are added to "*tags*" list. The use of tags allows including in the count the sequences that had poor polymerase activity, but were produced from correct hybridization of cleaved staple with the carrier. Sequences shorter than 70 nucleotides cannot contain both tags and are not counted.
5. **Search for one tag in the sequence**
   The sequence with only one tag is excluded from the analysis and added to "*only 3'* " or "*only 5'* " list.

6. **Search for cross tags**: sequences that contains two tags, but from different reference sequences are counted in the "*cross tags*" list. This is the case in which a staple hybridized to the wrong carrier, still forms a hybrid allowing the polymerization.
7. **Search for sequences with more than 5 unidentified (N) nucleotides**: these sequences are excluded from the analysis and counted in "*N threshold*" list.
8. **Search for local alignment**: a local alignment (allows the alignment in presence of gaps) is performed allowing 3 mismatches maximum. The sequences counted are added to "*mismatches*" list.

The steps are ordered in terms of time of consumption, from the fastest to the slowest.

The software samples the reads dataset, and the analysis on 300k, 1M, 5M, 10M, 20M reads was performed. Counts were found equivalent over 5 sampling replicates, while the value of 1M reads has been chosen as standard sampling value, corresponding to about one hour of work for in silico analysis.

Chapter 4 – Materials and Methods Project 2

4.1. Target sequence selection

Alteration of microRNA expression have been highlighted in several pathologies and there's effort for validation as biomarkers. One microRNA can be found altered in more than one disease therefore the analysis of only one miRNA is not specific. One of the most represented miRNA among several pathologies in miRNA-21, which have been chosen as target for this PhD project (67). Nevertheless, combinations of these molecules have been found to correlate with disease diagnosis and prognosis in specific way (68).

The diagnosis of non-small cell lung carcinoma (NSCLC) at an early stage, as well as better prediction of outcome remains clinically challenging due to the lack of specific and robust non-invasive markers. The scientific community is evaluating the possibility to use microRNAs, particularly those found in the bloodstream, for tumor diagnosis and prognosis. A panel of miRNAs should be able to discriminate between NSCLC patients and controls, and also might predict the prognosis of resectable NSCLC patients (69). Among tens of micro RNAs reported to be related to this pathology, miR223 and miR25 were chosen as target in this thesis (68-69-70). In the experiments performed for the thesis, DNA target molecules were used, for costs and biochemical stability reasons (DNA is much more cost effective than RNA and less susceptible to degradation). DNA target molecules have the same sequence of miRNA target, with the disadvantage that DNA-DNA binding is weaker than RNA-DNA binding. Nanosensor sensitivity will be likely higher for RNA target molecules.

4.2. Restriction enzymes selection

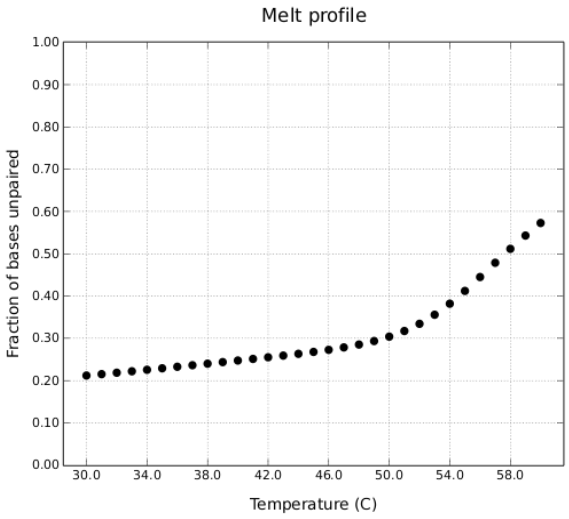
Two endonucleases with palindromic specificity were selected. These enzymes can be monomeric or dimeric. Dimeric enzymes typically have restriction sites about 6-8 pb long

that they can cleave in one step with the two subunits operating independently on one DNA strand. Monomeric enzymes usually have shorter restriction site (4 bp) that is cleaved in two steps with little accumulation of nicked intermediate molecules cleaved only on one strand. Moreover, each restriction enzyme has maximum efficiency in particular conditions of temperature and salt. (73) We chose SwaI (NewEngland BioLabs) that has 8 bp long restriction site (5'- ATTTAAAT -3'), and maximum activity at 25°C. The buffer recommended is Neb Buffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA, pH 7.9 at 25°C). The second enzyme we chose is DraI (Thermo Scientific) which has shorter restriction site (5'- TTTAAA -3') and reaches highest efficiency at 37°C. The buffer recommended is Tango Buffer Thermo (33 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL BSA.)

4.3. Nanosensor design

Starting from target sequence and enzymatic restriction site sequence, scaffold and probes sequences were elaborated using Nupack and IDT Oligoanalyzer (the latter already described in section 3.2). Nupack can be used in two ways: to design DNA or RNA sequences with specific characteristics, and to analyze secondary structures formed by one or more sequences. To design a new sequence, I used the Dot-Parents-Plus notation: each unpaired base is depicted with a dot, each base pair with matching parentheses, and each nick between strands with a plus. NUPACK calculates free energies and equilibrium concentrations of ordered complexes as described in the work by Dirks (71). Corrections for salt concentration are based on SantaLucia (72) parameters: I set NaCl and Mg^{2+} concentration to 100 mM and 10 mM respectively to fit the restriction enzyme buffer composition. I also set the temperature at 25°C which is the temperature of choice for enzyme SwaI. The output is expressed in terms of equilibrium probability, which depends on temperature and the Gibbs free energy of formation (71). The sequence produced by Nupack usually needs to be modified because it is rich in nucleotides repetition that can negatively affect oligo synthesis. In the analysis mode, one can choose the temperature, the nucleic acid type in input and salts concentrations, while the software performs a computation on the possible secondary structures in output. By inputting more than one sequence, the software predicts the formation of complexes (the number of molecules forming the complex is set by the operator). In the simulations performed the concentration

of probe, scaffold and target were 100 nM, 100 nM and 10 nM respectively. A full calculation of a system comprised of all the three probes in combination with a (long) scaffold was omitted to overcome issues associated with software performance (arising in simulating complexes formed by more than 3 strands). To estimate the melting temperature of a complex, Nupack displays the percentage of unpaired base pairs at equilibrium in a range of temperatures (fig. 4.1). Since the sensor is comprised of two feet that hybridize to the scaffold and a single stranded, target-complementary sequence (depicted with a loop in fig. 2.4), respectively, in target absence the loop sequence remains in the single-stranded form. Such a ssDNA structural contribution, leads to a minimum value of unpaired nucleotides in each probe of approx. 38-41% at 54-57°C (see the plot in fig. 4.1 below, relative to probe A, and the table in the same figure for all the three probes). For temperature values lower than 54-57°C, the residual fraction of unpaired nucleotides slightly changes (see the plot in figure 4.1 for probe A) due to the stabilization of mild secondary structure motifs in the target-complementary sequence in each probe.



Probe	Length	Min unpaired pb %
A	68	38
B	66	39
C	64	41

Figure 4.1. Melting profile from 30 to 60 °C, 1°C/step for probe A-scaffold A at 100 nM. The total nt length and the fraction of unpaired nucleotides at approx. 54-57°C are provided for each probe. The melting temperature (not shown) is calculated against the fraction of nucleotides involved in base pairing with the scaffold.

4.4. Nanosensor folding

For the first experiment, I set up an annealing protocol beginning with a denaturation step at 95°C, 3 minutes long, followed by a thermal ramp from 75°C to 25°C with 1°C/minutes rate. The whole protocol takes about 1 hour to be completed. The buffer used to anneal the nanosensor is the same used for enzymatic degradation with SwaI (Neb Buffer 3.1). If not specified the folding was performed at 15 nM of scaffold and 150 nM of probe. Any condition variation is indicated Appendix II.

Nanosensor folding protocol		
REAL TIME AMPLIFICATION		
1	95°C	03:00
2	75->25 °C	1°/min

Table 4.1. Nanosensor folding protocol.

4.5. Agarose gel electrophoresis (AGE)

Agarose gel was prepared and run using Bio-Rad Mini-Sub® Cell GT System. Agarose gel is prepared by heating a solution of TAE 10 X (Tris-acetate 400 mM, EDTA 10 mM) containing the specific percentage (w/v) of agarose powder (1.5 %). To load samples (25 μL) the gel is immersed running buffer (TAE) and the addition of glycerol containing loading buffer (5 μL) facilitates the deposition of samples on the bottom of the wells. Gel was run at constant voltage. The 100 base pairs molecular weight marker GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific) was used (1.5-2 μL).

4.6. Nanosensor purification from probe excess

Amicon® Ultra 0.5mL Filters have a vertical membrane with specific pore dimensions to *i)* concentrate biological samples containing antigens, antibodies, enzymes, nucleic acids, or microorganisms, purify macromolecular components found in tissue culture extracts or cell lysates, *ii)* clean PCR products from primers or proteins, *iii)* to perform desalting, buffer exchange and protein dialysis. For most of the DNA-related applications, it is preferred to use the 30K device that can be serve for PCR products separation from

primers up to 48 nucleotides in length. In our case, we need to separate the scaffold (about 200 nt) from the excess of probes (about 70 nt). This small difference in molecular weight cannot be properly handled using the 30K (probes would be retained by the filter as well as the scaffold), and therefore we utilized the 100 K device, which is typically applied to protein concentration with molecular weights above 150000 Da.

I purified 80 μL of sample diluted in 420 μL of water to a final volume of 500 μL (maximum volume of the filter). The device is placed in the microcentrifuge tube supplied by the kit, and the solution is spun at 14000 g for 5 minutes. The waste solution containing probe in excess flows through the filter reaching the bottom of the tube, and is then discharged. A washing step is performed by adding 480 μL of water on the filter to re-suspend the sample. The solution is mixed by slow pipetting and spun again at 14000 g for 5 minutes. To recover the nanosensor that is retained by the filter, the latter is turned upside down in a clean tube and it is spun at 1000 g for 2 minutes. To reconstitute a solution with the starting concentration, the recovered sample is brought to a final volume of 20 μL with MQ water.

4.7. Enzymatic reaction

The reaction is carried out at enzyme-specific temperature (25°C for SwaI and 37°C for DraI), for 2 hours or overnight. The incubations were performed in Binder Oven FD 115, Binder GmbH or in water thermal bath. The reaction volume used is 25-50 μL of 1 fold enzyme buffer to which we added 1-2 μL of enzyme. If not specified differently the nanosensor is diluted to 300 pM. Any condition variation is indicated in Appendix II.

4.7.1. Thermal equilibration and enzyme inactivation

This step has been used for three main reasons throughout the experimental setup, i.e. for ensuring enzyme thermal equilibration, nanosensor thermal equilibration, and as incubation step of the target for detection. I found that to obtain reproducible enzymatic reactions, a thermal equilibration of the enzyme in the reaction buffer before being added to the sample was needed. Therefore, the reaction buffer Tango (Thermo Scientific) was diluted to 1 fold in MilliQ water and incubated for 1 hour at 37°C. The enzyme was then added to the warm buffer and incubated at 37°C for 15 minutes more. In this way, the accuracy increases also because the enzyme is added to the sample using high volumes of

diluted solution. The nanosensor is diluted in 1-fold Tango Buffer to 600 pM concentration and incubated at 37°C for 1 hour before performing the enzymatic reaction. For performing detection experiments, the target is added to the solution containing the nanosensor in 1-fold Tango buffer, and the mix is incubated at 37°C for 1 hour before enzymatic reaction. In this way, target incubation is performed in the same conditions of the enzymatic reaction, without altering restriction site stability. DraI and SwaI enzymes are thermally inactivated with an incubation at 65 °C for 20 minutes.

4.8. Target incubation

The target is denatured for 5 minutes at 95 °C and is added to nanosensor solution in 20-fold excess (the nanosensor concentration is 600 pM). The incubation is carried out for 90 minutes at 37 °C, in the buffer specific for the enzyme reaction.

4.9. Quantitative PCR (qPCR)

SsoAdvanced™ Universal SYBR® Green Supermix (BioRad) is the mix containing the polymerase, the nucleotides and salts (Mg highly affects polymerase activity). For each qPCR assay, I used the specific primer couple at final concentration of 250 nM. The reaction is carried out in 22 μL final volume containing 2,5 μL of the sample (a different setup was used in some of the experiments. Details are reported in appendix II.

4.9.1. qPCR protocol

qPCR protocol was designed following BioRad directions: at 98°C the enzyme is activated, at 95°C the DNA is denatured and at 60°C annealing and extension are performed. For bingo qPCR, with respect to "Site Specific qPCR assay", I increased elongation and annealing time from 15 seconds to 1 minute. The cycle of amplification is repeated 36 or 41 times and, at the completion of each step the instrument measures the emitted fluorescence intensity. The Bingo-qPCR protocol has more repeats because the template is less concentrated. The melting curve is performed from 65°C to 95°C. The temperature is held constant for 5 seconds, and then incremented of 0.5°C (table 4.2). At every step the instrument measures the emitted fluorescence intensity.

SITE SPECIFIC		
	REAL TIME AMPLIFICATION	
1	98°C	01:00
2	95°C	00:15
3	60°C	00:15
	PLATE READ	
4	GO TO 2	35 more times
5	95°C	00:05
	MELT CURVE	
6	65°C	00:05
	PLATE READ	
	increment 0.5°C TO 95°C	
	END	

BINGO		
	REAL TIME AMPLIFICATION	
1	98°C	01:00
2	95°C	00:15
3	60°C	01:00
	PLATE READ	
4	GO TO 2	40 more times
5	95°C	00:05
	MELT CURVE	
6	65°C	00:05
	PLATE READ	
	increment 0.5°C TO 95°C	
	END	

Table 4.2. qPCR protocols

I always performed two measures on the same sample. To compare different samples, I used ΔCt values and not Ct values, this avoiding the need to set threshold at the same value.

4.9.2. Primer concentration optimization

The standard sample was obtained by folding the scaffold at 15 nM with probes in 10-fold excess, using the thermal ramp from 75°C to 25°C with 1°C/minutes rate. After folding, the sample was incubated at different temperatures (65°C for 20 minutes, 50 °C for 1 hour, 37 °C overnight, 4 °C for 2 days), and then diluted to 3 pM (final concentration in qPCR reaction). Possible alterations to the nanosensor-conformation induced by these incubation steps are not relevant to the qPCR process, while DNA degradation remains a concern. In this way, it ensured that the control sample has the same (or slightly worse) quality of the experimental samples.

Chapter 5 – Quantification of enzymatic restriction reaction products of DNA-nanostructures

5.1. Experimental design

5.1.1. Carrier length

Carriers have to be *i)* longer than the staples to allow linear polymerization, *ii)* long enough to be visualized with acceptable resolution on the gel used for purification and *iii)* well separated both from staples and scaffold bands. They also should not exceed in length to minimize the costs of the analysis. The length of choice was determined empirically analyzing with gel electrophoresis the products of triangular DNA origami treated with two thermal protocols for Linear PCR (L-PCR, protocols A and B, see section 5.4.). In both cases, the results of gel analysis indicated lack of reaction products in the range between 50 and 100 bp. In turn, carrier length was set to 70 nucleotides (fig. 5.1).

5.1.2. Carrier design

The design started by selecting all staples with a restriction site for MspI and obtaining their cleaved sequence. The palindromic restriction site sequence is 5'-C*CGG-3', with * indicating the cleaved phosphodiester bond between the two cytosine nucleobases. In our protocol the cleaved staple serves as a PCR primer for amplifying the carrier, and optimal elongation temperature should be 72°C. I chose to use as primer the longer fragment produced by cleavage, and obtained a melting temperature (Tm) ranging between a minimum of 42°C and a maximum of 68°C, with an average value of 59°C. L-PCR thermal ramps were set to obtain stable staple hybridization on the carrier (see section 5.4.). Having defined the staple sequences, I moved on to design the corresponding carriers that includes four different adjacent regions (from 3' to 5'):

1) a common 10-nt-long region for carrier barcoding during high-throughput sequencing experiments.
2) a carrier/staple-specific region of 6 nt (termed "index") to allow staple identification by limiting the analysis of sequencing results by reading only the first 50 nt starting from a fragment terminus;
3) a region of variable length allowing all carriers to reach the optimal length of 70 nt;
4) a cleaved-staple-complementary region with staple-dependent length ranging between 13 and 35 nucleotides (nt);

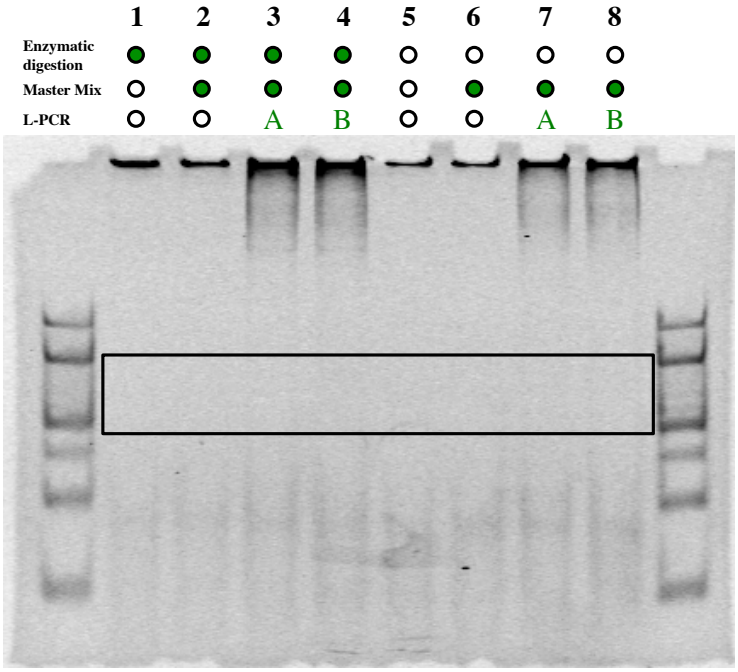


Figure 5.1. Digested (1-4) and not-digested (5-8) triangular DNA nanostructure samples were analyzed before (1-2-5-6) and after (3-4-7-8) L-PCR. The reaction was performed without carriers to determine the molecular weight of unspecific products. The gel shows that the area marked by the black square (50-100 bp) does not contain unspecific products therefore the length of the carrier was set to 70 bp.

I designed the secondary structures to have melting temperatures (Tm) lower than 50°C. Among all the secondary structures tested hairpins have the highest Tm, but it is still lower than 50°. Carrier for staple A11 (C_A11) forms hairpin with the contribution of two separated hybridization steps, one at 3' end and one in the middle of the sequence (fig. 5.2). The Tm of the whole structure is 50° C but the Tm associated with single hairpins is lower. So, the DNA staple should be able to displace both of them by hybridizing to the carrier. I checked for the presence of unwanted carrier-scaffold dimers, and found that only one carrier (C_A6) forms a mild secondary structure with melting temperature higher than 50°C (53.2

°C). The hybridization occurs within staple complementary region that cannot be easily modified by design, but the expected ΔG associated with the staple-carrier hybridization (-134.38 kcal/mol) is 7-fold higher than that associated with unwanted carrier-scaffold hybridization (-18.84 kcal/mol). All carrier sequences and melting temperature values of secondary structures are reported in appendix I.

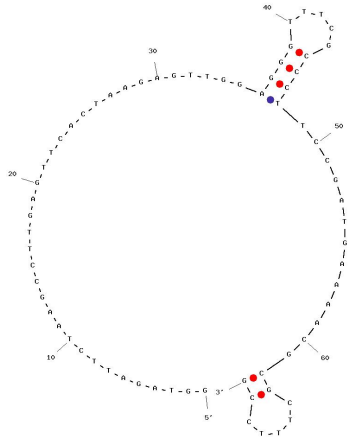


Figure 5.2. Hairpin formed by C_A11 calculated by IDT Oligoanalyzer.

5.2. dsDNA external control design

To compare NGS results from different samples, I introduced an external control consisting in a 70 bp dsDNA sequence formed by the hybridization of 70-nt-long complementary ssDNA molecules. The design of the sequences was performed starting from the vector pUC19, which is typically used for gene cloning to avoid strong secondary structures or difficulties in synthesis. To exclude the formation of unwanted secondary structures or cross-talks between the DNA molecules involved in the experiment, the pUC19-derived sequence was consistently changed using standard software for nucleic acids analysis. Figure 5.3 reports the starting sequence of the a 70-bp-long pUC19 fragment and the optimized reference sequence.

Fragment from pUC19

```
5' TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3' AAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTG
```

External CTRL sequence

```
5' TTTAGAAAAATACAATCCTGCCGTACCTTGAAGCATTTGCGACGCTATGGGCACCTGACGAACTATAGAC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3' AAATCTTTTTATGTTAGGACGGCATGGAACTTCGTAAACGCTGCGATACCCGTGGACTGCTTGATATCTG
```

Figure 5.3. Sequence of the template (pUC19) used to design the external CTRL and the resulting sequence with the modification highlighted in red.

5.3. Positive control design

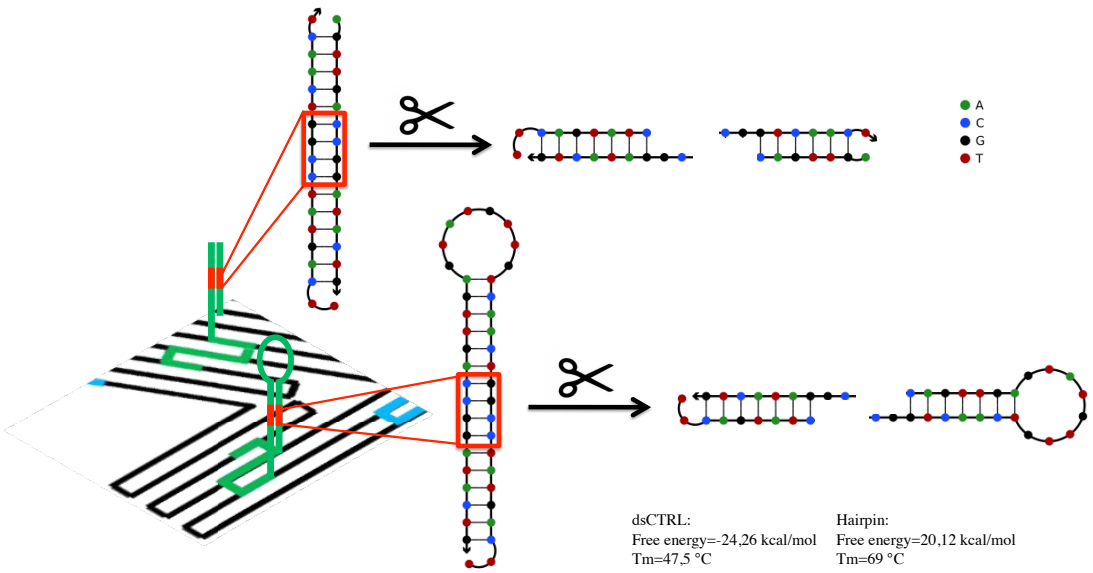


Figure 5.4. Schematic of positive control hairpin and dsDNA.

5.3.1. Hairpin design

To generate a reference DNA motif that can be digested with maximum efficiency during restriction reaction, thus serving as internal standard, I modified a staple (C10) that lies in the core of the structure (trapezoid C) and to elongated it sequence at one terminus. The added sequence is designed to form a stable hairpin that do not interfere with the DNA origami self-assembly process. The stem of the hairpin is projected on the surface of the

planar DNA nanostructure, and contains a MspI restriction site. To facilitate the enzymatic reaction, such a restriction site is flanked by 6 nucleotides at each side, while the size of the loop is minimized to 8nt (fig. 5.4).

5.3.2. Hairpin function evaluation

After equilibrating the folded structure of the modified C10 staple and performing the enzymatic digestion I analyzed the DNA products with gel electrophoresis. In figure 5.4, it can be seen that the band of the untreated hairpin (sample 2) is absent in the digested sample (3), which presents instead two bands corresponding to enzymatic products.

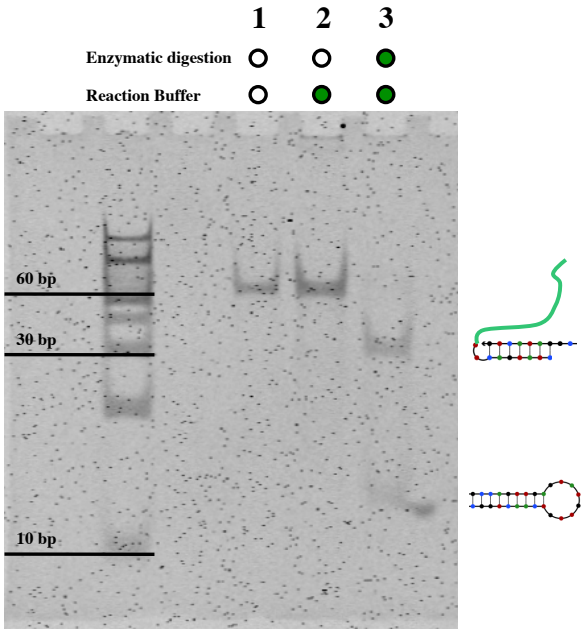


Figure. 5.5. Digested (3) and not-digested (1-2) modified C10 staple analyzed with gel electrophoresis. In sample 3 the control band matching a MW of 60 bp and corresponding to the uncleaved hairpin is absent, while two distinct bands identify distinct enzymatic digestion products.

5.3.3. dsDNA positive control design

As an alternative, positive control, I chose staple A10 from an inner helix contained in the trapezoid A of the DNA triangle. Following the same procedure described above, I elongated its sequence at one terminus adding the sequence of the sole hairpin stem. The strand protruding from the surface of the structure forms a duplex by hybridizing to a complementary ssDNA added in solution during the DNA origami self-assembly process.

5.3.4. dsDNA positive control functional evaluation

After equilibrating the folded structure of the modified A10 staple and performing the enzymatic digestion, I analyzed the enzymatic products with gel electrophoresis. In figure 5.6, it can be seen that the band of the unmodified conformer lacks in the digested sample (2), while the main band associated with digestion products approximately matches 27 bp, as observed for the truncated stem-loop structure associated with the digestion of the modified C10 staple (fig. 5.5). The second fragment with lower molecular weight is not visible in the gel.

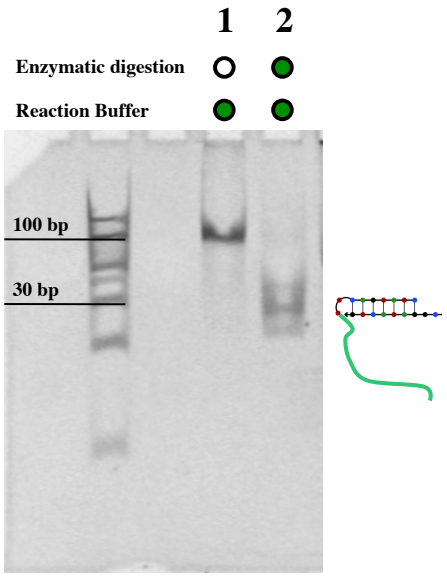


Figure 5.6. The duplex DNA control is analyzed in enzyme treated (1) and non-treated (2) versions. The band visible in sample 1 lacks in sample 2, indicating complete enzymatic digestion of the restriction site.

