UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

CORSO DI LAUREA MAGISTRALE IN CHIMICA

TESI DI LAUREA MAGISTRALE

Encapsulation of a DNA-protein conjugate in a DNA origami nanocage

Relatore: Prof. Fernando Formaggio

Correlatrice: Prof.ssa Barbara Saccà

Controrelatrice: Prof.ssa Chiara Maccato

Laureanda: Silvia Zampini

2025161

Anno Accademico 2021/2022

Index

1. Introduction

1.1. DNA Nanotechnology

The importance of nanotechnology has been recognised back in 1959, when the Nobel Laureate Richard Feynman gave a pioneering talk called "*There's plenty of room at the bottom*"[1]. In the 1980s, nanotechnology rapidly expanded and became a discipline at the cross-section of engineering and natural sciences. Nanotechnology can be generally divided into top-down and bottom-up approaches. Top-down approaches are typical of physicists and aim to reach the nanosized world starting from a macroscopic material and progressively reducing its dimensions. Contrarily, bottom-up methods are based on the orderly association of smaller units, mostly molecules, through a self-assembly process. This second approach is the one used by nature. In DNA nanotechnology, the self-recognition properties of the DNA molecule are used to construct larger supramolecular structures, making this method a classic bottom-up approach based on self-assembly.

DNA nanotechnology was invented in 1982 by Ned Seeman: his revolutionary idea was to use DNA as the building block of nanosized objects with geometrical precision[2,3]. In 1953, James Watson and Francis Crick published their pioneering studies[4] on the three-dimensional structure of DNA. The DNA is a biopolymer made of nucleotides. A nucleotide is composed of a five-carbon sugar, the 2'-deoxyribose, a phosphate group and one nucleobase. There are four possible nitrogen-containing nucleobases: two purines, adenine (A) or guanine (G), and two pyrimidines, cytosine (C) or thymine (T). The nucleotides are chained together through a phosphate group, forming phosphodiester bonds. The terminal residue whose C5' is not linked to another nucleotide is called the 5' end, and the terminal residue whose C3' is not linked to another nucleotide is called the 3' end. By convention, the sequence of nucleotide residues in a nucleic acid is written, left to right, from the 5' end to the 3' end. The nucleobases recognize each other with a specific scheme of hydrogen bonds, where a guanine and a cytosine are linked by three hydrogen bonds, whereas a thymine and an adenine are linked by two hydrogen bonds (*Figure 1*). This scheme of hydrogen bonds defines the Watson-Crick complementarity rule and enables the link between two so-called complementary single strands, that are coiled one around each other in a double helix[5].

In nature, DNA is mostly present in a right-handed double helix called B-form: the helical structure is parted in a major (22 Å wide) and a minor groove (12 Å wide). The helix has a

diameter of about 2 nm and a helical pitch of about 3.46 nm (ca. 0.34 nm/bp), meaning that every 10.5 base pairs, the strands perform a complete turnaround the helical axis.

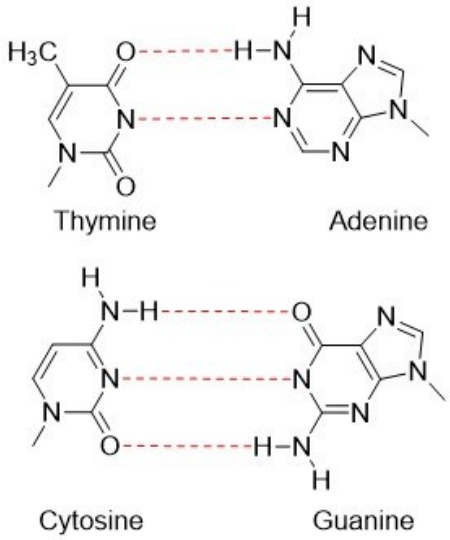

Figure 1: The base pairing scheme of the four nucleobases

The predictable self-recognition properties of the double helix make the DNA molecule an ideal construction material for nanometer-sized objects. Furthermore, because of its relatively high physical and chemical stability and its low costs, DNA can be conveniently modified with standard chemical conjugation procedures.

DNA nanotechnology can be divided into two main categories: structural and dynamic DNA nanotechnology. The first is about building complex static structures using the design rules of self-assembly, while the second enables these objects to modify their shape in response to an external stimulus. Structural DNA nanotechnology can be further divided into two classes, depending on the type of structures that can be attained: the multi-stranded approach and the DNA origami approach.

The first example of the multi-stranded approach was the Holliday junction[2] developed by Seeman. This motif was the simplest artificially designed DNA tile. Tiles are multi-branched motifs formed by the intertwining of two or more DNA duplexes in (at least) one common branch point, also called junction or crossover. The initial tile developed by Seeman was composed by four distinct 16 bp-long DNA strands, which intertwine in a common crossover, thus generating four double helical segments of 8 bp each. The tile was designed to not undergo any topological isomerization[6,7]. The sequences participating in the formation of the motif were

chosen to be maximally different from one another, such that competitive formation of alternative secondary structures was minimized. Although appearing to be rather simple, this design is indeed the conclusion of several years of theoretical and experimental work, which eventually signed the beginning of structural DNA nanotechnology. The initial idea of Seeman relied on the fact that such junctions constitute the building blocks of larger crystal-like structures. To enable that, junctions are usually connected through hybridization of complementary single stranded segments protruding from the 5' or 3' termini of the junction's arms (sticky-ends). This method requires a high control over the stoichiometry and purity of the constituent oligonucleotides, thus resulting in error-prone and lengthy synthetic processes.

In 2006 Paul Rothemund introduced the scaffolded DNA origami method[8]. Here, a long single-stranded DNA (called scaffold) is folded into a desired shape by the help of hundreds of short oligonucleotides, called staples, in a way that recalls the Japanese art of paper folding. The scaffold is usually the viral single-stranded genome of the M13 bacteriophage. This means that it's circular and so must be the folding path: the scaffold therefore runs back and forth along the first half of the shape till the top of it and then runs back along the second half of the shape, thus returning to the initial position. The virtual line that separates the two halves of the structure and along which the scaffold inverts its direction is called seam. The staples are designed to hybridize to different regions of the scaffold, through formation of Holliday junctions (or crossovers). Basically, a DNA origami is a pattern of DNA crossovers, whose relative orientation defines the final geometry of the structure. To obtain a planar origami structure, crossovers must be placed at 180° one another, such to connect adjacent helices along the same plane. By imposing staple crossovers every 1½ helical turn, the helical domains of a given helix will be connected to the helical domains of two adjacent helices (above and below the given helix) through crossovers established alternatively at 0° and 180°: as 1 helical turn = 10.5 bp; 1½ turns will be exactly 15.75 bp, in practice 16 bp (the small deviation can accumulate along the structure and cause a global twist).

The DNA origami method is more efficient than the multi-stranded approach, mainly because of the entropic advantage in using a long scaffold strand for folding[9]. In the origami strategy, the stoichiometric ratio of the staples is less relevant than in the multi-stranded approach, because the initial correct arrangement of the scaffold favours the correct binding of the remaining staples, such that possibly existing wrong or truncated sequences can be easily displaced. Consequently, experimental errors and time of synthesis are dramatically reduced,

too. This, together with the capability to generate nanoobjects with complex shapes of pre-defined dimensions and full molecular addressability, makes the DNA origami a very robust and powerful tool for construction of DNA-based architectures[10,11]. To help the DNA designer, software tools have been developed and are in constant evolution to assist the user in the engineering of both classes of DNA nanostructures, drastically reducing the time required. Nevertheless, fundamental design knowledge and informatic skills are still very important for the researcher working in this field[12]. Both multi-stranded and DNA origami structures are the result of a controlled thermal annealing of all participating strands into a well-defined and possibly unique structure. The self-assembly process is in most cases a thermodynamically driven process that leads to formation of the most stable structure, that is the shape with the lowest free energy.

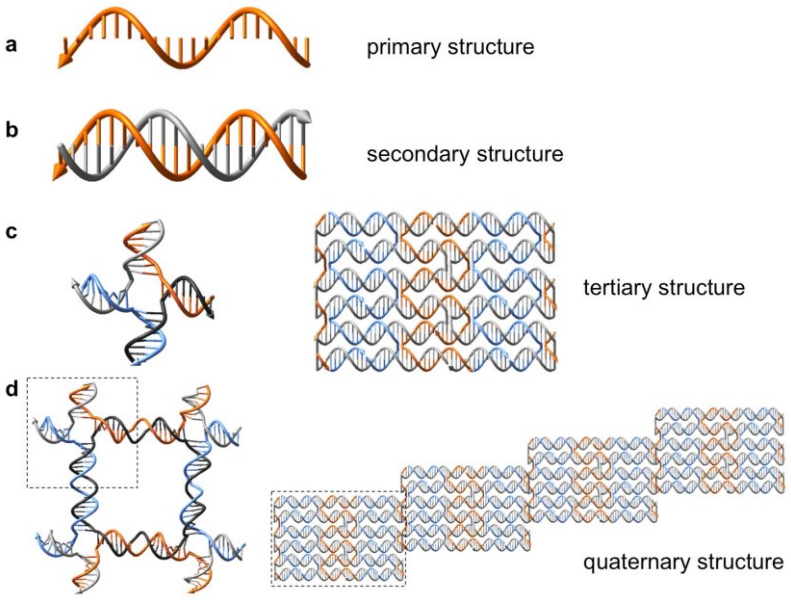

Figure 2: The four levels of structural hierarchy in DNA self-assembly.

DNA structures can be tentatively divided into four levels of complexity, following the same principles as in proteins, that is, a primary, secondary, tertiary, and quaternary structure[13]. According to this classification, each level of the hierarchy is populated by a well-defined DNA structure, which is used as building block for the next level. The first level, or primary structure, is simply the base sequence of a single-stranded DNA; the secondary structure is the double helix generated by hybridization of a single stranded DNA to a complementary strand; the three-dimensional organization of double helical domains into branched motifs and DNA origami shapes is described as the tertiary structure. Finally, self-association of those structures leads to

the fourth and highest level of structural organization, including large crystalline arrays and complex 2D and 3D objects.

The main advantage of the DNA origami method is the fact that each component staple, and therefore each nucleobase, occupies a predictable position within the final architecture, which is dictated by the helical structure of the double helix. As a consequence, any DNA origami surface can be in principle used as a molecular pegboard to create patterns of chemical groups with a resolution of only few nanometres. This unique addressability is given by the possibility to attach a desired chemical function to virtually every nucleobase. As the design map of the whole origami structure is known (because it is pre-designed by the user), every coordinate of the map, that is, every nucleobase of the origami structure, can be precisely addressed. These chemical groups can be further used to link various types of molecules, for diverse biological and physical applications: in this case an enzyme, the HTRA1 protease, has been encapsulated inside a DNA origami nanocage. For the encapsulation of the protein, selected staples on the inside surface of the DNA origami cage are elongated at one end. In this way, single stranded "protruding arms" point towards the inner cavity of the origami cage at desired positions. Those handles will catch the molecule of interest either through hybridization to a complementary DNA tagged protein (also referred to as conjugate, *Figure 4a*), or through specific recognition between a previously hybridized DNA-tagged ligand and the target species (*Figure 4b*).

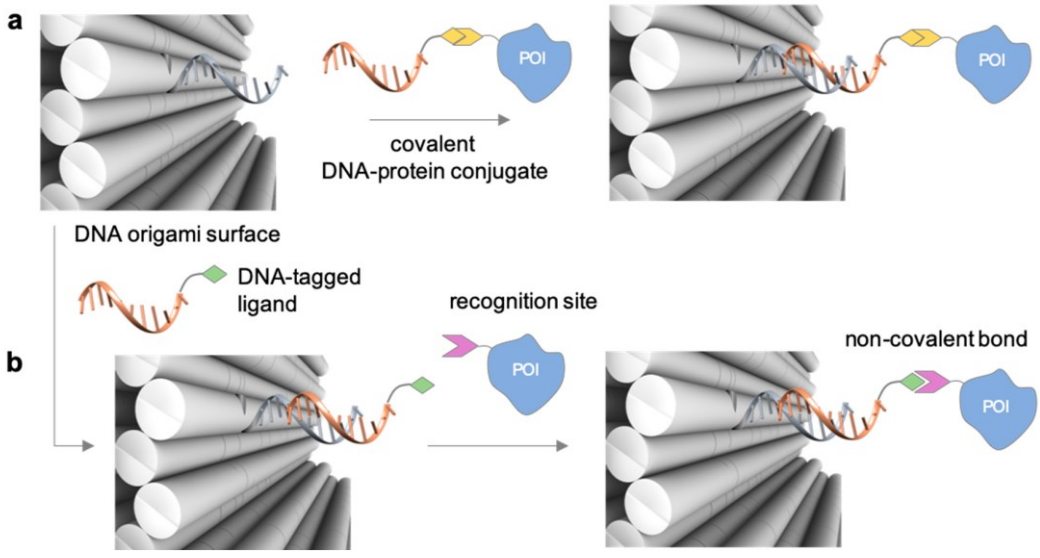

Figure 3: Binding strategies for the attachment of a Protein of Interest (POI) to a nanostructured DNA template. The surface displays a protruding arm (grey strand) to which the complementary DNA tag of the conjugate (orange strand) hybridizes. The protein can be organized either (a) through hybridization

of a covalent DNA-tagged conjugate or (b) through non-covalent binding to a previously hybridized DNA-tagged ligand.

The unique features of DNA nanostructures have revealed several interesting applications in materials science, physics, engineering, and biology. In particular they have been used to assemble complex biomolecular systems, such as multienzyme complexes, protein confinement and biomimetic channelling[14].

1.2. HTRA1

In this project the Protein of Interest is HTRA1 (high temperature requirement A), a homo-oligomeric and ATP-independent serine protease. This class of enzymes digests peptide bonds, primarily hydrophobic stretches in selected or misfolded substrate proteins. The protein has an important role in protein quality control because it removes misfolded polypeptides through degradation and plays a regulatory role by inactivating proteolytically negative regulators[15]. The loss of HTRA1 is correlated with many different severe diseases, including Alzheimer's disease[16].

The HTRA1 protein is composed of three domains: a central core, a C-domain and an N-domain. The core contains the active site responsible for enzymatic activity and most reported functions, while the C-terminal region holds a PDZ1 domain that facilitates protein-protein interactions. The structure and function of the N-domain is still not completely known, it contains 16 cystines, that are all part of an intensive disulphide bond network. In solution, the HTRA1 monomer associates into a funnel-shaped trimer, the most stable oligomer, with a central pore and outward-protruding PDZ1 domains. The active site contains the catalytic triad (His220, Asp250, Ser328), the oxyanion hole and the substrate-specificity pocket. In the catalytic triad, the Ser residue is responsible for the cleavage reaction of the amide bond and its reactivity is modulated by His and Asp, two polar side chains that determine its nucleophilic character. During protein cleavage, an intermediate with a negatively charged carbonyl oxygen occurs and this oxyanion is stabilized by two hydrogen bonds in the oxyanion hole of the protease, which is an amide nitrogen cradle formed by residues preceding the catalytic Ser residue. The residue preceding the bond that gets broken is accommodated into the S1 substrate specificity pocket, that accounts for the specificity of the protease.

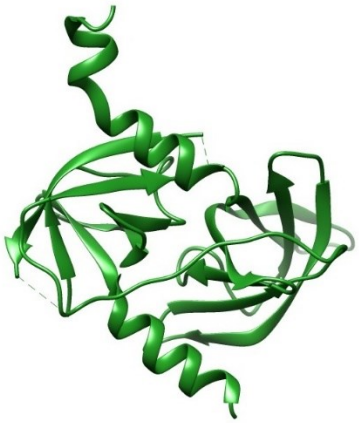

Figure 4: 3D representation of monomeric HTRA1. The image was elaborated with the Chimera software from the pdb file 3NUM of the Protein Data Bank[17].

The HTRA1 produced for this project is genetically modified at the N-terminus with a cysteine residue for further chemical conjugation. The protein was produced both in the active and in the inactive form: in the inactive HTRA1, Ser328 of the catalytic triad was replaced by an alanine. Both proteins have 334 amino acids and a molecular weight of 36 kDa.

1.3. DNA-Protein conjugate

For the functionalization of DNA origami is therefore fundamental to bind first the protein to a DNA oligonucleotide; the DNA-protein conjugate will be then later hybridized to the complementary handle protruding from the origami surface at a desired position. DNA-protein bioconjugation can be performed using several approaches. In all these strategies, the reaction takes place among a specific group attached to the DNA molecule and another group of the protein molecule, thus all these methods are chemo selective. Furthermore, they can be classified according to the strength and selectivity of the newly established interaction. Indeed, the DNA-protein interaction can be either covalent, and thus mainly strong and irreversible, or non-covalent and therefore weaker and reversible. The bond can be regioselective, that is directed exclusively to one specific site of the protein, or address many sites, which are chemically identical but spatially indistinguishable.

One of the most largely used approaches for DNA-protein conjugation is a covalent and non-regioselective method that uses heterobifunctional crosslinkers. These linkers bear an active ester for reaction with an amino group and a maleimide moiety for orthogonal binding to a thiol

group. For example, using the crosslinker sulfo-SMMC, the Lys residue on the surface of a protein can be bound to a thiol-modified DNA[18]. Considering that there is typically more than one accessible lysine residue, this method can lead to a heterogeneous mixture of products with an undefined number of DNA attached to an unknown point of the protein.

Other covalent methods require genetic manipulation of the protein to introduce a specific chemical handle. An example of this kind relies on the azido- alkyne cycloaddition. To apply this strategy, the protein must be previously modified with an azide group and can be further linked to an alkyne-modified ssDNA, either in presence[19] or absence of Cu(I) ions[20]. These covalent strategies require the incorporation of non-natural amino acid in the protein sequence and the effect of such alterations on the conformational structure and enzymatic activity might be relevant and difficult to predict.

A non-covalent method uses small supramolecular ligands that specifically bind to selected amino acid residues exposed on the protein surface[18]. If the protein displays several of these binding sites, all of them will have the same probability to be targeted by the corresponding DNA-tagged ligands and their spatial discrimination will be extremely challenging. In other words, this strategy will be non-regioselective.

Non-covalent and regioselective approaches combines two desirable features of DNA-protein conjugates: a non-covalent nature of their interaction and a specifically addressable region of binding, thus mimicking natural strategies of molecular recognition. Most non-covalent methods (both regio- or non regio- selective) are based on ligand-protein recognition. A typical example is the strong biotin-streptavidin interaction which is harnessed to attach a biotinylated DNA to a streptavidin or streptavidin-tagged enzyme[21]. In another approach, the protein of interest can be genetically modified with a poly(histidine)$_6$ tag on either the N- or C-terminus to bind a nitrilotriacetic acid (NTA)-modified oligonucleotide in the presence of nickel ions[22]. Other possible ligand-protein recognitions that can be exploited are small peptide analogues of natural substrates as well as specific antigens of antibodies. Oligonucleotide aptamers, short sequences of artificial DNA that bind a specific target protein, can be adapted to immobilize proteins onto DNA nanostructures by site-specifically incorporating aptamer sequences into DNA nanostructures[23].

The success in the application of these methods is highly dependent on the affinity of the ligand-protein interaction and the availability of synthetic or natural tags that can be appended to the DNA strand. If an enzyme needs to be stably bound to a cofactor to be active, it is possible to

link the ssDNA with a covalent bond to the cofactor and reinsert it into the apoenzyme to repristinate its full activity. This technique resulted in recovered activities, that were in some cases significantly different from the unmodified enzymes (the observed enzymatic activities appeared to be dependent on the length and content of the nucleobase sequence)[24].

In this project two different strategies were explored. In the first strategy, a covalent bond was established between the HTRA1 and a single-stranded oligonucleotide using AMAS as the heterobifunctional crosslinker (*Figure 6*). This crosslinker contains an N-hydroxy-succinimide ester that reacts with the amino group of the modified DNA chain and a maleimide moiety that reacts via Michael addition to the thiol group of the cysteine residue, genetically engineered on the N-terminus of the protein. In the trimeric form, the HTRA1 will display 3 Cys residues, which can be individually linked to a DNA strand, leading to a maximum of 3 DNA strands per trimer. Once the conjugate is synthetized, it is encaged inside the DNA origami structure through hybridization with the single stranded complementary handles that extend from the inside walls of the cage.

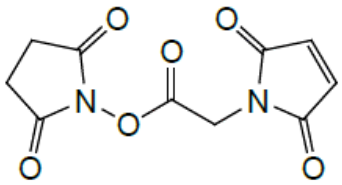

Figure 5: N-[α-maleimidoacetoxy] succinimide ester (AMAS)

The second approach used in this Thesis relies on the non-covalent interaction between the PDZ1 domains of the HTRA1 protein and a peptide ligand that mimics its specific substrate[25]. The ligand consists of a heptapeptide of sequence DPMFKLV. This peptide was found to bind to the PDZ1 domains of the DegP protein, a bacterial protease that belongs to the same family of the HTRA1 protease, with a K_d = circa 5 μM[26]. Thus, a 16-bases long DNA oligonucleotide was linked to the peptide and the resulting DNA-peptide conjugate was finally hybridized to the complementary strands protruding out of the inner surface of the DNA nanocage. The protein is linked to the cage by binding on the inner-oriented peptides that target the multiple recognition sites on the protein surface.

In both cases, the non-origami component (i.e., either the DNA-protein conjugate or the free protein) is added in stoichiometric excess to favour the formation of the product. Therefore, a

purification step is necessary to remove the unreacted component and isolate that target DNA-protein complex. Several purification strategies have been reported to achieve this task[27]. In this work, we have applied PEG-induced precipitation and ultrafiltration with molecular weight cut-off membranes.

1.4. Compartmentalization

In this work, I describe how the HTRA1 protein can be anchored to the inner walls of a densely packed DNA origami structure, resulting in its compartmentalization. The main features of the DNA origami compartment are a high structural stability and an enveloping capability.

In natural system, compartmentalisation is used as a general strategy to gate physical interactions between enzymes and their substrates. This affects biochemical reactions in many ways: by increasing local concentrations of enzyme and substrate, facilitating mass transport of reaction intermediates through enzymatic cascades, reducing toxicity of intermediates and protecting encapsulated contents from competing pathways[28]. Taking inspiration from nature, scientists have developed several artificial systems composed of compartmentalized enzymes. Among them, rationally designed DNA nano-compartments for large biomolecules raise a particular interest as an ideal tool to mimic and thus better understand some central aspects of living matter[29].

Recently, DNA nanostructure vault has been constructed to enclose a single protease within its internal cavity, thus shielding it from the substrate in solution and blocking the formation of the product. The reversible opening\closing of the vault was triggered by the addition of specific DNA strands[30]. However, binding of the enzyme into the cavity of the DNA vault was achieved through covalent and non-regioselective conjugation methods, raising some concerns about the homogeneity of the sample and its intrinsic activity prior to encapsulation within the DNA cavity. An alternative and promising method involves the use of specific recognition motifs on the inner side of the cavity for selective and non-covalent immobilization of a target protein[25]. When the protein is composed of multiple identical units (i.e. is a homo-oligomer), this method may take advantage of the multivalency of ligand interactions.

In general, studies on DNA nano-compartments can contribute to the advancement of our theoretical understanding of the effect of nano-confinement on the equilibrium state of the system and help in the formulation of novel hypotheses for the – sometimes unexpected –

phenomena observed. In small systems (typically of nanoscale size) the interactions between the system and its surrounding solvent cannot be ignored. The application of single-molecule force methods to 3D DNA origami compartments contributed to a first quantification of intriguing physical phenomena occurring within nanosized reaction vessels.

