THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 17 juin 2022 par

Sevde Nur KARATAS

Controlled covalent protein-DNA coupling for building DNA encoded protein assemblies

Discipline
Biochimie Structurale

École doctorale
ED 62- SCIENCES DE LA VIE ET DE LA SANTE

Laboratoire/Partenaires de recherche
Laboratoire d'Ingénierie des Systèmes Macromoléculaires (LISM UMR 7255)

Composition du jury

Thierry Michon	Rapporteur
(INRAE) UMR 1332	
Isabelle Mus-Veteau	Rapporteure
(CNRS) UMR 7275	
Gaetan Bellot	Examinateur
(INSERM) UMR 5048	
Kheya Sengupta	Examinatrice/Presidénte de Jury
(CNRS) UMR 7325	
James N. STURGIS	Directeur de thèse
(CNRS) UMR 7255	

Affidavit

Je soussigné, Sevde Nur Karatas, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de James N. Sturgis, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Marseille, le 13/04/2022

Liste de publications et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse :

1. Controlled covalent protein-DNA coupling for building DNA encoded protein assemblies (to be submitted).
   Authors: Sevde N. Karatas, Valérie Prima and James N. Sturgis

Participation aux conférences et écoles d'été au cours de la période de thèse :

1. 26-28 September 2018, October 15-19,2018 Montpellier**:**

   Inserm Workshops, 252, Bordeaux/France, Workshop: "DNA/RNA

   Nanotechnology biological applications

   Phase I - Critical Assessment:

   Phase II- Technical Workshop

Résumé

La nanoconstruction avec ADN est utilisée comme un outil puissant pour étudier le fonctionnement de systèmes biologiques comlpexes. Dans cette étude, nous avons utilisé et dévéloppé une nouvelle méthode pour construire des assemblages en utilisant des protéines couplées à l'ADN avec des réactifs bifonctionnels. J'ai montré la fonctionalité de ces méthodes avec le coulage et l'assemblage de deux protéines fluorescentes la sfGFP et la mRuby3. Nous avons pu confirmer la formation des assenblages souhaités par transfert d'énergie de fluorescence et par electrophorèse. Notre méthode nous permet de concevoir positionner plusieurs differentes molécules, telles que des protéines, dans assemblages complexes guidées par l'ADN. Ceci offre une plateforme nanotechnologique pour construire des nouveaux metamateriaux mais également permet de tester des hypothèses sur le role de l'organisation dans le fonctionnement de systèmes biologiques telles que les voies de signalisation, ou bien les membranes biologiques. Dans l'ensemble, cette méthode offre flexibilité, polyvalence et contrôlabilité à la conception en raison des caractéristiques de l'ADN lui-même et des modifications chimiques pour les protéines de manipulation sont simples et peu coûteuses. Afin de permettre l'utilisation dans les constructions complexes j'ai également étudié la formation d'assemblages d'ADN dite origami en conditions isothermes. Cette methode basé sur la gestion de la concentration d'ions promet la formation d'assemblages complexes a des temperatures ambiantes sans recours a une chaufage susceptible a denaturer des protéines.

Mots clés : ADN, nano-assemblage, réactifs bi-fonctionnels, assemblages macromoléculaires guidés par l'ADN.

Abstract

DNA nanoassembly offers a powerful tool for the study of complex biological systems. In this study I have used and developed a new method for the construction of complex assemblies using DNA attached to proteins by bi-functional reagents. I have demonstrated the fitness of these methods with the linkage and assembly of two fluorescent proteins sfGFP and mRuby3. I was able to confirm the formation of the expected and designed assemblies by fluorescence energy transfer and gel electrophoresis. My method allows to imagine the accurate positioning of multiple different molecules, such as proteins, in DNA guided macromolecular assemblies. This opens the possibility of using DNA guided assembly as a platform for the construction of complex metamaterials and for the investigation of the role of organization in the function of biological systems such as signaling pathways or membranes. Altogether this method offers flexibility, adaptability, and controllability in the design of complex assemblies based on DNA properties and the ease and controllability of the chemistries used. To pave the way for their use in the formation of complex structures I also studied the formation of, so called, origami DNA assemblies under isothermal conditions. This technique based on the control of ionic strength ammowl the assembly of complex structures at ambient temperatures without using a heating step that could well denature proteins.

Key Words: DNA, nano assembly, bi-functional reagents, DNA guided macromolecular assemblies.