5.4. Linear PCR protocol evaluation

L-PCR begins with a denaturation step that prepares dsDNA inputs for primers annealing and polymerase activation. Annealing temperature is typically set to be near primer the melting temperature of the amplicon to ensure hybridization of the DNA components. Our setup, however, includes more than 20 staples hybridizing to their carriers, with melting temperature ranging from 42°C to 68°C. For this reason, I optimized a protocol "A" with an annealing thermal ramp from 72 to 50 °C, followed by extension reaction performed in three steps at 60, 65 and 72 °C, and a protocol "B" based on 20 cycles of thermal annealing followed by temperature reduction from 70 to 50 °C and subsequent enzymatic extension at a fixed temperature (70 °C) (tab. 5.1).

A		
Deaturation	95°C	3 minutes
Annealing	72°C	30 seconds
	-1°C/min	to 50°C
Extention	60°C	1 minute
	65°C	1 minute
	72°C	1 minute

B		
Deaturation	95°C	3 minutes
Annealing	70°C	30 seconds
Extention	70°C	15 seconds
GO TO Annealing	-1°C/Cycle	20 cycles

Table. 5.1. L-PCR protocols are schematically represented. In protocol A annealing is performed by decreasing the temperature 1°C/minute to allow hybridization of all staples with different melting temperatures. Extension is performed in a separate step by increasing the temperature to 60-65 and 72 °C. In protocol B, annealing and extension reactions are cyclically performed in the same step. At every cycle the annealing temperature is reduced by 1°C. This protocol might favor the extension of staples provided with higher melting temperature because they have more chances to form completely hybridize to their carriers.

5.4.1. Cleaved and un-cleaved staples

To evaluate L-PCR protocol efficiency and specificity I chose the shortest cleaved staple and we ran 2 different PCR protocols with the cleaved and un-cleaved form of the staple and the respective carrier. I performed L-PCR at high concentration to ease the visualization of DNA in gel. Figure 5.7 compares the L-PCR products band at 70 bp of enzyme reacted and non-reacted staples: the band intensity increases in samples containing the cleaved form of the staple, meaning that the presence of un-cleaved staple inhibits polymerase action (as expected). The results also indicate no significant deviations between protocols A and B in terms of efficiency.

5.4.2. Negative controls

The performance of the L-PCR protocol A on negative controls A, B and C is evaluated in figure 5.8. Here, the ability of the protocol to convert ssDNA staples into 70-bp-long dsDNAs is evidenced. In fact, bands relative to samples after L-PCR are darker and match the position of the 70 bp marker. Differences in bands height might reflect a minor sequence dependency in migration, while differences in band intensity might be due to sample handling. A putative lower efficiency of control A with respect to controls B and C needs to be confirmed with a different setup, suggesting the need of accurate L-PCR calibration for each staple-carrier construct. Finally, the gel was loaded with the equivalent of two L-PCR reactions (for each of the 3 carriers) to ensure that the gel analysis is sufficiently accurate to

permit detection of traces of DNA corresponding to the case that the examined staples within the nanostructure are minimally processed by the restriction enzyme.

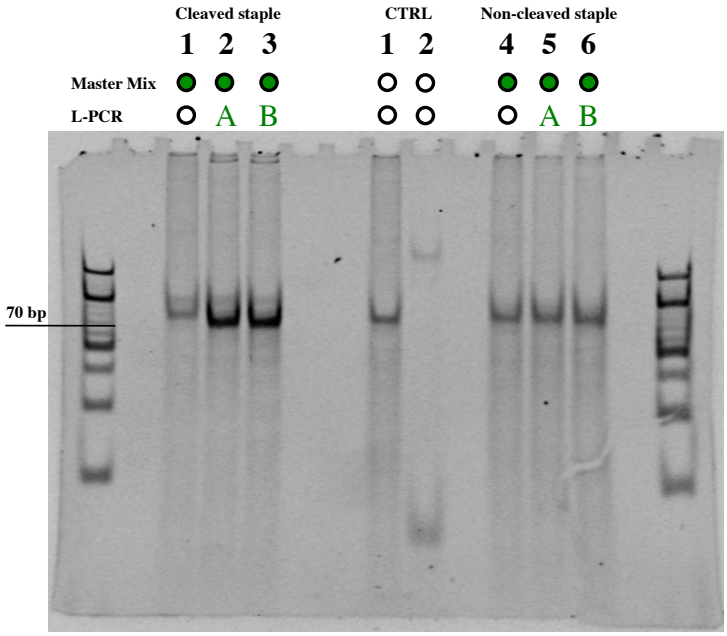


Figure. 5.7. Enzyme treated (1-3) and non-treated (4-6) staples were processed with the L-PCR protocols A and B. The L-PCR product (dark band at 70 bp) is present only in samples 2 and 3 meaning that the assay can discriminate between enzyme-cleaved an unmodified DNA staples. CTRL 1 is the sole carrier and CTRL 2 is the sole staple (the upper band might correspond to secondary structures containing multi-strands).

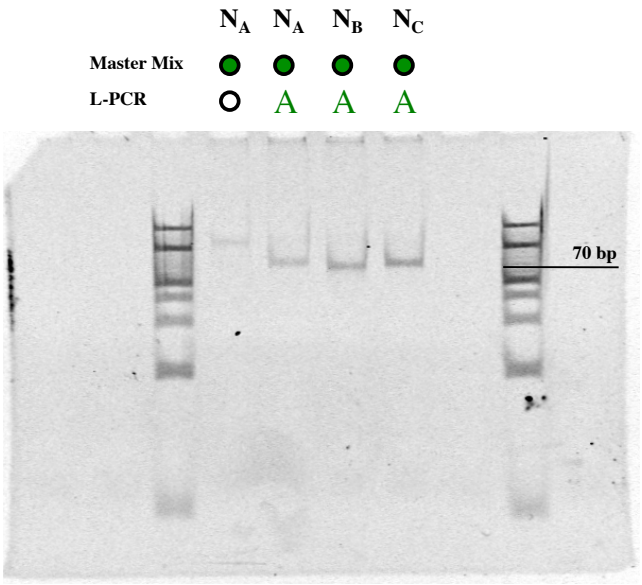


Figure 5.8. L-PCR products of negative controls A, B and C. The control sample contains staple A and carrier at the same concentration of the sample, but didn't undergo L-PCR reaction. With these conditions we can clearly see the 70 bp band for each sample.

The previous experiments were carried out in relatively simplified conditions, i.e. in which only the DNA involved in the L-PCR reaction is present. In the following experiment, L-PCR was performed on the fully formed triangular DNA origami, i.e. in a more complex DNA mixture containing carriers for the 3 negative control staples and for the positive control hairpin staple (see section 5.4.3), the scaffold along with all the other staples not directly involved in the PCR reaction. The gel below (fig. 5.9) shows that the band of sample containing only the carriers pool (X), is darker after L-PCR in sample 2-3-6-7, meaning that the reaction efficiently converts the carrier-staple hybrid in duplex DNA. Even in not-optimal experimental conditions the process appears to be extremely powerful: a small excess of carriers over the staples used in this experiment (2.5 fold) does not suppress the elongation reaction that can be detected even with poor gel resolution.

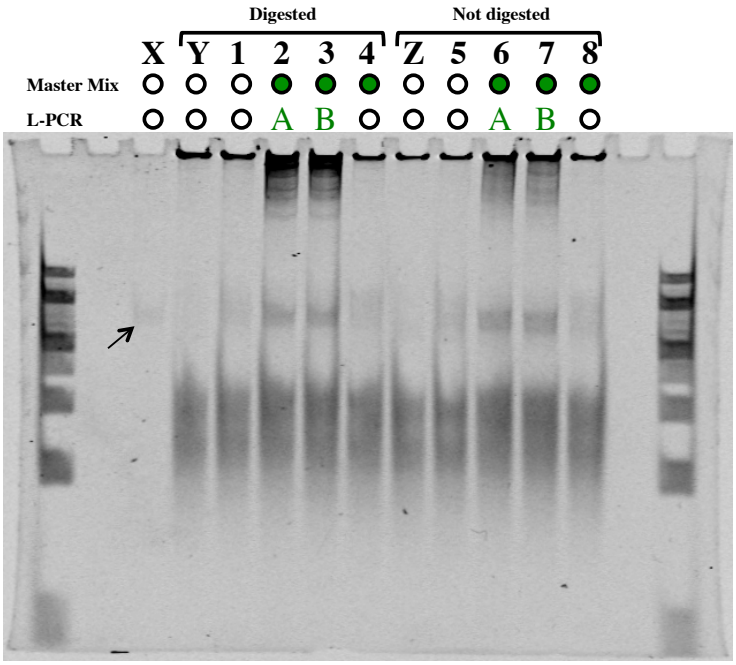


Figure 5.9. L-PCR test on the restriction enzyme products of the triangular DNA nanostructure. Sample X contains the carriers pool, sample Y and Z contain the digested and not-digested triangular DNA nanostructure (without carriers), respectively. Digested and not-digested samples were treated with L-PCR protocol for negative control staples A, B and C and positive control hairpin (samples 2-3-6-7), showing a darker band at the same height of carrier band (arrow in sample X). Sample 1-4-5-8 are controls containing the nanostructure.

5.4.3. Hairpin positive control

Digested and not digested hairpin samples were treated with L-PCR. The gel (fig. 5.10.) shows that digested sample band profile contains one product, corresponding to a complete

conversion of the cleaved hairpin in the 70 bp duplex DNA strand. In the not-digested samples, the polymerase acts on the 3' end of the un-cleaved hairpin, elongating the stem segment. A product that corresponds to the dark band at 60 bp in the gel below. To allow the accurate separation (with gel extraction) of the L-PCR product derived from fragmented stems from that (similar in molecular weight) derived from the unmodified stem-loop structure, I used higher acrylamide gel concentration (20 %).

I evaluated the efficiency of the restriction enzyme processing the hairpin inserted in the triangular DNA nanostructure by carrying out 10 L-PCR reactions and concentrating the samples and pooling the reaction products following EtOH-Na-Acetate-based DNA precipitation. Gel analysis results presented in fig. 5.11 show the presence of unspecific products in the digested sample (1), although the gel does separate well the different bands, thus allowing subsequent gel extraction.

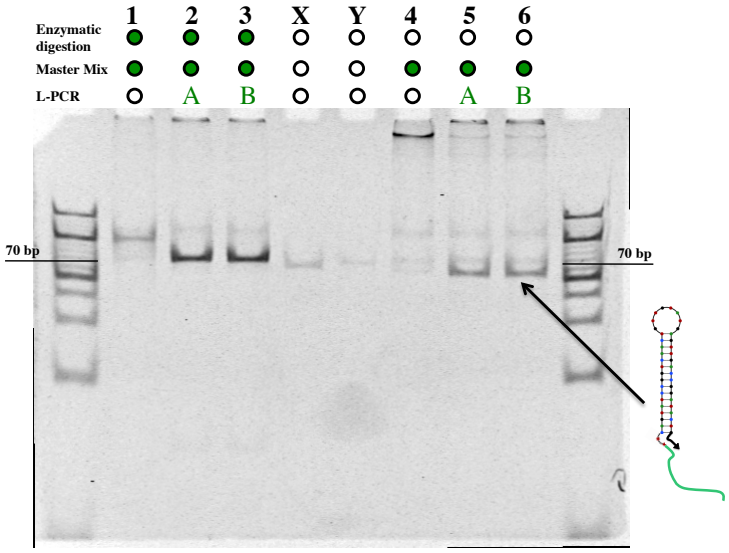


Figure 5.10. Digested (1-2-3) and not digested (4-5-6) hairpin samples were treated with L-PCR protocol A (2-5) and B (3-6). Sample X contains the hairpin in the enzymatic reaction solution, while sample Y contains the carrier. The difference in band intensity between sample 1 and samples 2-3 corresponds to the conversion of the ssDNA carrier in the dsDNA product of L-PCR. The same is found for sample 5-6 with respect to sample 4, despite the darker band correspond to the formation of unspecific product (as the band runs a little faster than in samples 2-3). Such an unspecific product is likely due to the polymerase processing at the 3' end of the intact hairpin, thus elongating the full ssDNA region of the staple.

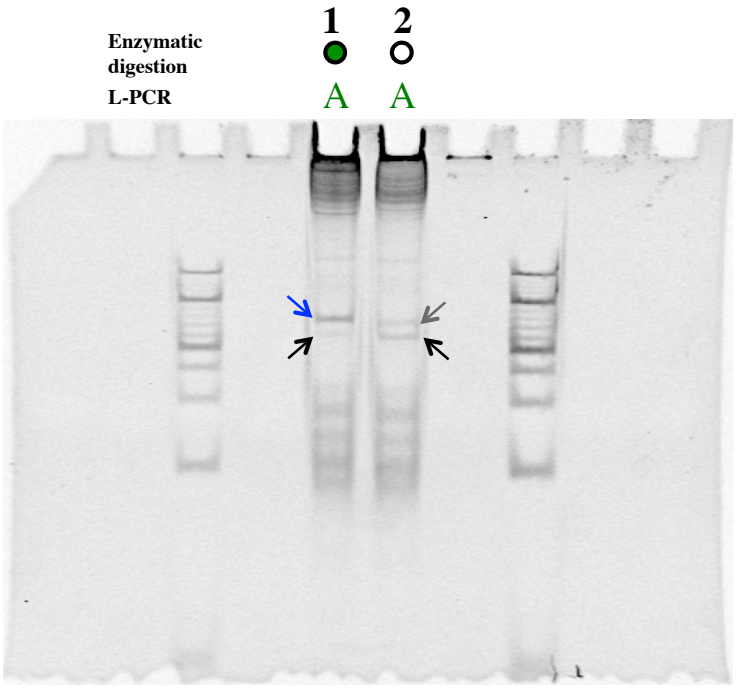


Figure 5.11. Hairpin L-PCR results within DNA nanostructure. The enzymatic processing leads to two bands corresponding to the L-PCR product (blue arrow), and the unspecific stem-loop product (black arrows), respectively. The latter indicates incomplete digestion of the hairpin immobilized within the DNA nanostructure. Also, the unmodified sample provides two distinct bands, one likely corresponding to the ssDNA carrier (grey arrow), and the other likely corresponding to the stem-loop hairpin.

5.4.4. dsDNA Positive Control

L-PCR was performed on a sample containing the DNA molecules corresponding to the cleaved and the unmodified DNA positive control staples. The gel in figure 5.12 shows that a L-PCR product is evident only if the cleaved-like DNA staple is present in solution (sample 3, dark band at 70 bp), while the unmodified-like staple forms high molecular weight aggregates (sample 4). In this experiment, we also formed the dsDNA positive control and digested it with the usual protocol. The inherent L-PCR product obtained (sample 2) is comparable with the one obtained using the synthetic cleaved-like sequence, Moreover, the presence of the enzymatic digestion product in the first two samples in evident in the gel. The results clearly demonstrate that the L-PCR effectively converts a cleaved staple into the 70-bp-long duplex output starting from both enzymatically cleaved or synthetic, cleaved-like staples.

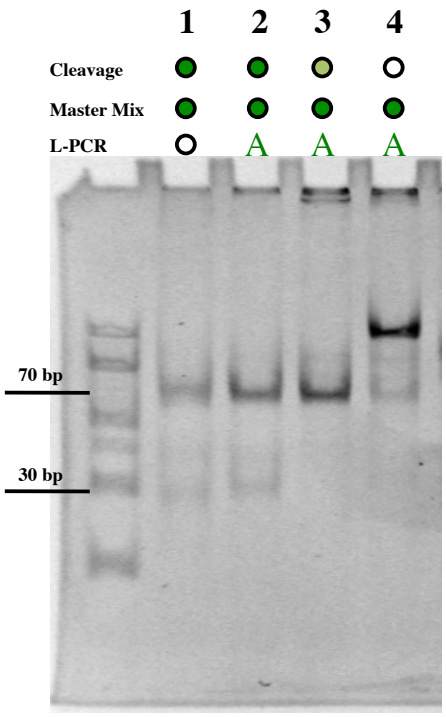


Figure 5.12. dsDNA positive controls employed to test the L-PCR protocol. In this experiment, I used an enzymatically-digested, positive control (2) and its corresponding cleaved-like sequence (3). The gel presents the analysis of the cleaved control without undergoing L-PCR (1), and the L-PCR product of the unmodified staple (4). Gel bands relative to sample 2 and 3 have comparable intensity and height while sample 4 presents high molecular weight bands. In sample 1 the band at 30 bp corresponds to the dsDNA product of enzymatic degradation (29 bp).

5.5. Gel extraction protocol

5.5.1. Nanostructure free sample

In this experiment we performed L-PCR protocol A on negative control staples and we loaded in the gel a DNA amount corresponding to 2 L-PCR reaction per well (5.13). To evaluate gel extraction efficiency not-purified L-PCR product was used as control: decreasing amount of DNA was loaded into the gel to simulate 100%, 50%, 25%, 12.5 % efficiency. Comparing sample band with 50% and 25% control bands with ImageJ an extraction yield of 28-31% is estimated.

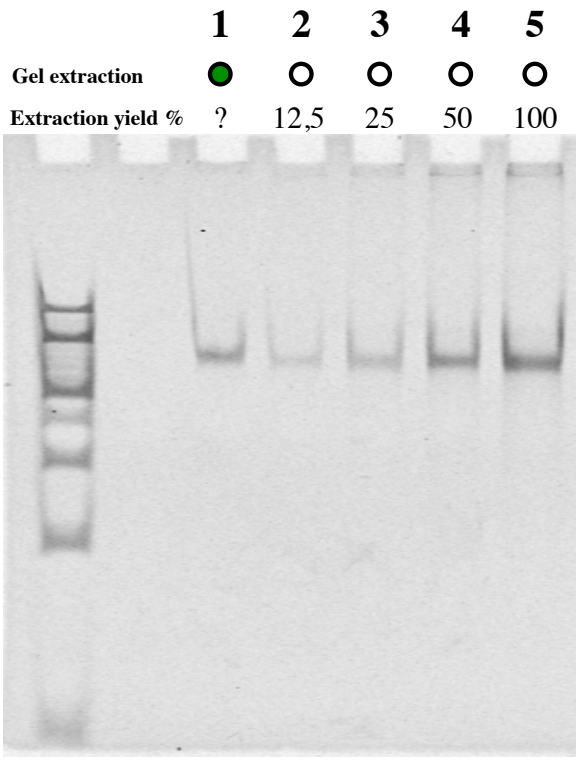


Figure 5.13. Evaluation of gel extraction efficiency of L-PCR product. The extracted DNA (1) is compared to progressively increasing DNA amounts (2 to 5) corresponding to different extraction efficiency percentages.

5.5.2. Nanostructure sample

Gel extraction procedure was tested with L-PCR products derived from DNA origami-containing samples, treated with the carriers for analyzing the negative controls and hairpin positive controls. The hairpin-containing staple was added to the annealing solution. The gel in figure 5.14 shows the sample profile before gel extraction. Extracted samples were analyzed by gel electrophoresis and compared to control samples that contain non-purified L-PCR products at progressive dilutions, as in the previous experiment. An extraction yield of 27-31% is estimated by comparing the sample band with control bands using ImageJ (fig. 5.15).

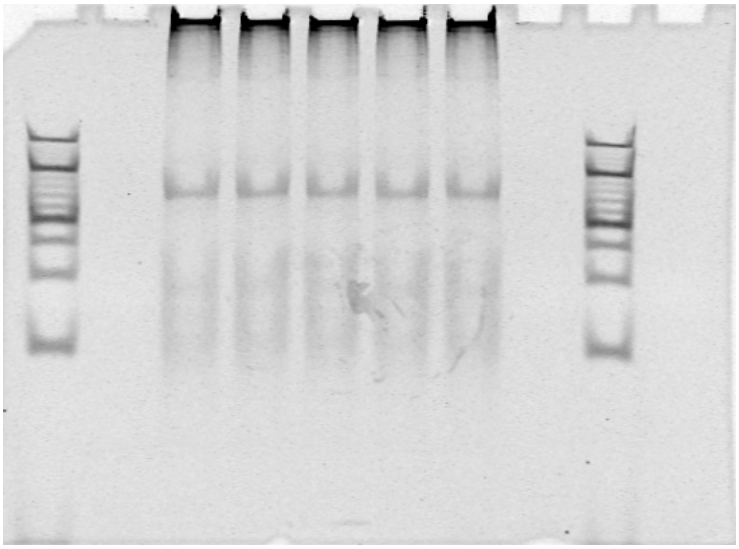


Figure 5.14. L-PCR product of triangular DNA nanostructure with negative controls and hairpin positive control before gel extraction.

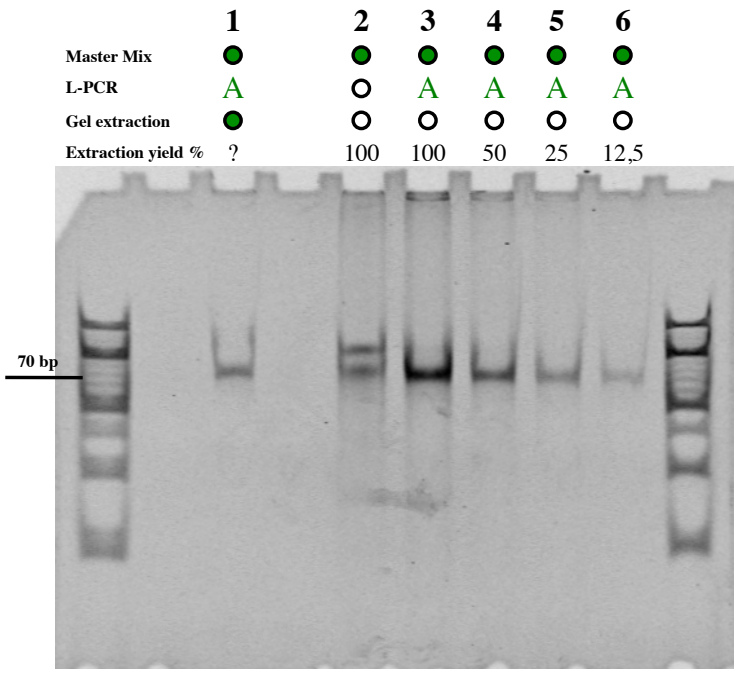


Figure 5.15. Evaluation of gel extraction efficiency of the L-PCR product derived form a triangular DNA nanostructure. The extracted DNA (1) is compared to progressively decreasing DNA amount (2 to 6) corresponding to different extraction efficiency percentages. All bands correspond to 70 bp ssDNAs and are analyzed with ImageJ to band intensity quantification.

5.6. Standard samples for NGS calibration

5.6.1. Duplex folding protocol setup

Gel analysis (fig. 5.16) of the duplex DNA, L-PCR products shows that all the 5 different molecules are correctly formed before and after EtOH purification (used to remove salts before NGS library preparation). Nevertheless, some high molecular weight aggregates are present in gel wells and there is also some unspecific background in the gel lanes. The DNA concentration was quantified with Qubit before sample pooling. We tried variations in DNA amount of 3%, 4% and 7% (if all staples are amplified in a sample containing the nanostructure, each sequence will be about 4%).

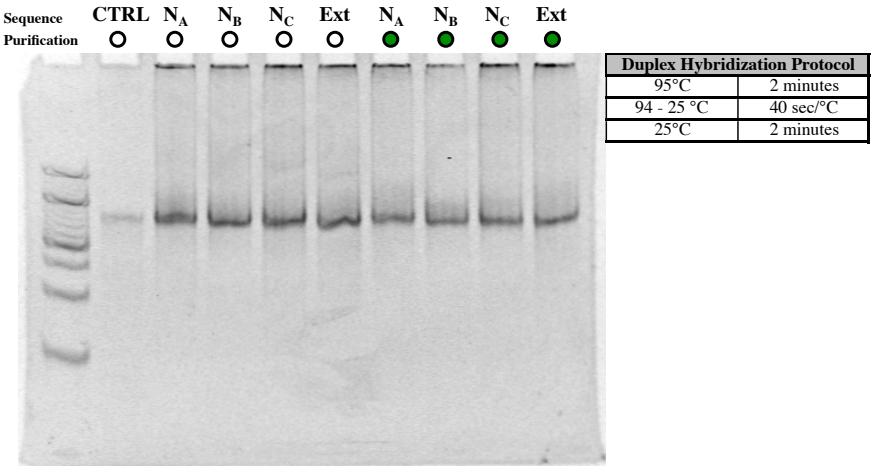


Figure 5.16. Gel analysis of the duplex hybridization product of negative control carriers and external control. The CTRL sample contains the ssDNA carrier for quantitative control A at the same concentration used in the sample.

Pool composition		
Oligo name	**DNA amout (ng)**	**%**
ds Ext CTRL	29.0	58
ds Qt A	5.3	11
ds Qt B	7.0	14
ds Qt C	8.8	18

Table 5.2. Composition of the sample prepared as standard for NGS calibration

Caliper profile of the sample had a peak around 80 bp as expected, while the library profile shows several peaks of spurious products in molecular weight spectrum between 250 and 400 bp. For this reason we decided to change the folding protocol of the sequences (fig.

5.17). As DNA concentration can affect the formation of DNA duplexes during the folding process, I prepared the sample at different DNA dilutions: 50, 25, 15 and 10 ng/μL. For each sample, an equal amount of DNA was loaded in the gel. The results highlight no significant difference in DNA quality as a function of DNA concentration or thermal treatment utilized for the folding process (fig. 5.18). Some unspecific product seems to be present in these samples as it happens with samples obtained using the previous protocol (fig. 5.16), and so I set the thermal treatment to reach a maximum value of 62°C with a DNA concentration of 15 ng/μL. At the end of duplex folding thermal protocol the 70 bp DNA fragments were purified by agarose gel extraction, and then pooled.

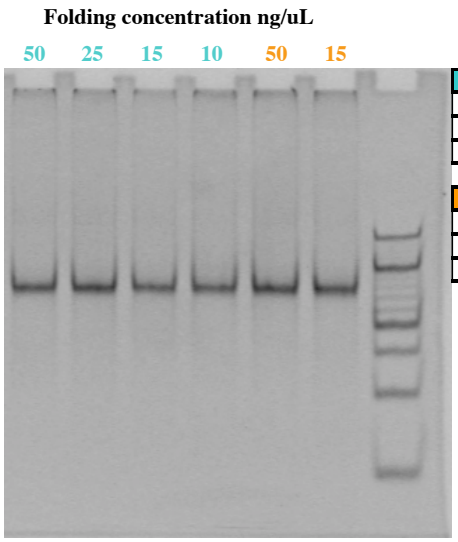


Figure 5.17. Two protocols are tested for folding DNA duplexes (orange and blue tables) at different sample concentrations. In the gel, the same amount of DNA is loaded in each well (50 ng). The intensities of the hybridization products indicate no significant difference in samples preparation.

5.6.2. NGS of Standard Sample

In pool composition External CTRL was used only to increase total DNA amount, for the other sequences we chose % values (tab. 5.3) similar to those used in previous experiment (tab. 5.2).