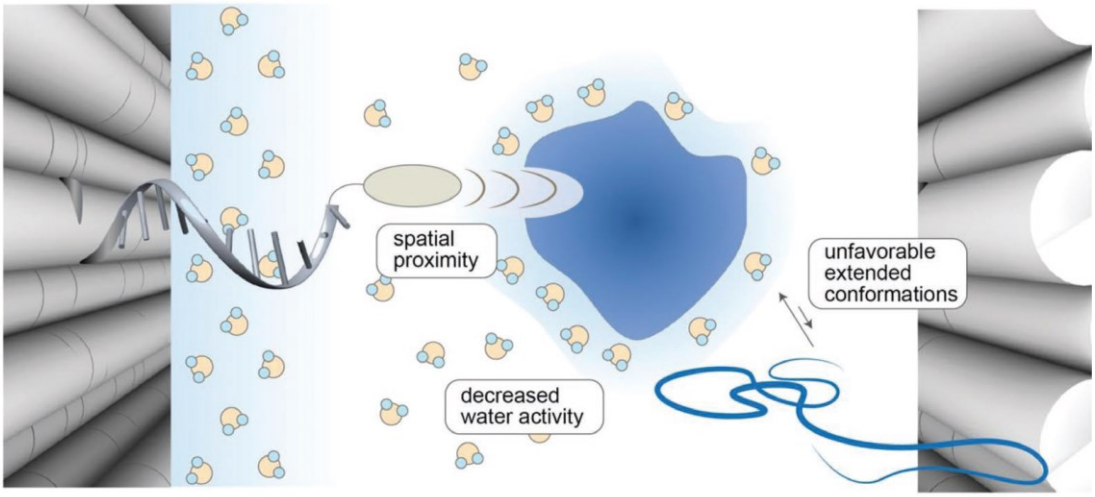

Figure 6: Schematic illustration of the main contributions to enhanced binding affinity within nanoconfined spaces.[29]

A study on the effect of the size of the cage on the mechanical stability of a G-quadruplex (G4) showed a phenomenon called "decreased water activity"[31]. According to this mechanism, the entropic cost for folding the structure is accompanied by the release of water molecules in the surrounding environment. However, the water molecules inside the DNA host are more ordered than in the bulk solution, due to the restricted space within the origami cage and the polyelectrolyte nature of its DNA walls. This leads to a decrease in water activity, which is expected to result in a reduction of the entropic cost for folding biomolecules inside a confined space. The mechanical stability of the guest molecule therefore increases when it is confined into a smaller environment.

In a recent study, the combined effect of nanoconfinement and proximity between binding components has been examined in terms of entropic changes[32]. In nanoconfined systems, binding events are entropically favoured. From the one side, the limited space available decreases the number of different conformations that can be assumed by the receptor, i.e., the extended conformation has lower probability to occur. This mechanism is similar to the one that is supposed to explain the chaperone-assisted folding of linear polypeptides into active

tertiary and quaternary conformations[33]. From the other side, the limited space within a nano compartment enhances the spatial proximity among the binding partners, thus increasing their chance to meet and bind.

1.5. Multivalency

The DNA origami nanocage produced in this work contains six protruding arms that point towards the inner cavity. Thus, it may act as a multivalent ligand in the binding to the HTRA1 protein. Furthermore, the HTRA1 protein is a homo-trimer, so every protein will carry three identical binding sites for binding to the DNA cage. This is true for both conjugation methods, the covalent and the non-covalent one. The DNA origami structure can thus be interpreted as a multivalent ligand system, or more precisely, it is a scaffold which multiple identical ligands are attached to. Therefore, binding will be the result of multiple forces applied simultaneously. This system can be used to better understand the mechanisms of multivalent binding in confined spaces and helps us to improve the theoretical models formulated until now to describe it.

Multivalency is an important phenomenon in biology and chemistry that has been studied for more than 20 years[34]. In this case study, the binding events occur over the entire surface of the protein and are mediated by the ligands anchored to the inner walls of the DNA origami cage. Multivalency, here in particularly homo-multivalency, is defined "as the operation of multiple molecular recognition events of the same kind occurring simultaneously between two entities." Such two entities are named, respectively, the receptor and the ligand. In this case the receptor is the HTRA1 protein that displays a number of binding sites or pockets on its surface to accept the ligands. Essentially, the ligands are covalently connected together by the DNA origami scaffold. In the non-covalent approach, the ligands are PDZ-binding peptides, previously conjugated to a DNA strand and hybridized to the inner protruding arms of the cage. In a covalent approach, DNA strands are first covalently bound to the protein and the resulting DNA-protein conjugate is finally encapsulated through hybridization to the complementary protruding arms.

To describe the system in quantitative terms, we indicate with m the total number of binding sites on the surface of the receptor and with n the total number of identical arms (or branches) of the multivalent ligand. The number of receptor-ligand interactions between the two species will be given by the number i. Thus, for example, assuming to have a receptor with two binding

sites (m = 2) that interacts with a ligand with two identical arms (n = 2), at equilibrium, the system will be partitioned among three possible states: the fully unbound state (i.e. no interactions between the two partners take place, i = 0), the partially bound state (i = 1) and the fully bound state (i = 2). As all ligand/receptor pairs are identical and therefore undistinguishable, every bound species can be indeed obtained in a number of different ways, resulting from all possible permutations of one single ligand to one of the receptor binding sites. In other words, each state of the system defined by $i > 0$ will be populated by a distribution of microscopically distinct but energetically equivalent complexes that share the same number i of interactions. For example, in the bivalent interaction described above, the partially bound form will have four possible microstates. The high symmetry of the system generates identical receptor/ligand pairs whose permutation results in a high microscopic degeneracy and this statistical factor becomes substantial as the number of binding sites and the number of ligands increase.

The binding strength of a multivalent system, called *avidity*, is difficult to measure quantitatively and it's related to the monovalent interaction between the ligand\receptor pair, named *affinity*. This is defined by the dissociation constant K_d, but the relationship between these two properties depends on the type of interaction.

A thermodynamic model of multivalency was developed by Kitov and Bundle[35]. The model assumes that only one ligand can bind to one receptor at a time and that all ligand/receptor binding pairs act independently and have the same binding energy. The interaction is assumed to be relatively weak and therefore the statistical relevance of the partially bound forms at equilibrium becomes significant and has to be considered.

The avidity constant is defined taking into consideration all possible (partially and fully) bound species. The formation of a multivalent system can be viewed as the result of two sequential processes: an initial inter-molecular interaction between one branch of the ligand and one binding site on the receptor ($\Delta G°_{inter}$) and an intra-molecular contribution given by the binding of the remaining branches (i – 1) of the already bound ligand with the nearby accessible binding sites of the receptor ($\Delta G°_{intra}$). Again, every complex characterized by a certain number of established bonds (i) is an ensemble of energetically equivalent microscopic arrangements that differ only in the way the ligands are engaged on the protein binding sites. The statistical distribution of all microscopic states is given by the degeneracy (or multiplicity) factor Ω_i.

$$\Omega_i = \frac{n!\,m!}{(n-i)!\ (m-i)!\ i!}$$

$$\Delta G_i^\circ = \Delta G_{inter}^\circ + (i-1)\Delta G_{intra}^\circ - RTln\left(\frac{\Omega_i}{\Omega_0}\right)$$

The first term of the equation ($\Delta G°_{inter}$) is similar to the free energy of the corresponding monovalent interaction and is typically very low. The second and third term represent instead two aspects of the multivalent effect: the energetic advantage given by additional specific interactions and the entropic advantage provided by the increased disorder of the system. For a multivalent ligand, the value of $\Delta G°_{intra}$ is typically low and very often lower than $\Delta G°_{inter}$. This means that although the first binding event will cost a certain amount of free energy, the binding of each of the following ligands will be facilitated by their spatial proximity. The total free energy cost for formation of a multivalent complex will be then less than the sum of all monovalent contributions. Moreover, the higher is the number of ligands, the more favourable will be the interaction.

Multivalent strategies may therefore allow even very weak ligands to become strong binders by simply linking them together: this increases their local concentration in the vicinity of the target and gives them the chance to act simultaneously.

1.6. Catalytic activity

Many studies have investigated the effect of DNA on the activity and the reaction rate of an enzyme or enzyme cascade. The extent of this effect, which may be either positive or negative, seems to be highly variable and dependent on the type of enzymes tested, the length and base sequence of the DNA strand linked (or simply mixed) to the protein and the topology of the scaffold to which the enzymes are bound. Sometimes, the kinetic parameters reported by different research groups greatly differ, even if they are referred to the same enzyme system. Indeed, different factors contribute to the catalytic activity of the DNA-enzyme system and the identity and weight of all these factors have not yet been completely elucidated.

A recent study on the impact of DNA on a single enzyme showed an enhanced catalysis[36]. The reason for this effect seems to be based on unspecific interactions between the substrate and the DNA, that is responsible for bringing the substrate in close proximity to the enzyme. This

proximity effect seems to be dependent on the net charge of the substrate as well as on the extent of enzyme tethering to the DNA scaffold.

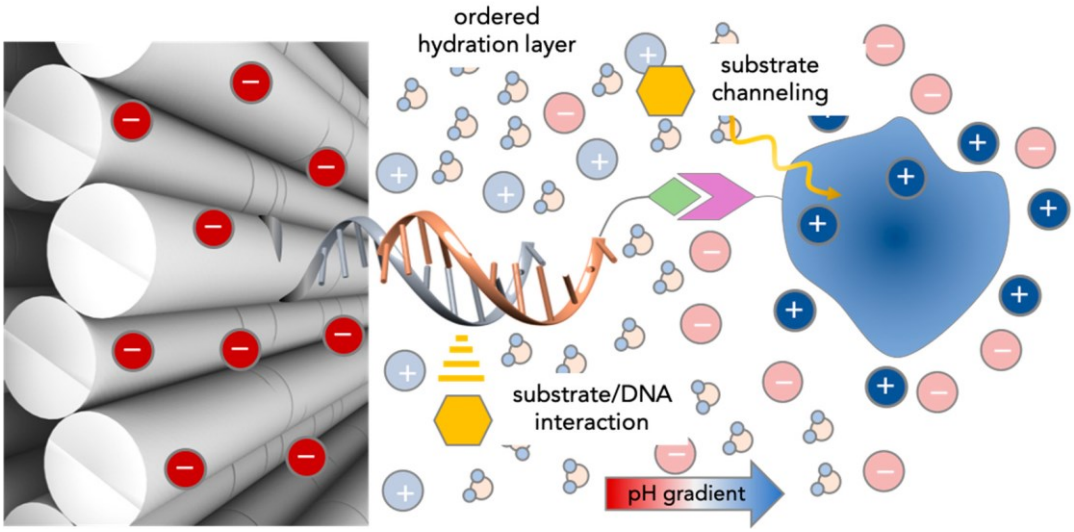

Figure 7: Current hypotheses on the observed increase of enzymatic activity for DNA-scaffolded enzymes.[18]

In case of enzyme cascades, it has been often observed that the catalytic activity of the cascade increases for shorter inter-enzyme distances. This result was attributed to a facilitated diffusion process of the intermediate species from one enzyme to another (substrate channelling). This hypothesis has been later debated: a recent study did not report any remarkable increase in the enzymatic activity of the cascade upon connecting the enzymes with a non-DNA linker. In addition, the estimation of the time required by a substrate molecule to diffuse from one enzyme to the other pinpoints that the linker provides almost no benefit over free-floating enzymes.

The effect has been explained as the consequence of a pH gradient between the DNA layer and the nearby protein surface, with a pH next to the DNA surface that is lower than in bulk solution. This is supposed to affect (mostly enhance) the catalytic activity of the enzyme[37]. In these studies, however, the DNA structure has been modelled as an infinite planar surface with constant charge density and this is far from a realistic situation.

A further hypothesis suggests that the highly dense and negatively charged DNA surface and the (mostly positively) charged protein surface generate an ordered hydration layer, which is supposed to stabilize the conformation of the enzyme and accelerate the reactions occurring at its surface[38]. This argument has been questioned by the extremely small thickness of the water

layer (few angstroms): in fact, the "cloud" of counterions that covers a uniformly charged molecule can be estimated by the Debye screening length (λ_D):

$$\lambda_D \approx \sqrt{\frac{1}{4\pi l_B C_{bulk}}}$$

where l_B is the Bjerrum length (that for a molecule in water at room temperature, is around 0,7 nm) and C_{bulk} is the concentration of counterions in the bulk solution. This means that, in a solution with a C_{bulk} of about 100 mM, the Debye screening length is about 1 nm, beyond this distance the charge on the protein surface will not be "felt" by other charges, the electric potential will become essentially zero and the charge distribution will be uniform.

Clearly, some theoretical aspects of the topic are still missing or should be reconsidered in order to be able to explain all these partially contradictory experimental data. Using simpler constructs, the role played by the different parameters of the system can be more easily identified and quantified.

2. Aim of the thesis

The main goal of this thesis is to develop a strategy for the selective encapsulation of a protein guest into a synthetic DNA host.

The protein chosen as guest is HTRA1. This protein has a short lifetime and a complex activity profile. Being able to prolong the lifetime of unstable proteins and control their enzymatic activity would be an enormous step forward in the realization of artificial protein machines.

HTRA1 is encaged inside a hexagonal prism nanocage, obtained using DNA origami-based nanotechnology.

The encapsulation is performed with two different methods. One method uses a heterobifunctional crosslinker for the synthesis of a covalent DNA-protein conjugate. This conjugate is then hybridized to complementary protruding arms appended to the inner side of the DNA host structure. In the second approach, a DNA-tagged ligand that binds specifically to the recognition domain of HTRA1 is hybridized on the inner surface of the cavity. The HTRA1 recognises the ligand and is encapsulated inside the DNA host. The two techniques are developed, optimised, and then compared.

This novel DNA-protein complex can be used to increase the understanding of complex biological processes, like the effect of multivalency and confinement on host-guest interactions taking place within DNA origami structures. The catalytic activity of the encapsulated enzyme will be probably also affected by the presence of the surrounding DNA nanocage.; however, in a way that is still unclear.

3. Results and discussion

3.1. NE_6 nanocage self-assembly and characterization

The DNA origami cage used in this project is called "*Nemesis"* (in this thesis is referred to as NE). It has the shape of a hollow hexaprism, made of a double layer of DNA helices. Actually, this object is not a single origami, but a hierarchical post-assembly association of two DNA origami structures, the half-cages "*Narcissus"* (N) and "*Echo"* (E). The cage is a relatively large nanosized object: the hexagonal section has an inner radius of 17 nm and an external radius of 29 nm; the prism is 41 nm long. The cage is designed to display a defined number of protruding arms (PAs) at selected positions of the inner surface, which are used for encapsulation of the target protein. The full cage contains six protruding arms and is called NE_6, while the half-cages contain each 3 PA and they are referred to as N_3 and E_3 respectively (for a control experiment it is possible to also produce the cage without protruding arms NE_0).

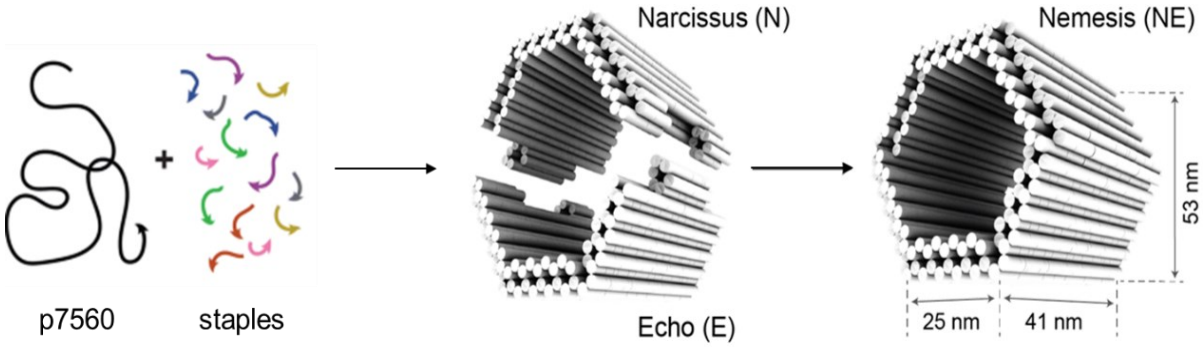

Figure 8: Principle of DNA origami assembly. The long single- stranded p 7560 scaffold is annealed with multiple short staples. The staples bring together different regions of the scaffold via base pairing to form the desired structure. The two half-cages are formed and in a second step the full cage is assembled.

To translate the desired final shape into the folding route of a given scaffold and generate corresponding staple sequences the caDNAno software was used, one of the most used software tools for designing 2D and 3D origami[39]. The scaffold used is the p7560, a circular 7560 nucleobases-long single-stranded strand derived from the m13mp18 viral genome of the M13 phage.

The buffer used for the assembly is TEMg20, containing Tris, EDTA and magnesium ions. The conjugate acid of tris has a pK_a of 8.07 at 25 °C, which implies that the buffer has an effective pH range between 7.1 and 9.1 at room temperature.

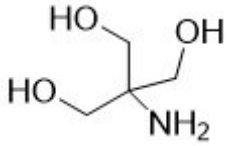

Figure 9: Tris-(hydroxymethyl) aminomethane Tris

The yield of DNA origami in terms of fraction of correct structures is highly sensitive to Mg^{2+} concentration because magnesium ions (and to a certain extent also sodium ions) are known to assist the folding and steadiness of densely packed DNA nanostructures, through screening of inter-helical repulsions and stabilization of DNA Holliday junctions. The optimal concentration of Mg^{2+} varies with the complexity of the DNA origami, in this case it was found to be 20 mM[12].

The half-cages are folded via one-pot assembly. The concentration of staples is 50 times higher than the concentration of the scaffold. The mixture undergoes a thermal annealing process, in which it is heated to high temperatures for a short time and then gradually cooled down to allow a spontaneous self-assembly of the DNA origami structure. Alternatively, as in this project, assembly occurs isothermally at a temperature that is slightly below the global melting temperature of the structure. This strategy exhibits a high yield, robustness, and the ability to produce complex shapes, that arise from the high cooperativity of multiple scaffold-staples interactions during origami folding. A second step at lower temperature is required for the hierarchical assembly of the two origami units.[40]

To link DNA origami structures together various possible strategies can be used. One possibility relies on the base hybridization between the staples of one structure and the scaffold of the other. Of course, each structure could display both extended staples and unpaired to scaffold to act simultaneously as a donor and acceptor of base-pairs. Another strategy relies on base stacking between facing helices. The stacking is an interaction between the exposed nucleobases of different origami unit. In the NE_6 cage selected staples are designed to run from the edge of one structure to the side of the other structure, thus joining the two halves by hybridization[12]. To enable the formation of a full prism, the two halves are designed to be shape-complementary to one another.

In addition to the desired products, the final reaction contains excess DNA staple strands and some by-products, like misfolded objects and higher order aggregates (these are limited as much as possible by optimizing the self-assembly conditions). Thus purification and enrichment steps are crucial for the application of the cage. The method chosen to purify the final product is the precipitation by use of a depleting agent like polyethylene glycol (here PEG-800 is chosen): this technique is based on the capability of the polymer to attract water molecules away from the solvation layer around the origami, thus causing its precipitation. This method enables the samples to be concentrated and the buffer adjusted, but it introduces residual PEG molecules[41].

The nanocage has been characterized using both ensemble and single-molecule techniques. Agarose gel electrophoresis is used to assess the self-assembly process. The agarose gel is formed by bundles of polysaccharide molecules interconnected in a 3D network by hydrogen bonds, it is characterized by larger pore sizes and a lower resolution capability than polyacrylamide gels, which makes it particularly suitable for separating large DNA origami from small staple oligonucleotides. Since the gel is stabilized by weaker non-covalent bonds, it is mechanically fragile and must be casted in horizontal chambers and kept in an ice-water bath. Upon application of an electric field, the DNA molecule migrates along the gel matrix according to its size, charge, and shape. At the working pH, DNA origami possesses a net negative charge given by the phosphate groups, so it will run towards the anode (the positive electrode). DNA is then visualized by staining the gel with Ethidium bromide (EtBr), an intercalating UV-fluorescent dye. It's an aromatic compound that, when exposed to ultraviolet light, will emit fluorescent light at 605 nm. Ethidium bromide intensifies its fluorescence almost 20 times after binding with DNA, because water acts as an efficient quencher for this dye, so in solution its emission is very low, while when it binds with DNA and intercalates between the base pairs, the hydrophobic environment increases its fluorescence.

In *Figure 10*: Agarose gel electrophoresis of the NE6 DNA origami nanocage the correct production of the NE_6 full cage has been assessed: the two half-cages (E_3 and N_3) and the full cage NE_6, before and after the purification, have been run on the gel to check their correct assembly. The excess of staples migrates faster and can be seen at the bottom of the gel. These staples are completely absent in the purified product (lane 4) while they are still present in the other lanes, meaning that they were successfully eliminated by the PEG precipitation step. There are some high molecular weight aggregates in the samples (the pale stains above the cage) that are only partially removed by the purification, but they don't affect the suitability of the cage for further

reaction and conjugation. These bands are typically associated to dimeric forms of the loaded origami and often result by unspecific stacking at the high concentration used to prepare the samples. The shade at the very bottom of the gel is the running front. Altogether, the gel shows that the assembly was successful.

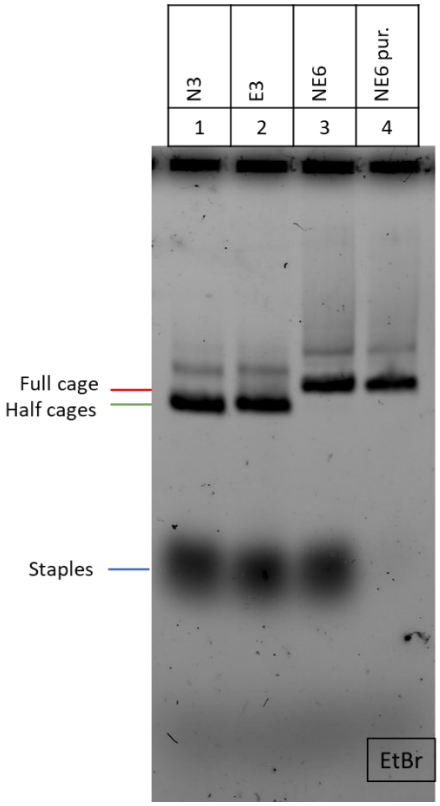

Figure 10: Agarose gel electrophoresis of the NE_6 DNA origami nanocage

In general, gel electrophoresis is a great tool to gather a global picture of the assembly process through the characterization of the average molecular weight of the loaded species. It is also a simple and resource-friendly analysis method. Nevertheless, its main disadvantage is that it cannot provide in-depth structural insights of the structure. For this purpose, single molecule techniques, such as atomic force microscopy (AFM) and transmission electron microscopy (TEM), are implemented.

AFM is a high-resolution technique, which allows to analyse the topography of a sample with nanometre resolution. The instrument has a mechanical probe able to sense the intermolecular forces occurring between the tip and the sample deposited on an atomically flat surface. The AFM image of the half cage E_3 in *Figure 11*: AFM images of the E3 DNA origami half cage was

performed in air. The structure deformed during imaging but it's still possible to appreciate the three-faced feature of the half prism. Formation of the correct structure was almost quantitative.

For characterizing 3D origami structures, the TEM instrument is more suitable. Uranyl formate is commonly used to produce negative stain contrast in TEM micrographs because it's excluded from the densely packed DNA structure, which instead becomes bright and visible. The images are used to assess the successful formation of the nanocage. Figure *12*: TEM image of the NE6 DNA origami cage illustrates a representative micrograph area of negatively stained NE_6 DNA origami. The cage appeared as a rectangular shape with three sections: a darker one flanked on both sides by a brighter region, as expected by a 2D projection of the full cage when observed from the side (long axis). From these images one can assess the correct and complete self-assembly process of the cages, the cages are quite stable, they don't lead to the formation of higher order aggregates and there are no half cages left uncoupled. This synthesis is thus considered to be well-established and characterised.