Acknowledgements

The completion of this dissertation would not be possible without the support of many people. First, I would like to thank to my thesis advisor Prof. James Sturgis is an amazing scientist and mentor, and he provided me the best training environment I could have imagined. I was pushed to come up with my own ideas and learned how to write and speak like a scientist. His willingness to share his knowledge and expertise in every single step and progress of the ongoing projects at any time day and night was very fruitful for shaping up my ideas scientifically. It was an honor and great experience to work with him. I would like to thank Research Engineer Valérie Prima who provided me her generous guidance, constant positive attitude, and excellent training during my PhD. She taught me how to be strong even in the worst scenarios.

I would like to thank the entire Sturgis Lab, past and present members; Prof. Jean-Pierre Duneau, Dr. Lucie Bergdoll, Dr. Marcel Giansily and Batiste Thienpot for their help with designing experiment, their friendship and support.

I would also thank to my best friends because things are never quite as scary when you have best friends. I have overcome many challenges about my academic and personal life thanks to them. I would like to thank Sinem Usluer, Nazlı Kocatuğ, Aurely Cébe, Kevin Johan, Nur Su Çakır and Bülent Uraz for sincere friendship.

Without my family, I would not be where I am today. I would like to thank my parents Şirvan Savsun and Kemal Karataş, my sister Yağmur Karataş and Denat Pavli for supporting me in everything that I do. They always pushed me to do better, and my drive to succeed was instilled from a young age. Thank you from the bottom of my heart.

Abbreviations and Acronyms

2-ME	2-mercaptoethanol
BChl	Bacteriochlorophyll
CC	Core Complex
CuAAC	Copper(I)-catalyzed azide-alkyne 1,3-dipolar cycloaddition
DBCO	Dibenzyl Cyclooctyne
DTT	Dithiothreitol
ER	Endoplasmic Reticulum
HPLC	High Performance Liquid Chromatography
ICM	Intra Cytoplasmic Membrane
LH	Light-harvesting
NHS-ester	N-hydroxysuccinimide Ester
PSU	Photosynthetic Unit
PTM	Post-translational Modification
RC	Reaction Center
Rsb.	Roseobacter
Rb.	Rhodobacter
SPAAC	Strain-promoted Alkyne-Azide Cycloaddition
TCEP	Tris (2-carboxyethyl) phosphine
EM	Electron Microscope
EtBr	Ethidium Bromide
FRET	Förster Resonance Energy Transfer
SST	Single-Stranded Tiles
TAEMg	Tris - Acetate - EDTA + MgCl2 buffer
TANa	Tris - Acetate + 100 mM NaCl (for 1x) buffer

TAENa	**Tris - Acetate - EDTA + 100 mM NaCl (for 1x) buffer**
(T)EM	**(Transmission) Electron Microscopy**
TEMg	**Tris-EDTA buffer**
Tm	**Melting Temperature**

TABLE OF CONTENTS

Chapter 1 DNA nanotechnology

The object of supramolecular chemistry is understanding and controlling non-covalent interactions which exist between molecules in a defined environment. Supramolecular chemistry has a capability of creating highly complex and functional molecules, ranging from catenanes to molecular muscles or nano robots. This field allowed to open new perspectives on how to use molecules not only as reactants, but also as building blocks for larger, more complex, and reconfigurable structures. With the development of supramolecular chemistry, the perspective of scientists on some molecules changed. In the beginning of 80's, DNA started to be regarded as a potential building block for nanomaterials. This eventually led to the birth of DNA nanotechnology which has the potential to be a valuable tool for the creation of nano objects with an unprecedented control over their shapes and sizes.

In the first section of chapter 1, I will provide a general overview about the DNA structure and its potential as material for nanofabrication. In the second section, I will present the more important technologies in this field: DNA origami and DNA Single-Stranded Tiles (SST) as well as their most common applications. Also, it will be described the thermal assembly and isothermal self-assembly process of DNA origami. The final part of the section describes application of DNA nanotechnologies.

1.1. Structure of DNA

Deoxyribonucleic acid (DNA) is the main macromolecule which encodes genetic information in all living organisms. DNA was discovered for the first time as nuclein in 1869 by Friedrich Miescher and its name changed from the original nucleic acid to deoxyribonucleic acid [1]. The chemical composition of DNA was demonstrated by Albrecht Kossel, Phoebus Levene [2] and Erwin Chargaff [3]. In 1953, Francis H.C Crick, Rosalind Franklin, James D. Watson, and Maurice H.F. Wilkins published the structure of this key molecule, and they were awarded the Nobel Prize in Physiology and Medicine in 1962.