Pool composition		
Oligo name	**DNA amout (ng)**	**%**
ds Ext CTRL	26.6	56
ds Qt A	5.3	11
ds Qt B	8.8	18
ds Qt C	7.0	15

Table 5.3. Composition of the sample used as standard for NGS calibration

Capillary electrophoresis analysis (with Bioanalyzer) was utilized to confirm the purity of the sample. The curve in fig. 5.18-input presents a peak at 77 bp and evidences a moderate peak broadening that can be ascribed to sample heterogeneity. The sample, in fact, contains 70-bp-long fragments carrying different sequences potentially responsible for uneven migration in the capillary electrophoresis device. This result confirms the interpretation of the effect found in testing the gel migration of negative controls depicted in fig. 5.8. Here, the bands relative to sample A and C similarly migrate in the gel, while sample B migrate slightly faster. Capillary electrophoresis analysis results combined with Qubit measurements set the final concentration at 3.6 ng/μL.

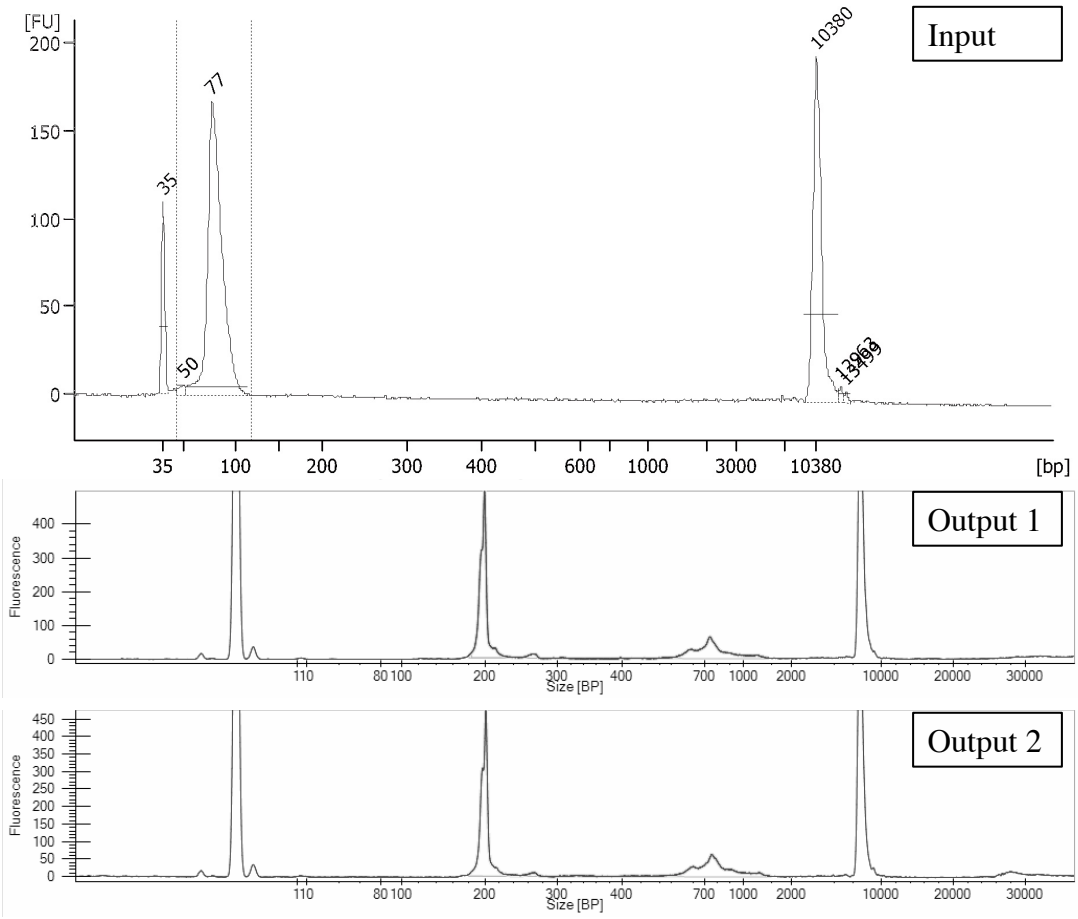
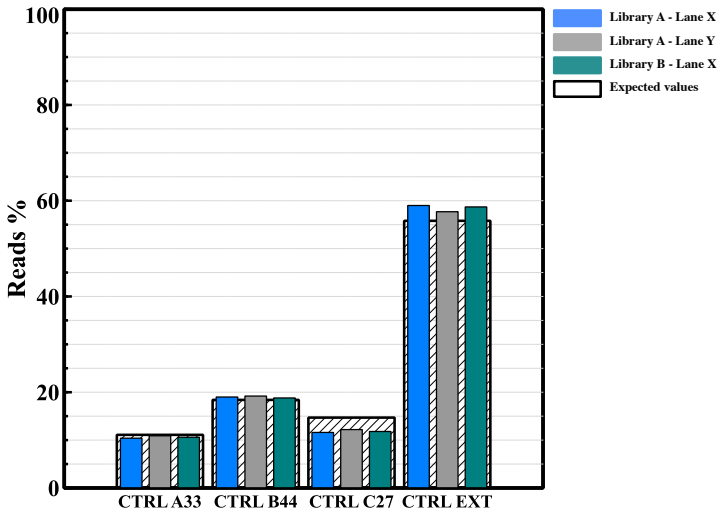


Figure 5.18. Molecular weight profiles of the sample used to calibrate NGS. ***Input.*** Capillary electrophoresis profile of the sample used as input for library preparation obtained with a Bionalyzer instrument. After purification by gel extraction the sample has a peak at 77 bp. ***Output 1-2.*** The input sample has been used to prepare 2 libraries. Caliper profile shows a very sharp peak at 200 bp. Not mentioned peaks correspond to molecular weight markers.

Following sample pooling, I obtained two libraries that shared a very sharp peak at 200 bp in the Caliper profiles relative to their capillary electrophoresis analysis (fig. 5.18-output). One library was split to be loaded in two different NGS experiments (performed using different microfluidic chips). Plot 5.1 reports expected and experimental results. Differences among results obtained from separate lanes and libraries range between 0.19-1.35 %, and differ from those expected by 0.17-3.26 %, respectively.

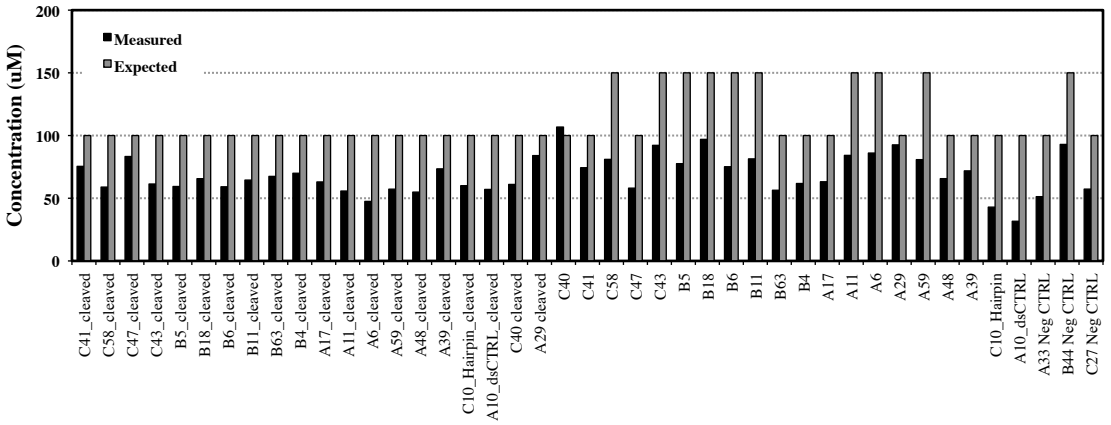


Plot 5.1. NGS results of the control samples. I obtained two libraries from the same DNA pool, one of which was loaded in two different NGS microchips for high-throughput sequencing. The quantification differences are between 0.19-1.35 %. The plot also reports the expected values differing from the experimental by 0.17-3.26 %.

5.7. Linear PCR efficiency

5.7.1. Oligos quantification

To evaluate L-PCR efficiency, I quantified the concentration of staples with the spectrophotometer Varian Cary 5000. The measured values are generally lower than expected (plot. 5.2).



Plot. 5.2. The histogram presents measured and expected concentrations of the ssDNA staples.

5.7.2. NGS results

To measure the L-PCR efficiency, I pooled cleaved-like staples at equimolar concentration, unmodifiable staples (i.e. lacking the restriction site) and their relative carriers (in excess). Qubit and Bionalyzer quantification estimated the mean DNA concentration value of 18.6 ng/μL in the purified sample, associated with a sharp peak at 76 bp in DNA length distribution. The Caliper profile of the library produced from this sample is shown in fig. 5.19, and shows a dominant peak at around 200 bp, which corresponds to the expected library product, while a second peak around 270 bp might be associated with the formation of a double carrier molecule during library preparation. Such an unwanted peak, however, is over 6-fold lower in intensity than the peak associated with the exact L-PCR product, thus allowing NGS sequencing analysis of the sample.

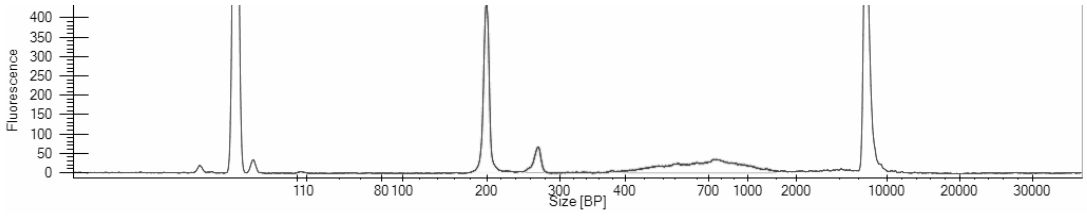
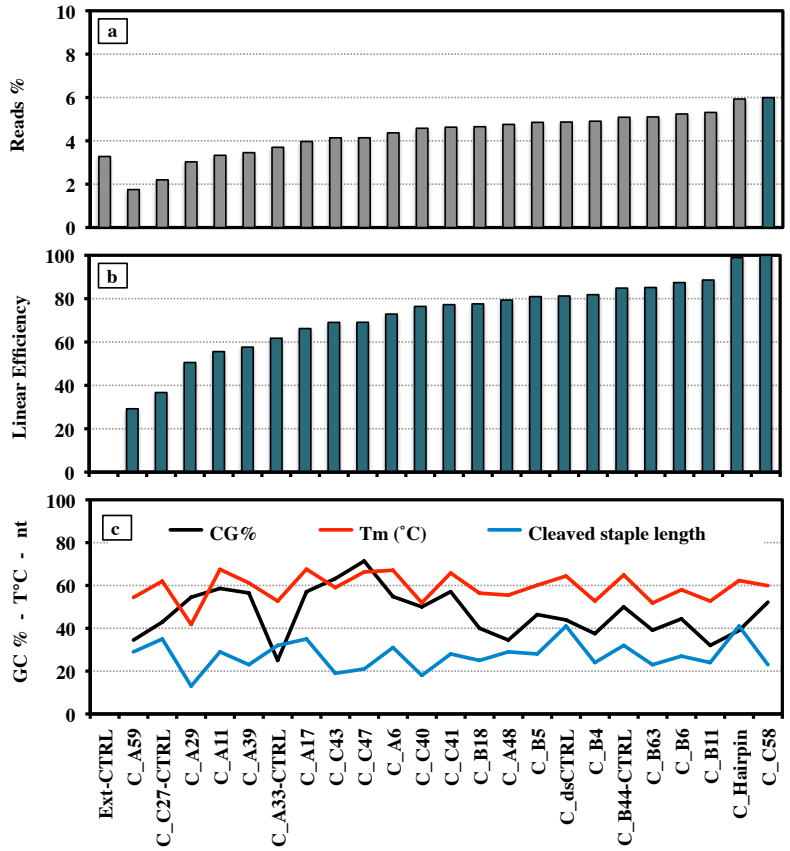


Figure 5.19. Caliper library output profile of the sample for L-PCR efficiency evaluation. The main peak around 200 bp corresponding to the right L-PCR product is flanked by a second peak at around 270 bp, which might origin from the formation of carrier pairs.

The NGS results in plot 5.3-a show that L-PCR is staple-dependent. The C58 carrier (C_{C58}) has the maximum reads value, and as such it was chosen as reference for data normalization. Normalized linear efficiency data (% of C_{C58} counts) are presented in plot 5.3-b.



Plot 5.3. L-PCR efficiency determined from NGS results. ***a.*** For each sequence, the carrier read counts are measures of the sequence amount outputted by L-PCR. The C58 carrier (C_C58) has the maximum reads value (in green) and was chosen as reference sequence to normalize L-PCR efficiency values. ***b.*** Normalized L-PCR efficiency values expressed as % of C_C58 counts. ***c.*** The GC % content, the melting temperature (°C) and the length for each cleaved-like staples are reported.

5.8. Products quantification with NGS analysis of the enzymatic cleavage of the triangular DNA nanostructure

5.8.1. L-PCR product

The triangular DNA origami was self-assembled to present positive control hairpins protruding from its surface. Following enzymatic restriction reaction, the L-PCR step was performed with five carriers, three of which relative to the negative controls, one relative to the positive control hairpin, and one relative to the shortest staple incorporated within the triangular DNA origami (i.e. A29 that leads to a 13-nt-long enzymatic cleavage product). An intermediate verification step ensured that the L-PCR protocol A employed successfully converted ssDNA inputs into dsDNA outputs (fig. 5.20). Total DNA concentration in the sample obtained after gel purification was measured with Qubit to be 4.9 ng/μL in 18 μL within instrumental accuracies. The yield of the post-reaction sample purification was near 15%.

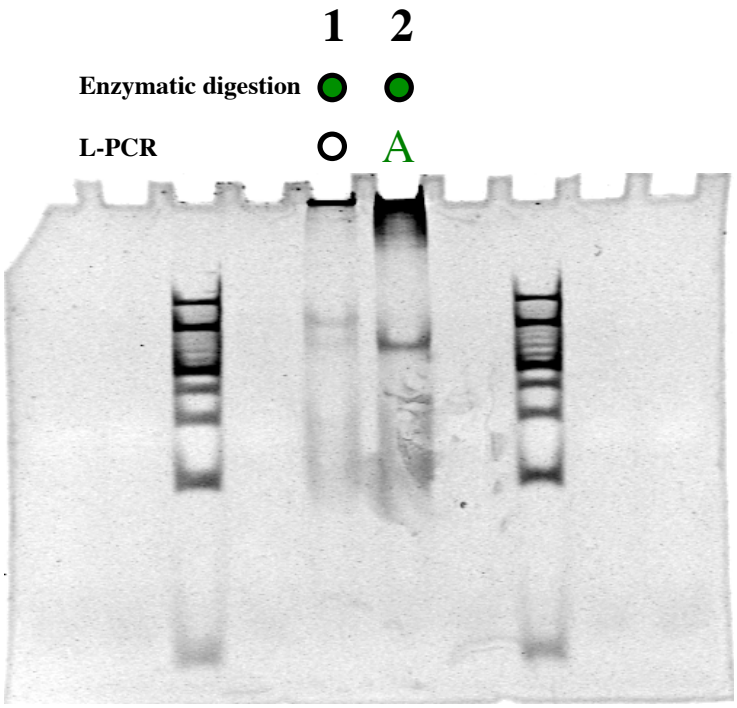


Figure 5.20. The gel clearly shows the presence of a dark band in sample 2 that corresponds to the expected L-PCR product, which is absent in sample 1 according to that fact that it lacked L-PCR treatment.

Before library preparation the sample was analyzed with Bionalyzer for quality control. The results confirmed the Qubit concentration quantification, while the evidenced peak at 78 bp is compatible with the expected 70 bp L-PCR product (fig. 5.21-a).

A second sample was prepared following the workflow of sample 1, but a lyophilization step was added after the enzymatic reaction to concentrate the sample before the L-PCR treatment, as well as to limit the reaction volumes of the subsequent steps, reduce required reagents volumes and number of gel bands extractions. The total DNA concentration in the sample after purification was estimated to be 12.8 ng/μL (Qubit) (fig. 5.21-b).

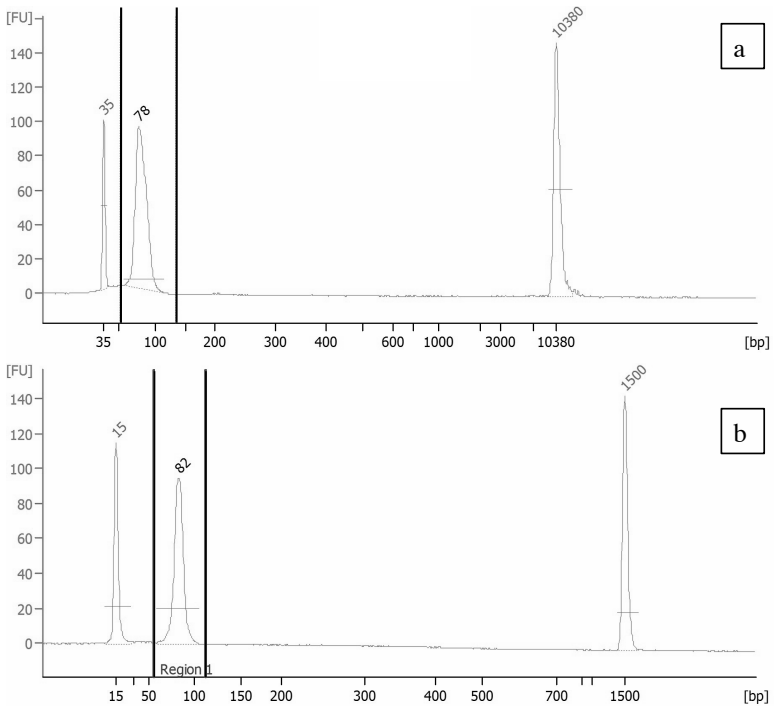


Figure 5.21. Bioanalyzer profile of the two DNA origami prepared for NGS analysis post-restriction reaction treatment with (b) and without (a) the addition of a lyophilization step before the L-PCR treatment.

5.8.2 Library product

The library product relative to sample 1 contains several DNA species, while a predominant peak is found at 195 bp, which almost perfectly corresponds to the proper length (190 bp) of the length of the fully formed dsDNA carrier (70 bp) plus the length of the terminal dsDNA adapters (60 bp each) (fig. 5.22-a).

For sample 2, two libraries were prepared (Fig. 5.22-b and c), while in one of them (Fig. 5.24-c) the DNA-end repair step was skipped and thus the library preparation protocol was

set to start from the carrier-adapter ligation step. Bionalayzer analysis indicates that the sample for which the end-repair process was performed successfully produced a high-quality library (see the sharp peak in fig. 5.22-b), while omitting the end-repair led to an insufficiently populated library (see no peak found near 190 bp in fig. 5.22-c).

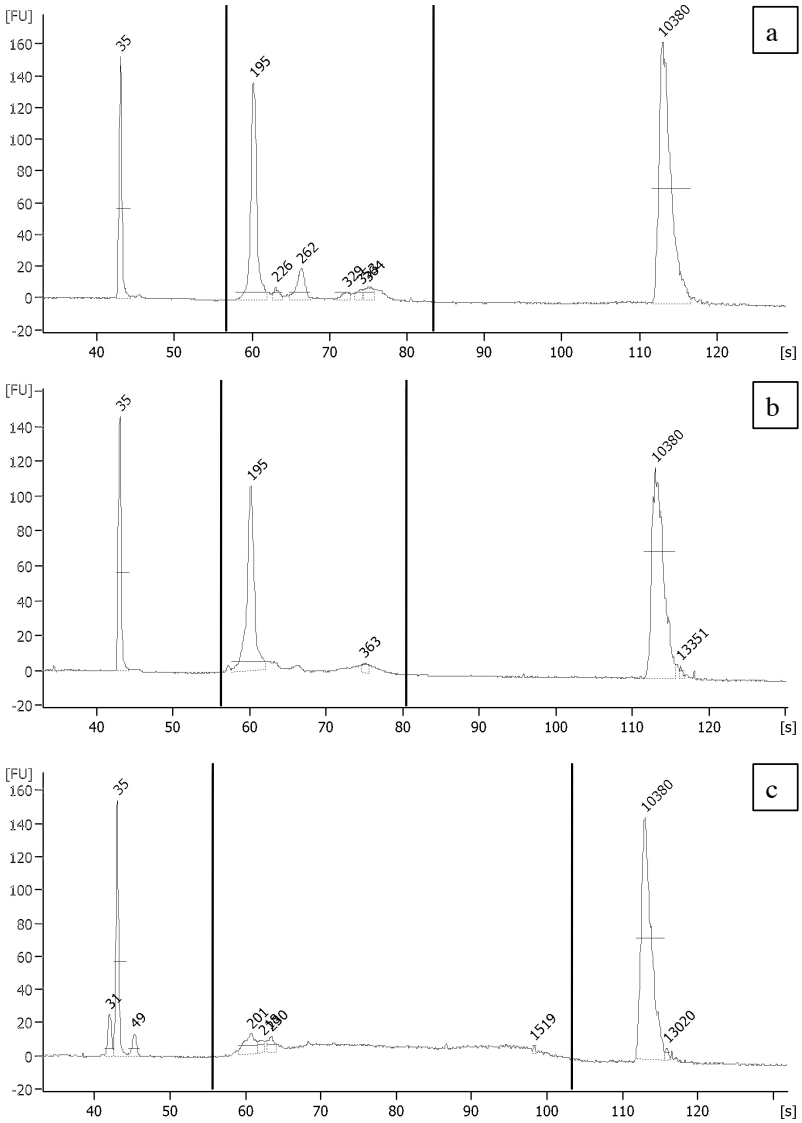
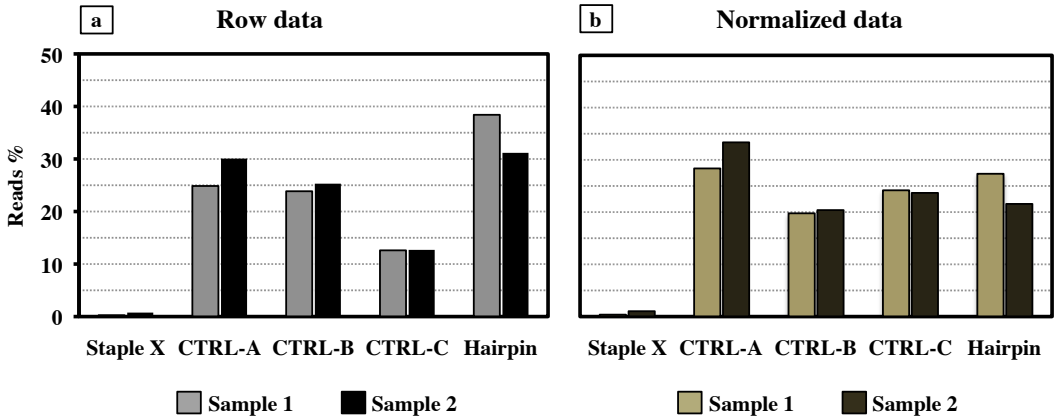


Figure 5.22. ***a.*** The NGS library product contains several DNA species, although the predominant peak is at 195 bp, which nearly perfectly corresponds to the length of the dsDNA carrier-derived molecule (70 bp) plus the length of the adapters (120 bp in total). ***b.*** NGS library product of the second sample (lyophilized) processed with a DNA-end repair step. ***c.*** Library product of the second sample (lyophilized) omitting the DNA-end repair step.

5.8.3. NGS data analysis

Illumina NGS data were analyzed with a customized algorithm able to count the number of molecules (reads) present in the sample. Plot 5.4-a depicts the reads values as % of the total amount of the molecules sequenced, while plot 5.4-b presents reads % are calibrated against L-PCR efficiency evaluated previously (see plot 5.3). My results clearly show that the staple X (A29, containing a MspI restriction site) lacks to generate its L-PCR amplification product in both the examined samples, strongly suggesting protection from MspI enzyme cleavage in the triangular DNA nanostructure (being unmodified, such DNA staple cannot fully hybridize to its designed carrier, thus inhibiting polymerase action during L-PCR). In contrast, negative and positive controls lead to the expected L-PCR products in both samples.



Plot 5.4. ***a.*** Reads values are indicated as % of the total amount of the molecules sequenced. ***b.*** % reads are calibrated against L-PCR efficiency values provided in plot 5.3.

Chapter 6 – Results of DNA based nanosensor for accurate nucleic acid quantification (Bingo-qPCR assay)

6.1. Bingo-qPCR nanosensor design using Nupack

Components of such sensor are three "foot-loop" DNA probes sequentially assembled over a long ssDNA sequence termed "scaffold". Each probe carries a target-complementary sequence in the central loop and scaffold complementary sequence at both edges ("feet"). One of the feet (anchor) keeps the probe hybridized to the scaffold (in green in figure 6.1) while the other contains the restriction site (in orange in figure 6.1). For each target molecule, I designed 3 different probes with different anchor feet to allow the best control of restriction reaction during the experiments.

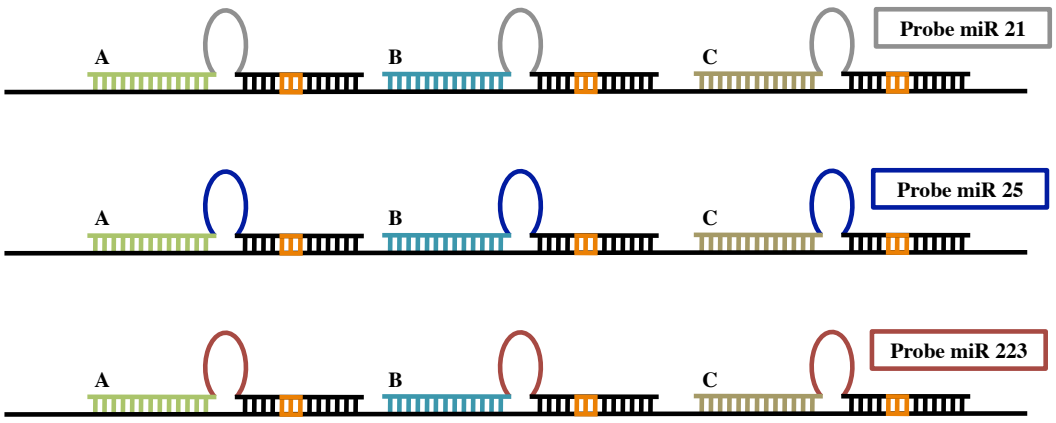


Figure 6.1. Schematics of different nanosensors, one for each target molecule (miRNA). The different versions differ in the loop, that has target-specific sequence, while scaffold and probe feet sequences are the same. On one nanosensor the three probes share the same sequence for the foot with the restriction site, but have different sequences for the other one that serves as anchor.