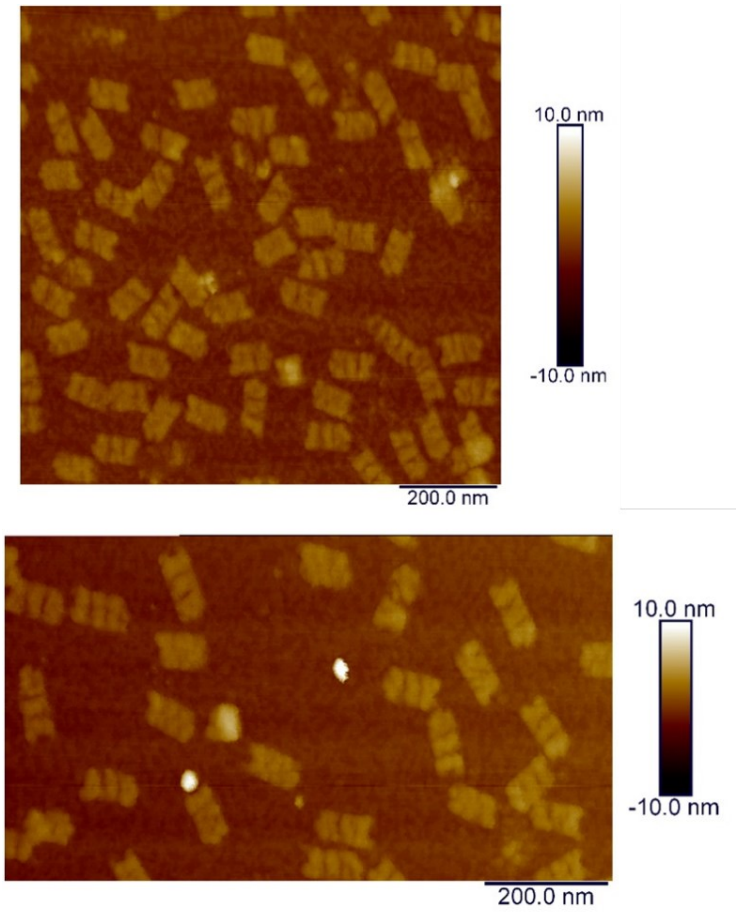

Figure 11: AFM images of the E_3 DNA origami half cage

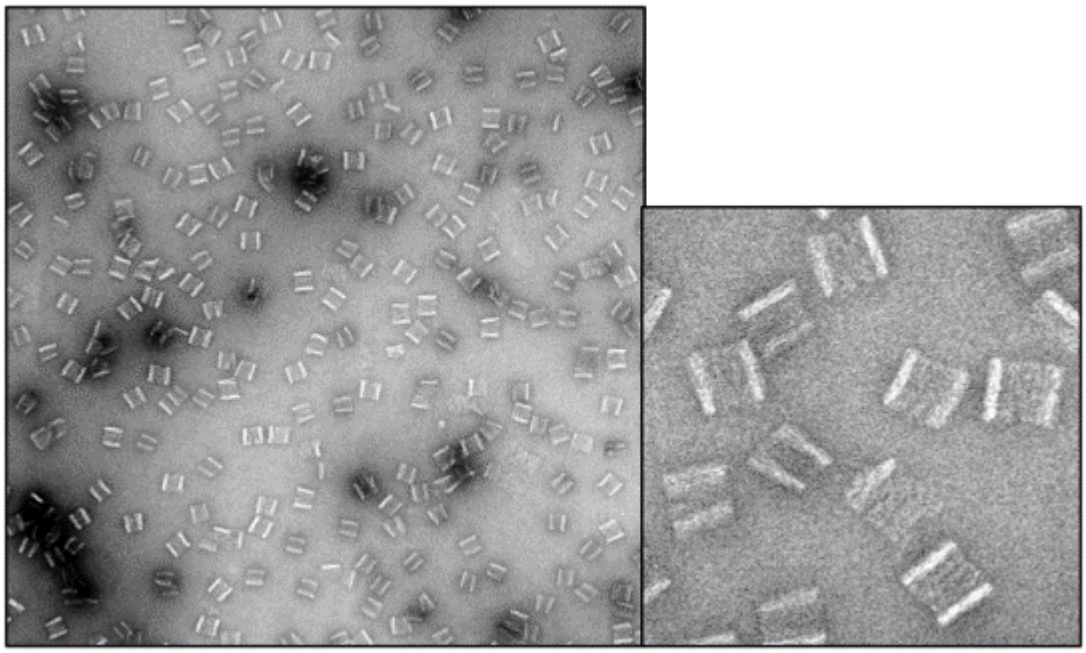

Figure 12: TEM image of the NE_6 DNA origami cage and a zoom in.

3.2. Synthesis of a covalent DNA-protein conjugate through the hetero-crosslinker AMAS

To enable the binding of the HTRA1 protein to the NE_6 DNA nanocage a covalent link is established between a single-stranded DNA and the protein. Different strategies for the synthesis of DNA-protein conjugates are nowadays available. This conjugation is based on the crosslinker N-[α-maleimidoacetoxy] succinimide ester (AMAS). The conjugation is a stepwise process: the first step is the coupling of the NH_2F9^{FAM} to the AMAS, then the functionalized oligo is conjugated to HTRA1, that is pre-treated with TCEP (a reducing agent for the removal of disulphide bonds).

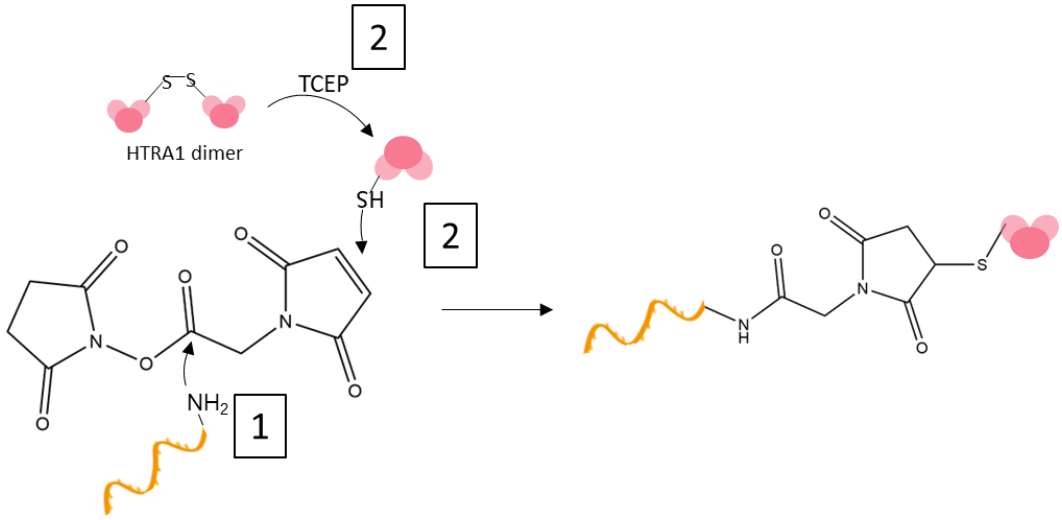

Figure 13: Reaction scheme of the conjugation

A bifunctional hetero-crosslinker is a small molecule, which bears two orthogonal chemical groups reactive to two different functional targets, one is typically placed on a protein of interest and the other in the macromolecule that is intended to be linked. The AMAS linker contains an amine reactive N-hydroxy-succinimide (NHS) ester group at one end (this reacts with the amino-modified F9 oligonucleotide), and a maleimide function at the other end (this can be utilized to react specifically with the cysteine residues genetically introduced in the HTRA1). The conjugation has been performed using both active HTRA1 and inactive HTRA1.

The N-hydroxy-succinimide (NHS) ester is perhaps the most commonly used amine-reactive group. NHS-containing reagents react with nucleophiles with release of the NHS leaving group to form an acylated product. In the presence of amine nucleophiles that can attack at the electron

deficient carbonyl of the active ester, the NHS group rapidly leaves, creating a stable amide linkage with the amine compound. Sulfhydryl and hydroxyl groups may also react with such active esters, but the products of such reactions, thioesters and esters, are unstable in aqueous environments or in the presence of amine nucleophiles. NHS esters have a half-life on the order of hours under physiological pH conditions, however, hydrolysis and amine reactivity both increase with increasing pH (at 0°C and at pH 7.0, the half-life is typically 4 to 5 h, but already at pH 8.6 and 4°C the half-life is only 10 min).

Maleimides are derivatives of the reaction of maleic anhydride and an amine derivative. The double bond of maleimides may undergo an alkylation reaction with sulfhydryl groups to form stable thioether bonds. The reaction that takes place is a Michael's addition, in which one of the carbons adjacent to the maleimide double bond undergoes nucleophilic attack by the thiolate anion to generate the addition product. The high reactivity of the maleimide olefin is due primarily to the ring strain arising from the bond angle distortion and the positioning of the carbonyl groups in the cis conformation[42]. Maleimide reactions are specific for thiols in the pH range of 6.5 to 7.5, but at higher pH values some cross-reactivity with amino groups takes place. Although the thioether bond formed between a sulfhydryl and a maleimide group is reversible under certain conditions, the maleimide moiety itself is relatively stable to degradation.

For these reasons, conjugation experiments involving this type of heterobifunctional crosslinker are usually performed at pH 7.2 – 7.5, with the NHS-ester (amine-targeted) reaction being accomplished before the maleimide (sulfhydryl-targeted) reaction.

The buffer used is PBS, a phosphate-buffered saline solution, with a pH of 7.2. This buffer was chosen because it doesn't contain primary amines that may compete with the NHS ester coupling.

The F9 oligo is labelled with a FAM (fluorescein) chromophore at the 5' terminus for monitoring the reaction via a spectrophotometer. Fluorescein represents one of the most popular among all fluorescent labelling agents. Its fluorescent properties are due to the presence of a multi-ring aromatic xanthene core structure, which creates the rigid planar nature of its upper, fused three-ring system. Fluorescein has an effective excitation wavelength range of about 488 to 495 nm, and the emission spectrum occurs between 518 and 525 nm[43]. Since we are working with fluorescent labels, the reagents have to be kept in the dark as much as possible.

The amino-modified F9 oligo purchased from IDT was initially purified from small impurities and primary amines contaminates present in the crude product. This operation is done by

isopropanol precipitation. In this method, the DNA is precipitated by adding sodium acetate (CH_3COONa) and isopropanol. The salt neutralizes the charge of the phosphate backbone of DNA, making it less hydrophilic and therefore less soluble in water. Isopropanol has a lower dielectric constant than water, which enables better interaction of phosphate residues with salt, so that the charge of DNA can be better shielded. As a consequence, the DNA will precipitate. The benefit of this process is that one can, in one step, purify and concentrate our sample to a very small volume. After the precipitation, F9 is washed with ethanol to remove isopropanol, dried and resuspended in PBS buffer.

The F9 oligo is reacted with a 50-fold excess of AMAS for 2 hours at room temperature. A UREA-PAGE was performed to check the successful coupling of the oligo with AMAS, but the difference in the molecular size was too small to be detectable by gel electrophoresis. Another method to evaluate the reaction could have been MALDI-TOF mass spectroscopy. Unfortunately, for this spectroscopy the sample must be pure from salt, and even if the oligo was desalted with a spin column and rebuffered in double distilled water from PBS buffer, the residual of salts was still too high to obtain a reliable mass spectrum. The analysis showed multiple peaks, with m\z ratios different from what expected, probably due to the presence of various types of counterions that crystallized with the oligo during sample preparation.

In the meantime, the HTRA1 protein is reacted with tris(2-carboxyethyl) phosphine (TCEP) to reduce the disulphide bond formed between the terminal cysteine residues of the protein. TCEP is a phosphine derivative that is very stable in aqueous solution, and it doesn't react with other common reactive groups of biomolecules. The reaction of TCEP with biological disulphides proceeds with initial cleavage of the S–S bond followed by oxidation of the phosphine (
Figure **14:** Chemical structure of tris (2-carboxyethyl) phosphine (TCEP) and its mechanism for reducing disulphide bonds.). The stability of the phosphine oxide bond that is formed in this process is great enough to prevent reversal of the reaction. Since this reaction is performed without any added –SH compounds, subsequent conjugation with the generated sulfhydryl groups can be carried out without removal of excess TCEP or reaction by-products[44].

200-fold excess of TCEP was added to the HTRA1 protein to ensure the complete breakage of the disulphide bond that causes the aggregation of the protein. Such a high amount was found to be necessary to avoid the formation of dimers.

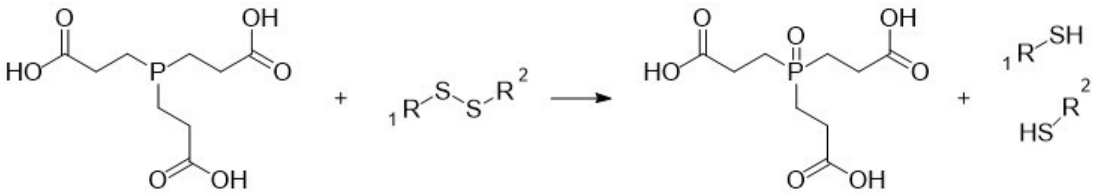

Figure 14: Chemical structure of tris (2-carboxyethyl) phosphine (TCEP) and its mechanism for reducing disulphide bonds.

The AMAS-modified F9 oligo is mixed with the TCEP and the protein in ratio of circa 1:1 and incubated for about 12 hours a 4°C in the fridge. The NH_2-$F9^{FAM}$ has to be purified from the unreacted AMAS with another isopropanol precipitation step before reacting with HTRA1 and TCEP.

Following conjugation, the purification of the DNA-conjugated protein is especially important to produce a high-quality sample. The presence of unconjugated DNA or free protein decreases the yield of encapsulated protein into the DNA nanostructure or may lead to unspecific aggregations. The purification method chosen for isolating the DNA-HTRA1 conjugate is the ion exchange chromatography (IEC). This chromatographic method relies on an anion exchanger optimised for separating DNA oligos from excess of unconjugated proteins. In this column, anions bind to cationic groups linked to the solid matrix of the stationary phase[45]. The eluant's salt concentration increases progressively during the elution time so that the species that bind tightly to the ion exchanger can be eluted by applying a higher salt concentration that reduces the affinity with which the matrix binds to them. The column effluent is monitored for the presence of DNA by measuring its absorbance under UV light illumination at 260 nm.

The three main components of the reaction mixture are the desired conjugate and the two possible unreacted species, the HTRA1 protein and the oligo. The protein is positively charged, because of the numerous lysines present in its primary sequence (that are charged at the working pH, which is circa 7.5). Thus, the protein is the first component of the reaction mixture that will be eluted from the column. The F9 oligo contains multiple negative charges (the phosphate groups of the backbone), so it is the last component to be revealed in the chromatogram, at around 22 minutes. The conjugate has both properties from the two species, and it will show an intermediate behaviour. In practice it will be present almost as a shoulder before the oligo elution peak.

Two conjugation reactions were performed in parallel: one with the active form of the HTRA1 protein and one with the inactive form. The very small elution peak corresponding to HTRA1 appeared at around 4 minutes. Then some peaks appeared at higher retention times: in the purification of the active HTRA1 there was only one peak at 16 min (*Figure 15*: IEC chromatogram of active HTRA1, with the peak at 16 min.*)* while for the inactive HTRA1 there were 3 peaks collected at 17 min, 19 min and 21 min respectively (*Figure 16*: IEC chromatogram of inactive HTRA1, with 3 peaks collected at 17 min, 19 min, 20 min.).

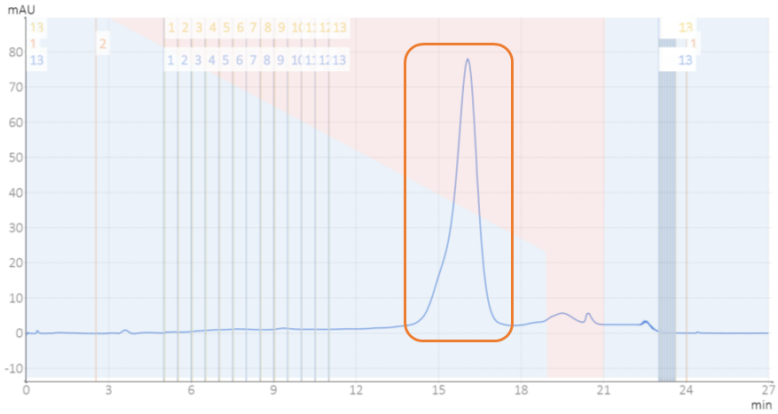

Figure 15: IEC chromatogram of active HTRA1, with the peak at 16 min.

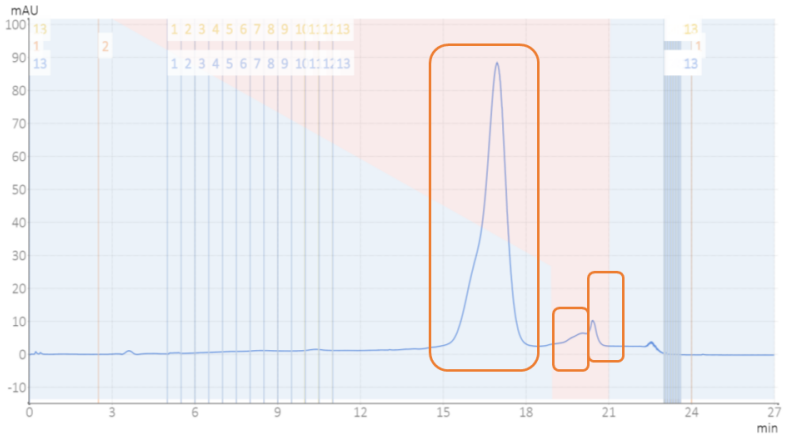

Figure 16: IEC chromatogram of inactive HTRA1, with 3 peaks collected at 17 min, 19 min, 20 min.

These peaks were collected in a series of fractions, and then concentrated and rebuffered in PBS with a 10 kDa MWCO filter (Ultrafree Amicon centrifugal filter devices from Millipore). The conjugate concentration was measured by absorbance at 260 nm with the Nanodrop. One should consider however that by measuring the absorbance at 260 nm, the protein component present

in solution will not be reliably estimated since it has a specific absorbance peak at 280 nm that may also partially overlap with the DNA spectrum. A better evaluation of the DNA-conjugate concentration should be carried out using the 260-nm and 280-nm absorbance and the specific extinction coefficients of the DNA and protein, individually. But since the conjugate will be added in stoichiometric excess to the origami cage to favour the formation of the product, an accurate estimation of the concentration of the DNA-protein conjugate at this stage is not critical, particularly because the resulting DNA-protein nanostructure will be later exposed to a further purification step.

The purified product is then analysed with SDS-PAGE. In this technique, molecular separation occurs by gel filtration (size and shape) as well as electrophoretic mobility (electric charge). A detergent, specifically sodium dodecyl sulphate (SDS) $[CH_3\text{-}(CH_2)_{10}\text{-}CH_2\text{-}O\text{-}SO_3^-]\ Na^+$, is used to denature the protein. Amphiphilic molecules such as SDS interfere with the hydrophobic interactions that normally stabilize proteins. The large negative charge of SDS masks the proteins' intrinsic charge, enabling the separation of proteins according to their molecular mass. In SDS-PAGE, the relative mobilities of proteins vary approximately linearly with the logarithm of their molecular masses (*Figure 18*). Because SDS disrupts noncovalent interactions between polypeptides, SDS-PAGE yields the molecular masses of the monomers of multi-subunit proteins[45].

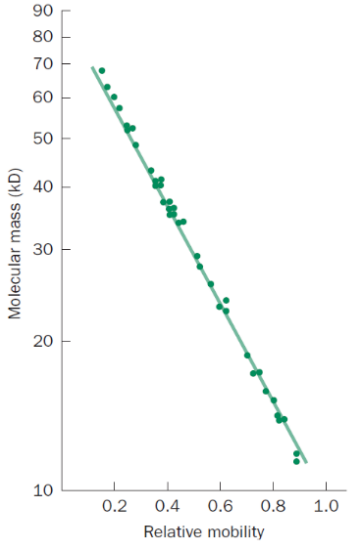

Figure 17: Logarithmic relationship between the molecular mass of a protein and its electrophoretic mobility in SDS-PAGE. The masses of 37 proteins ranging from 11 to 70 kDa are plotted.[46]

Depending on the pore size of the gel and the visualization technique used, samples containing less than a nanogram of protein can be separated and detected by gel electrophoresis.

The SDS-PAGE analysis is performed to assess the trend and success of the conjugation. For this reason, besides the samples isolated from IEC, the crude reaction mixture is loaded and the HTRA1 protein a the NH_2-$F9^{FAM}$ oligo are added as controls. A protein ladder is also added in the first lane as a MW reference. The protein ladder is a mixture of "marker" proteins of known molecular masses, including those of the species under investigation, and it makes possible to identify the molecular mass with a 5% to 10% accuracy. The gel bands can be visualized by different techniques: first of all, the gel is scanned under various conditions of illumination/emission to visualize the products containing fluorophore dyes (for example, FAM). Then the gel is soaked in a solution of SYBR GOLD stain, an asymmetrical cyanine dye, sensitive to DNA and 25 to 100 times more sensitive than ethidium bromide. Thus, this staining procedure is suitable even for the detection of small oligos. The proteins are detected by using. a second staining solution containing Coomassie brilliant blue that binds tightly to proteins.

The HTRA1 control shows two bands, the main band is the monomer, that can be identified by the molecular mass (36 kDa) and the fact that it's visible only in the Coomassie stain. The second band is the dimer formed by disulphide bonds. The NH_2-$F9^{FAM}$ control migrates at the very bottom of the gel as a consequence of its low molecular weight (5718,9 g\mol), and is visible in the FAM channel (thanks to the labelling) and after SYBR GOLD staining.

The conjugate is identified by its molecular mass, that is roughly the sum of the HTRA1 and the NH_2-$F9^{FAM}$, so around 41 kDa. This band is visible in all three channels and confirms the presence of both the protein and the labelled DNA oligo in the species. The gels performed in the first attempt of this synthesis are already indicative of some of the problems that characterized this conjugation. In the reaction mixture's lane, there are several bands that show the presence of numerous by-products. These by-products can be due to the presence of impurities in the adducts which contaminate the reaction. In particular, since the HTRA1 protein is a protease with auto-catalytic activity, it is able to digest itself, generating many protein fragments during the reaction. These fragments may also react with the AMAS-coupled oligo, resulting in low-molecular weight by-products, which are difficult to stain by Coomassie. The same problem occurs with the inactive protein but to a less extent. A protein band at around 75 kDa appears, and is attributed to a covalent HTRA1 dimer formed by two monomers linked

by a disulphide bond. This band is present in the protein control and in the reaction mix, although there is a high amount of reducing agent TCEP. Moreover, the reaction is not as regioselective as it is supposed to be. An explanatory example is the band in the FAM channel with the same molecular weight of the dimer. Here, the only available cysteine residues are occupied in the formation of a disulphide bond; therefore, the dimer shouldn't be able to react with the maleimide-tagged oligo. The presence of DNA in the dimer band can be thus explained only by assuming that the maleimide groups can react with one of the numerous amino groups on the surface of the protein. But the most important problem is that in the purified sample there is no trace of conjugate. Probably, most of the product got lost during the purification process, for reasons that are not yet clear and are currently under investigation.

The very low reaction yield (< 1%) may have different causes. Since the conjugation is a stepwise synthesis (with first the coupling of AMAS to the oligo and then the linking to the protein), it requires multiple purification phases that cause inevitably some loss. A part of the reagents degrades during the reaction time, in particular AMAS undergoes hydrolysis and the HTRA1 is subjected to autoproteolysis (a process in which the peptide bond is cleaved in a self-catalysed intramolecular reaction). In addition, alternative reaction routes may take place. Finally, the purification procedure may not be sufficiently accurate to resolve the various species in solution.