1.1.1. Composition of single strand DNA

A DNA molecule is composed of two polynucleotide chains which are known as DNA strands. DNA strands are copolymers consisting of four repetition of monomers called nucleotides (see figure 1.3). A nucleotide is a chemical unit constituted of three parts: a phosphate group, a deoxyribose sugar, and a nitrogenous base, as presented in Figure 1.1. The association between sugar and nitrogenous base is defined as nucleoside.
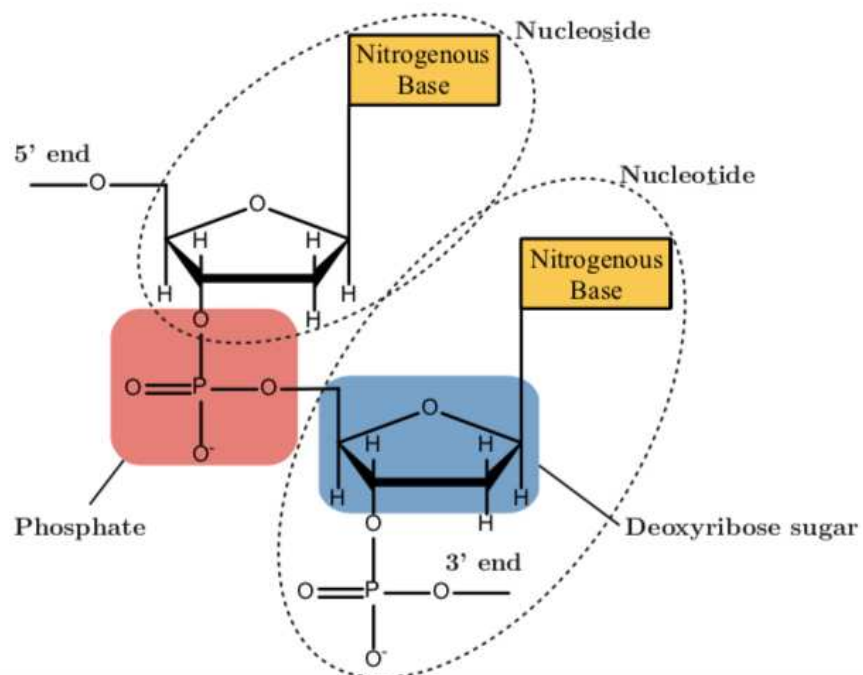
Figure 1.1. *A DNA strand is composed of a succession of nucleotides composed of a phosphate group (pink), deoxyribose sugar (blue) and nitrogenous base (yellow) Adapted from* [4]*.*

The attachment of the phosphate group to the nucleoside forms a nucleotide. The single nucleotides bind each other through a phosphodiester bond between the 5' primary alcohol group of one sugar and the 3' secondary alcohol group of the adjacent nucleoside's deoxyribose. Each polynucleotide chain of dsDNA is constructed by nucleotides through phosphodiester bonds. At this point, I need to mention two features of the DNA structure: firstly, the phosphate molecule in the DNA structure does not only allow the construction of DNA strands but is also responsible for the negative charge of DNA molecules, because of the negatively charged oxygen atoms

of the phosphate group. Second, the pentose moiety does not have an alcohol functional group on its 2' carbon, thus it is named as deoxyribose [4].

In the chemical composition of DNA in all living organisms, while the phosphate group and deoxyribose sugar are common, the combination of the nitrogenous bases makes DNA different for each organism. In the 19th century, the nitrogenous bases were discovered by Albrecht Kossel as four different bases: Adenine, Thymine, Guanine, Cytosine. These four bases are heterocycles containing nitrogen atoms as presented in Figure 1.2.