To design the sequences, I used the open-source, online-available software Nupack that calculates DNA or RNA sequences for custom input structures. As input for probes, I designed a strand long 70 nucleotides, with no self-complementarity. I tuned the calculations to avoid the sequences of the target molecule and the restriction site for SwaI (ATTTAAAT). The sequence obtained was a prototype that I further modified to introduce the target-complementary sequence in the middle part, and the restriction site sequence in one of the two feet, with 6 nucleotides at each side of the site to improve enzyme accessibility to the restriction site (1). Following modifications to the sequence were made to calibrate melting temperature (T_m) of the three segment of the probe. To cause the disruption of the restriction site upon hybridization of the target, the melting profile should be T_m(loop) > T_m(restriction site foot) < T_m(anchor foot). The limiting factor in this task is the length of the target. In fact, our ssDNA target has the sequence of a 22-nt-long miRNA molecule, and therefore the foot carrying the restriction site should be relatively short or have a lower GC% to reduce its melting temperature (2). These requirements are in contrast with restriction site stability. To come up with the final sequences (tab. 6.1 and Appendix II), I introduced several changes to the prototype sequence, and used Nupack to simulate the behavior of each single probe *i)* individually, *ii)* in the presence of the corresponding scaffold (sequence complementary to the feet of one probe) and *iii)* with the target. In Nupack, all the designed probes showed the ability to detach the foot carrying the restriction site in presence of the target as depicted in figure 6.2.

Oligo Name	Sequence 5'-3'
Probe A miR21	GAATTAATTTAAATAAATTATCTCAACATCAGTCTGATAAGCTACTAAAGAAAGGAAAGAGCGATAAA
Probe B miR21	GAATTAATTTAAATAAATTATCTCAACATCAGTCTGATAAGCTACTATCTCCACCCACACACAGAC
Probe C miR21	GAATTAATTTAAATAAATTATCTCAACATCAGTCTGATAAGCTACTCCTCCTCTTCACCCTCCC
Probe A miR25	GAATTAATTTAAATAAATTATCTCAGACCGAGACAAGTGCAATGCTAAAGAAAGGAAAGAGCGATAAA
Probe B miR25	GAATTAATTTAAATAAATTATCTCAGACCGAGACAAGTGCAATGCTATCTCCACCCACACACAGAC
Probe C miR25	GAATTAATTTAAATAAATTATCTCAGACCGAGACAAGTGCAATGCTCCTCCTCTTCACCCTCCC
Probe A miR223	GAATTAATTTAAATAAATTATCTGGGGTATTTGACAAACTGACACTAAAGAAAGGAAAGAGCGATAAA
Probe B miR223	GAATTAATTTAAATAAATTATCTGGGGTATTTGACAAACTGACACTATCTCCACCCACACACAGAC
Probe C miR223	GAATTAATTTAAATAAATTATCTGGGGTATTTGACAAACTGACACTCCTCCTCTTCACCCTCCC
miR21 DNA	TAGCTTATCAGACTGATGTTGA
miR25 DNA	CATTGCACTTGTCTCGGTCTGA
miR223 DNA	TGTCAGTTTGTCAAATACCCCA

Table 6.1. Probes and targets sequences.

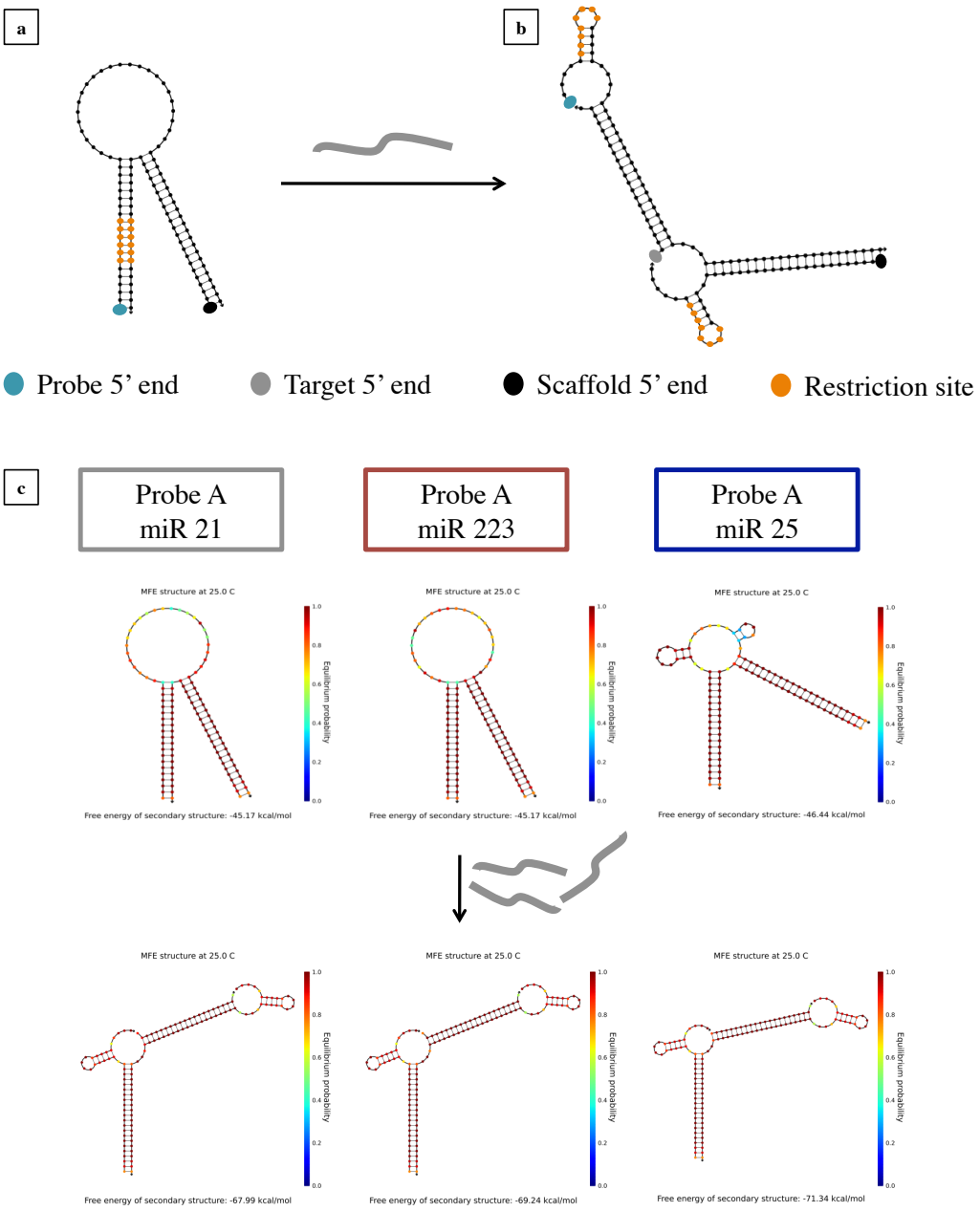


Figure 6.2. Nupack simulations of restriction site unfolding upon target hybridization. ***a.*** When the probe and its complementary scaffold sequences are present in solution, they hybridize as depicted forming the restriction site. ***b.*** In the presence of target, probe, target and scaffold hybridize as depicted forming a complex in which the duplex restriction site is disrupted. ***c.*** For the three A probes designed (one for each target sequence) the presence of the target leads to highly probable restriction site disruption.

To estimate the sensitivity of the probes I proportionally reduced the concentration of probe, scaffold and target, to the lowest value that still allows the complete hybridization of the target to the probe. In fact at low concentration, the probe-scaffold complex forms, but the target does not hybridize because too diluted. To respect the statistical requirements

of the assay, the concentration ratio for these simulations was target:probe:scaffold = 1:10:10. In the first run of simulations a single probe, the scaffold fragment complementary to it and target were used as input and the results show different detection limits for each miRNA from pM to fM sensitivity (tab. 6.2 simulation 1). Considering probes individually might alter the results, and therefore I performed a more comprehensive simulation (tab. 6.2 simulation 2) using probes A, B and C for the same target, along with the corresponding scaffold fragments and the target (with the software set to calculate complexes comprised of 3 strands). Minimum target concentration values also in this simulation are in the range of pico-femto molar. In a third simulation (tab. 6.2 simulation 3), I analyzed a simplified scaffold (142 nt long) with complementary region for 2 probes (A, B) separated by poly-T (20xT – scaffold A sequence – 20xT – Scaffold B sequence – 20xT), in combination with the respective probes, and the miR21 ssDNA sequence. The 6 pM concentration value obtained is compatible with results from simulation 1 and 2. In this simulation a small % of target-probe dimers is present, suggesting that the target not only hybridizes to the probe, but also induces complete detachment of the probe from the nanosensor.

		Simulation 1	**Simulation 2**	**Simulation 3**
Target	Probe	Target lowest concentration (pM)		
miR21	A	2	2	6
	B	2	2	
	C	2	2	
miR25	A	0.005	0.0046	
	B	0.003	0.004	
	C	0.005	0.0033	
miR223	A	0.02	0.024	
	B	0.009	0.02	
	C	0.02	0.017	

Table 6.2. Minimum concentration at which all target is hybridized to the probe-scaffold complex, allowing restriction site protection. ***Simulation 1*** involves one probe at a time, the complementary scaffold fragment and the target and allows the formation of 3-strands complexes. ***Simulation 2*** involves probes A, B and C for the same target, the corresponding scaffold fragments and the target, and allows the formation of 3-strands complexes. ***Simulation 3*** was performed only for miR21 and involves a simplified scaffold (142 nt long) with complementary region for 2 probes (A, B) separated by poly-T (20xT – scaffold A sequence – 20xT – Scaffold B sequence – 20xT), the respective probes, and the target.

I checked restriction site stability by evaluating scaffold-probe melting profile at a concentration of 100 fM, 100 pM and 100 nM. The melting temperature at 100 fM is 46-

47°C for probes A, B and C, which is well above the working temperature of DraI and SwaI. At higher concentrations the calculated temperatures increase.

Probe	T (°C) 100 fM	T (°C) 100 pM	T (°C) 100 nM
A	46	51	54
B	47	53	55
C	47	53	57

Table 6.3. Melting temperature of each probe-scaffold hybrid calculated with Nupack at different concentrations.

6.2. Bingo-qPCR nanosensor design check using IDT OligoAnalyzer

With IDT Oligoanalyzer I estimated melting temperature of nucleic acids strands, defined as the temperature at which half of the DNA molecules are present in solution in the form of a duplex. Melting temperature changes as a function of the concentration. To understand which one of the two feet of the probe is more prone to dissociate, I analyzed separately their melting temperatures at a DNA concentration of 100 nM and 100 pM (the lowest concentration accepted by the software). Melting temperature of restriction site foot is influenced by the presence of the anchor foot and vice versa, therefore these values are not intended to be used as reference, but just to evaluate differences in thermodynamic properties between the feet of the probes. Oligoanalyzer calculated higher melting temperature of the anchor foot with respect to the site foot, suggesting that the probe should open on the restriction site side upon hybridization of the target. Moreover, the target T_m is higher than restriction site-carrying foot T_m, confirming the behavior predicted by Nupack (table 6.4).

Oligo name	Tm (°C) 100 uM	Tm (°C) 100 nM
Foot with the restriction site	46.5	37.1
Anchor foot A	60.8	51.9
Anchor foot B	65.1	55.1
Anchor foot C	64.0	52.8
miR21 - DNA	58.1	51.9
miR25 - DNA	66.4	57.1
miR223 - DNA	62.6	54.9

Table 6.4. Melting temperature of probe feet and target calculated with IDT Oligoanalyzer.

6.3. Bingo-qPCR nanosensor design check using Nupack

6.3.1. Probe secondary structure

I analyzed the ability of probes to fold into secondary structures thus preventing the hybridization to the target or to the scaffold, or forming additional restriction sites in solution that might sequester restriction enzyme molecules. The free energy of the scaffold-probe hybrid calculated by Nupack is always 7-10 fold higher than the one showed by probe hairpins, self-dimer or hetero-dimer (table 6.5 and table 6.6).

	Scaffold hybrid	Hairpin	Homo-Dimer
Oligo name	ΔG (Kcal/mol) at 25°C, 100 nM	ΔG (Kcal/mol) at 25°C	ΔG (Kcal/mol) at 25°C, 100 nM
probe A mir21	-45.17	-5.24	-20.07
probe B mir21	-46.17	-5.73	-
probe C mir 21	-43.77	-4.53	-
probe A mir25	-45.17	-6.2	-
probe B mir25	-46.17	-6.48	-
probe C mir 25	-43.77	-5.81	-
probe A mir223	-45.17	-3.81	-
probe B mir223	-46.17	-5.58	-
probe C mir 223	-43.77	-4.78	-

Table 6.5. Secondary structures of probe and scaffold free energies calculated with Nupack.

	Etero-Dimer
Oligo name	ΔG (Kcal/mol) at 25°, 100 nM
Probe A miR21-B miR21	19.42
Probe B miR21-C miR21	-
Probe A miR21-C miR21	19.42
Probe A miR25-B miR25	-
Probe B miR25-C miR25	-
Probe A miR25-C miR25	-
Probe A miR223-B miR223	-
Probe B miR223-C miR223	-
Probe A miR223-C miR223	-

Table 6.6. Probes hetero-dimers free energies calculated with Nupack.

6.3.2. Target secondary structure

To evaluate the tendency of the target molecules to hybridize to the probe, I analyzed their eventual secondary structure. MiR21 and miR223 seem to be more prone to form hairpins, but, even if this aspect might reflect their lowest affinity to the probe compared to the affinity of miR25 as calculated by Nupack, hairpin ΔG values are so low that they can be ignored (table 6.7).

	Hairpin	Homo-Dimer
Oligo name	ΔG (Kcal/mol) at 25°C	ΔG (Kcal/mol) at 25°C, 100 nM
miR 21	-1.91	-
miR 25	-	-
miR 223	-0.11	-

Table 6.7. Target secondary structure free energy.

6.3.3. Probe localization on the scaffold

Previous results showed that probes perfectly hybridize to their specific scaffold fragment. However, the final scaffold sequence will host 3 probes with a common sequence (foot with the restriction site). Therefore, I tested the hybridization of one probe (A) in the presence of a simplified scaffold consisting of 3 probe-complementary sequences connected by poly-T. In this particular situation, the results show that probe A can hybridize to the right site, but also to partially to the B site (fig. 6.3-a). However, in the presence of all 3 probes, each one recognizes its corresponding, complementary sequence (fig. 6.3-b).

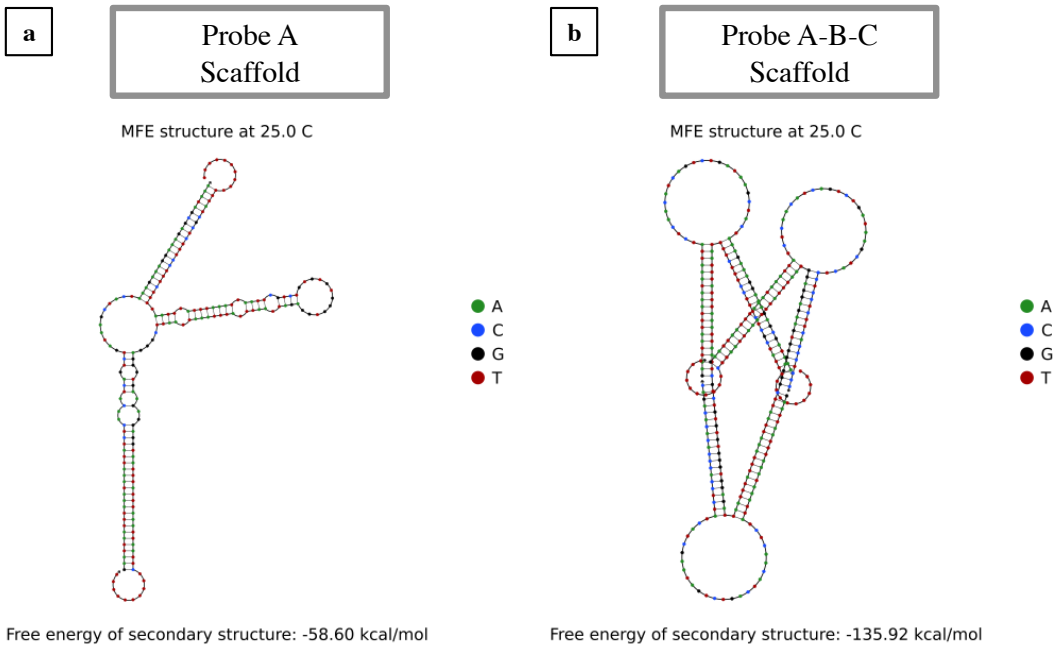


Figure 6.3. Nupack test of probe localization on the scaffold. ***a.*** Probe A for miR21 in presence of a simplified scaffold with 3 probe-complementary sequences connected by poly-T: probe A can hybridize to the right site, but also to the B site. ***b.*** Simulation containing probe A, B and C for miR21 and the simplified scaffold: the hybridization occurs in the expected position and the nanosensor forms correctly.

6.4. Scaffold design

The scaffold final length is 242 nucleotides. The three hybridized probes occupy 120 nucleotides, and I used the remaining nucleotides to accommodate the complementary sequences of 6 primers and 1 TaqMan probe for qPCR. As a guide, I chose a fragment of Lambda vector, I inserted the sequences of the three probes and I checked for homo-dimers

in BLAST (Basic Logical Alignment Search Tool). This open source program is used to check alignment of long sequences (for more details see section 3.2.2) and resulted useful also in scaffold analysis. I also analyzed the sequence in Nupack to check for secondary structure. To avoid unspecific hybridization longer than 10 base pairs, which is about half the length of each probe feet I modified the scaffold sequence along with probe sequences. Moreover, the scaffold was designed to meet the requirements for qPCR primers and TaqMan probe (see section 6.4). The ultimate scaffold sequence (Appendix II) was analyzed in Nupack for secondary structures at different temperatures. At 25°C the scaffold is predicted to form a secondary structure with a relevant free energy if compared to the free energy of probe-scaffold complex (fig. 6.4). This secondary structure is avoided at higher temperature therefore the protocol to fold the nanosensor will include a denaturation step, followed by a temperature decreasing thermal ramp.

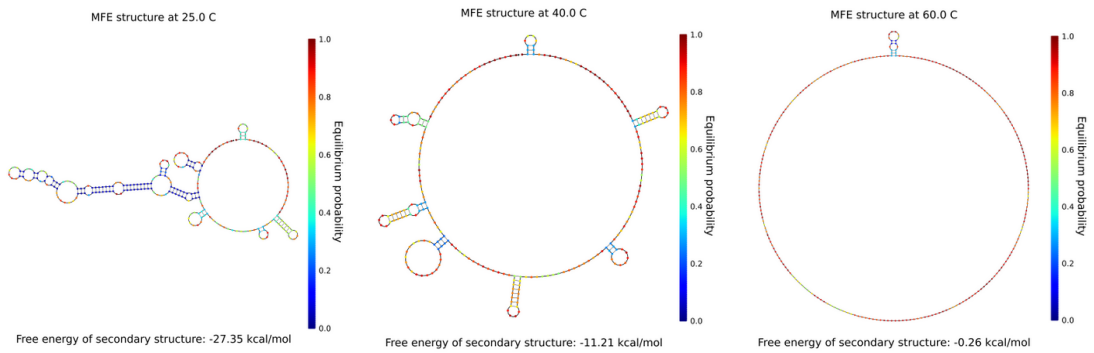


Figure 6.4. Nupack simulation of scaffold secondary structure at different temperatures.

6.5. qPCR primers and probe design

Basic requirements for primer design are the following:

- Design primers with a GC content of 50–60%.
- Maintain a melting temperature around 60 ºC: the polymerase used in Bingo-assay has annealing and extension temperature of 60 °C.
- Avoid secondary structure.
- Avoid primer-dimer formation: check sequence of forward and reverse primers to ensure no 3' complementarity.

- Place Gs and Cs on ends of primers to facilitate hybridization and polymerase binding.

Basic requirements for TaqMan probe design are the following:

- Melting Temperature 5-10 °C higher than the one of the primers.
- Less than 30 nucleotides in length.

Typically, these parameters are used to choose a primer complementary sequence on a non-modifiable template sequence. In the specific case of the nano sensor design, I could modify the scaffold and create a DNA sequence hosting primers and TaqMan probe complementary sequences with suitable characteristics (fig. 6.5). The latest design (sequences in Appendix II) shows that primers secondary structures analyzed with IDT Oligoanalyzer (DNA oligo 0.2 mM, Na^+ 50 mM, Mg 3 mM, dNTPs 0.8 mM) have 5-fold lower ΔG values with respect to the dimer with the scaffold.

Oligo name	Tm	GC %	Length	ΔG (kcal/mole)	ΔG Self Dimer	ΔG Hairpin	
Primer Taq For	61.7	50.0	20	-37.19	-5.19	-1.5	
Primer A Rev	62.2	47.6	21	-41.06	-3.61	-2.27	
Primer B For	61.7	50.0	20	-34.68	-	1.94	
Primer B Rev	62.3	52.4	21	-40.48	-3.61	0.46	
Primer C For	62.1	52.6	19	-35.34	-3.53	-0.18	
Primer Taq Rev	61.9	50.0	20	-38.63	-3.52	0.67	
TaqMan Probe	67.6	58.3	24	-44.56	-4.41	-0.66	

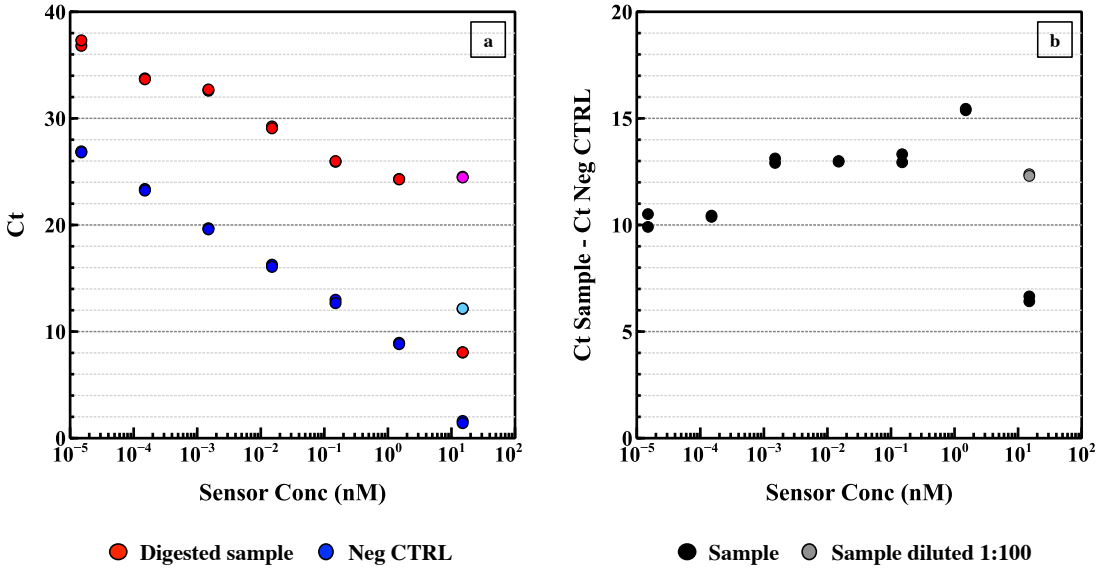
	Hetero-dimers ΔG (kcal/mole)						
Oligo name	Primer Taq For	Primer A Rev	Primer B For	Primer B Rev	Primer C For	Taq Probe Rev	Primer Taq Rev
Primer Taq For	-	-5.23	-4.89	-6.66	-3.53	-4.87	-5.53
Primer A Rev	-	-	-5.46	-6.68	-3.53	-6.68	-5.12
Primer B For	-	-	-	-3.07	-3.55	-6.73	-6.62
Primer B Rev	-6.66	-6.68	-3.07	-	-6.62	-5.13	-6.62
Primer C For	-	-	-	-	-	-4.41	-5.12
Primer Taq Rev	-	-	-	-	-	-	-
TaqMan Probe	-	-	-	-	-	-	-5.02

Figure 6.5. Primers secondary structures free energies.

6.6. Bingo-qPCR detects enzymatic degradation

The folding of the nanosensor was carried out with a one to one ratio between probes and scaffold, otherwise the target would hybridize to free probes excess without affecting nanosensor conformation, thus preventing target detection. In this experiment we tested different concentration of nanosensor in enzymatic reaction solution ranging from 15 nM to 15 fM to find the best condition for the enzymatic reaction (SwaI): low DNA concentration reduces DNA-enzyme binding efficiency, but high DNA concentration

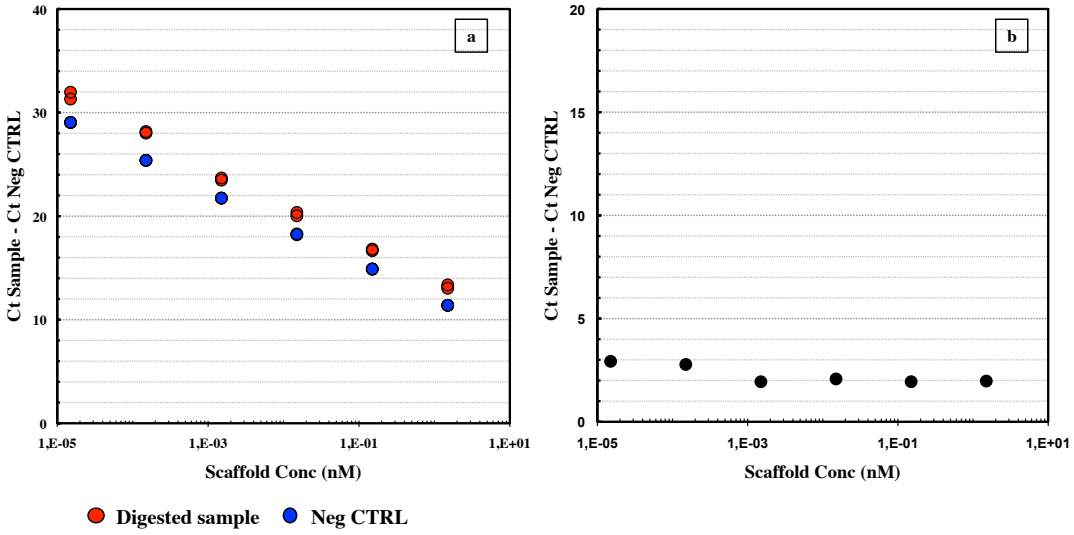
might require more enzyme units to efficiently cleave the well-formed restriction sites. For each digested sample, a negative control without enzyme in reaction solution was processed and all samples have been analyzed in double.



Plot 6.1. In ***a*** we plotted Ct values obtained with Bingo-assay of digested and not-digested (Neg CTRL) samples. Each point corresponds to a different concentration of the nanosensor in the enzymatic reaction. Higher Ct values for digested samples were expected as the bingo-scaffold concentration should be reduced by enzymatic restriction. 1.5 nM samples were also analyzed at 1:100 dilution (see pink and light blue dots) to fit qPCR concentration range. In ***b*** we plotted the ΔCt values of the same samples of plot a. The enzyme seems to be more active with 1.5 nM nanosensor concentration.

In plot 6.1-*a* we can see the difference between Neg CTRL and digested samples. When the enzyme is present the Ct values increases because the template (bingo-scaffold) is less abundant. ΔCt values (plot 6.1-*b*) range from 10 to 15 roughly, meaning that at all the concentrations tested the nanosensor are digested, although with variations. This behavior demonstrates the ability of the protocol to fold the nanosensor, and suggests 1.5 nM concentration of DNA in enzymatic reaction as best condition for enzyme reactivity. The ΔCt at 1.5 nM is 15.4 that corresponds to 2.8 % of non-cleaved restriction sites.