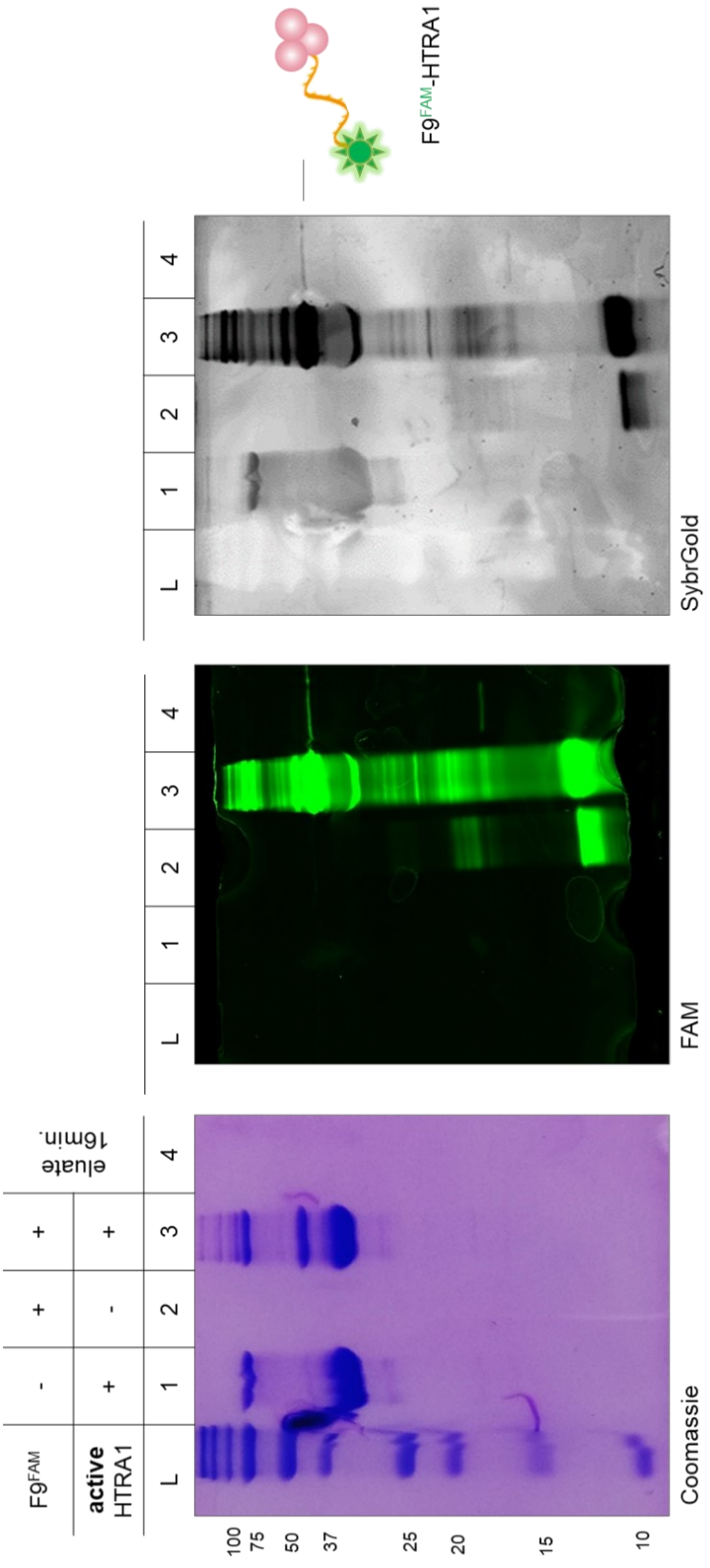

Figure 18: SDS-PAGE of active HTRA1-NH_2F9^{FAM} conjugate reaction.

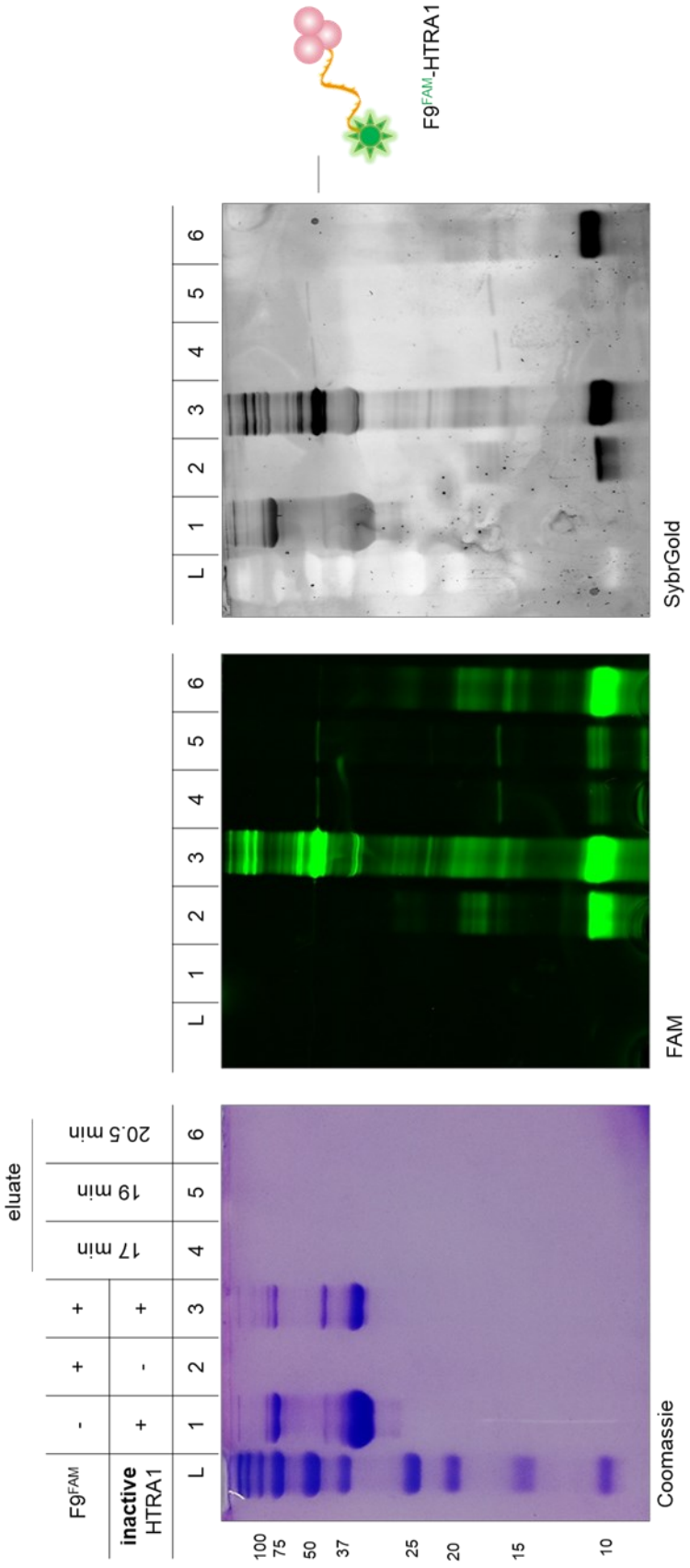

Figure 19: SDS-PAGE of inactive HTRA1-NH_2F9^{FAM} conjugate reaction.

3.2.1. Optimization

The HTRA1-NH_2F9^{FAM} conjugation synthesis required therefore an in-depth optimization to address all its problems. The challenges that prevent the formation of a conjugate with good yields concern both the purification method (that wasn't able to fully separate the conjugate and caused a big loss of product) and the synthesis itself, that should be made more selective and efficient.

The time-course of the reaction was examined to find the optimal reaction time for formation of the product. In order to do this, the reaction was stopped at different times by freezing the solution mix at -80°C with liquid nitrogen. The aliquots were then analysed by SDS-PAGE to see the amount of conjugate obtained over time. The gel in *Figure 21* shows that the best yield of conjugation is reached after 2 hours. Afterwards, there is no increase in product formation and instead some degradation seems to occur. Anyway, the intensity of the conjugate band is not sufficient to gain information about the reaction, because it may be influenced by sample preparation and the amount of sample loaded as well as by other experimental errors. A better way to interpret gel results is by using the ratio between the intensity of the conjugate band and the intensity of the other bands, like the F9 oligo band in the FAM channel or SYBR GOLD, and the HTRA1 band in the Coomassie.

The first attempt involved the modification of the buffer gradient in the ion exchange chromatography device. The eluant was flown at half the speed as in the standard elution program, so that the concentration changes more slowly, releasing the reaction mix components more gradually. The purification took double the time, but the shape of the chromatogram didn't change significantly. As shown in *Figure 22*, the conjugation peak is still not isolated from the F9 oligo peak. This is confirmed also by the SDS-PAGE analysis (*Figure 23*): the purified product contains the conjugate, but it also contains another side product, that is not the desired conjugate because it has a lower molar weight (between 15 and 20 kDa from the ladder in the Coomassie). This side product is even more present than the conjugate.

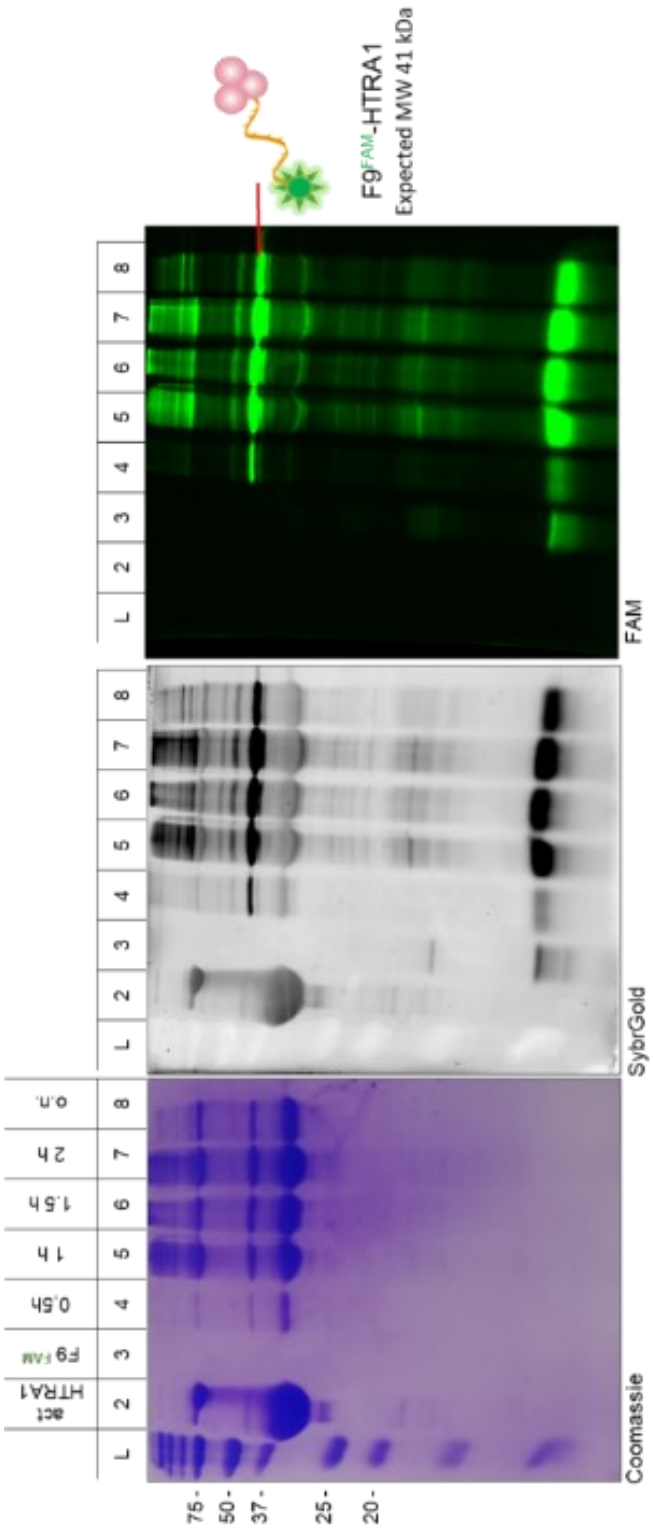

Figure 20: SDS-PAGE control of active HTRA1-NH_2F9^{FAM} reaction through time.

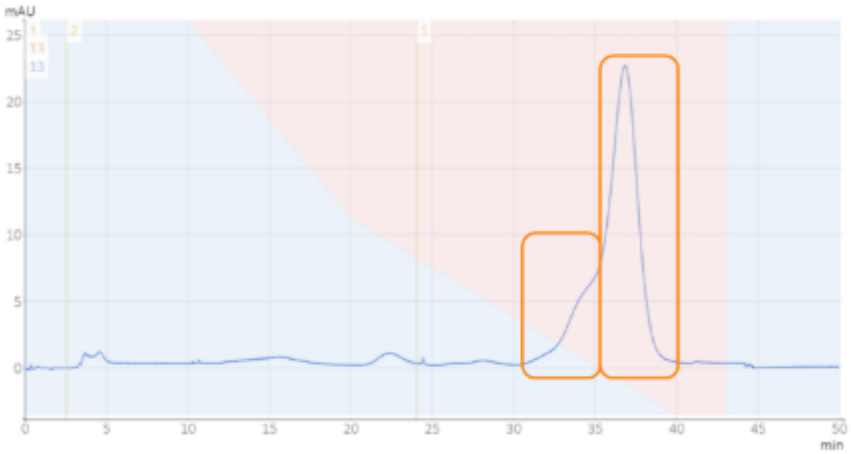

Figure 21: IEC chromatogram of active HTRA1, with the peak at 33,5 min and a second peak at 37 min, here the elution time has been doubled.

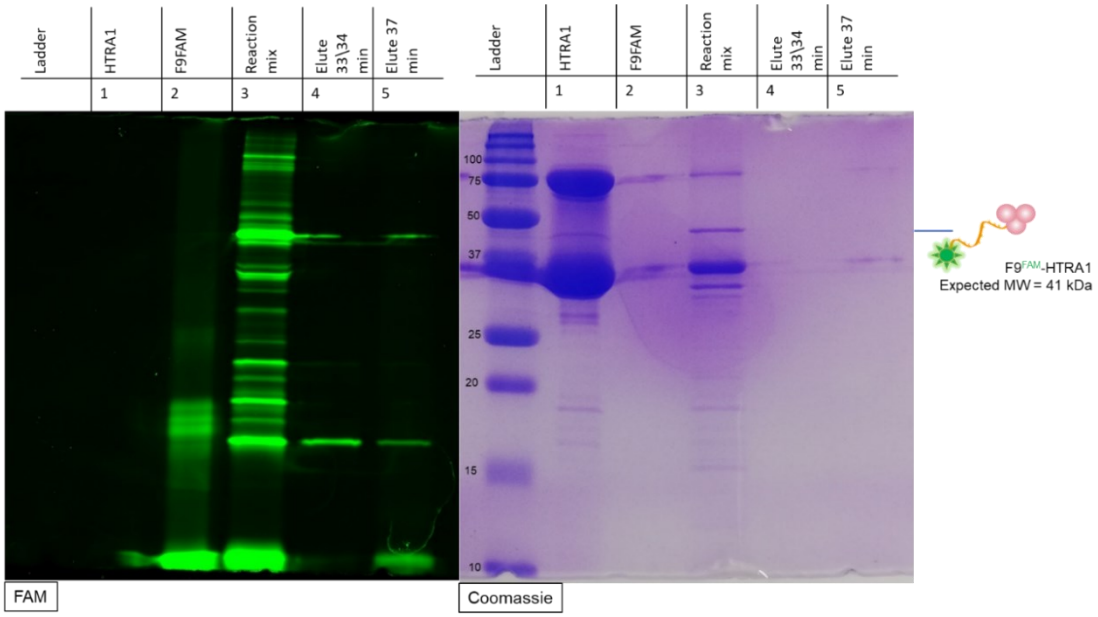

Figure 22: SDS-PAGE of active HTRA1-NH_2F9^{FAM} conjugate reaction purified with IEC.

Since IEC separation of the conjugation product from the starting materials was not successful, other purification strategies were tested. A method based on Zhou et al.' s approach[47] was tested, in which magnetic beads are used to purify a protein of interest via a biotin displacement strategy from streptavidin-coated magnetic beads (MB). The displacement relies on the much stronger affinity of biotin to streptavidin than a biotin analogue called D-desthiobiotin. In our approach, instead of using a biotin derivative, the displacement is accomplished through toehold mediated single-strand displacement (TMSD).

Biot-cF9(22), an oligo complementary to the F9 but elongated of 5 nucleobases and functionalized with biotin, was bound in solution to streptavidin coated magnetic beads (the two species display one of the strongest non-covalent bindings in nature, with a dissociation constant Kd < 10^{-12} M). The reaction mixture is added to the magnetic beads solution, so that the F9(16) recognizes and binds the complementary cF9(22) single strand, leaving 6 nucleobases unpaired. Washing the MB allows the elimination of all the species that are not attached to the beads. The F9(16) oligo is removed from the magnetic beads by F9(22) following the TMSD. The so-called "toehold" is the unpaired binding region on the duplex and represents the recognition and anchoring point for the removal of the F9 oligo. Clearly, recognition and binding are facilitated when the toehold is longer, typically, 4 bases-long toehold stretches are sufficient to start a TMSD reaction at room temperature, and in this case the toehold is 6 bases-long. F9(22), initially attached only on the toehold to the cF9(22), invades the nearby duplex region and migrates onto the complementary region occupied by F9. By doing that, the F9 is finally displaced and unbound from the complex. Despite the invasion and branch migration step should be viewed as a reversible process (some bases are hybridized and de-hybridized), the global direction of the reaction is eventually shifted towards the final product. The reason for that is simple: the final state of the system includes a longer (and fully hybridized) duplex made of 22 base pairs and such a duplex is thermodynamically more stable than the initial one, which has only 16 hybridized bases. This mechanism is used to dynamically reconfigure DNA nanostructures and, if suitably designed, can result in smart nanodevices and reaction networks for signal propagation, like the one firstly proposed by Yurke et al[48].

Figure 23: Schematic representation of the single stranded displacement mechanism during magnetic beads purification.

This purification method is quite simple and it's very efficient in separating the conjugate from the excess of protein, but it can't separate the conjugate from the excess of oligo and – unfortunately in our case - the unreacted oligo was always abundant in the reaction mixture. Furthermore, the washing steps required during this procedure caused every time a loss of magnetic beads and thus product. The SDS-PAGE analysis in

Figure 24: SDS-PAGE of HTRA1-NH2F9FAM conjugate synthesis purified with magnetic beads. shows that the purified product contains only a small amount of NH_2F9^{FAM}, without the conjugate. This result can also be attributed to the fact that. the DNA part of the conjugate is sterically hindered by the large HTRA1 protein and the recognition with the cF9(22) on the magnetic bead is more difficult and less statistically probable than it is for the oligo alone.

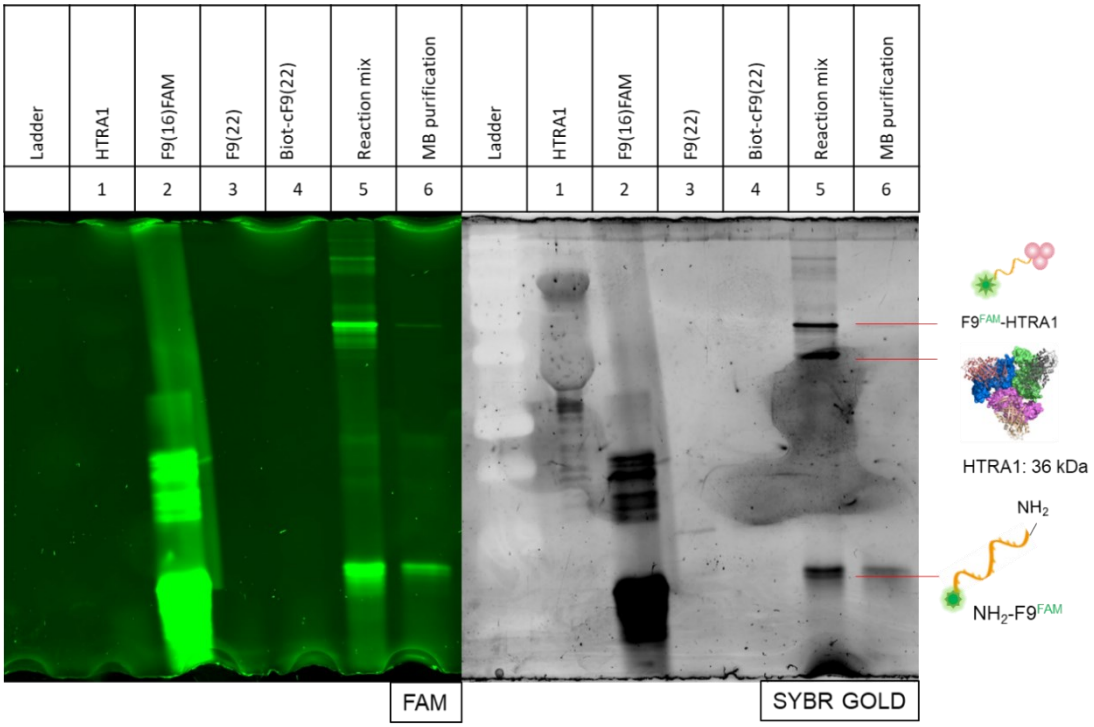

Figure 24: SDS-PAGE of HTRA1-NH_2F9^{FAM} conjugate synthesis purified with magnetic beads.

Finally, the synthesis that led to the product with the highest yield was obtained by optimizing and improving different aspects of the reaction.

First of all, the buffer of the solution was changed from PBS to HEPES buffer, which is the buffer used for HTRA1 synthesis. The 4-(2-hydroxyethyl)-1-piperazine-ethane-sulfonic acid (HEPES) is a zwitterionic buffering agent. Zwitterions are able to act as buffers because of their ability to act as both a base and an acid. This allows them to neutralize any base or acid added to the solution, keeping the pH relatively stable. Every zwitterion has a specific pH range and capacity that determines how much base or acid can be neutralized before the pH changes. At 25°C HEPES has a pK_{a1} = 3, and a pK_{a2} = 7.5, therefore the useful pH range in which this molecule has a buffering power is between 6.8 and 8.2.[49]

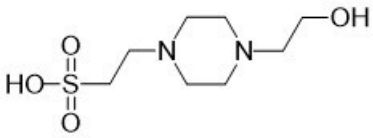

Figure 25: HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)

The conjugation was carried out using only the inactive HTRA1, that has a higher lifetime and therefore gives less auto-catalytic degradation by-products.

The F9 oligo was purified by denaturing PAGE extraction instead of isopropanol precipitation. Gel extraction allows a better removal of possible impurities from the purchased oligo (visible as fluorescently labelled by-products), even though this process has a much lower recovery yield (circa 60% compared to >90%) and it's time consuming (it takes three days to be completed, instead of around 45 minutes for the isopropanol precipitation). After this purification, a UREA-PAGE gel is performed (*Figure 26*: UREA PAGE of NH2F9FAM) to compare the purchased oligo with the purified oligo using the two purification methods. The gel extraction removed the low-weight fluorescent contaminants that created a shade under the band of the purchased and IP-purified oligo. The ubiquitous higher band is some unknown aggregation of the DNA molecule.

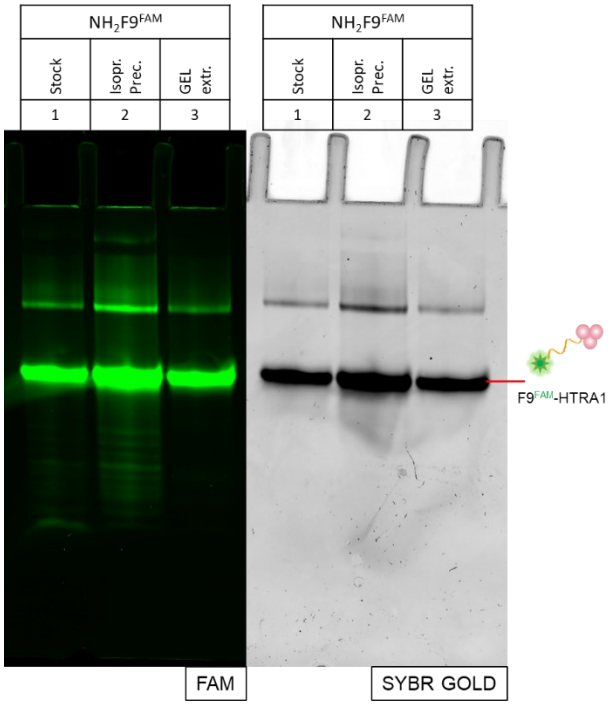

Figure 26: UREA PAGE of NH_2F9^{FAM}.

The conjugation reaction with the protein was performed using the oligo purified via gel extraction and the oligo purified via isopropanol precipitation to compare the effect of the two methods on the synthesis.

The reaction conditions of the last conjugation step were optimised to push the reaction: the AMAS-coupled oligo was concentrated as much as possible after the removal of excess crosslinker. The F9 oligo and the HTRA1 protein are both quite expensive reagents, but since part of the oligo is lost during the purification and the coupling with AMAS, it was chosen as the limiting reactant and an excess of protein was used.

Mixing AMAS-coupled oligo with HTRA1 and TCEP gives some precipitation issues. After some investigations, the instability was attributed to the presence of TCEP that drastically lowers the pH of the solution. This happens because TCEP is commercially available in the form of hydrochloride, that is the acid salt derived from the reaction of TCEP with hydrochloric acid, which means that its solution has a low pH (pH ≈ 1). Therefore, adding TCEP drastically alters the pH of the solution and causes the formation of precipitate. Interestingly, the precipitate forms when all three species are mixed together, oligo, HTRA1 and TCEP, but not for any other combination of two of them. This precipitate also contains the conjugate. This side effect makes the purification more difficult because the product may get stuck in the chromatographic column's matrix. The precipitation problem is limited by dissolving fresh aliquots of TCEP in buffer instead of milliQ water, and especially adjusting every time the pH and bringing it back to an ideal level with NaOH.

Despite the issues encountered, size exclusion chromatography was still considered for purification of the sample from the other reaction components. In size exclusion chromatography, molecules are separated according to their size and shape. The stationary phase consists of gel beads containing pores that span a relatively narrow size range. The pore size is determined by the extent of cross-linking between the polymers of the gel material. When molecules of various sizes are passed through a column containing such "molecular sieves," the molecules that are too large to pass through the pores are excluded from the solvent volume inside the gel beads. These large molecules therefore traverse the column more rapidly than small molecules, which instead pass through the pores. Within the size range of molecules separated by a particular pore size, there is a linear relationship between the relative elution volume of a substance and the logarithm of its molecular mass (assuming the molecules have similar shapes)[50].

Three different fractions have been collected from the chromatograms in *Figure 27*: SEC chromatograms of the HTRA1(SA)-NH2F9FAM reaction, with the oligo purified via isopropanol precipitation and gel extraction.. The latest peak has the lowest molecular weight and so it's assigned to the F9 oligo, the intermediate peak is the HTRA1 protein, and the conjugate elutes as a low broad curve next to the protein peak.

The SDS-PAGE analysis of the fractions collected from the two reactions are run together (*Figure 28*: SDS-PAGE of HTRA1(SA)-NH2F9FAM conjugate reaction purified with SEC.). The protein and oligo control are present. The fractions are just numbered in order of exit from the SEC column. Lane 5 (labelled as Is. Prec. 1) contains the conjugate in the highest amount and with the best purity so far. An additional DNA-containing band is present, which emits in the FAM channel but apparently doesn't contain protein components, since it's not visible in the Coomassie: this side product has a molar weight of approximately 20 kDa and its origin is unknown. As expected, the second collected fraction of each chromatogram is the one that contains most protein and the third one is mostly the F9 oligo.

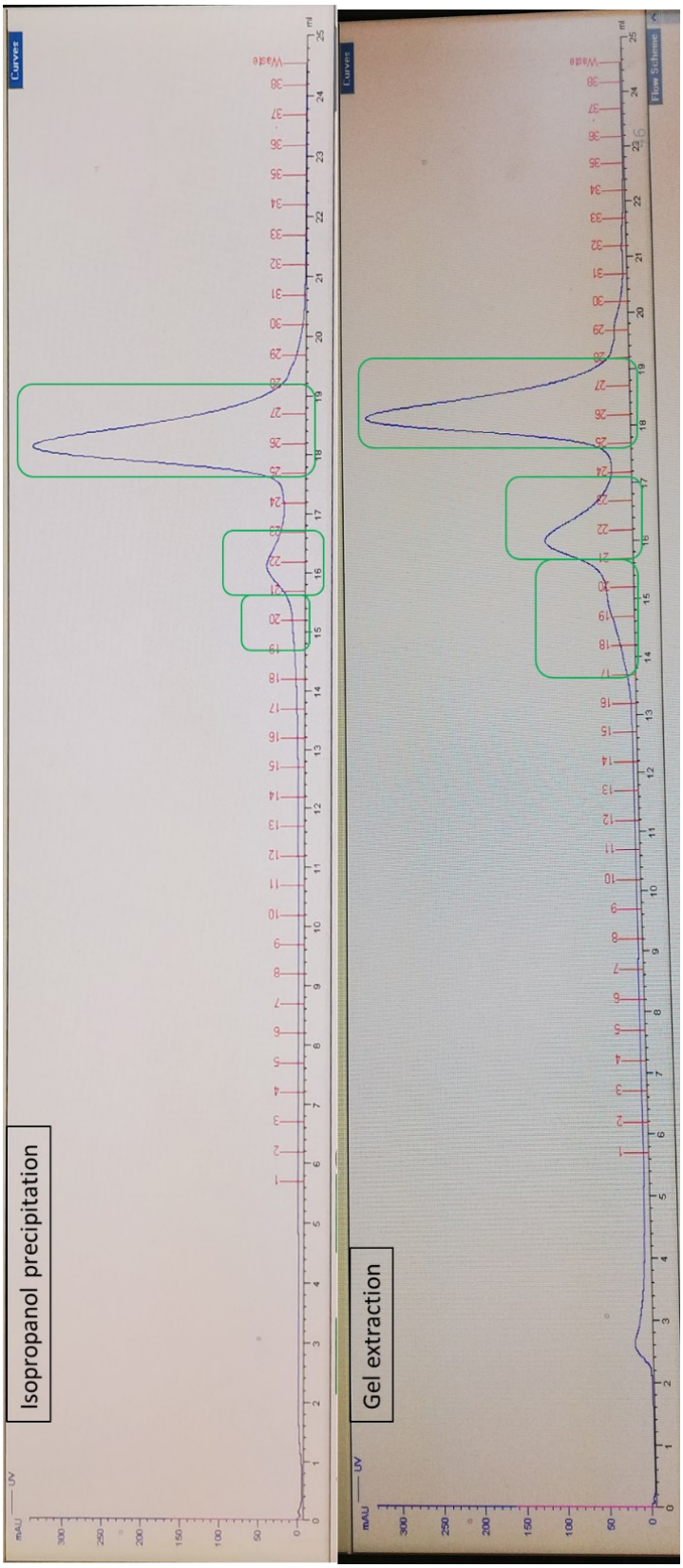

Figure 27: SEC chromatograms of the HTRA1(SA)-NH_2F9^{FAM} reaction, with the oligo purified via isopropanol precipitation and gel extraction.

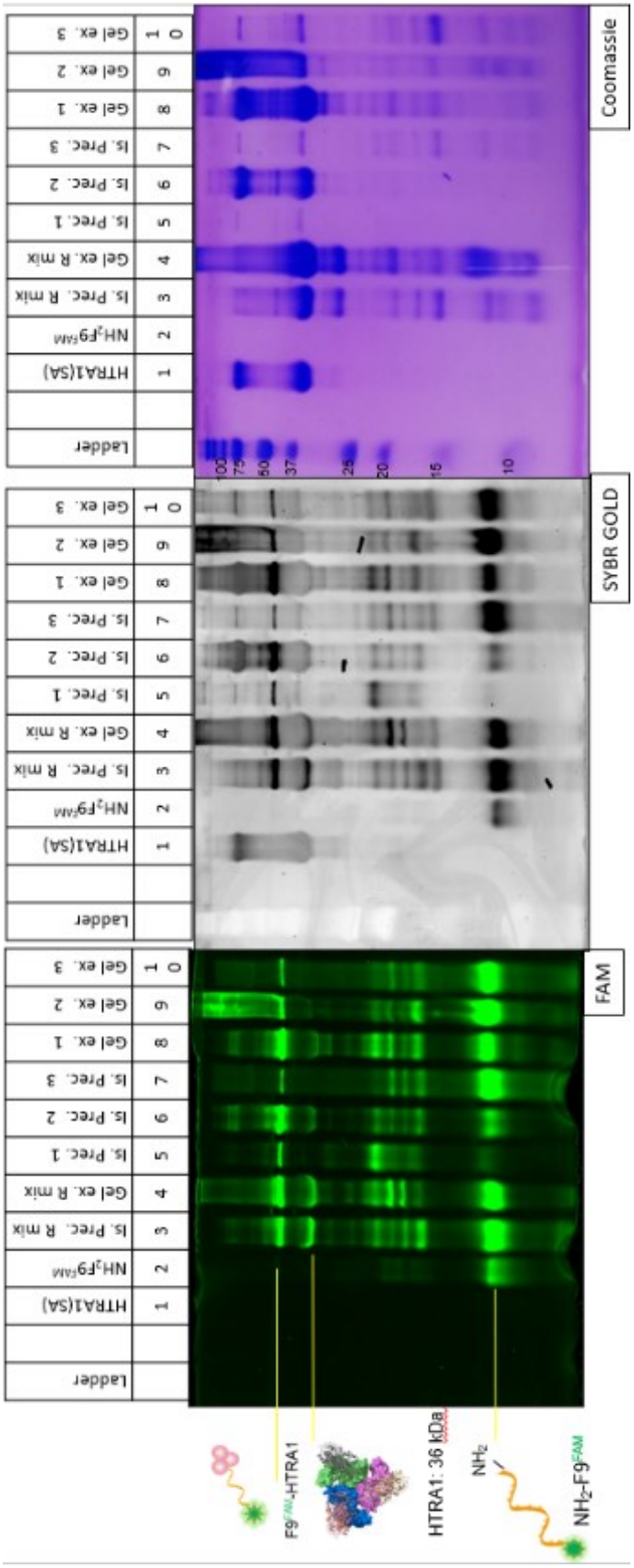

Figure 28: SDS-PAGE of HTRA1(SA)-NH_2F9^{FAM} conjugate reaction purified with SEC.

3.3. Encapsulation of the conjugate inside the NE_6 DNA origami nanocage

The HTRA1(SA)-NH_2F9^{FAM} conjugate isolated from the SEC is then used for encaging inside the NE_6 DNA origami previously prepared. The encapsulation is performed simply by incubating the conjugate with the nanocage in a ratio of 2:1 (conjugate:PA). Since the NE_6 cage contains six protruding arms, the concentration of the conjugate is 12-fold that of the DNA origami.

The encapsulation is assessed with an agarose gel. In this gel the HTRA1 protein (in lane 7) and the nanocage alone (in lanes 1, 2 and 3) are added as a control. Hybridization of the FAM-labelled conjugate to the complementary PAs present into the cage will result in the appearance of a FAM signal in correspondence to the DNA origami band. However, one should consider that also the FAM-labelled oligo alone (with no protein attached to it) will in principle give the same result. The gel in *Figure 30* shows FAM emission in correspondence of the NE_6 band. The low-MW fluorescent bands in lanes 4, 5, and 6 are given by the excess of conjugate in solution.

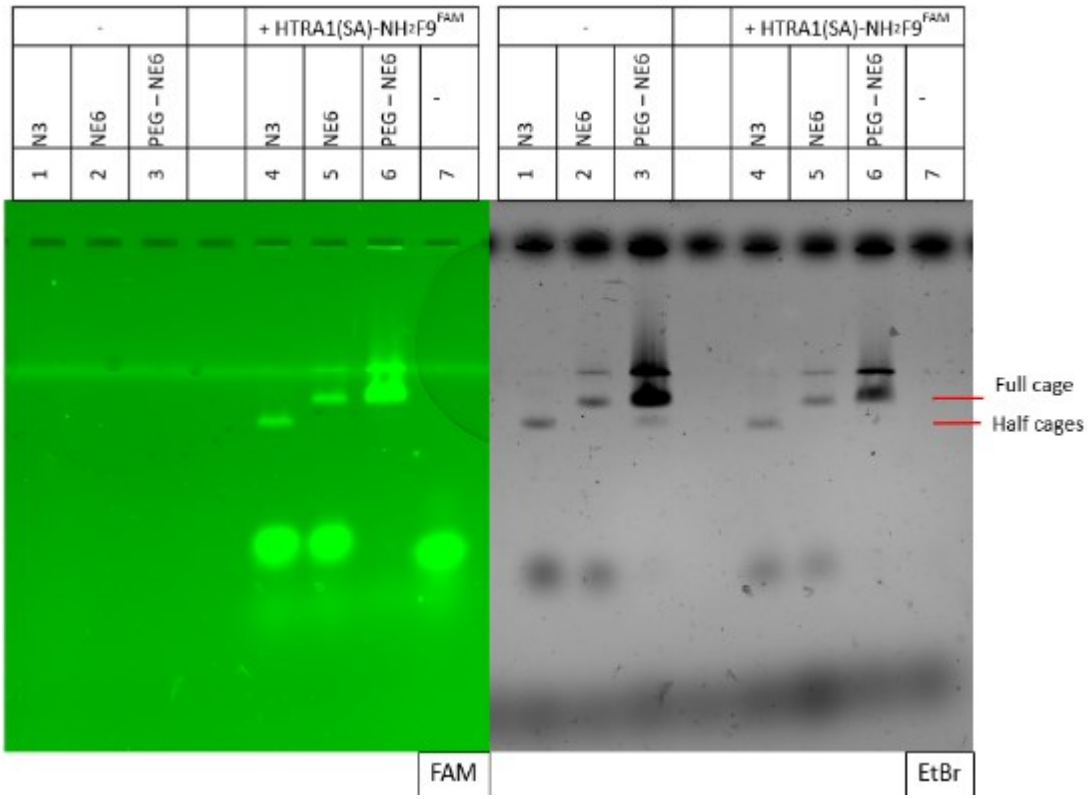

Figure 29: AGE of HTRA1(SA) encapsulation in the NE_6 DNA origami nanocage.

Once the encapsulation process is analysed with AGE, the next step is to purify the complex from the excess of conjugate. This can be performed using different methods[27]: here PEG purification was chosen for its reliability and high recovery yield. The procedure is identical to the purification of the NE_6 nanocage from the excess of staples.

To really confirm the presence of HTRA1 inside the walls of the nanocage, the best technique is transmission electron microscopy TEM. The protein should be visualised as a bright and small spot inside the DNA structure because it absorbs fewer electrons (leading to more transmitted electrons) than the uranyl-stained DNA helices, which instead absorb or scatter more electrons (fewer transmitted electrons). However, negative-staining TEM generally produces a low-resolution image of enzyme complexes, which can make it difficult to quantitatively evaluate the encapsulated enzymes. In *Figure 30*: TEM images of the HTRA1SA-NE6 complex with all empty cages. there is no sign of protein, which suggests that the encaging process actually failed.

The fluorescence at the nanocage seen in the agarose gel may be caused by the contaminant present in the conjugate (as previously characterised in *Figure 29*), that maybe contains the FAM-labelled F9 oligo but not the HTRA1 protein and it may have hybridised the DNA handles on the inner surface of the NE_6 nanocage in place of the desired conjugate.

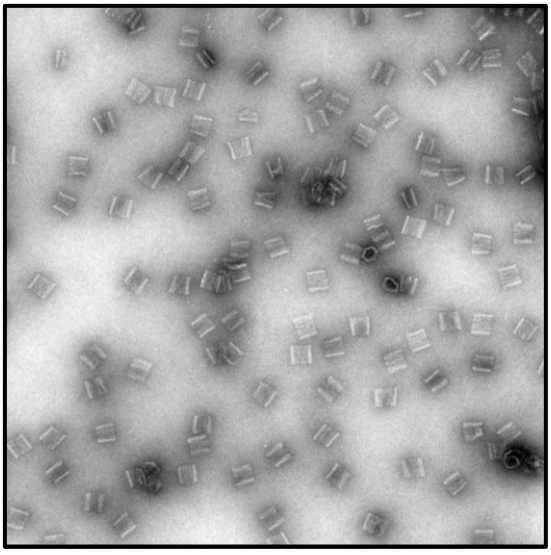

Figure 30: TEM images of the HTRA1SA-NE_6 complex with all empty cages.

3.4. Problems of the covalent conjugation chemistry

The failed encapsulation of the conjugate raised more questions about the conjugation method itself. Even if the reaction between a thiol and a maleimide to generate a thiol-succinimide product is one of the most popular methods for site-selective modification of cysteine residues in bioconjugation technology, in this work the downsides of this bioconjugation method made it is not suitable for the reaction with the HTRA1 protein.

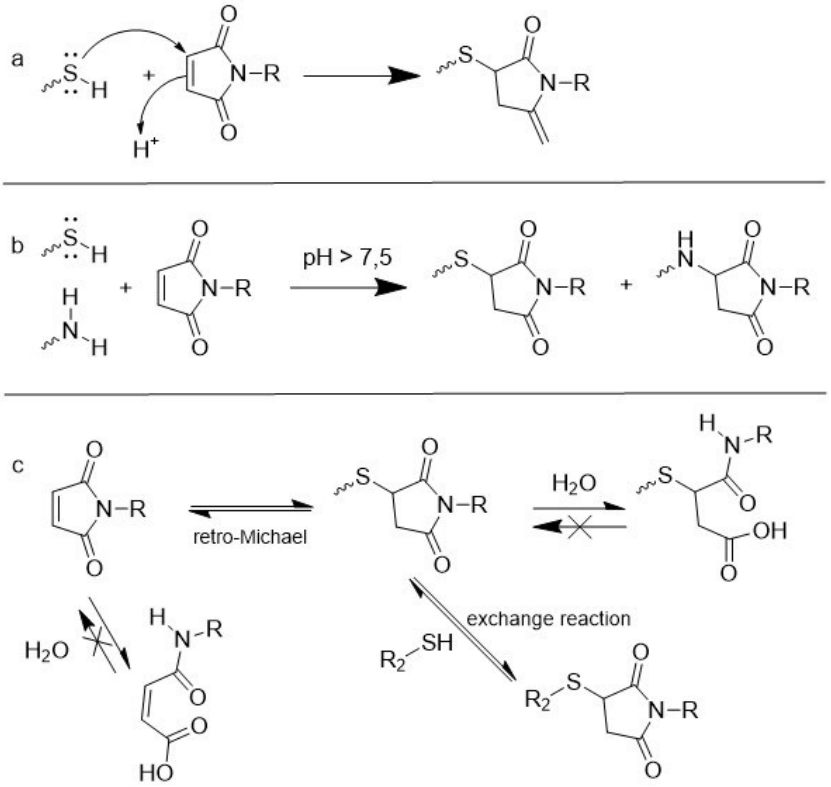

Figure 31: The reactions of maleimide: (a) the general mechanism of the thiol-maleimide reaction, (b) the competition between thiols and amines in the reaction with the maleimide group at basic pH, (c) some possible side-reactions of the thiol-maleimide reaction.

Maleimide reactions are specific for thiols in the pH range of 6.5 to 7.5, and even if at pH 7.0 the reaction of the maleimide with sulfhydryl groups proceeds at a rate 1000 times greater than its reaction with amines, at higher pH values some cross-reactivity with amino groups takes place. The control of the pH in the reaction seemed not to be sufficient to prevent the competitive reaction of free primary amines with thiols at the maleimide double bond, since

there are some by-products that contain the AMAS-modified oligo and the protein fraction with the thiol groups occupied in a disulphide bond.

In aqueous solution the maleimide ring can be opened by hydrolysis to an open maleamic acid form that is unreactive toward thiol groups: hydrolysis may also occur after sulfhydryl coupling to the maleimide, but in this case the opened ring is stable. This ring opening reaction typically happens faster the higher the pH becomes.[51] The unreacted oligo, that was always present in the reaction mixture, even if in sub-stoichiometric ratio, indicated that the degradation of maleimide in solution before reacting with the HTRA1 thiol group was significant.

Another big drawback is caused by the retro-Michael addition: the newly formed thioether bond can regenerate the maleimide in the presence of external thiols, thus making the reaction reversible. It is possible that the thiol group on the cysteine residue of HTRA1 is not easily accessible to the crosslinker, although the crystal structure of the protein shows that the N-terminus is located on the protein surface.

The issues coupled to the maleimide-chemistry described above are known in the literature and have led to the development of alternative Cys selective modification techniques that exploit the unique nucleophilic properties of the sulfhydryl group[45]. Palladium (II) complexes can be used for efficient and highly selective cysteine conjugation reactions to modify biomolecules[52]. The radical-mediated thiol–ene reaction is a simple and highly efficient chemistry, implemented in biofunctionalization and polymer modification[53]. Another method showed that ethynyl-phosphonamidates display excellent cysteine-selective reactivity combined with superior stability of the thiol adducts, when compared to classical maleimide linkage[54].

A recent strategy for the bioconjugation of DNA to the cysteine residues of proteins or antibodies has been developed which yields to homogenous and well-defined products[55]. Here, oligonucleotides are labelled with a benzoyl acrylic acid moiety and then conjugated to proteins through cysteine selective thiol-Michael addition reaction.

3.5. Non-covalent approach for the HTRA1-DNA origami complex

Since the covalent crosslinker-based method was troublesome and is indeed non-regioselective, it was decided to choose a completely new approach.

Our goal is to build a model to gain a deeper insight into the interesting properties of DNA-enzyme nano-constructs. Thus, the role played by the different parameters of our system should be ideally easily identified and quantified. Employing covalent and non-regioselective methods for the synthesis of DNA-enzyme conjugates introduces an unknown degree of variability in the activity assays, due to the intrinsic inhomogeneity of the product obtained. Even if an equimolar mixture of DNA and protein can be successfully isolated by chromatographic techniques, the final product is most probably a stochastic combination of distinct chemical species, where DNA strands are linked at different and unpredictable positions of the protein surface. Such species are extremely challenging, if not even impossible, to separate and may all contribute (however not equally) to the global kinetic properties of the DNA-scaffolded enzyme. The covalent interaction can also change the protein properties and behaviour in a non-predictable manner. An ideal model would, therefore, be constituted by an enzyme that can be linked in a non-covalent fashion to the walls of a DNA origami structure, ideally in a multivalent fashion.

A completely new approach was followed based on a non-covalent binding between the HTRA1 protein and the NE_6 nanocage. This strategy is inspired by the work of Sprengel at al.[25].

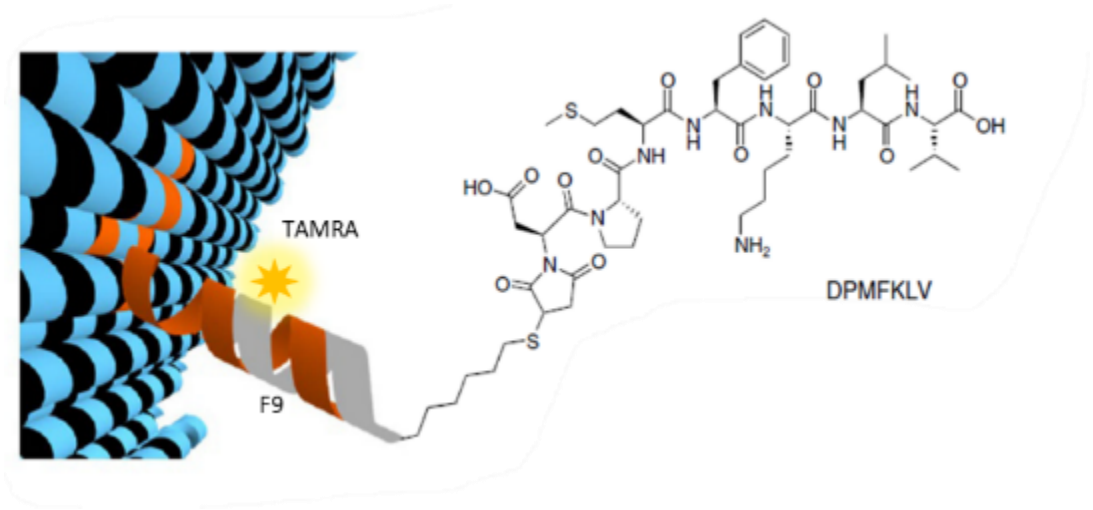

Figure 32: The DNA origami host is internally decorated with protruding DNA strands for further hybridization to complementary DNA–peptide conjugates.