To confirm nanosensor folding we performed the same experiment on the scaffold without probes. The presence of palindromic restriction sites in the sequence introduces the opportunity to form secondary structures containing restriction sites that can be recognized by the enzyme. The results (plot 6.2) show that there is scaffold cleavage, but at about 22-fold lower efficiency than in the case of the fully formed nanosensor.



Plot 6.2. In ***a*** we plotted Ct values obtained with Bingo-assay of digested and not-digested (Neg CTRL) scaffold. Each point corresponds to a different concentration of the scaffold in the enzymatic reaction. Higher Ct values for digested samples were expected as the bingo-scaffold concentration should be reduced by enzymatic restriction. In ***b*** we plotted the ΔCt values of the same samples of plot a with a scaffold concentration of 1.5 nM in enzymatic reaction, the digestion is more than 20 times less efficient than nanosensor degradation (ΔCt=1.97 compared to ΔCt=15.4).

6.7. Bingo-qPCR model Test

Samples were incubated at 40°C for 8 hours in enzymatic reaction buffer before adding the enzyme to simulate target incubation. To start, I performed on sample and CTRL (without enzyme in the digestion reaction) site-specific qPCR using 3 different primer couples in separate reaction solution. Then, I performed Bingo-qPCR on the same samples. Plot 6.3 compares bingo results with standard results showing the correspondence between the two assays. I used the ΔCt values to calculate the percentage of un-cleaved restriction sites A, B and C in digested sample. Results reported in table 6.8 show that bingo results are close to site-specific assays results. It is worth noting that the probability for each restriction site to be cleaved is not the same for sensing elements A, B and C, and therefore the bingo-scaffold probability should be calculated as

$$P = p_A \cdot p_B \cdot p_B$$

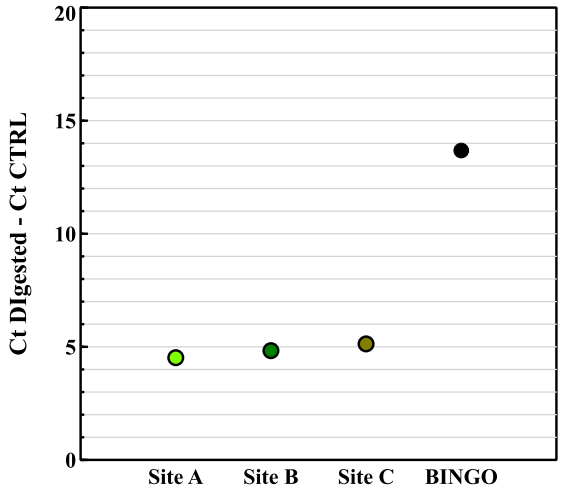
With this consideration, the expected value for ΔCt(Bingo) is 14.48 with

$$\Delta Ct(Bingo) = \Delta Ct(A) + \Delta Ct(B) + \Delta Ct(C)$$

.

	ΔCt = CTRL - Sample	Un-cleaved Restriction sites %
A	-4.52	4.36
B	-4.83	3.52
C	-5.13	2.86
BINGO	-13.7	4.24

Table 6.8. The reported ΔCt values between digested and not digested (CTRL) samples obtained by site-specific assays (A, B, C) and by Bingo-assay were used to calculate the un-cleaved restriction sites within the nanosensor after enzymatic degradation.



Plot 6.3. Digested nanosensor is analyzed with site-specific qPCR and Bingo-assay. The plot shows a ΔCt increment of about 3-fold with respect to ΔCt_A, ΔCt_B, ΔCt_C.

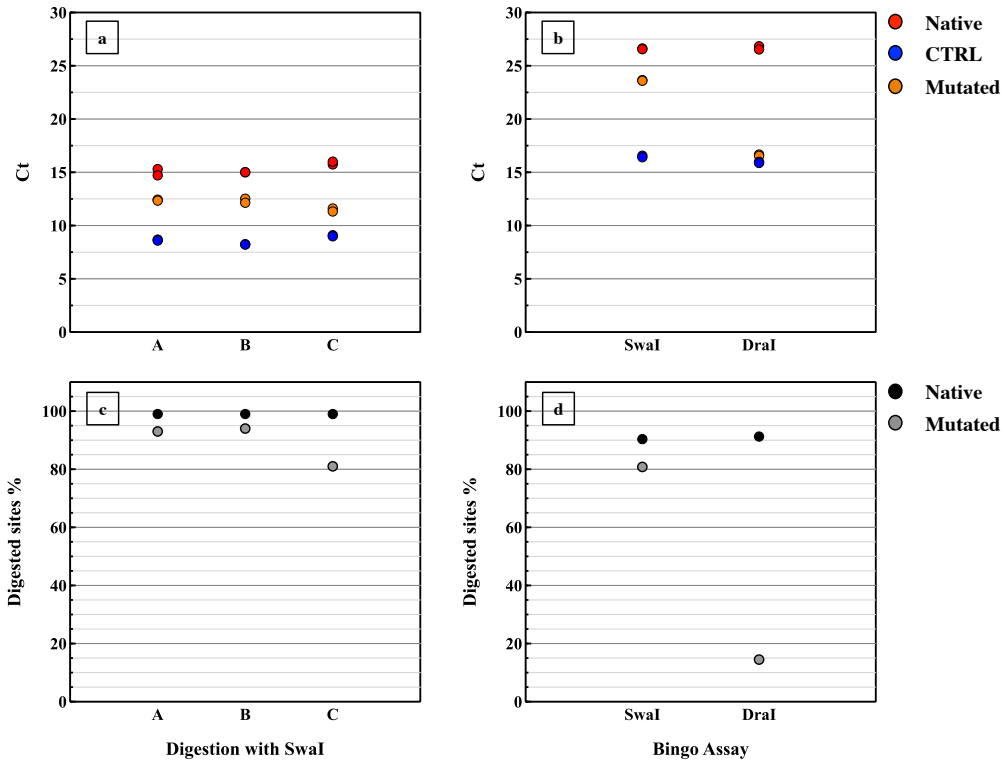
6.8. Bingo-qPCR nanosensor folding check

6.8.1. Negative control

To confirm nanosensor folding I designed a negative control of the system by mutating the restriction sites on the probes. To inhibit enzyme activity, yet allowing probe hybridization on the scaffold, I mutated only two bases in the middle of the restriction site. So, the sequence ATT-*TA*-AAT was converted into ATT-*CT*-AAT. In the first experiment the nanosensor was folded with 1:1 ratio between scaffold and probes. I performed qPCR on each restriction site and results are reported in plot 6.4-a. SwaI enzyme was able to digest the mutated restriction sites (with lower efficiency for site C) confirming the formation of the mutated restriction site: in terms of cleaved sites percentage, SwaI digested 99% of the native sites and 93-94% of mutated sites A and B and 81% of mutated sites C (plot 6.4-c).

To obtain a consistent negative control I should have increased the number of mutated bases, but this would have destabilized the probe structure. Hence, I changed the restriction enzyme and used DraI, which has shorter restriction site (TTTAAA) and might be more sensitive to small mutations.

In this second experiment, I folded the nanosensor with 10-fold excess of probes and digested it with DraI and SwaI. In plot 6.4-b-d Bingo-qPCR results show that while SwaI digests about 80% of mutated restriction sites, DraI is more sensitive to site mutation: its low efficiency in digesting the mutated nanosensor (15% of mutated sites are cleaved) is a demonstration of nanosensor folding.



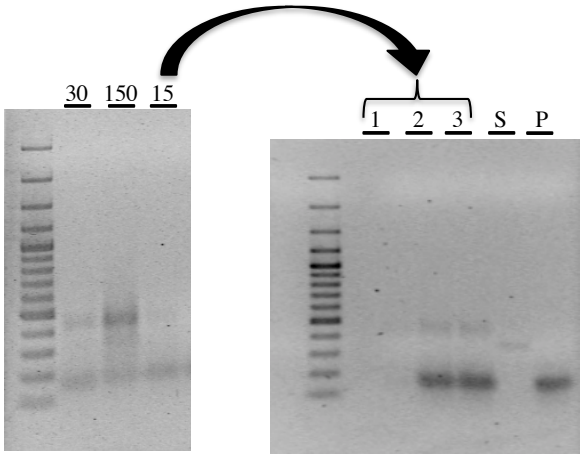
Plot 6.4. Mutated and native nanosensors were digested with SwaI. ***a*** Site-specific qPCR show that SwaI enzyme was able to digest the mutated restriction sites with lower efficiency for site C (yellow dots) confirming the formation of the mutated restriction site. ***b*** Bingo-qPCR show that while SwaI digests the mutated restriction sites, DraI is more sensitive to site mutation and behaves like the not digested control (blue dots). ***c-d*** Digested restriction site % in native and mutated nanosensor after digestion with SwaI and DraI corresponding to the ΔCt values obtained from plot a and b respectively.

6.8.2. Nanosensor folding optimization

I examined the nanosensor formation by folding probes at constant concentration (150 nM) with progressive concentration of the scaffold (15 nM, 30 nM, 150 nM), and analyzed the

samples with agarose gel (fig. 6.6-a). For samples 150, 30 and 15 nM of scaffold, two distinct bands are present in the gel: probe band at lower molecular weight that corresponds to the probe control band in gel 6.6-b, and nanosensor (higher molecular weight). In the sample with 150 nM scaffold (1:1 scaffold-probe ratio) there is also an additional light band, corresponding to the scaffold control band in gel 6.6-b, suggesting that the nanosensor is not completely folded. Moreover, the presence of the probe band confirms the partial folding. No scaffold band is present in the sample with 10-fold probe excess therefore we chose this condition to fold the nanosensor and we set up a protocol to purify the sensor afterwards (section 4.5). To evaluate the purification efficiency, gel analysis of the purified sample compared to the not purified sample was performed (fig. 6.6-b). In the not purified sample (2), probe band is clearly detectable, while in the purified sample is not present. Nevertheless, the nanosensor band in the purified sample has low intensity meaning that high percentage of the nanosensor is lost during purification. This gel gives also further information about sensor folding:

- The scaffold alone runs differently from scaffold-probe sensor: data confirm the formation of the nanosensor
- The non-purified samples treated with and without folding thermal ramp have very similar composition: with a 10-fold probe excess with respect to scaffold, the nanosensor folds with a certain efficiency even without thermal ramp.



a. Samples are nanosensor folded at different scaffold concentrations (nM). Probe concentration is 150 nM

b. Samples are

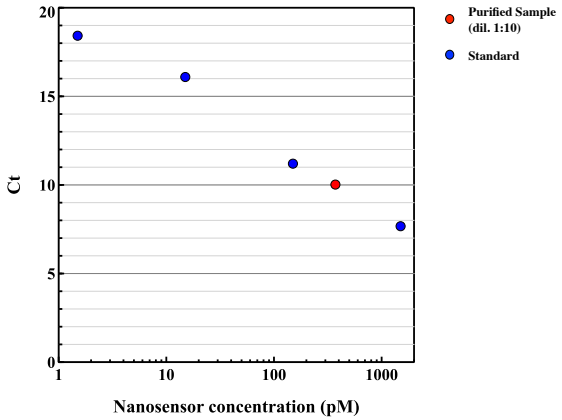
1. Purified
2. Not purified
3. Not folded

S. Scaffold
P. Probe Mix

Figure 6.6. ***a.*** The nanosensor is folded in different probe-scaffold ratios by reducing scaffold concentration (15 nM, 30 nM, 150 nM) at constant probe concentration. Samples with 150, 30 and 15 nM of scaffold show

probe band and nanosensor band. In the sample with 150 nM scaffold, an additional light band corresponds to the scaffold control band in gel 6.6-b, suggesting that the nanosensor is not completely folded. Moreover, the presence of the probe band confirms the partial folding. No scaffold band is present in the sample with 10-fold probe excess. ***b.*** Gel analysis of the purified sample compared to the not purified sample: in the not purified sample (2), probe band is clearly detectable, while in the purified sample is not present. Scaffold and nanosensor bands have different molecular weight confirming the formation of the probe scaffold complex. Sample 3 contains the scaffold probe mixture in 1:10 ratio without thermal treatment. The profile of sample 3 is almost the same of sample 2 that contains the folded nanosensor suggesting that with 10-fold probe excess the nanosensor folds even without thermal ramp.

We estimated the recovery yield of purification protocol with Bingo-qPCR. A standard curve was prepared using progressive dilutions of a sample containing the scaffold and the probes (1:1 ratio) at known concentration (plot 6.5). All the dilutions were analyzed with qPCR, and the Ct values were correlated with the concentration of the samples. Using this calibration curve, we calculated the concentration of the purified sample that resulted to be 75% lower than the initial concentration. Despite a recovery yield of 25% is rather low (it can be improved using other techniques such as magnetic beads), it was sufficient to for the protocol utilized in the following experiments.

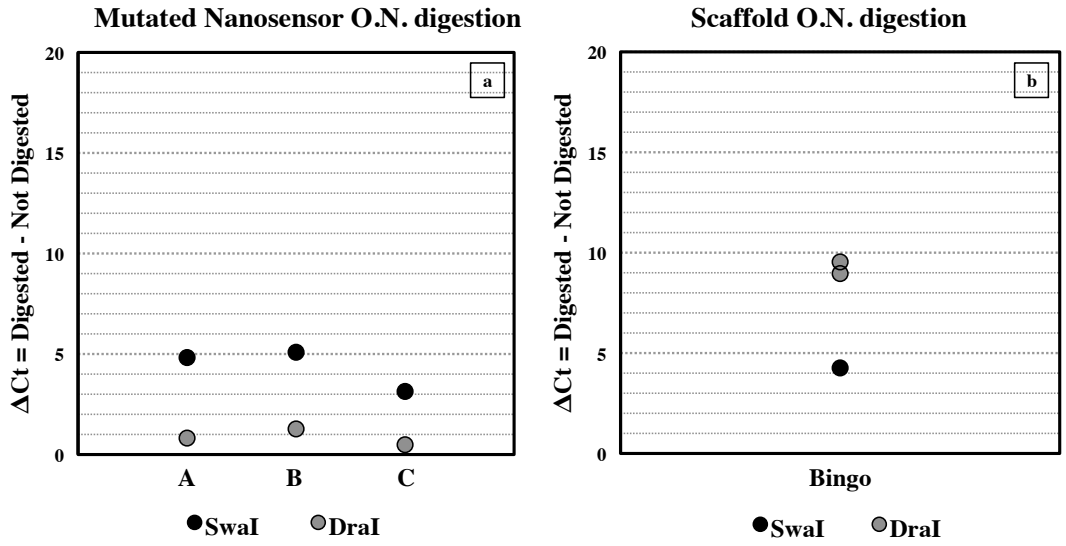


Plot 6.5. Estimation of nanosensor purification protocol recovery yield. The standard curve was prepared using progressive dilutions of a sample containing the scaffold and the probes (1:1 ratio) at known concentration. Each sample was analyzed with qPCR and the Ct values were correlated with the concentration of the samples. The estimated recovery yield is about 25%.

6.8.3. Overnight nanosensor stability

To evaluate nanosensor stability the digestion was performed overnight on native and mutated nanosensor. After more than 12 hours incubation the results confirmed those previously obtained: complete degradation of native sample with DraI, protection from

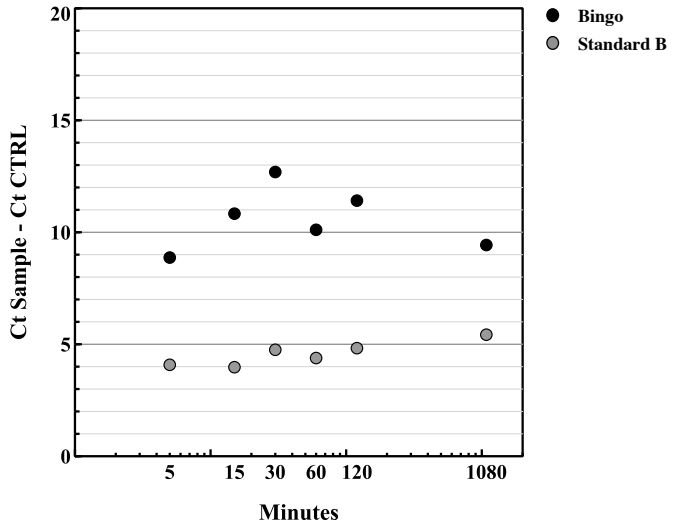
DraI activity on mutated nanosensor, and a partial degradation of mutated sample in site C by enzyme SwaI (Plot 6.6-a). In addition to nanosensor testing, we analyzed the scaffold alone with the same protocol. Results with the Bingo-qPCR show that DraI is not able to digest the mutated nanosensor, but can digest the scaffold alone. SwaI, instead, has lower efficiency in scaffold degradation. These findings demonstrate the formation of the mutated nanosensor and its overnight stability.



Plot 6.6. ***a.*** The mutated nanosensor was digested with SwaI and DraI overnight. ΔCt values obtained by site-specific qPCR of sample digested with DraI are close to 1, the accuracy limit of qPCR, indicating almost no degradation of the nanosensor. SwaI instead is able to digest the mutated nanosensor with high efficiency. ***b.*** The scaffold was digested with SwaI and DraI overnight. ΔCt values obtained by Bingo-qPCR of sample digested with DraI indicate a degradation of almost 90% of the restriction sites on the scaffold, while SwaI is not able to digest the scaffold alone (2.8 % of restriction sites on the scaffold are cleaved).

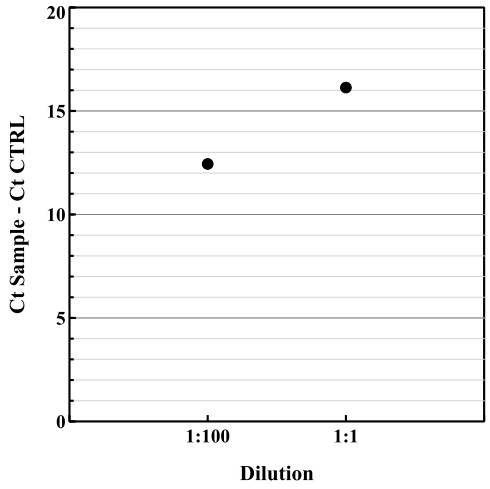
6.9. Bingo-qPCR detects enzymatic activity variations

We monitored DraI activity with standard (on site B) and Bingo-qPCR at different time (15, 30, 60, 240 minutes an over-night). Different tubes were prepared for each time point. The nanosensor was diluted with the buffer solution and split into distinct aliquots to which the enzyme was added separately, right before the reaction start. The enzyme activity was stopped by freezing the samples. Results in plot 6.7 clearly show that, while site-specific qPCR is quantifying the same degradation rate at all time points, Bingo-assay senses some kind of enzymatic fluctuation that has no correlation with the time. The kinetic experiment demonstrated that Bingo-assay is extremely sensitive to small variation in enzyme efficiency that in this case might be due to variation in enzyme concentration (pipetting error of enzyme solution in the reaction mixture).



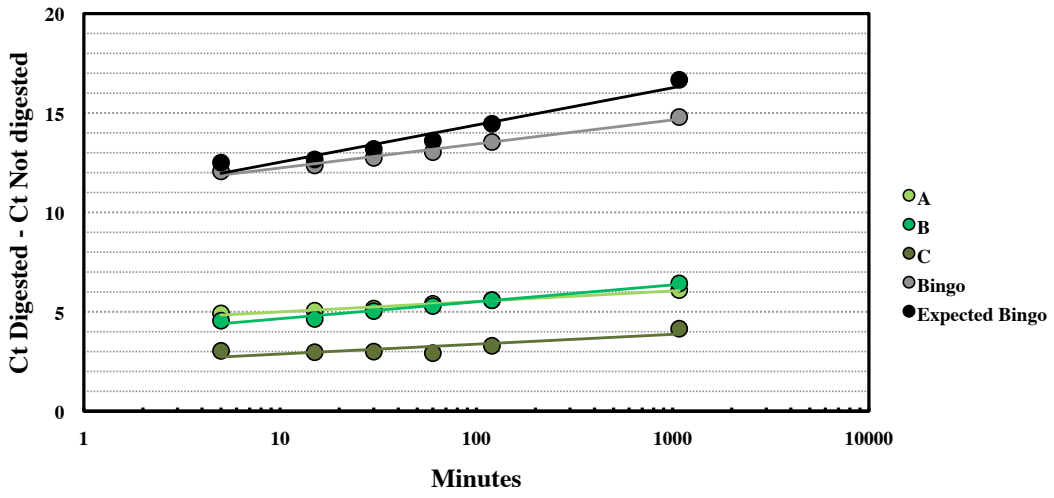
Plot 6.7. DraI kinetic on nanosensor monitored with Bingo-qPCR. While site-specific qPCR is quantifying almost the same degradation rate at all time points, Bingo-assay senses some kind of enzymatic fluctuation that has no correlation with the time.

In the next experiment, I tested the ability of Bingo-assay to detect variation of enzymatic activity as a consequence of thermal equilibration before reaction. Two samples of purified nanosensor were diluted at 1.5 nM and 15 pM, and incubated at 37°C for 1 hour to equilibrate enzymatic reaction conditions. Prior to enzymatic reaction, the 1.5 nM sample needed 100-fold dilution (to provide 15 pM nanosensor concentration) which was the defined enzymatic reaction concentration in this experiment. The dilution step could disturb equilibration, while the 15 pM sample was treated with the enzyme as it was. The data presented in plot 6.8 demonstrate a difference of about 4 Ct between the two conditions, evidencing lower enzymatic efficiency in the diluted sample.



Plot 6.8. We tested the ability of Bingo-assay to detect variation of enzymatic activity as a consequence of thermal equilibration before reaction. Both samples were incubated at 37°C for 1 hour before enzymatic reaction, but one of the two was diluted before adding the enzyme. This step could alter the equilibrium condition, likely responsible of a reduced efficiency of enzymatic cleavage (lower ΔCt value).

I confirmed the results presented above with a second kinetic experiment. After folding and purification, the sample was thermally equilibrated and digested with DraI. For each time point (5-15-30-60-120), site-specific qPCR and Bingo-qPCR were performed and compared. The results (plot 6.9) show linear correlation between time and digestion both with site-specific qPCR and Bingo-qPCR. Moreover, experimental bingo-related values (gray line) are compatible with those predicted by the model with a maximum error for the O.N. sample of 11.23 %. The expected bingo ΔCt values are calculated as the sum of the ΔCt values for A, B and C (black line).



Plot 6.9. Monitoring DraI activity on the nanosensor over the time. Linear correlation between time and DNA degradation is detected both with site-specific qPCR and Bingo-qPCR. Moreover, experimental bingo-related values (gray line) are compatible with those predicted by the model with a maximum error for the overnight reaction.

6.10. qPCR assay evaluation

The agarose gel analysis (fig. 6.7) shows that Bingo-qPCR products of digested and not digested nanosensors are about 250 bp long, which is compatible with the expected value of 238 bp of the amplicon.

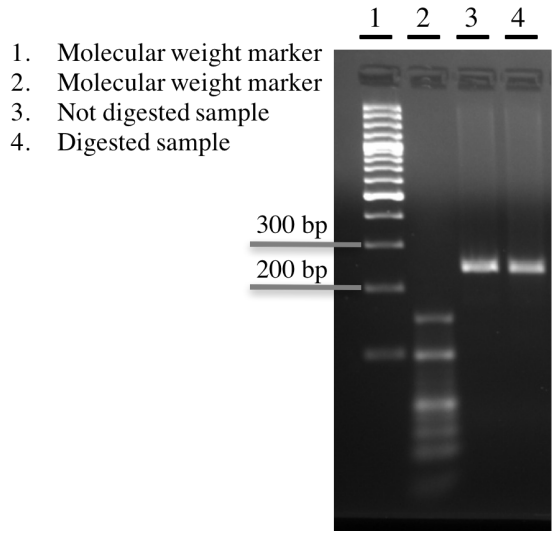


Figure 6.7. Agarose gel analysis (2% agarose gel) of Bingo-qPCR product. Not digested sample and digested sample produce identical bands of about 250 bp. The length of the amplicon is 238 bp.

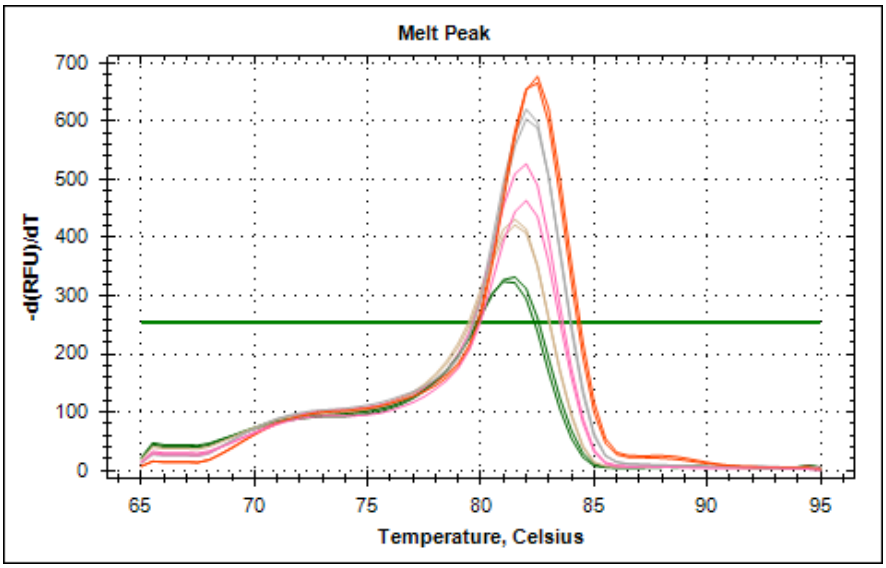
The melting curve profiles obtained for each qPCR assay product show one single peak corresponding to the specific DNA sequence amplified. The experimentally obtained melting temperatures values are comparable with those calculated with OligoAnalyzer (tab. 6.9), in fact they both show almost the same melting temperature for amplicons A and B, and for amplicons C and Bingo. The difference between calculated and expected values depend on DNA and salt concentration.

	Melting Temperature (°C)	
	OligoAnalyzer	Melting curve
Amplicon A	67	75
Amplicon B	67	75
Amplicon C	72	81
Amplicon BINGO	72	82

Table 6.9. Melting temperature of qPCR amplification product obtained with site-specific qPCR assay and with Bingo-qPCR assay.

6.10.1. Primer concentration optimization

The best primer-template concentration ratio depends on DNA characteristics and quality and should always be determined empirically. Typical primer concentration used in qPCR ranges between 250 and 500 nM and suits a large template DNA concentration range (template might also be fM). To optimize the experiment, we tested 5 primers concentrations (1 μM, 500 nM, 250 nM, 125 nM and 62.5 nM) to analyze a standard sample (section 4.9.1). Figure 6.8 shows that Ct value for bingo-scaffold is the same for all primers concentrations indicating that the sensitivity of the protocol is independent on primer concentration. Melting curve of PCR product show one single peak for each primer condition demonstrating that no unspecific products is present at the end of the reaction. Ct values vary from approx. 16.1 to 16.9 increasing primer concentration, and the behavior is confirmed by the variation in peak height of the inherent melting temperature curves. Nevertheless, ΔCt values <1 are not significant, and considered within the systematic error of the qPCR analytical process.