The interaction exploited in this approach is the specific binding between the PDZ1 domain of the HTRA1 protein and its natural substrate. A small peptide, the heptapeptide DPMFKLV (Aspartic acid - Proline - Methionine - Phenylalanine -0iLysine - Leucine - Valine) was found to be a good analogue of the substrate of the PDZ1 domain of DegP, a bacterial protease that belongs to the same family of HTRA1. The peptide was obtained by solid phase peptide synthesis (SPPS), using a Wang resin and applying the standard Fmoc/tBu chemistry. In the final coupling step, the peptide is functionalised with a C6 linker and a maleimide group. Finally, after deblocking and removal from the resin, the peptide is linked to a thiol-modified oligonucleotide (F9(16)), partially complementary to the protruding arms of the NE_6 DNA origami nanocage (see Supplementary information in Sprengel at al.'s article for more details)[25]. The C6 linker provides some degree of orientational freedom for easier positioning of the ligand into the protein-binding pocket, while the DNA spacer ensures stable ligand anchoring to the host DNA surface, resulting in a host-to guest bridge of circa 10 nm in length. The F9 oligo used is functionalised with the TAMRA chromophore at the 5' terminus; allowing to track the DNA-peptide conjugate by a bright, pH-insensitive orange-red fluorescence (with an approximate excitation maximum at 546 nm and an emission maximum at ~579 nm).

The peptide tagged F9 oligo is hybridized inside the cavity of the NE_6 nanocage in a one pot reaction mixture, using a molar ratio of p7560 scaffold: staples: peptide-F9 = 1:50:100. The self-assembly process follows the same steps as the usual preparation of the NE_6 cage previously developed and described. The only change in the procedure regards the purification method: instead of PEG precipitation, ultrafiltration was implemented. The purification with molecular weight cut-off (MWCO) filters gives residual-free separation but can give lower recovery yields than PEG precipitation. Using the 100 kDa cut-off filter it was possible to separate the DNA origami cage from the low molecular weight excess staple strands with multiple cycles of buffer exchange and filtration (at least six) in a quick and efficient way, enabling also to adjust buffer conditions[56].

An AGE was performed to assess the assembly of the NE_6 cage with the peptide (*Figure 34*). The gel was scanned at specific excitation/emission wavelengths to reveal the TAMRA fluorescence signal from the labelled oligo. Staining with EtBr allowed to detect the DNA-containing bands of the samples. The correct assembly of the two half-cages E_3 and N_3 and the full cage NE_6 before and after the purification was checked. The gel shows very clear bands that are visible both in the TAMRA and in the EtBr channel, indicating that the self-assembly

of the DNA origami nanostructure has been successful and the labelled oligo has correctly hybridised to the DNA strands partially complementary to the peptide tagged F9. The purification with MWCO membrane filters resulted to be as effective as the PEG precipitation and completely removed the excess of staples from the reaction mixture.

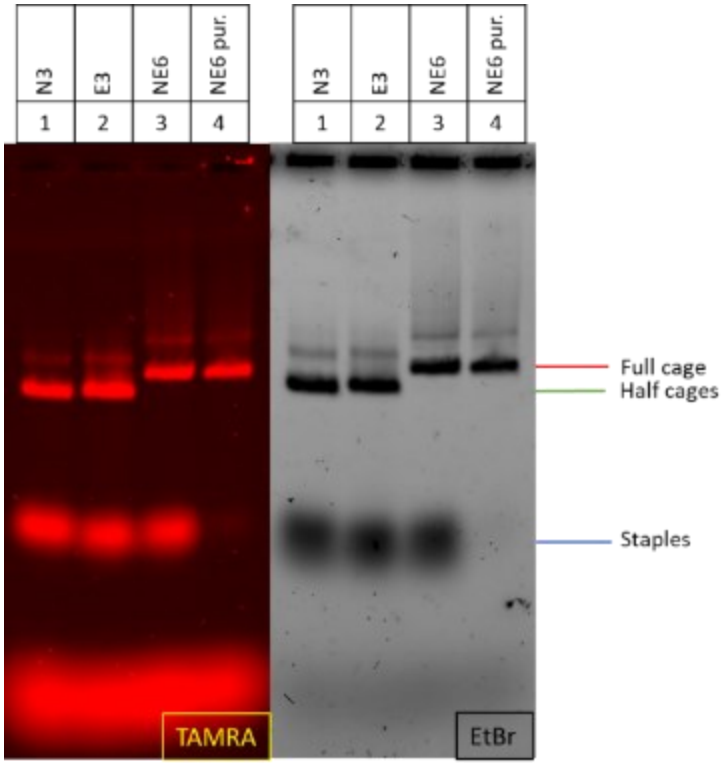

Figure 33: Agarose gel electrophoresis of the NE_6-TAMRA-F9-DPMFKL DNA origami nanocage.

The so-functionalized DNA origami structures were finally reacted with the HTRA1 protein. The encaging was done following the same exact procedure used for the encapsulation of the covalent HTRA1(SA)-NH_2F9^{FAM} conjugate. Three samples have been prepared to test the encapsulation with a molar ratio of protein to DNA cage being 6:1, 12:1 or 18:1 (a ratio of 1:1, 2:1, 3:1 to the number of protruding arms). The three mixtures were incubated in a thermocycler for 2h at 30°C and then purified from the excess of protein using PEG precipitation, as previously described with the other encaging attempts.

An agarose gel has been performed and gave the expected results (*Figure 35*). Interestingly, the DNA-protein bands in lane 6, lane 7 and lane 8 are broader and more blurred than the DNA origami only bands.

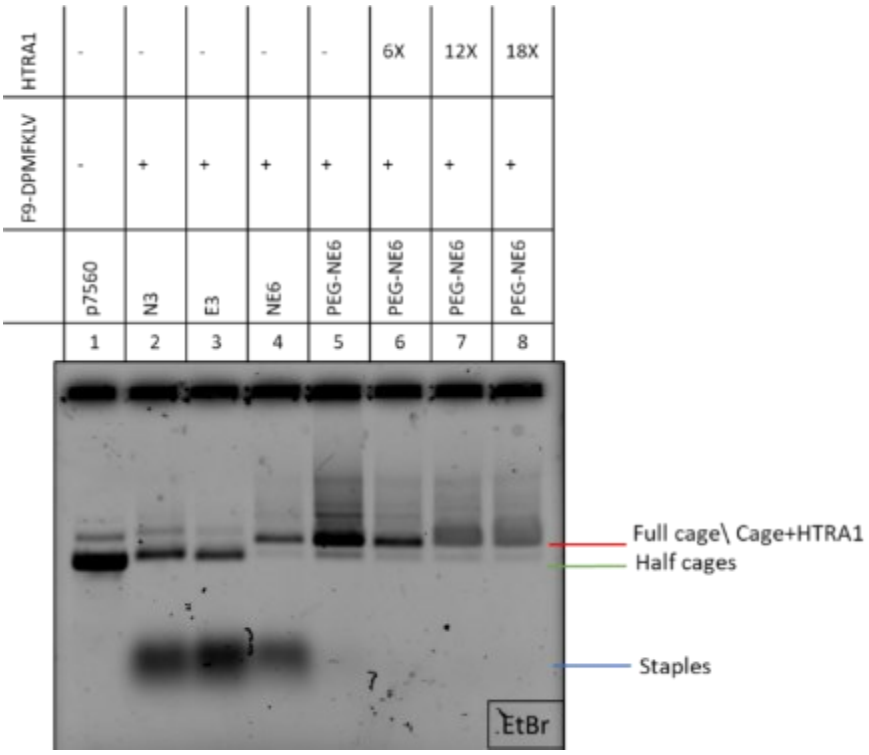

Figure 34: AGE of HTRA1SA encapsulation in the NE_6-TAMRA-F9-DPMFKL DNA origami cage.

But the TEM images in *Figure 36* showed no protein inside the cage. This could be due to a low-resolution power for identification of the protein, which is shielded by a relatively thick wall of DNA. All three samples however contain correctly assembled DNA nanocages.

Current studies in the laboratory demonstrate that the loading efficiency of a protein within a DNA-origami host may be highly dependent from the electrostatic interaction between the protein and the NE_6 nanocage. HTRA1 contains numerous lysine residues, that bear a positive charge on the side chain at the neutral working pH. This feature results in a net positive charge of the protein, particularly at its surface. The DNA origami nanostructure, instead, is made of a long sequence of nucleotides, that are composed of an aromatic nucleobase, a pentose sugar and a phosphate group pointed towards the outside of the DNA double helix. These phosphate groups bear a negative charge that gives a global negatively charged character to the DNA origami nanocage. A non-covalent electrostatic interaction is established between the two species. This non-specific electrostatic interaction can be compared to an ion-ion interaction, even if the two parts are not charge punctiform species, but large and very complex objects with an extended surface of interaction.

This unspecific binding competes with the specific host-guest interaction between the peptide ligand and the PDZ1 domains of the protein and may prevent encapsulation. Thus, the loading efficiency of the protein might be highly dependent from a series of different factors, which affect the electrostatic interactions of the two partners, for example pH.

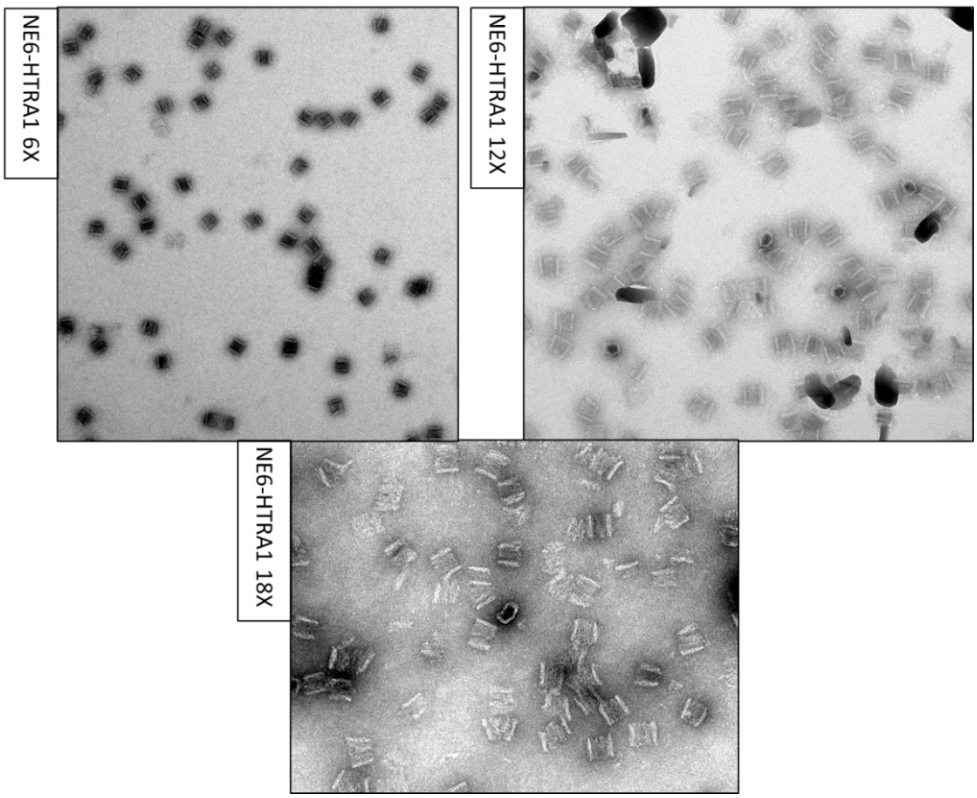

Figure 35: TEM images of the HTRA1SA-NE_6-TAMRA-F9-DPMFKL DNA origami cage.

3.5.1. HTRA1 labelling with NHS-ester A488 dye and encapsulation inside the NE_6-TAMRA-F9-DPMFKL DNA origami

Recent results in the laboratory have shown that shielding the positive charges exposed on the protein surface notably enhances the yield of protein encapsulation and simultaneously reduces unspecific electrostatic interactions with the negatively charged DNA host. This favors the specificity of the recognition binding event between the peptide and the protein. This effect can be achieved for example by increasing the pH of the reaction mixture, or by fluorescent labelling of the exposed amino acid residues or adding lysine-specific molecular tweezers. In this work a negatively charged fluorescent dye was exploited to partially screen the protein positive charge, thus lowering unspecific electrostatic interactions with the negatively charged DNA cage.

To circumvent the unspecific binding, the HTRA1 protein has been coated with different extents of the A488 dye. This dye is bound to the positively charged lysines and bears two negative charges at the working pH. Furthermore, this allows to better analyse the encaging process by following the specific fluorescence of the A488 dye.

The A488 dye is a chromophore that emits in the green part of the spectrum, the maximum of absorption is at 495 nm (with a molar extinction coefficient of 71800 $Lmol^{-1}cm^{-1}$), and the maximum of emission is at 519 nm (with a quantum yield of 0,91). The dye contains an NHS-ester function to bind the amino group of the lysines on the HTRA1 protein surface. The dye is soluble in organic solvents, so it is dissolved in DMSO in a quite high concentration. The A488 chromophore is chosen because it presents an emission spectrum far away from the absorption range of the TAMRA fluorophore, thus avoiding any possible FRET (Förster Resonance Energy Transfer) effect that would interfere with the interpretation of the signals.

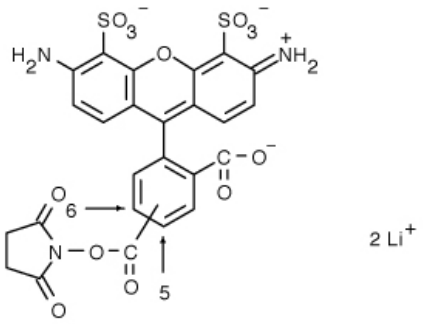

Figure 36: Molecular structure of NHS ester substituted Alexa A488 dye.

For the labelling, the NHS-ester functionalised A488 dye is mixed with the inactive HTRA1 protein in various combinations. The amount of A488 dye added is based on the number of lysines present in the protein. HTRA1 contains 14 lysine residues in the protein sequence, and they're all considered to be accessible to the NHS ester group of the fluorophore; so the A488 mixed is a percentage of the amino groups present in solution (in practice 14 times the quantity of the protein multiplied by a certain percentage). Many different labelling grades have been realized: 5%, 15%, 20%, 30%, 40% and 50% of the total amount of lysine residues.

The actual number of dye molecules linked to the HTRA1 surface remains unknown because it depends on the yield of the reaction between the NHS ester and the amino group of the protein. The percentage number with which all the labelled samples are named and classified has to be intended as indicative values of the reaction conditions and do not refer to the effective level of labelling.

The SDS gel in *Figure 38* shows the different labelled protein samples. The band intensities decrease when going from the protein sample labelled at about 5% of the exposed lysine residues to the sample labelled at 50%. This decrease in intensity of the protein bands is due to the fact that Coomassie binds to the amino groups of the protein (mainly lysine residues) and so the higher is the extent of protein fluorescent labelling at the lysine residues, the lower will be the possibility for the Coomassie to bind to the protein. In other words, the strongly labelled proteins will have a reduced number of available free amino groups. Therefore, the intensity of the staining indicates that the labelling is gradually increasing but can't be a good indicator of the amount of HTRA1 in the loaded sample.

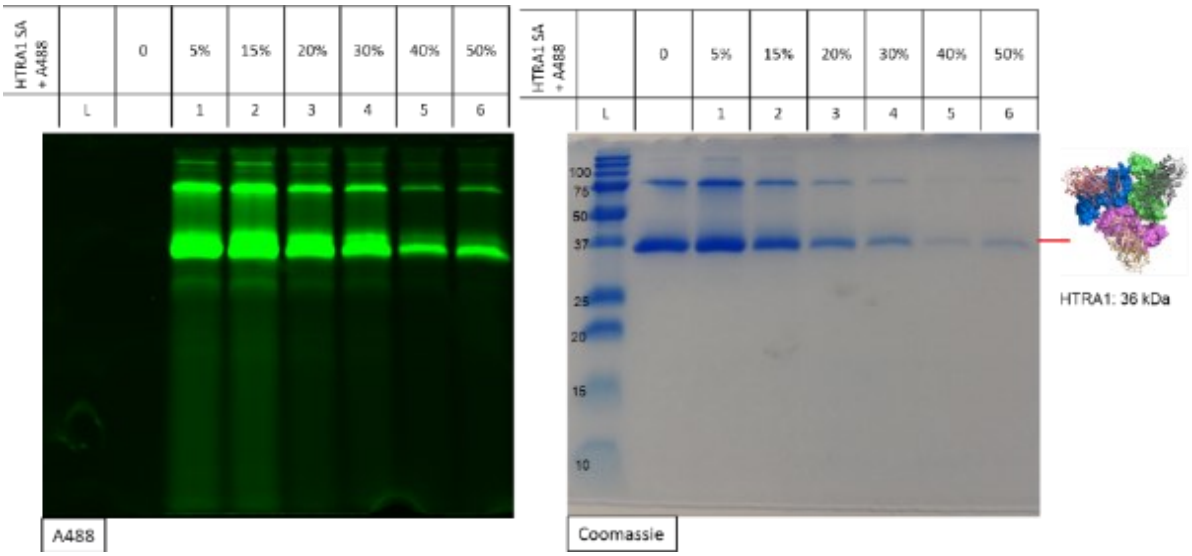

***Figure 37**:* SDS-PAGE of HTRA1 labelling with NHS-ester A488 chromophore

After the synthesis of the labelled protein samples, HTRA1(SA)-A488 is reacted for 2 hours at 30°C with the NE_6-TAMRA-F9-DPMFKL DNA nanocage in a molar ratio of the protein to the ligands (the complementary protruding arms) of 2:1. Six different encapsulation samples have been reacted using the six different labelled proteins prepared with A488. The mixtures are analysed with gel electrophoresis and visualised by TEM to check the encaging process.

In the agarose gel (*Figure 39*) the six encaged samples are run together with the six labelled proteins and the DNA nanocage alone as a control. The gel was scanned with a Typhoon FLA 9000 at different wavelengths to record the presence of HTRA1 protein (A488-labelled) and peptide ligand (TAMRA-labelled); DNA was visualized upon staining with ethidium bromide.

The six A488-labelled HTRA1 proteins show an interesting behaviour: the first lanes with less dye attached to the protein migrate above the loading pocket towards the cathode (the negative electrode) while the sample with the highest degree of labelling migrates in the opposite direction towards the anode. All samples follow a progressive and coherent pattern (except for lanes 4 and 5, which is probably due to an inversion in the sample loading into the gel). These bands are visible only upon illumination at the excitation wavelength of the fluorophore used for protein labelling (488 nm), and indicate that the different labelling degree changes the migration rate of the protein in the applied electric field. The HTRA1 protein alone has a positive net charge at the conditions used for gel separation (TEMg buffer pH 7.6), with a consequently migration rate in the direction opposite to the applied field. The fluorophore molecules on the protein surface modify the net charge of the HTRA1 by introducing two

negative charges in the place of a positive one, and thus affect the diffusion properties of the protein in an agarose gel matrix.

The progressive decrease in intensity of the protein bands in the A488 channel are probably due to a systematic error in the calculation of the concentration of the samples. The concentration of the protein is evaluated by measuring the absorbance at 280 nm, but the dye must have some kind of absorption even at that wavelength, and this has altered all the measurements, causing the insertion in the reaction mixture and in the gel of a lower amount of protein than expected, which decreases with the amount of A488 dye present at the protein surface.

The NE_6 nanocage doesn't present any remarkable gel mobility shift after the binding of the protein to the peptide-modified DNA origami structure, although their co-migration proved their mutual interaction.

The gel mobility of the DNA-origami protein complex (and in general of every molecular species) is mainly the result of two contributions: its net charge and total mass. The encapsulation of the HTRA1 protein within the DNA host will lead to a negligible variation of the negative surface charge density of the DNA origami structure, thus explaining its almost unaffected migration rate. The increase in the molecular weight of the DNA origami is quite small compared to the dimensions of the NE_6 nanocage alone and it's difficult to reveal by agarose gel electrophoresis. However, this is valid both for proteins bound within the cavity of the host and on its external surface.[25]

However, the TEM micrographs captured in *Figure 40* showed that the encapsulation wasn't effective. For example in samples 2 and 5 (where the NE_6 cage is mixed with the 15% A488-HTRA1 and the 40% A488-HTRA1 respectively) it is possible to see the HTRA1 as a white spot inside the nanocage (the cages with the protein inside are highlighted with an orange circle). Interestingly, the protein could be visualized only in the cages oriented vertically in respect to the grid, with the hexagonal face facing the microscope. No protein could be visualized in the cages oriented along their longitudinal axis, which were the majority of the structures in the TEM micrographs. Thus, it is possible that the resolution power of the microscope is too low to individuate the HTRA1 when surrounded by the DNA layer and that instead the protein becomes visible when no DNA covers it.

So, even if this new strategy is more promising than the other covalent approach, the protocol still requires some optimisation. The reaction conditions have been considered too mild and have been improved to move the equilibrium towards the complexed state. A large excess of

HTRA1 was used to push the reaction, using a protein concentration that is 20-fold the number of binding sites available (120 times the concentration of NE_6 nanocage in solution). The mixture was incubated for longer time and at a higher temperature. The excess of protein was removed with PEG precipitation, as always. In this experiment, only the A488-HTRA1 protein that displayed the best behaviour in the previous encapsulation test was employed, that is, the 40% A488-HTRA1.

In *Figure 41* this second encapsulation has been checked with an agarose gel and the results are as expected. The protein band is large and broad because too much sample was added in the gel pocket causing an anomalous running of the protein.

The final TEM image in *Figure 42* shows numerous HTRA1 proteins present on the surface of the TEM grid. The HTRA1 is not always segregated into the walls of the DNA origami nanocage, but it's often bound to the external surface or at the edges of the nanostructure. A hypothesis that can be formulated is that the sample preparation for TEM imaging may have caused the deformation of the origami complex.

Thus, the labelling with the negative charged dye did not lead to complete protein encapsulation, but resulted in reduced unspecific electrostatic interactions with the negatively charged DNA host, enabling the specificity of peptide ligand binding to emerge.

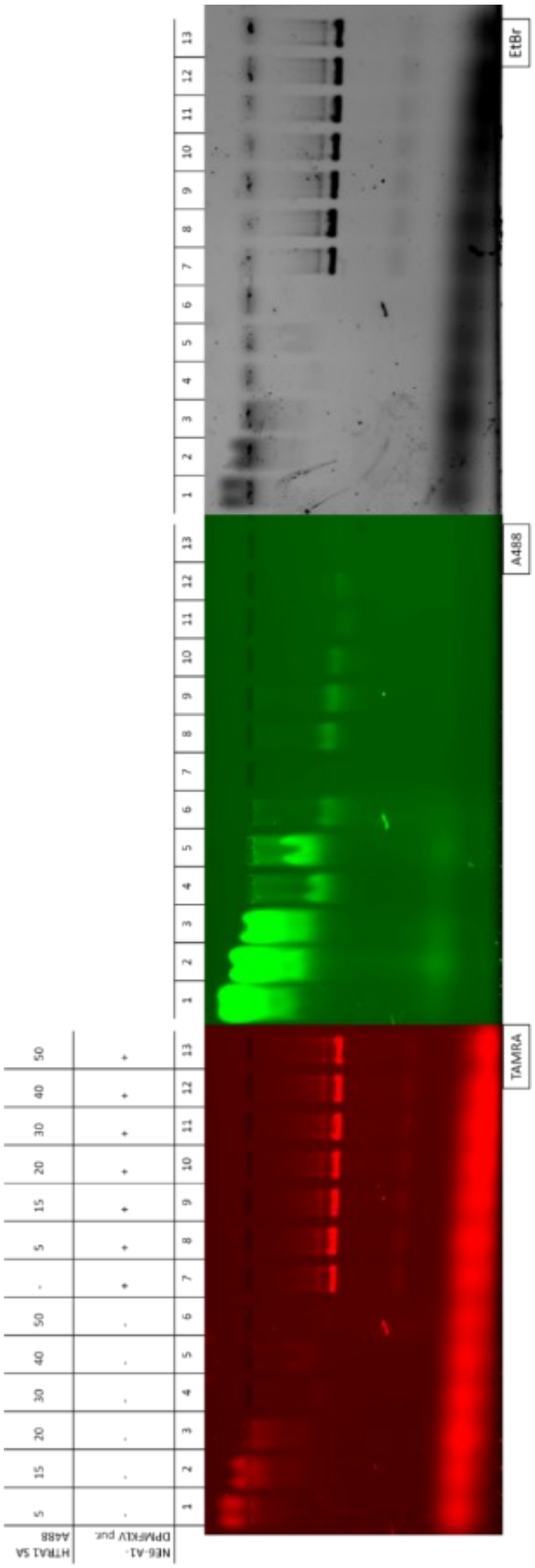

Figure 38: AGE of the encapsulation of HTRA1SA-A488 inside the NE_6-TAMRA-F9-DPMFKL DNA origami nanocage.

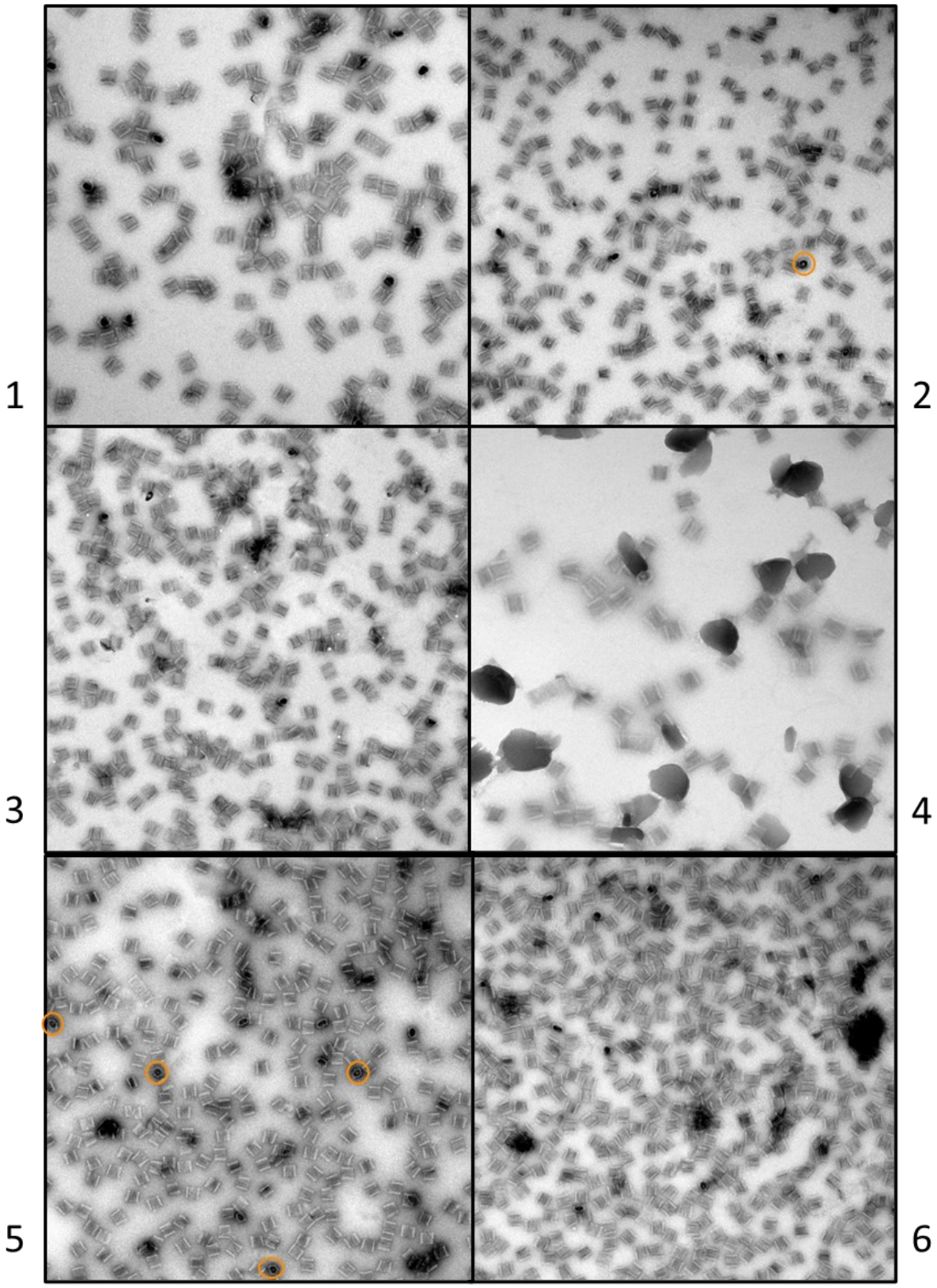

Figure 39: TEM images of the NE_6-TAMRA-F9-DPMFKLV HTRA1SA-A488 encage. The six images correspond to the six labelled protein samples.

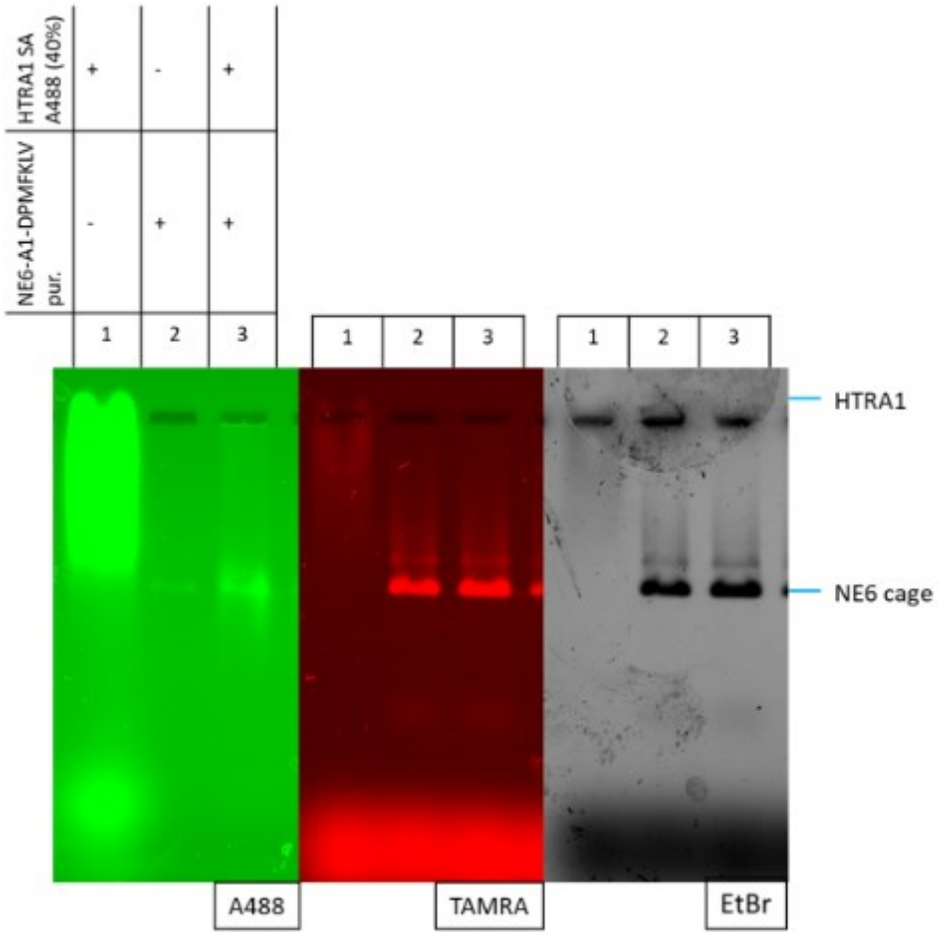

Figure 40: AGE of NE_6-TAMRA-F9-DPMFKLV HTRA1(SA)-A488 40% encage.

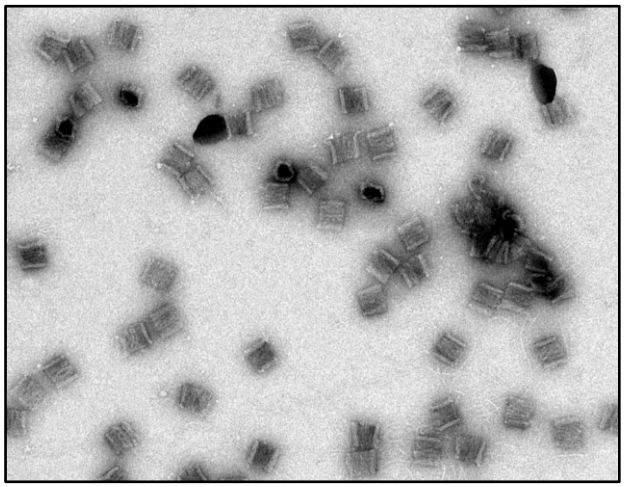

Figure 41: TEM images of NE_6-TAMRA-F9-DPMFKLV HTRA1(SA)-A488 40% encage.

4. Conclusions and outlook

The ambition and long-term goal of this work is to investigate the effect of DNA caging on the stability and enzymatic activity of the hosted protein.

At this purpose, proteins with a short lifetime and complex activity are ideal guest candidates. In fact, being able to prolong the lifetime of unstable proteins and to control their enzymatic activity would be an enormous step forward in the realization of artificial protein machines. The choice of the HTRA1 protein is in this sense very lucky, although challenging: it's a serine protease that displays an autoproteolytic activity.

The DNA host is a nanocage, made from the hierarchical assembly of two DNA origami half cages. Its design and self-assembly are fully established and considered successful.

HTRA1 has been firstly conjugated with a single-stranded DNA oligonucleotide using a heterobifunctional crosslinker and then encaged inside the inner walls of the nanocage by hybridization to a complementary protruding strand. This protocol revealed to be unsuitable to insert the HTRA1 inside the DNA nanocage. The conjugation synthesis was not efficient, and it was not possible to isolate the product in a sufficient amount.

A second strategy for the encapsulation was implemented by placing a set of peptide ligands on the inner surface of the cage. These peptides specifically bind to the PDZ1 domains of the HTRA1 protein with weak non-covalent interactions. These interactions can be modulated by controlling the geometry and stoichiometry of the ligands and their corresponding binding sites on the protein surface. This method doesn't require any previous covalent modification of the protein that can alter the enzyme properties and activity. Partial suppression of the DNA-protein electrostatic attractions through the labelling of lysine' s side chains on the HTRA1 protein surface with a A488 dye enhanced the specificity of host-guest recognition, thus confirming the important role of the molecular shell surrounding the protein surface for DNA-caging purposes.

This second method should be improved to increase the yield of HTRA1 loaded origami structures and the concentration of protein within the DNA origami nanocage should be exactly determined.

The next step forward would be to perform activity assays to test the effect of the encapsulation on the enzymatic activity of the HTRA1 protein (that could be in theory enhanced) and to test whether the encaging would lead to its stabilization and therefore also to a reduced

autoproteolytic activity. An ideal test should use a high affinity substrate in combination with a high sensitivity output signal. Single molecule FRET is a technique that could be used with promising results.[57]

In principle, the method proposed here is applicable to every protein-recognition motif and enables to take advantage of the chemical affinity and the geometric compatibility for efficient binding, using a multiplicity of recognition events occurring in a confined space. The DNA nanocage can be even modified with distinct ligands to arrange multiple enzymes into synthetic protein scaffolds, allowing, for example, to engineer and improve metabolic pathways in a programmable manner.[58]

5. Experimental section

5.1. Reagents and solvents

- Tris-(hydroxymethyl) aminomethane Tris (Rectapur)
- Boric acid $B(OH)_3$ (Sigma-Aldrich)
- Ethylenediamine tetra-acetic acid EDTA (Roth)
- Magnesium acetate tetrahydrate $Mg(CH_3COO)_2 \bullet 4H_2O$ (Merck)
- Magnesium chloride hexahydrate $MgCl_2 \bullet 6H_2O$ (Roth)
- Potassium chloride KCl (Sigma-Aldrich)
- Sodium chloride NaCl (Sigma-Aldrich)
- Disodium hydrogen phosphate Na_2HPO_4 (Sigma-Aldrich)
- Sodium phosphate monobasic NaH_2PO_4 (Sigma-Aldrich)
- Hydrochloric acid (Riedel-de Haën, 37% concentrated solution)
- Sodium hydroxide NaOH (Sigma-Aldrich)
- 4-(2-hydroxyethyl) -1-piperazineethanesulfonic acid HEPES (Sigma-Aldrich)
- Glycerol (Sigma-Aldrich)
- N-[α-maleimidoacetoxy] succinimide ester AMAS (Thermo Scientific)
- Dimethyl sulfoxide DMSO (Sigma-Aldrich)
- Tris-(2-carboxyethyl) - phosphine hydrochloride TCEP (Sigma-Aldrich)
- Ethanol EtOH (Merck)
- n-Butanol (Roth)
- Sodium acetate NaAc (Roth)
- Ammonium acetate NH_4OAc (Sigma-Aldrich)
- Formamide (Fluka)
- Polyethylene glycol PEG8000 (Ph.Eur.) (Promega)
- NHS-Alexa 488 fluorescent dye (Invitrogen)
- Streptavidin coated magnetic beads (Merck)
- Urea (AppliChem)
- GlycineH_2N-CH2-COOH (Sigma-Aldrich)
- Sodium dodecyl sulfate (CH_3-$(CH_2)_{10}$-CH_2-O-SO_3^-) Na^+ SDS (Sigma-Aldrich)
- 40% Acrylamide: bis-acrylamide (19:1) (Roth)
- N, N, N', N'-Tetramethylethylendiamine TEMED (Merck)

- Ammonium persulfate APS (Applichem)
- Agarose (Thermo Fisher Scientific)
- Bromophenol blue sodium salt (BPB; Merck)
- Precision Plus Protein™ All Blue Prestained Protein Standards (BioRad)
- SYBR Gold Nucleic Acid Gel Stain (Invitrogen)
- Ethidium bromide EtBr (1% solution, 10mg/mL; Roth)
- Coomassie Brilliant blue G250 (Sigma-Aldrich)
- Uranyl formate $UO_2(CHO_2)_2 \bullet H_2O$ (Electron Miscroscopy Sciences)

The DNA origami cage used in this work is called "*Nemesis*" and will be referred to as NE_6, where 6 stays for the number of protruding arms available in the inner cavity of the cage. For NE_6 DNA origami: the scaffold p7560 was already available in the laboratory, while the staples were purchased from IDT (Integrated DNA Technologies) in 96-well plates with a precise scheme. Both are kept frozen in small aliquots at -20 °C.

The single-stranded oligonucleotide used in this Thesis is called F9 and has the following sequence: 5'-GTGGAAAGTGGCAATC-3'. NH_2F9^{FAM}, the F9 with an amino group at the 3' terminus and the FAM (fluorescein) chromophore at the 5' terminus, came from IDT. It's kept frozen in small aliquots at -20 °C.

The HTRA1 and HTRA1(SA), or inactive HTRA, have been kindly provided by the laboratory of Prof. M. Ehrmann (ZMB, University Duisburg-Essen). The proteins are kept frozen in small aliquots at -80 °C.

For the magnetic beads purification: $F9(22)^{FAM}$ (sequence: GTGGAAAGTGGCAATCCACTTC), labelled with FAM at the 5' terminus, and Biot-$cF9(22)^{FAM}$, the same oligonucleotide functionalized with biotin at the 3' terminus, are bought from IDT.

TAMRA-F9-DPMFKLV, the F9 (sequence: 5'-GTGGAAAGTGGCAATC-3') with the TAMRA chromophore at the 5' terminus and the DPMFKLV peptide at the 3' terminus, was already available in the laboratory. They are kept frozen in small aliquots at -20 °C.

5.2.Instruments

Ion Exchange Chromatography (IEC)

The instrument is a proFIRE (Dynamic Biosensors) and it's an anion separating column. The eluent is composed of two buffers: buffer A with Na_2HPO_4\$NaH_2PO_4$ 50 mM (33.3 mM Na_2HPO_4, 16.7 mM NaH_2PO_4) NaCl 150 mM (pH 7.0) and buffer B with Na_2HPO_4\$NaH_2PO_4$ 50 mM (33.3 mM Na_2HPO_4, 16.7 mM NaH_2PO_4) NaCl 1 mM (pH 7.0). The elution program used for separation was: 4 min. buffer A, then buffer A and buffer B are mixed together to create a gradient of NaCl salt concentration that changes with a rate of 10 mM/min. After 15 minutes only buffer B is eluted for 2 min. Finally, buffer A is eluted for 6 min. to wash the column from all species that are more tightly bound to the matrix; this sums up to a total elution time of 27 min.

Size Exclusion Chromatography (SEC)

The size exclusion chromatographer is Äkta and the column used is a Superose 6 increase 10/300 equilibrated with buffer D (150 mM NaCl, 50 mM NaH_2PO_4, 1 mM TCEP, pH 8.0).

Nanodrop spectrophotometer

The instrument measures the absorbance and gives the concentration. The wavelength used for DNA is 260 nm. The theoretical extinction coefficients for oligonucleotides are obtained from the Integrated DNA Technology (IDT) Biophysics website (http://biophysics.idtdna.com/).

Only 2 µL of samples are needed. The result is given in ng/µL. For single-stranded DNA the molar concentration is calculated with the following equation, where MW is the molar weight in g\mol.

$$C(\mu M) = \frac{10^3 \times C(ng/\mu L)}{MW}$$

Atomic force microscope (AFM)

The sample (with a 5 nM concentration) was deposited on freshly cleaved mica surface (Plano GmbH) and adsorbed for 3 min at room temperature. After washing with milliQ H_2O, the sample was dried in air and scanned in ScanAsyst Mode using a MultiMode 8 atomic force microscope (Bruker). The images were analysed using the Nanoscope Analysis software.

Transmission electron microscope (TEM)

The samples were prepared by applying 4 µL of sample on glow discharged copper grids (Agar Scientific), covered by a carbon film. After 60 sec, excess of buffer was blotted from the side with a piece of filter paper. The grid was washed twice with two drops of milliQ water and subsequently stained with two drops of 1% solution of uranyl formate in milliQ water. All images were taken with a Jeol 1400PLUS transmission electron microscope. Digital micrographs of NE_6 DNA origami cage, alone and with the HTRA1 encapsulated, were recorded at a nominal magnification of 20.000x.

5.3. GELS

5.3.1. Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE (sodium dodecyl sulphate–polyacrylamide gel electrophoresis) is a method used to separate proteins in denaturing conditions

After assembling the BioRad gel apparatus, the two phases of the gel are prepared as follows: The resolving gel (lower part) is prepared by mixing the following components in a beaker: 2.5 mL of buffer A (375 mM Tris, 1% w/V SDS, pH 8.7), 5 mL Acrylamide/Bis solution (29% acrylamide 1% bisacrylamide, the commercially available solution, stored at 4 °C to reduce the health hazard), 2,5 mL of water, 50 µL of 10% APS (the APS is stored as 1 mL aliquot of 10% (w/w) water solution and frozen at -20 °C till use) and 20 µL of TEMED; this amount is sufficient for 2 gels. Pour the gel between the gel plates and add isopropanol (to avoid contact with air). It's important to be quick because once the TEMED has been added, the gel starts to polymerize within minutes.

When the resolving gel is polymerized (circa 30 min), remove the isopropanol and add the stacking gel (the upper part). The stacking gel is prepared with 2.5 mL of buffer B (120 mM Tris, 1% g/L SDS, pH 6.8), 1,5 mL Acrylamide/Bis solution, 6 mL of water, 70 µL of 10% APS and 35 µL of TEMED. Insert the comb and let it rest for 1 h RT.

Fill the chambers of the apparatus with the SDS running buffer (24.8 mM Tris, 200 mM glycine, 3.5 mM SDS). Make sure that the apparatus is leak-proof before proceeding. Prerun the gel for 20 min at 180 V.

Prepare the sample for the analysis: use circa 10 µL of sample and add 2.5 µL of SDS sample loading buffer 4X (400 mM Tris, 280 mM SDS, 4.3 M glycerol, 6 mM bromophenol blue, pH 6.8). Boil the samples for 10 min at 90 °C. Add the samples to the gel and run the gel for about 1 h at 180 V.

The labelled samples can be scanned with the Typhoon FLA9000 fluorescence scanner (from GE healthcare Life Sciences). DNA components can be visualized by using SYBR GOLD, proteins can be stained using a colloidal Coomassie stain. Gel images are analysed using the ImageJ software.

5.3.2. Denaturing Urea polyacrylamide gel electrophoresis UREA PAGE

For the denaturing PAGE prepare the gel mixture with 2.4 mL TBE 5X (445 mM tris base, 445 mM boric acid, 10 mM EDTA, pH 8.0), 7.5 mL acrylamide/bis solution, 5 g urea, 2.1 mL water, 100 µL APS, 10 µL TEMED. Allow for polymerization to occur for at least 20 min. Clamp the gel into the electrophoresis tank and equilibrate in TBE 1X buffer (89 mM tris base, 89 mM boric acid, 2 mM EDTA, pH 8.0) for about 20 min, at room temperature and 190 V voltage.

To each sample, add an equal volume of formamide loading and tracking buffer 2X (8 parts v/v formamide, 2 parts v/v TBE 5X buffer and few µg of bromophenol blue sodium salt). Load the samples and run the gel at room temperature and 190 V for approximately 45 min (until the tracking dye has reached 2/3 of the complete run).

The labelled samples can be scanned with the Typhoon FLA9000 fluorescence scanner (from GE healthcare Life Sciences), DNA components can be visualized by staining the gel with SYBR GOLD and scanning it with the Typhoon. Gel images are analysed using the ImageJ software.

5.3.3. Agarose gel electrophoresis (AGE)

The AGE is performed in a horizontal chamber. The gel is a 1% w/v agarose in TEMg buffer (20 mM Tris base, 2 mM EDTA, 12.5 MgCl2, pH 7.6). The gel is prepared and kept at 60 °C. Pour the solution onto the gel chamber till about 3 mm height. Place the comb on top of the gel and let cool at room temperature. When the gel is ready, remove the comb and fill up the chamber with TEMg as running buffer. Keep the chamber in ice. Prerun the gel at 80 V for 30 min.

Prepare 10 µL of sample by adding 2.5 µL of loading buffer for agarose 5X (40% glycerol, 45 mM Tris base, 45 mM boric acid, 1 mM EDTA, 11 mM $MgCl_2$ and few µg of bromophenol blue). Add the samples to the chamber and run the gel at 80 V for 2 h 30 min.

When the gel is ready (the blue dye has reached the bottom), extract the gel and analyse the possible fluorescent components using the Typhoon FLA9000 fluorescence scanner (from GE healthcare Life Sciences). Then image the DNA with EtBr staining: put the gel in an EtBr solution in TEMg and wait around 20 min, then scan it with the Typhoon. Gel images are analysed using the ImageJ software.

5.4. Syntheses and DNA origami assembly

5.4.1. Preparation of the DNA origami cage (NE_6)

The staples that belong to different parts of the origami, like core, lateral edges, shape-complementary edges (involved in the full cage assembly) and protruding arms (used for the protein encapsulation) are pre-mixed and dissolved in Millipore water at 100 µM concentration.