Primer concentration (nM)	Melting temperature (°C)	Ct
1000	82.5	16.06
500	82.0	16.17
250	82.0	16.33
125	81.5	16.50
62.5	81.5	16.86

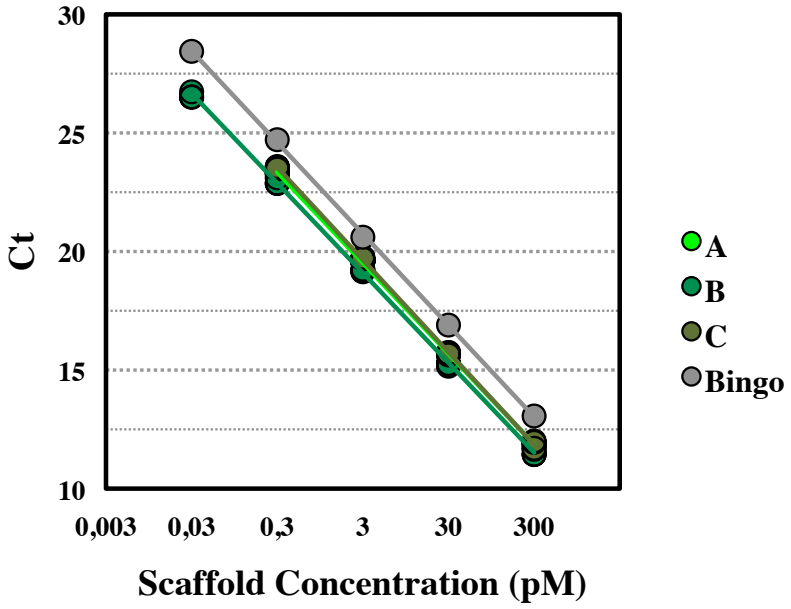
Figure 6.8. Melting curve of qPCR product show one single peak for each primer condition demonstrating that no unspecific products is present at the end of the reaction. The variation in peak height of melting temperature curves is mirrored by variation in Ct values (from 16.1 to 16.9 increasing primer concentration). Nevertheless, variations <1 Ct are not significant, and considered within the systematic error of the qPCR.

6.10.2. qPCR efficiency

Since qPCR quantification is based on the correlation between initial template amounts and the Ct values obtained during amplification, efficiency of qPCR assay is the basic parameter for accurate and reproducible quantification of the sample.

To test assay efficiency, sensitivity and concentration working range, serial dilutions (from 3 nM to 300 fM) of purified nanosensor were prepared and analyzed with site-specific qPCR assay and with Bingo-assay. After qPCR reaction, I plotted Ct versus the logarithm of the input nucleic acid concentration, and performed linear regression analysis.

E% and R^2 values for bingo and site-specific qPCR are reported in figure 6.9.



	A	B	C	BINGO
Efficiency (%)	**82**	**84**	**81**	**83**
Slope	-3.84	-3.79	-3.90	-3.82
R^2	1.00	1.00	1.00	1.00

Figure 6.9. In the plot are reported the Ct values of each qPCR assay for different nanosensor concentrations and the calculated regression linear trends. The table reports the efficiency values obtained from the regression analysis.

6.11. Target Detection

6.11.1. Target-probe binding test

In this section, I confirm target-probe hybridization with a polyacrylamide gel electrophoresis analysis. Each probe was allowed to recognize the target at 1:1 and 1:10 probe:target ratios, and probe concentration of 5 μM (in different tubes). The conditions favor DNA staining in gel with ethidium bromide.

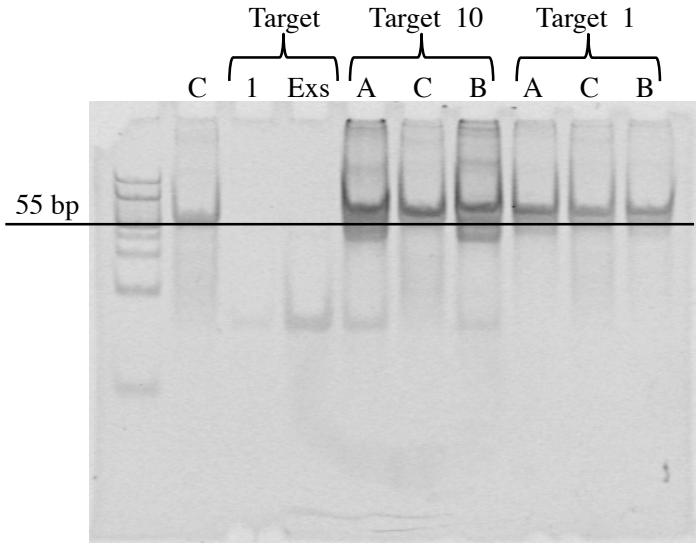
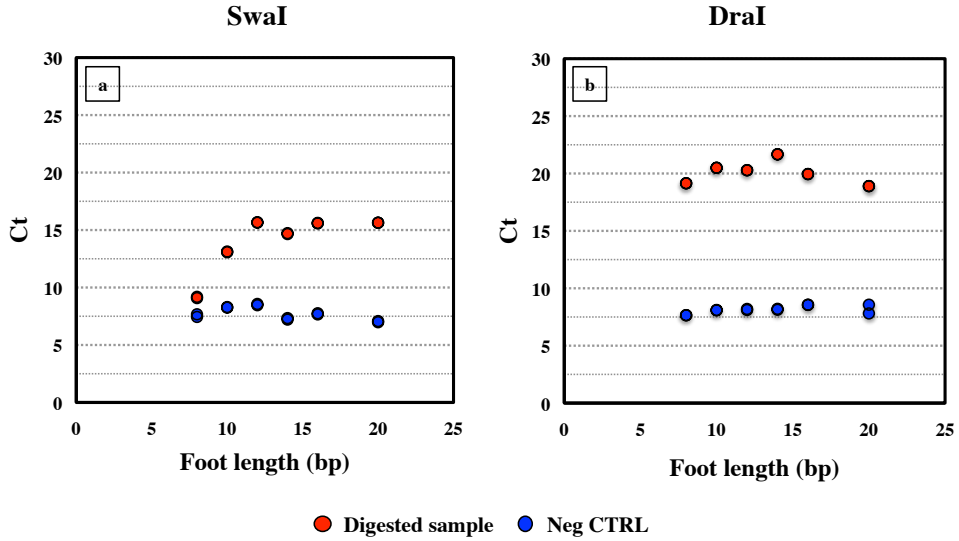


Figure 6.10. 20% polyacrylamide gel analysis of target-probe hybridization products. In this gel, different behaviors after target hybridization are noticed: free probe C runs at lower molecular weight than probe A, B and C hybridized to the target. Moreover, additional bands corresponding to target-probe secondary structures appear after hybridization. It is worth considering that the run of ssDNA- and dsDNA-containing motifs in the gel cannot be easily correlated to their molecular weight. The absence of the target-specific band in sample A, B, C combined with 1-fold diluted target confirms the specific formation of the complex.

In the gel (fig. 6.10) probe C (representative for all the three probes examined) runs at about 55 bp. The dominant band in the presence of the target runs a bit slower indicating that hybridization between target and probe occurs. In addition, the light band associated to the target at 1-fold dilution is not present in the probe-containing samples reflecting a complete hybridization. The presence of secondary structures at lower molecular weight indicates the formation of more compact hybrids.

6.11.2. Target hybridization to the nanosensor

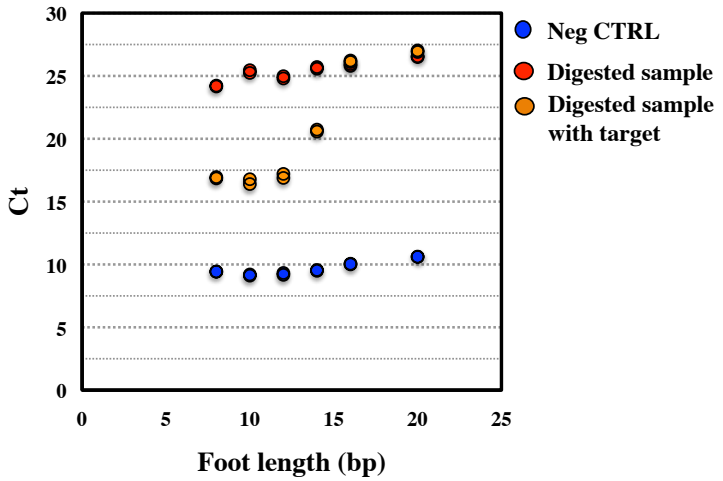
In the following experiment, I tested the nanosensor function as a detection system for short ssDNA strands. The sensor is designed to detect miRNA therefore a DNA sequence corresponding to miRNA25 has been used (RNA is extremely susceptible to degradation for proof-of-concept experiments). We folded different nanosensor versions using probes with decreasing length of the foot carrying the restriction site, i.e. 20, 16, 14, 12, 10, 8 nucleotides. Before the enzymatic reaction the nanosensor was incubated with the target. The first aspect to monitor in this experiment is the folding efficiency of the nanosensor as a function of the decreasing length of the critical foot. Despite the low efficiency of both enzymes used in this reaction (DraI and SwaI), we notice that SwaI is significantly affected by the change in foot length. In particular, SwaI has lower cleavage efficiency if the restriction site is flanked by less than 3 pb at each side. On the contrary, DraI cleavage efficiency does not change upon foot length variations.



Plot 6.10. Digestion profile of the nanosensor with decreasing length of the foot with the restriction site. ***a***. Digestion performed with SwaI shows reduction of cleavage efficiency as the restriction site is flanked by less than 3 pb at each side. ***b***. Digestion performed with DraI showing not dependence upon changes in the critical foot length.

The results of the last experiment performed with DraI are shown in plot 6.11. In this experiment we folded the 6 versions of the sensor with different foot length and we incubated them with the target. We can see that digested sample without target (red points) have comparable Ct values considering the qPCR accuracy limit of Ct +/- 1. The presence of the target in the samples (orange points) influences the enzymatic activity in a foot

length related sigmoid trend. The sensor is not protected from enzymatic degradation if the foot is 20 or 18 nucleotides long (as the probe-scaffold non-binding configuration is too stable for the target to displace the site-carrying probe foot). When the foot length is between 14 and 12 the target can hybridize, leading enzyme activity to c.a. 92 % and 84 % respectively. With foot length of 10 and 8 the enzyme activity is stabilized around 82 % in both cases.



Plot 6.11. DraI digestion profile of different versions of the nanosensor. Decreasing the length of the foot carrying the restriction site favors the hybridization of the target and partial nanosensor protection from enzymatic activity.

Chapter 7 – Discussion

7.1. Quantification of enzymatic restriction reactions in DNA nanostructures

The first attempt at quantitatively evaluating the efficiency of enzymatic restriction reactions in DNA nanostructures using a PCR-based technique in our laboratory was carried out by examining the enzymatically produced DNA scaffold fragmentation, with a scaffold of over 7000 nt. The pour accuracy of PCR amplification in quantifying the stoichiometry of different scaffold fragments in solution, however, oriented our research towards seeking a more quantitative approach that can provide higher accuracy than standard qPCR, but without compromising on the achievable throughput.

As discussed previously, the accuracy limit of qPCR applied to target concentration quantification is substantially of 2-fold concentration variations (103) while by evaluating restriction reaction products with next-generation sequencing (NGS) analysis we could detect concentration variations smaller than 10% between distinct DNA targets (plot 5.4-b). Such targets however, need to be much shorter than the scaffold fragments produced in our experiments. In turn, focusing on the analysis of enzymatic products of DNA staples fragmentation (with DNA staples of 20-40 nt) seems much more promising in terms of accuracy and throughput, and can therefore be applied to characterizing enzymatic reactions in nanostructures for fundamental studies as well as for sensing applications. Nevertheless, producing NGS libraries derived from DNA samples originating from purified DNA nanostructure requires a specifically designed and optimized protocol. In this thesis, I described an adaptation of the L-PCR-based protocol, and demonstrated that it is able to discriminate between cleaved and un-cleaved staple conformations (fig. 5.7, 5.10, 5.11, 5.12). Using such a method, I was able to obtain high quality NGS libraries, and demonstrated the protocol ability to exclude damaged or partially formed DNA duplexes from the active library inputs, which is fundamental for maximizing assay accuracy (fig. 5.22). To test reproducibility and accuracy of our NGS-based DNA quantification protocol

I pooled different dsDNA sequences at known concentrations, to generate two libraries of which one was split into two samples analyzed in different sequencing experiments. The results show that the error associated with the NGS analysis of different replicates of the same sample is lower than 1.5% (plot 5.1). Similarly, the error associated with the NGS analysis of the stoichiometry of DNA staples in the sample is less than 3.5% (plot 5.1). I also designed an external control to calibrate the quantification of DNA samples, focusing on the biases potentially associated with sample manipulation (fig. 5.3). Staples quantification might also be biased by L-PCR uneven amplification therefore, I experimentally obtained L-PCR efficiency for each staple studied. The results show that all the staples were successfully converted into duplex DNAs, but with a staple-dependent amplification coefficient (plot 5.3). The data analysis indicates, however, no evident correlation between the amplification coefficient and factors that typically affect the polymerase-chain amplification processes such as the GC content, the melting temperature or length of DNA staples (see plot 5.3-c). In turn, using this protocol requires that L-PCR efficiency is measured for each target in order to properly calibrate NGS-based results. Staple C58 has the highest L-PCR efficiency, staple A59 provides the lowest (with a value that is about 30% that of C58), and the average L-PCR efficiency across the different staples is about 70% that of C58 (plot 5.3). Finally, I analyzed the negative controls, the hairpin control, and one staple of the digested triangular DNA nanostructure. I found differences in reads % among controls: CTRL-A and Hairpin have higher reads % than CTRL-B and CTRL-C, and also higher variability among samples (plot 5.4). Differences among negative controls might be ascribed to incorporation efficiency or stability within the structure. CTRL-A is placed on an external DNA helix and, therefore, its incorporation is likely to be more efficient compared to staples confined within the inner DNA nanostructure. For each staple the analysis shows a difference in reads % between the two samples ranging from 0.5% to 7.2%. It is possible that staples are incorporated in the structure with certain variability due to sample manipulation, by which CTRL-A (5.2% variation) might be more influenced as it is located on an external helix (79, 80). The staple showing higher variability is the hairpin, but in this case the enzymatic activity might play a role in varying the amount of cleaved staple quantified after two independent enzymatic reactions, showing the ability of this positive control to monitor the efficiency of the enzyme. The highest reproducibility is reached with staple X (A29), which contains a restriction site. The corresponding staple fragment is not transformed by L-PCR and it lacks in the final library, thus demonstrating the complete inability of MspI enzyme to

access the inherent restriction site. The results demonstrate that it is possible to use staples as indicators of restriction enzyme reactions for subsequent use as probes for detecting biomarkers.

7.2. "Bingo-qPCR" self-assembled DNA nanosensor for the accurate quantification of short nucleic acid targets

In this project, I aimed at developing a nanosensor for detecting small variations of RNA amounts. RNA quantification strategies based on reverse transcription-quantitative polymerase chain reaction (RT-qPCR) are central to current genomic analyses (104). RT-qPCR also is the current gold standard for the challenging quantitation of RNA in single cells (16, 105). Major limits of RT-qPCR include the fact that the results can be biased by **(i)** non-linear behavior of reverse transcription **(ii)** the obligatory amplification step that is target-dependent, and **(iii)** (as explained above) the inherent 2-fold threshold for discrimination of target concentration variation (103).

The designed Bingo-qPCR nanosensor is designed to detect short RNA molecules without requiring prior enzymatic conversion of RNA into cDNA (reverse transcription). To do so, it contains k=3 interconnected probes (designed for a miRNA target sequence) formed by the hybridization of three ssDNA probes over consecutive regions of a single ssDNA scaffold sequence. Each probe contains a recognition site for an ancillary restriction enzyme that is disrupted upon probe-target hybridization, and therefore sequestered during the subsequent restriction enzyme reaction. Such a sensing approach reflects a positive biosensing assay, as the higher the concentration of target the higher the number of restriction sites in the scaffold protected from enzymatic cleavage. A PCR-based amplification of the fully protected DNA scaffold (termed "bingo-scaffold") is successful only in the case that all the adjacent probes recognize a copy of the (same) target. Therefore, after the restriction reaction, the relative concentration of integer copies of bingo-scaffold should be proportional to $\boldsymbol{P}=\boldsymbol{p}^3$, where $\boldsymbol{p}$ is proportional to the relative concentration of individual unmodified copies of the restriction site in solution. In terms of PCR quantification, for a Bingo-qPCR nanodevice carrying 3 probes the ΔCt values associated with bingo-scaffold variations should be 3-fold higher than the ΔCt values associated with variations of a bingo-scaffold designed to host a single probe. Similarly, the detection uncertainty associated with 3 probes should be 3-fold lower than with a

single probe. Sequence design and optimization performed with the open-access, nucleic acid computation software Nupack, suggest the presence of small variations in probe-target affinity among the three adjacent probes that might lead to non-negligible alteration in target capturing probability over the probe nanoarray. In this work, I optimized a self-assembling protocol for the nanosensor, and demonstrated its formation with gel electrophoresis analysis (fig. 6.6). As control, I designed mutated probes to form a nanosensor carrying inactive restriction sites, and evaluated the enzymatic digestion products with qPCR in the two nanostructures (plot 6.4, 6.6). I demonstrated that the Bingo-qPCR assay is able to increase ΔCt values with respect to standard qPCR (plot 6.1, 6.3, 6.9), but in the first two experiments I obtained differences in the digestion rate between samples without target (ΔCt(bingo) = 15.4, and 13.7) that might depend on structural instability or on sensor sensitivity for small enzymatic variations caused by the drift of environmental conditions during the reaction (plot 6.8). A weakness of this approach is in the fact that in the likely case that the enzyme efficiency is sub-optimal, a few copies of bingo-scaffold always remain in solution, thus generating a background signal that adds to the target-specific signal. To stabilize and maximize enzymatic restriction efficiency I formed the nanosensor using 10-fold excess of probes, and introduced a thermal equilibration step before performing the enzymatic reaction. In this way the ΔCt values associated with the bingo-scaffold increase and approximately reach the threshold expected by the model (plot 6.9). The correlation between ΔCt values obtained by Bingo-qPCR and standard qPCR also depends on the efficiency of the different assays. The hallmarks of an optimized assay are:

- Linear standard curve with $R^2 > 0.980$. The R^2 value of a standard describes linear correlation between two parameters. Linearity is influenced by variability across replicates, reaction dynamic range and reaction efficiency. If perfect doubling of DNA occurs at every cycle the curves of serial dilutions will be evenly spaced and the regression line will be closer to a linear trend.
- High amplification efficiency (80–110%). Amplification efficiency (E) is calculated from the slope of the standard curve:

  $E = 10^{-1/slope}$ % Efficiency = (E – 1) x 100%

  During an ideal amplification reaction, a 2-fold increase in DNA amount per each amplification cycle is expected, thus corresponding to $E = 2 = 10^{-1/slope}$. Ideally, the optimal slope of the standard curve will be –3.32. An efficiency close to 100% is

the best indicator of a robust, reproducible assay. Low efficiencies, in contrast, may be caused by suboptimal reaction conditions (poor primers design, inefficient thermal cycle). Reaction efficiencies >100% may indicate the presence of unspecific products, such as primer-dimers.

In my experiment, I obtained qPCR efficiencies about 80% (fig. 6.9), meaning that there is no perfect doubling of DNA concentration at each PCR cycle. The consequent amplification bias is equal in the all the different assays, allowing therefore a study of the relationship between Bingo-qPCR ΔCt values and standard-qPCR ΔCt values to test our modelled Bingo-qPCR mechanism. Nevertheless, the quantification error should more evident for low concentration PCR templates as in the case of the bingo-scaffold. In contrast, the kinetics of enzymatic cleavage evidenced that the amount of bingo-scaffold template left in solution as a function of time is moderately higher than expected. The result, indicate therefore that the probability $\boldsymbol{P}$ to find an unmodified bingo-scaffold is higher than $\boldsymbol{p}^3$. Site cleavage on a bingo-scaffold fragment is therefore more likely to occur than on an integer bingo-scaffold. Our interpretation is that DNA accessibility to restriction enzymes is likely to increase when the molecular weight of the Bingo-qPCR nanodevice decreases, as complex secondary structures of the nanodevice become less likely.

After nanosensor characterization, I carried out a proof-of-concept evaluation of its performance in detecting a ssDNA target molecule. Gel electrophoresis results (fig. 6.10) show that the target hybridizes to the probe in absence of the scaffold. In a qPCR experiment we created six different versions of the nanosensor by progressively reducing the length of the foot carrying the restriction site, and each one was incubated with the target in large excess. In the samples lacking the target, I noticed a reduction in the enzymatic activity of SwaI for shorter feet (8-12 nt, plot 6.10-a). This behavior can be explained by a progressive destabilization of the restriction site. Nevertheless, the experiment performed with DraI revealed the ability of the versions with a short foot (8-14 nt, plot 6.11) to sense the target. The results might be biased by the ability of the enzyme used to digest the probe-free scaffold. In fact, the hybridization of the target on the nanosensor might destabilize not only the restriction site, but also the anchor foot causing the removal of the probe from the scaffold. In turn, restriction sites localized in different regions of the single stranded scaffold sequence can hybridize forming duplex segments able to be recognized and digested by the enzyme. In this regard, SwaI has a significantly lower ability to digest the ssDNA scaffold, indicating a possible solution for implementing

an accurate sensor using a restriction-site carrying foot that hybridizes to the scaffold with a sequence of 12-14 nt. Regarding Bingo-qPCR nanodevice thermodynamics, the detection limit calculated by Nupack (pico-femto molar) is indicative because Nupack cannot provide a careful analysis of the apparent equilibrium constants relative to the association of the DNA strands involved (probes, scaffold and target copies). Basically, the sensing element of the Bingo-qPCR nanodevice is a DNA switch that is in equilibrium between two states: one is competent for target binding, while the other is non-competent (106). The sensing mechanism at the basis of DNA switches is an increase in the population of the target-binding conformers in the presence of the target in solution, accompanied by a proportional reduction in the population of non-binding conformers (two-state population shift mechanism) (106). Current experiments are underway in our lab to extend the experiment presented above by systematically varying the length of the restriction-site carrying foot and of its target ssDNA sequence towards achieving a consistent thermodynamic analysis of the nanosensor. In general, however, DNA switches exhibit nanomolar affinities for their nucleic acids targets (61, 106). This aspect does not pose a limitation for the applicability of the Bingo-qPCR nanosensor as its main scope is to provide high detection accuracy rather than high sensitivity. A strategy to overcome the detection limit of DNA switches is to replace the sensing element based on the DNA switch to a small DNA tile comprised of a few tens DNA staples, in which the restriction site in the scaffold sequence is flanked by one or more copies of a linear ssDNA probe that upon hybridization to target molecules, stiffens to a linear duplex that prevents the enzyme from cleaving the restriction site. Such as approach is in line with the results of enzymatic restriction reactions in DNA nanostructures described in Section 2.1 (see also Section 7.1). In this case, the lack of an equilibrium between target-unbound, stem-loop probes and target-bound, linear conformers will reduce the detection limit, as it occurs in DNA microarrays employing linear oligonucleotide probes.

Chapter 8 – Conclusions and future perspectives

Nucleic acid self-assembly and DNA-directed immobilization allow the design and generation of sensing nanoarrays comprised of target-specific probes immobilized over DNA nanostructures serving as water-soluble supports. The ability of such DNA nanostructures to properly interact with biomolecular components in complex biological environments is essential to their successful application in biology and medicine, and is therefore under current, intensive investigation. Studies have been performed to evaluate DNA nanostructures degradation in biological environments, demonstrating an increased stability than circular dsDNA plasmid in cell lysates or short RNA molecules in vivo.

In our laboratory, we specifically investigated the steric-regulated interaction between DNA nanostructures and restriction enzymes. Being able to control restriction reactions on the surface of a DNA Origami opens the door to the development of innovative sensing approaches. Probes can be immobilized on a DNA Origami surface in proximity of a restriction site to permit its protection from the action of an ancillary restriction enzyme upon a probe-target recognition event. In this way, the target-dependent enzymatic cleavage allows registering target-related information on the DNA molecules that are embedded within the DNA Origami support. Such an "DNA imprinting" method could serve to avoid the perturbative sample manipulation associated with the extraction of proteins or nucleic acids from a liquid sample for enabling their detection.

In this thesis, following two sub-projects, I focused on the development of accurate detection approaches for studying biochemical reactions with DNA nanostructures such as DNA hybridization and enzymatic cleavage by exploiting state-of-the-art biomedical instrumentation.

In the first sub-project, I developed a detection approach based on the application of the Illumina Next-Generation Sequencing (NGS) platform using linear PCR to generate reaction-dependent DNA libraries. My results demonstrate the power of this approach to quantify small differences in incorporation efficiency of short (approx. 30 nt) DNA

molecules (with an uncertainty of less than 10%) within a DNA nanostructure comprised of over 7000 base pairs. The protocol developed also highlighted the weakness of the nucleic acid quantification procedures based on DNA amplification or retro-transcription reactions. The lack of correlation between the measured changes of linear PCR efficiencies among over 20 DNA molecules and their thermodynamic-related features such GC % content, melting temperature and length, highlight the limits of NGS analysis for nucleic acids quantification. By optimizing and applying a suitable renormalization approach, I focused on the analysis of restriction reactions with the Rothemund's DNA triangle. The results demonstrate the ability of our detection approach to quantify the efficiency of cleavage of restriction sites within such nanostructure by focusing on two sites in particular, one of which resulted to be highly cleaved, and the other fully protected. The results therefore, provide a proof-of-concept that the proposed NGS-based analysis of DNA nanostructures can be optimized for the highly parallel quantification of enzymatic reactions under nanoscale confinement.

In the second sub-project, I developed a proof-of-concept approach for the amplification-free and retro-transcription-free quantification of short RNAs (such as miRNAs) termed "Bingo-qPCR" assay. The approach relies on the function of a small nanodevice comprised of approx. 500 nucleotides, designed to detect small concentration variations of a target sequence using standard qPCR instrumentation. The sensor is formed by three probes containing a restriction site that unfolds upon target binding. In target depletion regime only a few nanosensor copies in solution bind three target molecules, and being therefore fully protected from the subsequent action of an ancillary restriction enzyme (bingo-scaffold). The results confirm our detection model, for which target and bingo-scaffold concentration are correlated by a cubic function, thus converting a 2-fold variation of target concentration into a 8-fold variation of bingo-scaffold concentration. In addition, I applied the Bingo-qPCR nanosensor to quantify restriction enzyme efficiencies reaching an accuracy near 15% without serial dilutions.

Combined together, the results obtained in this thesis set the basis for successfully combining high-throughput, low-cost technologies for DNA quantification to the investigation of biochemical reactions with DNA molecules confined within highly packed nanostructures, and their application to implement innovative detection approaches, with which biomarker abundance is measured using DNA barcodes.

Appendix I

I.1. Staples sequences for the triangular DNA nanostructure

In bold the sequences containing the restriction site for MspI and those used as positive and negative controls.