Half-cage N_3 and E_3 assembly: The scaffold p7560 in the final concentration of 10 nM is mixed with 50-fold excess of staples. 3.6 µL of p7560 (1.39 µM), 25 µL of core staples (1 µM), 25 µL of edge staples (1 µM), 5 µL of hybridization staples (5 µM), 12.5 µL of PA staples (2 µM) are mixed together in a total volume of 500 µL in TEMg20 buffer (50 mM Tris, 10 mM EDTA, 20 mM $MgCl_2$, pH 7.6). The assembly is done by slow thermal annealing in a temperature controller: the program first heats the solution to 65 °C, then goes to 52 °C and keeps it at that temperature for 3 h and then cools down to 21 °C.

Full cage NE_6 assembly: The two half-cages are directly mixed in equal volume without any previous purification. Previous experiments have shown that the assembly is complete after 3 hours incubation at 40 °C, using a Thermocycler Mastercycler nexus gradient (Eppendorf).

PEG purification of DNA origami: Mix the NE_6 cage with equal volume of PEG (polyethylene glycol) buffer supplemented with the same concentration of Mg^{2+} in assembly buffer (15% w/V PEG8000, 5 mM Tris, 1 mM EDTA, 505 mM NaCl, 200 mM $MgCl_2$). Centrifuge at 28600 *g* for 30 min at room temperature (25 °C). Discard the supernatant and let the pellet resuspend in TEMg20 buffer overnight.

A small fraction of N_3 E_3 and unpurified NE_6 is kept aside for control. 100 µL of NE_6 cage 38 nM were obtained, with a reaction yield of 76%.

The reaction is checked with AGE and the final product is assessed with AFM and TEM.

5.4.2. HTRA1-NH_2F9^{FAM} conjugation

NH_2F9^{FAM} is stored at -20°C as a 100 µM solution in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2).

Purification via isopropanol precipitation: 100 µL of the oligonucleotide are mixed with 100 µL of NaOAc 3 M and 800 µL of isopropanol (1 eq. vol of salt and 8 eq. vol of isopropanol). The solution is centrifuged at 49300 *g* for 30 min at 4 °C, then washed 3 times with ethanol (quick centrifuge spins) and let dry for 5 min. The supernatant is discharged, and the pellet is resuspended in 50 uL PBS buffer.

Add 50-fold AMAS crosslinker to F9 oligo: 25 µL of AMAS crosslinker 20 mM (1.2 mg in 238 µL of DMSO) are added to the F9 oligo to a total volume of 75 µL. The crosslinker is sensitive to air and water, so it is weighed in small aliquots in a glove box and stored in the freezer at -20 °C. It reacts for 2 hours at room temperature. After that, the solution undergoes again isopropanol precipitation with the same procedure as seen before (this time 75 µL of NH_4OAc 5 M and 600 µL of isopropanol are added). The pellet from the precipitation is resuspended in 15 µL of HEPES buffer (25 mM HEPES, 300 mM NaCl, pH 7.0). The concentration of the solution was measured with the Nanodrop spectrophotometer and was 567 µM.

Cross-linking reaction and thiol reduction: This AMAS-F9 solution is mixed with 60 µL of HTRA1(SA) 880.5 µM and 50 µL of TCEP 100 mM (resulting in a circa 6.2-fold excess of HTRA1 to the oligo and 100-fold excess of TCEP to the protein). The pH is measured and brought to pH = 7 with NaOH. The mix reacts for 2 h at room temperature.

Purification of the conjugate: The solution is purified using the size exclusion chromatography. The eluant was concentrated and rebuffered in HEPES buffer using a 10 kDa MWCO filter. The fractions are analysed with SDS-PAGE to confirm the identity of conjugate.

50 uL of HTRA1-NH_2F9^{FAM} conjugate with a concentration of 1.13 µM were obtained. Yield: 0.56%.

5.4.3. Purification of NH_2F9^{FAM} with gel extraction

A denaturing Urea PAGE is performed as in Section 3.3.2., in which 50 µL of NH_2F9^{FAM} 100 µM are loaded in each well, filling 5 wells: this leads to a total volume of 250 µL purified. When the run is complete, it's placed on a black surface to see the fluorescent bands of the oligo. The bands are cut and put into a 15 mL Greiner tube containing 5 mL TE buffer (10 mM Tris base, 1 mM EDTA, pH 7.6), then it's mixed with a sharp-pointed glass bar and eluted overnight in a rotator.

The tube is centrifuged at 4000 *g* for 5 min. at 20 °C. The supernatant is carefully removed and placed into a clean tube with 5 mL of n-butanol, then it's vigorously shaken. When the two phases are well separated again (after 10-15 minutes), 200 µL of NaOAc 3 M are added and also 4 mL of absolute ethanol. The mixture is frozen at -80 °C overnight.

After that, the tube is centrifuged at 4000 *g* at 4 °C for 30 min. The supernatant is poured off. 2 mL of ethanol 70% v\v are added, vortexed well and centrifuged again at 4000 g at 4°C for 15 min. The supernatant is removed carefully, and the pellet is let dry in air for 20-30 min. The pellet is resuspended in 50 µL of PBS, vortexed, and centrifuged at 4000 *g* at room temperature for 1 min.

The concentration of NH_2F9^{FAM} is measured with the NanoDrop: 339.3 µM. The recovery yield resulted to be 67.9%.

5.4.4. Purification of HTRA1-NH_2F9^{FAM} conjugate with magnetic beads

2 mg of magnetic bead are dissolved in 250 µL of TEN binding buffer (20 mM Tris, 0.5 M NaCl 1mM EDTA pH 7.5). The solution is washed with TEN buffer 4 times and the incubated with 100 µL of Biotin-cF9(22) 100 µM at 4°C overnight. The excess is removed washing 3 times with 400 µL of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2). The beads are incubated with the NH_2F9^{FAM}-HTRA1 conjugate (400 µL in PBS) for 4 h at 4°C. The solution is washed 3 times with PBS to pull-down the excess of protein. 100 µL of F9(22) 100 µM is added and let react for 2 h. The solution is centrifuged at 10000 *g* for 3 minutes and the magnetic beads are removed.

5.4.5. Encapsulation of HTRA1-NH_2F9^{FAM} conjugate inside NE_6 cage

75 µL of NE_6 38 nM in TEMg20 buffer are mixed with 30.3 µL HTRA1-NH_2F9^{FAM} conjugate 1.13 µM in HEPES buffer. The sample is heated in the Thermocycler Mastercycler nexus gradient (Eppendorf) at 37 °C for 2h and then brought back to 4 °C.

The excess of protein is removed with PEG precipitation: 105 µL of PEG (polyethylene glycol) buffer (15% w/V PEG8000, 5 mM Tris, 1 mM EDTA, 505 mM NaCl, 200 mM $MgCl_2$) is added to the solution, then centrifuged at 28600 *g* for 30 min at room temperature (25°). The supernatant is carefully removed, and the pellet is resuspended in 25 µL of TEMg20 overnight.

The encaging is assessed with AGE and with TEM.

5.4.6. NE_6 cage DNA origami functionalization with TAMRA-F9-DPMFKLV

The NE_6 DNA origami cage self-assembly follows the same process as the previous NE_6 cage.

Half-cage N_3 and E_3 assembly: The scaffold p7560 in the final concentration of 10 nM is mixed with 50-fold excess of staples and 100-fold excess of the peptide-tagged F9. 3.6 µL of p7560 (1.39 µM), 25 µL of core staples (1 µM), 25 µL of edge staples (1 µM), 5 µL of hybridization staples (5 µM), 12.5 of PA staples (2 µM) and 1.92 µL TAMRA-F9- DPMFKLV (26,04 µM) are mixed together in a total volume of 500 µL in TEMg20 Buffer (50 mM Tris, 10 mM EDTA, 20 mM $MgCl_2$, pH 7.6). The assembly is done by slow thermal annealing in a temperature controller: the program first heats the solution to 65 °C, then goes to 52 °C and keeps it at that temperature for 3 h and then cools down to 21°C.

Full cage NE_6 assembly: The two half-cages are directly mixed in equal volume without any previous purification. Previous experiments have shown that the assembly is complete after 3 hours incubation at 40 °C, using a Thermocycler Mastercycler nexus gradient (Eppendorf).

Purification of DNA origami via ultrafiltration: A 0.5 µL 100 kDa MWCO Amicon ultrafiltration unit is pre-washed with TEMg20 buffer and centrifuged at 5500 *g* for 2 min. Then the NE_6 cage is added to the device and centrifuged at 5500 *g* for 2 min, the filtered solution is discarded, and the tube is filled again with buffer. This operation is repeated at least 8 times and then the concentrated product solution is transferred in a new tube centrifuging for 2 min at 3000 g.

A small fraction of N_3, E_3 and unpurified NE_6 is kept aside for control. 150 µL of NE_6 cage 20 nM were obtained, with a reaction yield of 60%.

The reaction is checked by AGE and the final product is assessed by AFM and TEM.

5.4.7. HTRA1(SA) labelling with NHS-ester A488 fluorophore

The HTRA1(SA) is modified at the lysine residues with Alexa 488, using the N-hydroxy-succinimide ester of the fluorescent dye.

A 6 mM solution of A488 dye in DMSO is prepared. HTRA1(SA) 507.5 µM is mixed with A488 in different quantities, and HEPES buffer is added to keep DMSO concentration to 5% V\V:

(1) 10 µL HTRA1(SA) + 0.592 µL A488,
(2) 5 µL HTRA1(SA) + 0.888 µL A488,
(3) 3 µL HTRA1(SA) + 0.711 µL A488,
(4) 3 µL HTRA1(SA) + 1.066 µL A488,
(5) 3 µL HTRA1(SA) + 1.421 µL A488,
(6) 3 µL HTRA1(SA) + 1.776 µL A488.

The labelled samples are protected from light by wrapping the reaction vessel in aluminium foil. The reaction is left on ice for 3 h. Excess of A488 label is removed with Pierce™ Dye Removal Columns (Thermofisher) according to manufacturer's manual.

5.4.8. Encapsulation of HTRA1 inside the NE_6-TAMRA-F9-DPMFKL DNA origami cage

25 µL of NE_6-TAMRA-F9-DPMFKL 20 nM in TEMg20 buffer are mixed with 1,94 µL HTRA1(SA)-A488 in HEPES buffer at a concentration of 30.55 µM (120 times excess of protein to the cage). The mixture is incubated for 6 h at 37 °C in a shaker.

The excess of protein is removed with PEG precipitation: 105 µL of PEG (polyethylene glycol) buffer (15% w/V PEG8000, 5 mM Tris, 1 mM EDTA, 505 mM NaCl, 200 mM $MgCl_2$) is added to the solution, then centrifuged at 28600 *g* for 30 min at room temperature (25 °C). The supernatant is discarded, and the pellet is resuspended in TEMg20 overnight.

The encaging is assessed by AGE and TEM.

6. References

(1) Feynman, R. There's Plenty of Room at the Bottom. In *Feynman and Computation*; CRC Press, Boca Raton (2002).

(2) Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, *99* (2), 237–247.

(3) Seeman, N. C. DNA in a Material World. *Nature* **2003**, *421* (6921), 427–431.

(4) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738.

(5) Bruice, P. Y. *Organic Chemistry*, Eighth edition.; Pearson, Upper Saddle River, NJ (2016).

(6) Seeman, N. C.; Kallenbach, N. R. DNA Branched Junctions. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23* (1), 53–86.

(7) Seeman, N. C.; Kallenbach, N. R. Design of Immobile Nucleic Acid Junctions. *Biophys. J.* **1983**, *44* (2), 201–209.

(8) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302.

(9) Yan, H.; LaBean, T. H.; Feng, L.; Reif, J. H. Directed Nucleation Assembly of DNA Tile Complexes for Barcode-Patterned Lattices. *Proc. Natl. Acad. Sci.* **2003**, *100* (14), 8103–8108.

(10) Saccà, B.; Niemeyer, C. M. DNA Origami: The Art of Folding DNA. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 58–66.

(11) Tørring, T.; Voigt, N. V.; Nangreave, J.; Yan, H.; Gothelf, K. V. DNA Origami: A Quantum Leap for Self-Assembly of Complex Structures. *Chem. Soc. Rev.* **2011**, *40* (12), 5636.

(12) Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C.; Yan, H.; Zhan, P. DNA Origami. *Nat. Rev. Methods Primer* **2021**, *1* (1), 1–24.

(13) Pfeifer, W.; Saccà, B. From Nano to Macro through Hierarchical Self-Assembly: The DNA Paradigm. *ChemBioChem* **2016**, *17* (12), 1063–1080.

(14) Fu, J.; Wang, Z.; Liang, X. H.; Oh, S. W.; Iago-McRae, E. St.; Zhang, T. DNA-Scaffolded Proximity Assembly and Confinement of Multienzyme Reactions. In *DNA Nanotechnology: From Structure to Functionality*; Fan, C., Ke, Y., Eds.;

Topics in Current Chemistry Collections; Springer International Publishing: Cham, 2020; pp 125–155.

(15) Clausen, T.; Kaiser, M.; Huber, R.; Ehrmann, M. HTRA Proteases: Regulated Proteolysis in Protein Quality Control. *Nat. Rev. Mol. Cell Biol.* **2011**, *12* (3), 152–162.

(16) Grau, S.; Baldi, A.; Bussani, R.; Tian, X.; Stefanescu, R.; Przybylski, M.; Richards, P.; Jones, S. A.; Shridhar, V.; Clausen, T.; Ehrmann, M. Implications of the Serine Protease HtrA1 in Amyloid Precursor Protein Processing. *Proc. Natl. Acad. Sci.* **2005**, *102* (17), 6021–6026.

(17) Truebestein, L.; Tennstaedt, A.; Mönig, T.; Krojer, T.; Canellas, F.; Kaiser, M.; Clausen, T.; Ehrmann, M. Substrate-Induced Remodeling of the Active Site Regulates Human HTRA1 Activity. *Nat. Struct. Mol. Biol.* **2011**, *18* (3), 386–388.

(18) Jaekel; Stegemann; Saccà. Manipulating Enzymes Properties with DNA Nanostructures. *Molecules* **2019**, *24* (20), 3694.

(19) Duckworth, B. P.; Chen, Y.; Wollack, J. W.; Sham, Y.; Mueller, J. D.; Taton, T. A.; Distefano, M. D. A Universal Method for the Preparation of Covalent Protein-DNA Conjugates for Use in Creating Protein Nanostructures. *Angew. Chem. Int. Ed Engl.* **2007**, *46* (46), 8819–8822.

(20) Khatwani, S. L.; Kang, J. S.; Mullen, D. G.; Hast, M. A.; Beese, L. S.; Distefano, M. D.; Taton, T. A. Covalent Protein–Oligonucleotide Conjugates by Copper-Free Click Reaction. *Bioorg. Med. Chem.* **2012**, *20* (14), 4532–4539.

(21) Li, H.; Park, S. H.; Reif, J. H.; LaBean, T. H.; Yan, H. DNA-Templated Self-Assembly of Protein and Nanoparticle Linear Arrays. *J. Am. Chem. Soc.* **2004**, *126* (2), 418–419.

(22) Shen, W.; Zhong, H.; Neff, D.; Norton, M. L. NTA Directed Protein Nanopatterning on DNA Origami Nanoconstructs. *J. Am. Chem. Soc.* **2009**, *131* (19), 6660–6661.

(23) Williams, B. A. R.; Diehnelt, C. W.; Belcher, P.; Greving, M.; Woodbury, N. W.; Johnston, S. A.; Chaput, J. C. Creating Protein Affinity Reagents by Combining Peptide Ligands on Synthetic DNA Scaffolds. *J. Am. Chem. Soc.* **2009**, *131* (47), 17233–17241.

(24) Yang, Y. R.; Liu, Y.; Yan, H. DNA Nanostructures as Programmable Biomolecular Scaffolds. *Bioconjug. Chem.* **2015**, *26* (8), 1381–1395.

(25) Sprengel, A.; Lill, P.; Stegemann, P.; Bravo-Rodriguez, K.; Schöneweiß, E.-C.; Merdanovic, M.; Gudnason, D.; Aznauryan, M.; Gamrad, L.; Barcikowski, S.; Sanchez-Garcia, E.; Birkedal, V.; Gatsogiannis, C.; Ehrmann, M.; Saccà, B. Tailored Protein Encapsulation into a DNA Host Using Geometrically Organized Supramolecular Interactions. *Nat. Commun.* **2017**, *8* (1), 14472.

(26) Merdanovic, M.; Mamant, N.; Meltzer, M.; Poepsel, S.; Auckenthaler, A.; Melgaard, R.; Hauske, P.; Nagel-Steger, L.; Clarke, A. R.; Kaiser, M.; Huber, R.; Ehrmann, M. Determinants of Structural and Functional Plasticity of a Widely Conserved Protease Chaperone Complex. *Nat. Struct. Mol. Biol.* **2010**, *17* (7), 837–843.

(27) Shaw, A.; Benson, E.; Högberg, B. Purification of Functionalized DNA Origami Nanostructures. *ACS Nano* **2015**, *9* (5), 4968–4975.

(28) Chen, A. H.; Silver, P. A. Designing Biological Compartmentalization. *Trends Cell Biol.* **2012**, *22* (12), 662–670.

(29) Huang, J.; Gambietz, S.; Saccà, B. Self-Assembled Artificial DNA Nanocompartments and Their Bioapplications. *Small* **2022**, 2202253.

(30) Grossi, G.; Dalgaard Ebbesen Jepsen, M.; Kjems, J.; Andersen, E. S. Control of Enzyme Reactions by a Reconfigurable DNA Nanovault. *Nat. Commun.* **2017**, *8* (1), 992.

(31) Shrestha, P.; Jonchhe, S.; Emura, T.; Hidaka, K.; Endo, M.; Sugiyama, H.; Mao, H. Confined Space Facilitates G-Quadruplex Formation. *Nat. Nanotechnol.* **2017**, *12* (6), 582–588.

(32) Jonchhe, S.; Pandey, S.; Beneze, C.; Emura, T.; Sugiyama, H.; Endo, M.; Mao, H. Dissection of Nanoconfinement and Proximity Effects on the Binding Events in DNA Origami Nanocavity. *Nucleic Acids Res.* **2022**, *50* (2), 697–703.

(33) Hartl, F. U. Molecular Chaperones in Cellular Protein Folding. *Nature* **1996**, *381* (6583), 571–580.

(34) Mammen, M.; Choi, S.-K.; Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed.* **1998**, *37* (20), 2754–2794.

(35) Kitov, P. I.; Bundle, D. R. On the Nature of the Multivalency Effect: A Thermodynamic Model. *J. Am. Chem. Soc.* **2003**, *125* (52), 16271–16284.

(36) Gao, Y.; Roberts, C. C.; Zhu, J.; Lin, J.-L.; Chang, C. A.; Wheeldon, I. Tuning Enzyme Kinetics through Designed Intermolecular Interactions Far from the Active Site. *ACS Catal.* **2015**, *5* (4), 2149–2153.

(37) Idan, O.; Hess, H. Origins of Activity Enhancement in Enzyme Cascades on Scaffolds. *ACS Nano* **2013**, *7* (10), 8658–8665.

(38) Zhao, Z.; Fu, J.; Dhakal, S.; Johnson-Buck, A.; Liu, M.; Zhang, T.; Woodbury, N. W.; Liu, Y.; Walter, N. G.; Yan, H. Nanocaged Enzymes with Enhanced Catalytic Activity and Increased Stability against Protease Digestion. *Nat. Commun.* **2016**, *7* (1), 10619.

(39) Douglas, S. M.; Marblestone, A. H.; Teerapittayanon, S.; Vazquez, A.; Church, G. M.; Shih, W. M. Rapid Prototyping of 3D DNA-Origami Shapes with CaDNAno. *Nucleic Acids Res.* **2009**, *37* (15), 5001–5006.

(40) Dunn, K. E.; Dannenberg, F.; Ouldridge, T. E.; Kwiatkowska, M.; Turberfield, A. J.; Bath, J. Guiding the Folding Pathway of DNA Origami. *Nature* **2015**, *525* (7567), 82–86.

(41) Stahl, E.; Martin, T. G.; Praetorius, F.; Dietz, H. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angew. Chem.* **2014**, *126* (47), 12949–12954.

(42) Northrop, B. H.; Frayne, S. H.; Choudhary, U. Thiol–Maleimide "Click" Chemistry: Evaluating the Influence of Solvent, Initiator, and Thiol on the Reaction Mechanism, Kinetics, and Selectivity. *Polym. Chem.* **2015**, *6* (18), 3415–3430.

(43) Hermanson, G. T. *Bioconjugate Techniques*, Third edition.; Elsevier/AP, London (2013).

(44) Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. Selective Reduction of Disulfides by Tris(2-Carboxyethyl) Phosphine. *J. Org. Chem.* **1991**, *56* (8), 2648–2650.

(45) Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level*, 3rd ed.; Wiley, Hoboken, NJ (2008).

(46) Weber, K.; Osborn, M. The Reliability of Molecular Weight Determinations by Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. *J. Biol. Chem.* **1969**, *244* (16), 4406–4412.

(47) Zhou, Z.; Xiang, Y.; Tong, A.; Lu, Y. Simple and Efficient Method to Purify DNA–Protein Conjugates and Its Sensing Applications. *Anal. Chem.* **2014**, *86* (8), 3869–3875.

(48) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. A DNA-Fuelled Molecular Machine Made of DNA. *Nature* **2000**, *406* (6796), 605–608.

(49) Good, N. E.; Winget, G. D.; Winter, W.; Connolly, T. N.; Izawa, S.; Singh, M. Hydrogen Ion Buffers for Biological Research. *Biochemistry* **1966**, *5* (2), 467–477.

(50) Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*, 6th ed.; Thomson Brooks/Cole, Belmont, CA (2007).

(51) Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. – Eur. J.* **2019**, *25* (1), 43–59.

(52) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526* (7575), 687–691.

(53) Hoyle, C. E.; Bowman, C. N. Thiol-Ene Click Chemistry. *Angew. Chem. Int. Ed.* **2010**, *49* (9), 1540–1573.

(54) Kasper, M.; Glanz, M.; Stengl, A.; Penkert, M.; Klenk, S.; Sauer, T.; Schumacher, D.; Helma, J.; Krause, E.; Cardoso, M. C.; Leonhardt, H.; Hackenberger, C. P. R. Cysteine-Selective Phosphonamidate Electrophiles for Modular Protein Bioconjugations. *Angew. Chem. Int. Ed.* **2019**, *58* (34), 11625–11630.

(55) Konč, J.; Brown, L.; Whiten, D. R.; Zuo, Y.; Ravn, P.; Klenerman, D.; Bernardes, G. J. L. A Platform for Site-Specific DNA-Antibody Bioconjugation by Using Benzoylacrylic-Labelled Oligonucleotides. *Angew. Chem. Int. Ed.* **2021**, *60* (49), 25905–25913.

(56) Wagenbauer, K. F.; Engelhardt, F. A. S.; Stahl, E.; Hechtl, V. K.; Stömmer, P.; Seebacher, F.; Meregalli, L.; Ketterer, P.; Gerling, T.; Dietz, H. How We Make DNA Origami. *ChemBioChem* **2017**, *18* (19), 1873–1885.

(57) Thomsen, R. P.; Malle, M. G.; Okholm, A. H.; Krishnan, S.; Bohr, S. S.-R.; Sørensen, R. S.; Ries, O.; Vogel, S.; Simmel, F. C.; Hatzakis, N. S.; Kjems, J. A Large Size-Selective DNA Nanopore with Sensing Applications. *Nat. Commun.* **2019**, *10* (1), 5655.

(58) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux. *Nat. Biotechnol.* **2009**, *27* (8), 753–759.

7. Acknowledgments

I would like to thank my supervisor at Padova university Prof. Fernando Formaggio and Prof. Chiara Maccato, my supervisor at the university of Duisburg-Essen, Prof. Barbara Saccà. I want to thank all the members of AG Saccà and AG Kaiser research groups, especially Jing Huang.

I would like to also thank my friends and my family, who supported me.