A1 cggggtttcctcaagagaaggattttgaatta
A2 agcgtcatgtctctgaatttaccgactacctt
A3 ttcataatccccttattagcgtttttcttacc
A4 atggtttatgtcacaatcaatagatattaaac
A5 tttgatgattaagaggctgagacttgctcagtaccaggcg
A6 ccggaacccagaatggaaagcgcaacatggct
A7 aaagacaacattttcggtcatagccaaaatca
A8 gacgggagaattaactcggaataagtttatttccagcgct
A9 gataagtgccgtcgagctgaaacatgaaagtatacaggag
A10 tgtactggggatcttcattaaagcagagccac
A10 tgtactggggatcttcattaaagcagagccacttcagtatccggtcaact (dsDNA CTRL)
A10-Rev agttgaccggatactg (dsDNA CTRL complementary sequence)
A11 caccggaaagcgcgttttcatcggaagggcga
A12 cattcaacaaacgcaaagacaccagaacaccctgaacaaa
A13 tttaacggttcggaacctattattagggttgatataagta
A14 ctcagagcatattcacaaacgaattaataagt
A15 ggagggaatttagcgtcagactgtccgcctcc
A16 gtcagagggtaattgatggcaacatataaaagcgattgag
A17 tagcccggaataggtgaatgccccctgcctatggtcagtg
A18 ccttgagtcagacgattggccttgcgccaccc
A19 tcagaacccagaatcaagtttgccggtaaata
A20 ttgacggaaatacatacataaagggcgctaatatcagaga
A21 cagagccaggaggttgaggcaggtaacagtgcccg
A22 attaaaggccgtaatcagtagcgagccaccct
A23 gataacccacaagaatgttagcaaacgtagaaaattattc
A24 gccgccagcattgacaccaccctc
A25 agagccgcaccatcgatagcagcatgaattat
A26 caccgtcaccttattacgcagtattgagttaagcccaata
A27 agccatttaaacgtcaccaatgaacaccagaacca
A28 ataagagcaagaaacatggcatgattaagactccgacttg
A29 ccattagcaaggccgggggaatta
A30 gagccagcgaatacccaaaagaacatgaaatagcaatagc
A31 tatcttaccgaagcccaaacgcaataataacgaaaatcaccag
A32 cagaaggaaaccgaggtttttaagaaaagtaagcagatagccg
A33 cctttttttcatttaacaatttcataggattag (CTRL)
A34 tttaacctatcataggtctgagagttccagta

A35 agtataaaatatgcgttatacaaagccatctt
A36 caagtacctcattccaagaacgggaaattcat
A37 agagaataacataaaaacagggaagcgcatta
A38 aaaacaaaattaattaaatggaaacagtacattagtgaat
A39 ttatcaaaccggcttaggttgggtaagcctgt
A40 ttagtatcgccaacgctcaacagtcggctgtc
A41 tttccttagcactcatcgagaacaatagcagcctttacag
A42 agagtcaaaaatcaatatatgtgatgaaacaaacatcaag
A43 actagaaatatataactatatgtacgctgaga
A44 tcaataatagggcttaattgagaatcataatt
A45 aacgtcaaaaatgaaaagcaagccgtttttatgaaaccaa
A46 gagcaaaagaagatgagtgaataaccttgcttatagctta
A47 gattaagaaatgctgatgcaaatcagaataaa
A48 caccggaatcgccatatttaacaaaatttacg
A49 agcatgtatttcatcgtaggaatcaaacgatttttgttt
A50 acatagcgctgtaaatcgtcgctattcatttcaattacct
A51 gttaaatacaatcgcaagacaaagccttgaaa
A52 cccatcctcgccaacatgtaatttaataaggc
A53 tcccaatccaaataagattaccgcgcccaataaataatat
A54 tcccttagaataacgcgagaaaacttttaccgacc
A55 gtgtgataaggcagaggcattttcagtcctga
A56 acaagaaagcaagcaaatcagataacagccatattattta
A57 gtttgaaattcaaatatattttag
A58 aatagatagagccagtaataagagatttaatg
A59 gccagttacaaaataatagaaggcttatccggttatcaac
A60 ttctgacctaaaatataaagtaccgactgcagaac
A61 gcgcctgttattctaagaacgcgattccagagcctaattt
A62 tcagctaaaaaaggtaaagtaatt
A63 acgctaacgagcgtctggcgttttagcgaacccaacatgt
A64 acgacaataaatcccgacttgcgggagatcctgaatcttacca
A65 tgctattttgcacccagctacaattttgttttgaagccttaaa
B1 tcatatgtgtaatcgtaaaactagtcattttc
B2 gtgagaaaatgtgtaggtaaagatacaacttt
B3 ggcatcaaatttggggcgcgagctgagttaaa
B4 ttcgagctaagacttcaaatatcgcggaacga
B5 acagtcaaagagaatcgatgaacgaccccggttgataatc
B6 atagtagtatgcaatgcctgagtaggccggag
B7 aaccagacgtttagctatattttcttctacta
B8 gaataccacattcaacttaagaggaagcccgatcaaagcg
B9 agaaaagccccaaaaagagtctggagcaaacaatcaccat
B10 caatatgaccctcatatattttaaagcattaa
B11 catccaataaatggtcaataacctcggaagca
B12 aactccaagattgcatcaaaaagataatgcagatacataa
B13 cgttctagtcaggtcattgcctgacaggaagattgtataa
B14 caggcaagataaaaatttttagaatattcaac
B15 gattagagattagatacatttcgcaaatcata
B16 cgccaaaaggaattacagtcagaagcaaagcgcaggtcag
B17 gcaaatatttaaattgagatctacaaaggctactgataaa
B18 ttaatgccttatttcaacgcaagggcaaagaa
B19 ttagcaaatagatttagtttgaccagtacctt

B20 taattgctttaccctgactattatgaggcatagtaagagc
B21 ataaagcctttgcgggagaagcctggagagggtag
B22 taagaggtcaattctgcgaacgagattaagca
B23 aacactatcataacccatcaaaaatcaggtctccttttga
B24 atgaccctgtaatacttcagagca
B25 taaagctatgtaacagttgattcccatttttg
B26 cggatggcacgagaatgaccataatcgtttaccagacgac
B27 taattgcttggaagtttcattccaaatcggttgta
B28 gataaaaaccaaaatattaaacagttcagaaattagagct
B29 actaaagtacggtgtcgaatctgg
B30 tgctgtagatccccctcaaatgctgcgagaggcttttgca
B31 aaagaagttttgccagcataaatattcattgactcaacatgtt
B32 aatactgcggaatcgtaggggtaatagtaaaatgtttagact
B33 agggatagctcagagccaccaccccatgtcaa
B34 caacagtttatggggttttgctaatcaaaagg
B35 ggccgcttcgctgaggcttgcagggaaaaggt
B36 ggcgcagactccatgttacttagccgttttaa
B37 acaggtagaaagattcatcagttgagatttag
B38 cctcagaaccgccacccaagcccaataggaacgtaaatga
B39 attttctgtcagcggagtgagaataaccgata
B40 tattcggtttgcgggatcgtcacccgaaatcc
B41 gcgacctgcggtcaatcataagggacggaacaacattatt
B42 agacgttaccatgtaccgtaacacccctcagaaccgccac
B43 ccacgcatagaaaggaacaactaagtctttcc
B44 aattgtgtctcagcagcgaaagacaccatcgc (CTRL)
B45 ttaataaaacgaactaaaccgaactgaccaacgcctgata
B46 aggtttagtaccgccatgagtttcgtcaccaggatctaaa
B47 gttttgtcaggaattgcgaataatgccgacaa
B48 tgacaacaagcatcggaacgaggggagattt
B49 gtatcatctttgaaagaggacagaggaagaaaaatctacg
B50 agcgtaactacaaactacaacgcctatcaccgtactcagg
B51 atagttgcaatttttcacgttgatcatagtt
B52 agtacaactagcaacggctacagatgataccg
B53 accagtcaggacgttgtgaacggtgtacagaccgaaacaa
B54 acagacaacccaaatctccaaaaaaaaaatttctt
B55 aaacagctggctttgaggactaaagcgattat
B56 accaagcgcaggcgcataggctggagaactggctcattat
B57 tcgaggtgaggctccaaaaggagc
B58 gacccccagacttttcatgaggagcttgctt
B59 accttatgcgattttactgaccttcatcaagatcatcttt
B60 tcggtttatcaagtttccattaaacgggaatacac
B61 taaaacacgtaatcttgacaagaattaatcattgtgaatt
B62 aggcgaaagtaaaatacgtaatgc
B63 tggtttaatttcaactc<u>cggatattcattaccaacgaaag</u>
B64 caccaacctaacaaatcaacgtaacaataaattgggcttgaga
B65 ccctgacgagaacaccagaacgagtagagctgctcattcagtg
C1 atcgggagatatacagtaacagtacaaataatt
C2 cctgattaaaggagcggaattatctcggcctc
C3 ggcaaatcacctcaatcaatatctgcaggtcga
C4 cgaccagtacattggcagattcacctgattgc

C5 tggcaatttttaacgtcagatgaaaacaataacggattcg
C6 aaggaattacaaagaaaccaccagtcagatga
C7 ggacattcacctcaaatatcaaacacagtaga
C8 ttgacgagcacgtatactgaaatggattatttaataaaag
C9 cctgattgctttgaattgcgtagattttcaggcatcaata
C10 taatcctgattatcattttgcggagaggaagg
C10-(hairpin)
taatcctgattatcattttgcggagaggaaggttcagtatccggtcaactgttgtatgagttgaccggatactg
C11 ttatctaaagcatcaccttgctgatggccaac
C12 agagatagtttgacgctcaatcgtacgtgctttcctcgtt
C13 gattatacacagaaataaagaaataccaagttacaaaatc
C14 taggtgcataaaagtttgagtaacattgtttg
C15 tgacctgacaaatgaaaaatctaaaatatctt
C16 ggaatcagagcgggagatggaaatacctacataacccttc
C17 gcgcagaggcgaattaattattagcacgtaaattctgaat
C18 tatggaagcgaacgttattaatttctaacaac
C19 taatagatcgctgagagccagcagaagcgtaa
C20 gaatacgtaacaggaaaaacgctcctaaacaggaggccga
C21 tcaatagatattaaatcctttgccgaattgaacca
C22 caatatttgcctgcaacagtgccatagagccg
C23 ttaaagggattttagataccgccagccattgcggcacaga
C24 acaattcgacaactcgtaatacat
C25 ttgaggatggtcagtattaacaccttgaatgg
C26 ctattagtatatccagaacaatatcaggaacggtacgcca
C27 cgcgaactaaaacagaggtgaggcttagaagtatt (CTRL)
C28 gaatcttgagaagtgtatcggccttgctggtactttaatg
C29 accaccagcagaagatgatagcc
C30 ctaaaacattagaagaactcaaacttttataatcagtgag
C31 gccaccgagtaaaagaacatcacttgcctgagcgccattaaaa
C32 tctttgattagtaatagtctgtccatcacgcaaattaaccgtt
C33 cgcgtctgataggaacgccatcaacttttac
C34 aggaagatggggacgacgaccgtaatcatatt
C35 ctctagagcaagcttgcatgcctggtcagtt
C36 ccttcaccgcgagacgggcaacagcagtcaca
C37 cgagaaaggaagggaagcgtactatggttgct
C38 gctcatttttttaaccagccttcctgtagccaggcatctgc
C39 cagtttgacgcactccagccagctaaacgacg
C40 gccagtgcgatcccc<u>gggtaccgagtttttct</u>
C41 <u>tttcaccagcctggccctgagagaaaagc</u>cggcgaacgtgg
C42 gtaaccgtctttcatcaacattaaaatttttgttaaatca
C43 acgttgtattc<u>cggcaccgcttctggcgcatc</u>
C44 ccagggtggctcgaattcgtaatccagtcacg
C45 tagagcttgacggggagttgcagcaagcggtcattgggcg
C46 gttaaaattcgcgttaatgtgagcgagtaacatacgttgg
C47 tgtagatgggtgc<u>cggaaaccaggaacgccag</u>
C48 ggttttccatggtcatagctgtttgagaggcg
C49 gtttgcgtcacgctggtttgccccaagggagcccccgatt
C50 ggataggtacccgtcggattctcctaaacgttaatatttt
C51 agttgggtcaaagcgccattcgccccgtaatg
C52 cgcgcgggcctgtgtgaaattgttggcgatta

C53 ctaaatcggaaccctaagcaggcgaaaatccttcggccaa
C54 cggcggattgaattcaggctgcgcaacgggggatg
C55 tgctgcaaatccgctcacaattcccagctgca
C56 ttaatgaagtttgatggtggttccgaggtgccgtaaagca
C57 tggcgaaatgttgggaagggcgat
C58 tgtcgtgcacacaacatacgagccacgccagc
C59 caagttttttggggtcgaaatcggcaaaatccgggaaacc
C60 tcttcgctattggaagcataaagtgtatgcccgct
C61 ttccagtccttataaatcaaaagagaaccatcacccaaat
C62 gcgctcacaagcctggggtgccta
C63 cgatggcccactacgtatagcccgagatagggattgcgtt
C64 aactcacattattgagtgttgttccagaaaccgtctatcaggg
C65 acgtggactccaacgtcaaagggcgaatttggaacaagagtcc
L-B1A tgtagcattccttttataaacagtt
L-B2A ctttaattgtattccaccagagcc
L-B3A cactacgaaggttagcaccatta
L-B4A aataaggcttgaacaaagttac
L-A1C ttaattaatttttttaccatatcaaa
L-A2C ttaatttcatcttagactttacaa
L-A3C ctgtccagacgtataccgaacga
L-A4C tcaagattagtgtagcaatact
L-C1B gtgggaacaaatttctatttttgag
L-C2B cggtgcgggccttccaaaaacatt
L-C3B atgagtgagcttttaaatatgca
L-C4B actattaaagaggatagcgtcc

I.2. Carrier sequences for the staples of triangular DNA nanostructure containing restriction site of MspI

C_C41	ggtagattctggtaattgagttcactaagagttggaaagtaagctttctctcagggccaggctggtgaaa
C_C58	ggtagattctctcatatgagttcactaagagttggaggcttaactaagctcgtatgttgtgtgcacgaca
C_C40	ggtagattctaccgtttgagttcactaagagttggagggttcactaagagttagaaaaactcggtacccg
C_C47	ggtagattctttaggttgagttcactaagagttggagggttcactaagagtctggcgttcctggtttccg
C_C43	ggtagattcttatcactgagttcactaagagttggagggttcactaagagatgcgccagaagcggtgccg
C_B5	ggtagattctcaacggtgagttcactaagacttggagggttcgggtcgttcatcgattctctttgactgt
C_B18	ggtagattctcttcaatgagttcactaagagttggagggttcactttctttgcccttgcgttgaaataag
C_B6	ggtagattctgaatcctgagttcactaagagttggagtgataagcctactcaggcattgcatactactat
C_B11	ggtagattctgtgatatgagttcactaagagttggagggtcgaataaggttattgaccatttattggatg
C_B63	ggtagattctatacgctgagttcactaagagttggagggttcactaactttcgttggtaatgaatatccg
C_B4	ggtagattcttttaaatgagttcactaagagttggagggttacatacgatatttgaagtcttagctcgaa
C_A17	ggtagattctcggtaatgagttcactaagagttggcactgaccataggcaggggcattcacctattccg
C_A11	ggtagattctaagccttgagttcactaagagttggagggtttcgcccttccgatgaaaacgcgctttccg
C_A6	ggtagattcttgtagatgagttcactaagagttggagggagccatgttgcgctttccattctgggttccg
C_A29	ggtagattctagactatgagttcactaagagttggagggttcactacgagtaggattgccttgctaatgg
C_A59	ggtagattctctagtatgagttcactaagacatggagggttgataagccttctattattttgtaactggc
C_A48	ggtagattctaatccttgagttcactaagagttggagggttcgtaaattttgttaaatatggcgattccg
C_A39	ggtagattcttagtagtgagttcactaagagttggagggttcactaaacaggcttacccaacctaagccg
C_A33	ggtagattctgcctaatgagttcactaagagttggaggctaatcctatgaaattgttaaatgaaaaaagg
C_B44	ggtagattcttccgaatgagttcactaagagttggagggcgatggtgtctttcgctgctgagacacaatt

C_C27 ggtagattctttccaatgagttcactaagagttggaatacttctaagcctcacctctgttttagttcgcg
C_C10-hairpin ggtagattctgcgtattgagatcactatagatactgaaccttcctctccgcaaaatgataatcaggatta
C_A10-dsCTRL ggtagattctgcgtattgagatcactatagatactgaagtggctctgctttaatgaagatccccagtaca

I.4. Scaffold sequence (M13mp18)

TTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA
CTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTT
CGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACT
GGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTG
CCGATTTCGGAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGC
GTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCT
GTTGCCCGTCTCGCTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAA
CCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTT
CCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAC
TCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGG
AATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACG
AATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAG
CTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTA
CCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAA
TGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGA
GTGCGATCTTCCTGAGGCCGATACGGTCGTCGTCCCCTCAAACTGGCAGATGC
ACGGTTACGATGCGCCCATCTACACCAACGTAACCTATCCCATTACGGTCAAT
CCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAAT
GTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGT
TCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAA
CAAAATATTAACGTTTACAATTTAAATATTTGCTTATACAATCTTCCTGTTTTTG
GGGCTTTTCTGATTATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGA
TTACCGTTCATCGATTCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTGATA
GCCTTTGTAGATCTCTCAAAAATAGCTACCCTCTCCGGCATTAATTTATCAGCT
AGAACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTCTCAC
CCTTTTGAATCTTTACCTACACATTACTCAGGCATTGCATTTAAAATATATGAG
GGTTCTAAAAATTTTTATCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTA
TTACAGGGTCATAATGTTTTTGGTACAACCGATTTAGCTTTATGCTCTGAGGCT
TTATTGCTTAATTTTGCTAATTCTTTGCCTTGCCTGTATGATTTATTGGATGTTA
ATGCTACTACTATTAGTAGAATTGATGCCACCTTTTCAGCTCGCGCCCCAAATG
AAAATATAGCTAAACAGGTTATTGACCATTTGCGAAATGTATCTAATGGTCAA
ACTAAATCTACTCGTTCGCAGAATTGGGAATCAACTGTTACATGGAATGAAAC
TTCCAGACACCGTACTTTAGTTGCATATTTAAAACATGTTGAGCTACAGCACCA
GATTCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGG
AGCAATTAAAGGTACTCTCTAATCCTGACCTGTTGGAGTTTGCTTCCGGTCTGG
TTCGCTTTGAAGCTCGAATTAAAACGCGATATTTGAAGTCTTTCGGGCTTCCTC
TTAATCTTTTTGATGCAATCCGCTTTGCTTCTGACTATAATAGTCAGGGTAAAG
ACCTGATTTTTGATTTATGGTCATTCTCGTTTTCTGAACTGTTTAAAGCATTTGA
GGGGGATTCAATGAATATTTATGACGATTCCGCAGTATTGGACGCTATCCAGT
CTAAACATTTTACTATTACCCCCTCTGGCAAAACTTCTTTTGCAAAAGCCTCTC
GCTATTTTGGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTC

TTACTATGCCTCGTAATTCCTTTTGGCGTTATGTATCTGCATTAGTTGAATGTGG
TATTCCTAAATCTCAACTGATGAATCTTTCTACCTGTAATAATGTTGTTCCGTTA
GTTCGTTTTATTAACGTAGATTTTTCTTCCCAACGTCCTGACTGGTATAATGAG
CCAGTTCTTAAAATCGCATAAGGTAATTCACAATGATTAAAGTTGAAATTAAA
CCATCTCAAGCCCAATTTACTACTCGTTCTGGTGTTCTCGTCAGGGCAAGCCTT
ATTCACTGAATGAGCAGCTTTGTTACGTTGATTTGGGTAATGAATATCCGGTTC
TTGTCAAGATTACTCTTGATGAAGGTCAGCCAGCCTATGCGCCTGGTCTGTACA
CCGTTCATCTGTCCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTGACCG
TCTGCGCCTCGTTCCGGCTAAGTAACATGGAGCAGGTCGCGGATTTCGACACA
ATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGTTTCGCGCTTGGTATA
ATCGCTGGGGGTCAAAGATGAGTGTTTTAGTGTATTCTTTCGCCTCTTTCGTTTT
AGGTTGGTGCCTTCGTAGTGGCATTACGTATTTTACCCGTTTAATGGAAACTTC
CTCATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGTT
CCGATGCTGTCTTTCGCTGCTGAGGGTGACGATCCCGCAAAAGCGGCCTTTAA
CTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTG
TTGTCATTGTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAAATTCACCTCGA
AAGCAAGCTGATAAACCGATACAATTAAAGGCTCCTTTTGGAGCCTTTTTTTTT
GGAGATTTTCAACGTGAAAAAATTATTATTCGCAATTCCTTTAGTTGTTCCTTT
CTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAACCCCATACAGA
AAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTTACGCTA
ACTATGAGGGTTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGAC
GAAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAAT
GAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGG
CGGTACTAAACCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATAT
CAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATC
CTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTTTCAGAATA
ATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACT
CAAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAAA
AGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATT
CTGGCTTTAATGAAGATCCATTCGTTTGTGAATATCAAGGCCAATCGTCTGACC
TGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTCTGGTGGCG
GCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAG
GGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAGAT
GGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTA
CAGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCT
ATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACT
GGTGATTTTGCTGGCTCTAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAAT
TCACCTTTAATGAATAATTTCCGTCAATATTTACCTTCCCTCCCTCAATCGGTTG
AATGTCGCCCTTTTGTCTTTAGCGCTGGTAAACCATATGAATTTTCTATTGATT
GTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATATGTTGCCA
CCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTT
AATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTTTCCTT
CTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAG
ATAGCTATTGCTATTTCATTGTTTCTTGCTCTTATTATTGGGCTTAACTCAATTC
TTGTGGGTTATCTCTCTGATATTAGCGCTCAATTACCCTCTGACTTTGTTCAGG
GTGTTCAGTTAATTCTCCCGTCTAATGCGCTTCCCTGTTTTTATGTTATTCTCTC
TGTAAAGGCTGCTATTTTCATTTTTGACGTTAAACAAAAAATCGTTTCTTATTT
GGATTGGGATAAATAATATGGCTGTTTATTTTGTAACTGGCAAATTAGGCTCTG
GAAAGACGCTCGTTAGCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGGTGC
AAAATAGCAACTAATCTTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAG

GTTCGCTAAAACGCCTCGCGTTCTTAGAATACCGGATAAGCCTTCTATATCTGA
TTTGCTTGCTATTGGGCGCGGTAATGATTCCTACGATGAAAATAAAAACGGCT
TGCTTGTTCTCGATGAGTGCGGTACTTGGTTTAATACCCGTTCTTGGAATGATA
AGGAAAGACAGCCGATTATTGATTGGTTTCTACATGCTCGTAAATTAGGATGG
GATATTATTTTTCTTGTTCAGGACTTATCTATTGTTGATAAACAGGCGCGTTCT
GCATTAGCTGAACATGTTGTTTATTGTCGTCGTCTGGACAGAATTACTTTACCT
TTTGTCGGTACTTTATATTCTCTTATTACTGGCTCGAAAATGCCTCTGCCTAAAT
TACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGC
GTTGGCTTTATACTGGTAAGAATTTGTATAACGCATATGATACTAAACAGGCTT
TTTCTAGTAATTATGATTCCGGTGTTTATTCTTATTTAACGCCTTATTTATCACA
CGGTCGGTATTTCAAACCATTAAATTTAGGTCAGAAGATGAAATTAACTAAAA
TATATTTGAAAAAGTTTTCTCGCGTTCTTTGTCTTGCGATTGGATTTGCATCAGC
ATTTACATATAGTTATATAACCCAACCTAAGCCGGAGGTTAAAAAGGTAGTCT
CTCAGACCTATGATTTTGATAAATTCACTATTGACTCTTCTCAGCGTCTTAATCT
AAGCTATCGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACG
ATTTACAGAAGCAAGGTTATTCACTCACATATATTGATTTATGTACTGTTTCCA
TTAAAAAAGGTAATTCAAATGAAATTGTTAAATGTAATTAATTTTGTTTTCTTG
ATGTTTGTTTCATCATCTTCTTTTGCTCAGGTAATTGAAATGAATAATTCGCCTC
TGCGCGATTTTGTAACTTGGTATTCAAAGCAATCAGGCGAATCCGTTATTGTTT
CTCCCGATGTAAAAGGTACTGTTACTGTATATTCATCTGACGTTAAACCTGAAA
ATCTACGCAATTTCTTTATTTCTGTTTTACGTGCTAATAATTTTGATATGGTTGG
TTCAATTCCTTCCATAATTCAGAAGTATAATCCAAACAATCAGGATTATATTGA
TGAATTGCCATCATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGG
TGGTTTCTTTGTTCCGCAAAATGATAATGTTACTCAAACTTTTAAAATTAATAA
CGTTCGGGCAAAGGATTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATA
CTTCTAAATCCTCAAATGTATTATCTATTGACGGCTCTAATCTATTAGTTGTTA
GTGCACCTAAAGATATTTTAGATAACCTTCCTCAATTCCTTTCTACTGTTGATTT
GCCAACTGACCAGATATTGATTGAGGGTTTGATATTTGAGGTTCAGCAAGGTG
ATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACTGTTGCAGGCG
GTGTTAATACTGACCGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTTCGG
TATTTTTAATGGCGATGTTTTAGGGCTATCAGTTCGCGCATTAAAGACTAATAG
CCATTCAAAAATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGAAGGG
TTCTATCTCTGTTGGCCAGAATGTCCCTTTTATTACTGGTCGTGTGACTGGTGA
ATCTGCCAATGTAAATAATCCATTTCAGACGATTGAGCGTCAAAATGTAGGTA
TTTCCATGAGCGTTTTTCCTGTTGCAATGGCTGGCGGTAATATTGTTCTGGATA
TTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTA
CTAATCAAAGAAGTATTGCTACAACGGTTAATTTGCGTGATGGACAGACTCTT
TTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAAGATTCTGGCGTACCG
TTCCTGTCTAAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCCA
ACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTA
CACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTC

I.3. Carrier design

	Self dimer	Hairpin	Cleaved staple unspecific binding	Structural staples	Scaffold	Other carriers	Other staple with restriction site	Other cleaved staple with restriction site
Carrier name	Melting temperature (°C)							
C_A33		46.3	<10		34.9	<10	<10	<10
C_B44		46.9	<10		39.3	<10	10.7	10.7
C_C27	<10	41.8	<10		<10	/	<10	<10
C_A29		42.2	<10		8	27.6	/	<10
C_C10_Pos 2.0		30.3	<10		10	10.6	/	/
C_C41		40.3	<10	<10	23.1	10.7	<10	<10
C_C58	18.1	42.4	<10		30.2	14.7	<10	<10
C_C40	<34.5	45.9	<10		11.2	31.2	/	/
C_C47		29.7	<10	<10	48.6		20.3	/
C_C43	<18	44.5	<10		30.8	<10	/	/
C_B5		43.4	<10		24.9	<10	/	/
C_B18		37.9	<10		21.7		13.9	13.9
C_B6		44.6	<10		37.6	10.6	32.4	32.4
C_B11		41.7	<10		26.9		<10	<10
C_B63		30.6	<10		34.1		/	/
C_B4		36.6	<10		<10	<10	<10	<10
C_A17		42.9	<10		<10		<10	<10
C_A11		50.2	<10		42.6		24.9	24.9
C_A6		32.7	<10		53.2		32.6	32.6
C_A59	<10	39.2	<10		29.5	<10	<10	<10
C_A48		29.1	<10		26.9	<10	<10	<10
C_A39	<10	44.3	<10		17.4	32.5	<10	<10

Table I.1.
Cleaved staple un-specific: incorrect hybridization of the clevaed staple on the specific carrier.
Structural staples: all the staples forming the triangular DNA nanostructure that don't contain a restriction site

I.4. Gel Electrophoresis

Figure 5.1.
Acrylamyde: 16%
Sample: 25 μL
Purified DNA nanostructure: 10 μL
Loading buffer: 6 μL
Glycerol: 5 μL
10 bp ladder: 1,5 μL

Figure 5.5.
Acrylamyde: 16%
Sample: 25 μL
Purified hairpin folded at 1,5 μM: 0,6 μL
Carrier 1 μM: 4,5 μL
Loading buffer: 6 μL
Glycerol: 5 μL
10 bp ladder: 1,5 μL

Figure 5.6.
Acrylamyde: 16%
Sample: 50 μL
Loading buffer (20% glycerol): 15 μL
10 bp ladder: 2,5 μL

Figure 5.7.
Acrylamyde: 16%
Sample: 12,5 μL
Staple 1,5 μM: 6,3 μL
Carrier 5 μM: 1,9 μL
Loading buffer (20% glycerol): 6 μL
10 bp ladder: 1,5 μL

Figure 5.8.
Acrylamyde: 20%
Sample: 2,8 μL
10 bp ladder: 3 μL

Figure 5.9.
Acrylamyde: 16%
Sample: 25 μL
Purified Origami (CTRL): 10 μL
Carrier pool 750 nM: 1 μL
Loading buffer 6X: 6 μL
Glycerol: 5 μL
10 bp ladder: 2 μL

Figure 5.10.
Acrylamyde: 20%
Sample: 25 μL
Carrier 5 μM: 0,9 μL
Hairpin 150 nM: 6 μL
Loading buffer (20% glycerol): 11 μL
10 bp ladder: 1,5 μL

Figure 5.11.
Acrylamyde: 20%
Sample: 15 μL
Loading buffer (20% glycerol): 16 μL
10 bp ladder: 1,5 μL

Figure 5.12.
Acrylamyde: 16%
Sample: 25 μL
Loading buffer (20% glycerol): 11 μL
10 bp ladder: 1,5 μL

Figure 5.13.
Acrylamyde: 16%
Sample: 40 μL
CTRL 100%: 25 μL
CTRL 50%: 12,5 μL
CTRL 25%: 6,3 μL
CTRL 12,5%: 3,2 μL
Loading buffer (20% glycerol): 16 μL
10 bp ladder: 2,5 μL

Figure 5.14.
Acrylamyde: 20%
Sample: 50 μL
Loading buffer (20% glycerol): 16 μL
10 bp ladder: 2,5 μL

Figure 5.15.
Acrylamyde: 16%
Sample: 40 μL
CTRL 100%: 25 μL
CTRL 50%: 12,5 μL
CTRL 25%: 6,3 μL
CTRL 12,5%: 3,2 μL
Loading buffer (20% glycerol): 16 μL
10 bp ladder: 2,5 μL

Figure 5.16.
Acrylamyde: 20%

Sample: 1 μL
CTRL: 1 μL
Loading buffer (20% glycerol): 2 μL
10 bp ladder: 2,5 μL

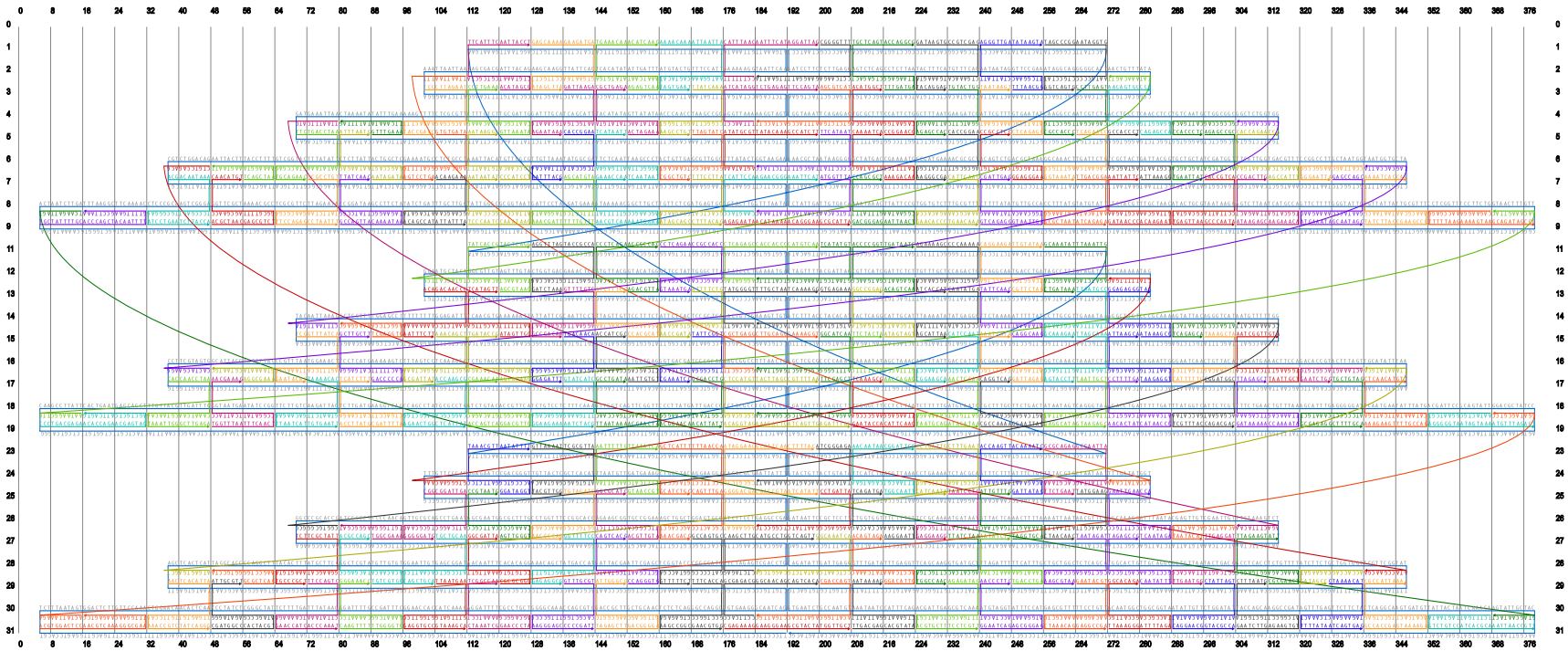
Figure 5.17.
Acrylamyde: 16%
Sample: 50 ng
Loading buffer (20% glycerol): 2 μL
10 bp ladder: 5 μL

Figure 5.20.
Acrylamyde: 20%
Sample: 50 μL
CTRL: 50 μL
Loading buffer (20% glycerol): 16 μL
10 bp ladder: 2,5 μL

I.5. Folding and L-PCR conditions

	Section	Oligo	Concentration
folding	5.3.2.	Hairpin	90 nM
	5.3.4.	ds CTRL	90 nM
L-PCR	5.4.1.	Staple	150 nM
	5.4.2.	Staple	116 nM
	5.4.3.	Staple	36 nM
	5.5.1.	Staple	21.4 nM (carrier 6 fold)
	5.5.2.	Triangle	2 nM (carrier 6 fold)

We used Cadnano software downloaded from http://cadnano.org, to generate and reproduce the design of the triangular DNA origami published by Rothemund P.W.K. (65). The long blue line corresponds to the scaffold while colored short lines are the staples.



Appendix II

II.1. Sequences

Probes MiR21

Probe A	gaattaatttaaataaattatctcaacatcagtctgataagctactaaagaaaggaaagagcgataaa
Probe B	gaattaatttaaataaattatctcaacatcagtctgataagctactatctccacccacacacagac
Probe C	gaattaatttaaataaattatctcaacatcagtctgataagctactcctcctcttcaccctccc

Probes MiR25

Probe A	gaattaatttaaataaattatctcagaccgagacaagtgcaatgctaaagaaaggaaagagcgataaa
Probe B	gaattaatttaaataaattatctcagaccgagacaagtgcaatgctatctccacccacacacagac
Probe C	gaattaatttaaataaattatctcagaccgagacaagtgcaatgctcctcctcttcaccctccc

Probes MiR223

Probe A	gaattaatttaaataaattatctggggtatttgacaaactgacactaaagaaaggaaagagcgataaa
Probe B	gaattaatttaaataaattatctggggtatttgacaaactgacactatctccacccacacacagac
Probe C	gaattaatttaaataaattatctggggtatttgacaaactgacactcctcctcttcaccctccc

Targets

miR21 DNA	tagcttatcagactgatgttga
miR25 DNA	cattgcacttgtctcggtctga
miR223 DNA	tgtcagtttgtcaaatacccca

Scaffold

TCTGACGACAGGTTTATCGCTCTTTCCTTTCTTTTAATTTATTTAAATTAATTCG
TTTCCGTTCTTCCGTCATACCTTTCTCTGGTCTGTGTGTGGGTGGAGATTAATTT
ATTTAAATTAATTCAGAATAATCGGCACCAGGAGTTACATCAGGGAGGGTGAA
GAGGAGGTAATTTATTTAAATTAATTCACGGTACTGTTGACCCACCACTCTGGG
CGGGAAGTTGGCTGGCTGTTTTCGT

Primers

Taq For	tgacgacaggtttatcgctc
A Rev	tatgacggaagaacggaaacg
B For	tttctctggtctgtgtgtgg
B Rev	ctcctggtgccgattattctg
C For	acatcagggagggtgaaga
Taq Rev	gaaaacagccagccaacttc

Taqman Probe

cagagtggtgggtcaacagtaccg

II.2. Gel Electrophoresis

Figure 6.6-a

Agarose: 1.5%
Sample: 25 μL
Loading buff: 5 μL
10 bp ladder: 2 μL
Final Volume: 25 μL

Figure 6.6-b

Agarose: 1.5%
Sample: 25 μL
CTRL: 25 μL (probe mix 300 nM, scaffold 250 nM)
Loading buff: 5 μL
100 bp plus ladder: 1.5 μL
Final Volume: 25 μL

Figure 6.10.
Acrylamyde: 20%
Sample: 6 μL (concentration of probe is 5 μM).
CTRL: 6 μL (concentration of probe and one-fold target is 5 μM, concentration of target excess is 60 μM).
Loading buffer (20% glycerol): 8 μL
10 bp ladder: 2,5 μL
Final Volume: 25 μL

II.3. Folding conditions and enzymatic reaction conditions

Experiment section	Scaffold	Probe	Initial denaturation	Enzymatic reaction
6.6	150 nM	150 nM	yes	2 h
6.7	50 nM	50 nM	no	3 h 30 min (1 nM)
6.8.1.	150 nM	150 nM	yes	2 h (1.5 nM)
6.8.1.	15 nM	150 nM	yes	2 h (15 pM)

II.4. qPCR (RT-PCR)

Experiment section	Sample volume (μL)	Final RT-PCR volume (μL)
6.6	5	25
6.7	4	20
6.8.1	4.6	23
6.8.1	4.6	23
6.8.3	2.5 (1:10 dilution for site-specific qPCR)	22

Bibliography

1. Nature.com. Subject: Genetics. http://www.nature.com/subjects/genetics
2. Lodish, H. F. *et al*. Molecular Cell Biology. New York: W.H. Freeman and Co (2003).
3. Gregory, T. R. Synergy between sequence and size in Large-scale genomics. *Nat Rev Genet* **6**, 699–708 (2005).
4. Pray, L. Transposons, or jumping genes: Not junk DNA? http://www.nature.com/scitable/topicpage/transposons-or-jumping-genes-not-junk-dna-1211 (2008)
5. Nature.com. Definition: Transcriptome. http://www.nature.com/scitable/definition/transcriptome-296 (2014)
6. Sealfon S. C., Chu T. T. RNA and DNA Microarrays, in Biological Microarrays: Methods and Protocols. *Methods in Molecular Biology* **671**, 3-34 (2011)
7. Cheng, G. Circulating miRNAs: Roles in cancer diagnosis, prognosis and therapy. *Advanced Drug Delivery Reviews* **81**, 75–93 (2015).
8. Pasquinelli, A. E. Non-coding RNA: MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Publishing Group* **13**, 271–282 (2012).
9. Loewe, L. Genetic mutation. http://www.nature.com/scitable/topicpage/genetic-mutation-1127 (2008).
10. Shearer, A. E. *et al*. Copy number variants are a common cause of non-syndromic hearing loss. **6,** 1–10 (2014).
11. Nature.com. Subject: Genomics. http://www.nature.com/subjects/genomics
12. Macaulay, I. C. & Voet, T. Single Cell Genomics: Advances and Future Perspectives. *PLoS Genet* **10,** e1004126–9 (2014).
13. National Institute of Health. Genetics Home Reference: Sickle Cell Disease. https://ghr.nlm.nih.gov/condition/sickle-cell-disease
14. Coughlin, C. R., Scharer, G. H., & Shaikh, T. H. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Med* **4**, 80–12 (2012).

15. Lan, H., Lu, H., Wang, X. & Jin, H. Review Article MicroRNAs as Potential Biomarkers in Cancer: Opportunities and Challenges. *BioMed Research International*, 125094 (2015).
16. de Planell-Saguer, M. & Rodicio, M. C. Detection methods for microRNAs in clinic practice. *Clinical Biochemistry* **46,** 869–878 (2013).
17. Li, Y. *et al*. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Research* **42,** 1070-1074 (2013).
18. Gim, J.-A., Ha, H.-S., Ahn, K., Kim, D.-S. & Kim, H.-S. Genome-Wide Identification and Classification of MicroRNAs Derived from Repetitive Elements. *Genomics Inform* **12,** 261–7 (2014).
19. Wang, Y., Li, X. & Tao, B. Improving classification of mature microRNA by solving class imbalance problem. *Scientific Report* **6**, 25941 (2016).
20. Saiki R.K. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354 (1985).
21. Higuchi, R., Fockler C., Dollinger G., Watson R. Kinetic PCR Analysis: Real-Time Monitoring of DNA Amplification Reactions. *Biotechnology* **11**, 1026-1030 (1993).
22. Kubista, M. *et al*. The real-time polymerase chain reaction. *Molecular Aspects of Medicine* **27,** 95–125 (2006).
23. BioRad. Real-Time PCR Application Guide.
24. Dymond, J.S. Explanatory Chapter: quantitative PCR. *Methods in Enzymology* **529**, 279-289 (2013)
25. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments.*Biomolecular Detection and Quantification* **3**, 9–16 (2015).
26. Bubner, B., Gase, K. & Baldwin, I. T. Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *BMC Biotechnology* **4**, 14–11 (2004).
27. Bustin, S. A. *et al*. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* **55**, 611–622 (2009).
28. Loftis, A. D., Reeves, W. K. Principles of Real-Time PCR. *Veterinary PCR diagnostics*, 3-17. Bentham Science Publisher (2012).

29. Vaerman, J. L., Saussoy, P., Ingargiola, I. Evaluation of real-time PCR data. *Journal of Biological Regulators and Homeostatic Agents* **18**, 212-214 (2004).
30. Baker, M. Digital PCR hits its stride. *Nature Methods* **9**, 541–544 (2012).
31. Whale, A. S. *et al*. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Research* **40**, 82-82 (2012).
32. Hospodsky, D., Yamamoto, N. & Peccia, J. Accuracy, Precision, and Method Detection Limits of Quantitative PCR for Airborne Bacteria and Fungi. *Applied and Environmental Microbiology* **76**, 7004–7012 (2010).
33. Ma, L. & Chung, W. K. Quantitative Analysis of Copy Number Variants Based on Real-Time LightCycler PCR. *Current Protocols in Human Genetics* **7.21**, 1-8 (2014).
34. Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K. & Weinhäusel, A. A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification* **1**, 23–33 (2014).
35. D'haene, B., Vandesompele, J. & Hellemans, J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50**, 262–270 (2010).
36. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature Protocols* **1**, 1559–1582 (2006).
37. Nature.com. Poly-A Tail. http://www.nature.com/scitable/definition/poly-a-tail-276
38. Head, S. R. *et al*. Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques* **56,** 1–13 (2014).
39. Metzker, M. L. Sequencing technologies — the next generation. *Nature Publishing Group* **11,** 31–46 (2009).
40. Bustin, S. A. *et al*. The need for transparency and good practices in the qPCR literature. *Nature Methods* **10**, 1063–1067 (2013).
41. Moltzahn, F. *et al*. Microfluidic-Based Multiplex qRT-PCR Identifies Diagnostic and Prognostic microRNA Signatures in the Sera of Prostate Cancer Patients. *Cancer Research* **71**, 550–560 (2011).
42. Pinto, D. *et al*. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology* **29**, 512-520 (2011)

43. Bumgarner, R. Overview of DNA Microarrays: Types, Applications, and Their Future. *Current protocol in Molecular Biology* **22.1**, 1–17 (2013).
44. Duan, D. *et al*. Label-free high-throughput microRNA expression profiling from total RNA. *Nucleic Acids Research* **39,** e154–e154 (2011).
45. Volokhov D. V., Kong H., Herold K., Chizhikov V. E., Rasooly A. Oligonucleotide Microarrays for Identification of Microbial Pathogens and Detection of Their Virulence-Associated or Drug-Resistance Determinants, in Biological Microarrays: Methods and Protocols. *Methods in Molecular Biology* **671**, 3-34 (2011).
46. Mardis, E. R. Next-Generation Sequencing Platforms. *Annual Rev. Anal. Chem.* **6,** 287–303 (2013).
47. Thermo Fisher Scientific. Sanger Sequencing Instruments and Accessories. https://www.thermofisher.com/it/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-technology-accessories.html
48. Illumina. Technology Spotlight (2010).
49. Illumina. Sequencing Platform Comparison Tool. http://www.illumina.com/systems/sequencing-platform-comparison.html
50. Thermo Fisher Scientific. Ion Torrent™ Next-Generation Sequencing Technology. https://www.thermofisher.com/it/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html
51. Thermo Fisher Scientific. Ion Torrent Application Note, The Ion PGMTM System, with 400-base read length chemistry, enables routine high-quality de novo assembly of small genomes. https://tools.thermofisher.com/content/sfs/brochures/Small-Genome-Ecoli-De-Novo-App-Note.pdf
52. Fluidigm. www.fluidigm.com/products/c1-system
53. Kan, Z.-Y. *et al*. Molecular Crowding Induces Telomere G-Quadruplex Formation under Salt-Deficient Conditions and Enhances its Competition with Duplex Formation. *Angewandte Chemie International Edition* **45**, 1629–1632 (2006).
54. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Science* **95**, 1460-1465 (1998).

55. Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Walder, J. A. Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations. *Biochemistry* **47**, 5336–5353 (2008).
56. OpenStax College, Biology. OpenStax CNX. http://philschatz.com/biology-book/contents/m44486.html
57. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *Journal of American Chemical Society* **131**, 17303–17314 (2009).
58. Wang, D. *et al*. Molecular Logic Gates on DNA Origami Nanostructures for MicroRNA Diagnostics. *Analytical Chemistry* **86**, 1932–1936 (2014).
59. Sanfiorenzo, C. *et al*. Two Panels of Plasma MicroRNAs as Non-Invasive Biomarkers for Prediction of Recurrence in Resectable NSCLC. *PLoS ONE* **8,** 54596–8 (2013).
60. Biorad. Introduction to PCR Primer & Probe Chemistries. http://www.bio-rad.com/it-it/applications-technologies/introduction-pcr-primer-probe-chemistries
61. Zheng, J. *et al*. Rationally designed molecular beacons for bioanalytical and biomedical applications. *Chemical Society Review* **44**, 3036-3055 (2015).
62. Krishnan, Y. & Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angewandte Chemie International Edition* **50**, 3124–3156 (2011).
63. Liu, M. *et al*. A DNA tweezer-actuated enzyme nanoreactor. *Nature Communications* **4,** 2127 (2013).
64. Feldkamp, U., Niemeyer, C. M. Rational Design of DNA Nanoarchitectures. *Angewandte Chemie International Edition* **45**, 1856-1876 (2006).
65. Rothemund, P. W. K. Folding DNA to Create Nanoscale, Shapes and Patterns. *Nature*, **440**, 297-302 (2006).
66. Woo, S., Rothemund, P. W. K. Programmable Molecular Recognition Based on the Geometry of DNA Nanostructures. *Nature Chemistry* **3**, 620-627 (2011).
67. Ke, Y., Ong, L. L., Shih, W. M., Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **338**, 1177-1183 (2012).
68. Rinker, S., Ke, Y., Liu, Y., Chhabra, R., Yan, H. Self-assembled DNA Nanostructures for Distance-dependent Multivalent Ligand-protein Binding. *Nature Nanotechnology*. **3**, 418-422 (2008).

69. Kuzyk, A., Laitinen, K. T., Torma, P. DNA Origami as a Nanoscale, Template for Protein Assembly. *Nanotechnology* **20**, 1-5 (2009).
70. Ghosh, D. *et al*. M13-templated Magnetic Nanoparticles for Targeted in Vivo Imaging of Prostate Cancer. *Nature Nanotechnology* **7**, 677-682 (2012).
71. Sacca, B., Niemeyer, C. M. DNA Origami: The Art of Folding DNA. *Angewandte Chemie International Edition* **51**, 58-66 (2012).
72. Douglas, S. M., Bachelet, I., Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831-834 (2012).
73. Meyer, R., Niemeyer, C. M. Orthogonal Protein Decoration of DNA Nanostructures. *Small* **7**, 3211-3218 (2011).
74. Kuzyk A. *et al*. DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, 311–314 (2012).
75. Acuna G.P. *et al*. Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **338**, 506–510 (2012).
76. Kukolka F., Schoeps O,. Woggon U,. Niemeyer C.M. DNA-Directed Assembly of Supramolecular Fluorescent Protein Energy Transfer Systems. *Bioconjugate Chemistry* **18**, 621–627 (2007).
77. Stein I.H., Steinhauer C., Tinnefeld P. Single-Molecule Four-Color FRET Visualizes Energy-Transfer Paths on DNA Origami. *Journal of American Chemical Society* **133**, 4193–4195 (2011).
78. Pinheiro A.V., Nangreave J., Jiang S., Yan H., Liu Y. Steric Crowding and the Kinetics of DNA Hybridization within a DNA Nanostructure System. *ACS Nano* **6**, 5521–5530 (2012).
79. Wei, X., Nangreave, J., Jiang, S., Yan, H. & Liu, Y. Mapping the Thermal Behavior of DNA Origami Nanostructures. *J. Am. Chem. Soc.* **135**, 6165–6176 (2013).
80. Ke, Y., Voigt, N. V., Gothelf, K. V. & Shih, W. M. Multilayer DNA Origami Packed on Hexagonal and Hybrid Lattices. *J. Am. Chem. Soc.* **134**, 1770–1774 (2012).
81. Peterson A.W., Heaton R.J., Georgiadis R.M. The effect of surface probe density on DNA hybridization. *Nucleic Acids Research* **29**, 5163–5168 (2001).
82. Ke Y. *et al*. Multilayer DNA Origami Packed on a Square Lattice. *Journal of American Chemical Society* **131**, 15903–15908 (2009).

83. Nanguneri S., Flottmann B., Horstmann H., Heilemann M., Kuner T. Three-Dimensional, Tomographic Super-Resolution Fluorescence Imaging of Serially Sectioned Thick Samples. *PLoS ONE* **7**, 38098 (2012)
84. Bates M., Huang B., Dempsey GT., Zhuang X. Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science* **317**, 1749–1753 (2007).
85. Lin C. *et al*. Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA. *Nature Chemistry* **4**, 832–839 (2012).
86. Jungmann R. *et al*. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods* **11**, 313–318 (2014).
87. Pinheiro A.V., Han D., Shih W.M., Yan H. Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology* **6**, 763–772 (2011).
88. Hubbell J.A., Chilkoti A. Nanomaterials for Drug Delivery. *Science* **337**, 303–305 (2012)
89. Zhao Y-X, Shaw A, Zeng X, Benson E, Nyström AM, Högberg B (2012) DNA Origami Delivery System for Cancer Therapy with Tunable Release Properties. ACS Nano 6:8684–8691
90. Walsh AS, Yin H, Erben CM, Wood MJA, Turberfield AJ (2011) DNA Cage Delivery to Mammalian Cells. ACS Nano 5:5427–5432
91. Jiang Q. *et al*. DNA Origami as a Carrier for Circumvention of Drug Resistance. *Journal of American Chemical Society* **134**, 13396–13403 (2012).
92. Shaw A. *et al*. Spatial control of membrane receptor function using ligand nanocalipers. *Nature Methods* **11**, 841–846 (2014).
93. Ke Y.G., Lindsay S., Chang Y., Liu Y., Yan H. Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays. *Science* **319**, 180–183 (2008).
94. Mei, Q.; Wei, X.; Su, F.; Liu, Y.; Youngbull, C.; Johnson, R.; Lindsay, S.; Yan, H.; Meldrum, D. Stability of DNA Origami Nanoarrays in Cell Lysate. Nano Lett., 2011, 11, 1477-1482.
95. Nishikawa, M.; Rattanakiat, S.; Takakura, Y. DNA-based Nano-sized Systems for Pharmaceutical and Biomedical Applications. Adv. Drug Deliv. Rev., 2010, 62, 626-632.

96. Castro C.E. *et al*. A primer to scaffolded DNA origami. *Nature Methods* **8**, 221–229 (2011)
97. Mei Q. *et al*. Stability of DNA Origami Nanoarrays in Cell Lysate. *Nano Letters* **11**, 1477–1482 (2011).
98. Lee H. *et al*. Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nature Nanotechnology* **7**, 389–393 (2012).
99. Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angewandte Chemie International Edition* **53,** 12735–12740 (2014).
100. Integrated DNA Technologies. OligoAnalyzer 3.1. https://eu.idtdna.com/calc/analyzer
101. National Center for Biotechnology Information (NCBI). Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi
102. Ovation® Ultralow Library Systems, Userguide PART NOS. 0303, 0304, 0305, 0330, 0331, Nugen
103. Ludlow, A. T. *et al*. Quantitative telomerase enzyme activity determination using droplet digital PCR with single cell resolution. *Nucleic Acids Research* **42,** e104–e104 (2014).
104. Mestdagh, P. *et al*. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nature Methods* **11,** 809–815 (2014).
105. Eastburn D.J., Sciambi A., Abate A.R. Ultrahigh-throughput Mammalian single-cell reverse-transcriptase polymerase chain reaction in microfluidic drops. *Analytical Chemistry* **85**, 8016-21 (2013).
106. Ricci, F., Vallée-Bélisle, A., Porchetta, A. & Plaxco, K. W. Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *Journal of American Chemical Society* **134,** 15177–15180 (2012).