Adenine

Thymine
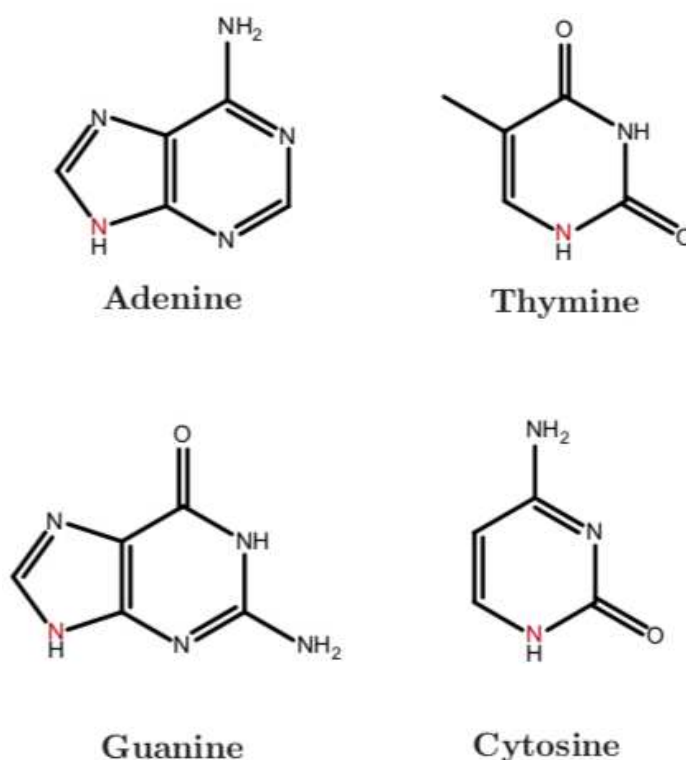
Guanine

Cytosine

Figure 1.2. Structures of the four nitrogenous bases adenine, thymine, guanine, and cytosine. Adenine and Thymine are known as Purines Guanine and cytosine is known as Pyrimidines. The red nitrogen atom representation of the attachment points between the bases and the 1' carbon of the ribose. Adapted from [4].

The nitrogenous bases are categorized as purines and pyrimidines according to their structures. While pyrimidines (thymine and cytosine) are heterocyclic molecules composed of a six-atom cycle, the purines (guanine and adenine) are bi-heterocyclic consisting of a six-atom cycle conjugated with a five atoms cycle as molecules [5].

1.1.2. Formation of double helix

The DNA molecules are formed as a double helix which means that two strands interact with each other. The interaction between two strands of DNA forms by the hydrogen bonds between the nucleotides. There are A, B and Z forms of double helix. In the A form, since there is less distance between bases, it makes the turn every 11 base pairs and it is a right-handed double helix. The Z form of double helix is left-handed and has lots of purines and pyrimidines that make alterations consecutively. However, among these three forms of double helix forms, the most common form adopted by DNA in solution is the right-handed double helix (named B-DNA) but depending on environmental conditions such as pH or ions present in solution, a left-handed Z-DNA helix or an A-DNA conformation can be observed. In living organisms, only B- and Z-DNA have been observed [6].
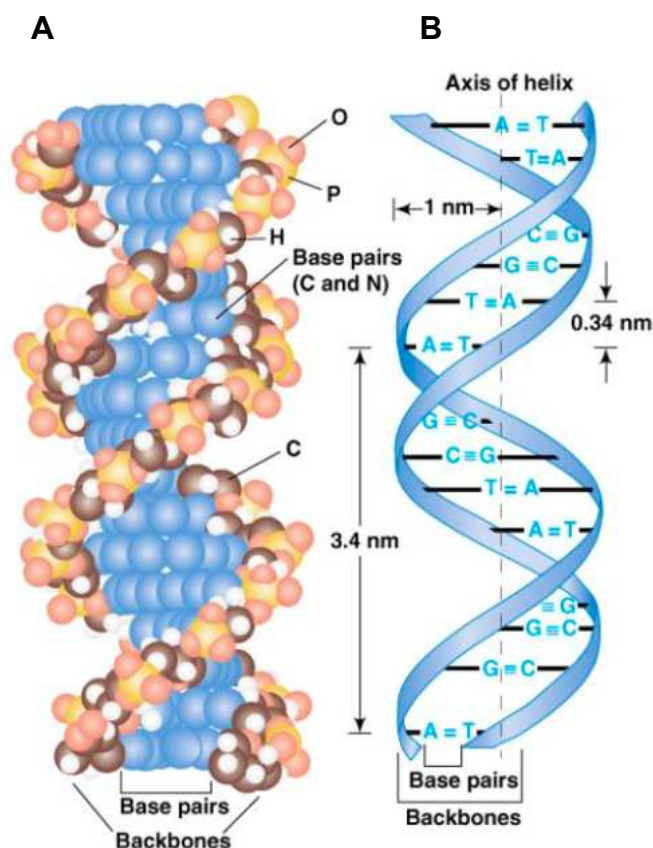
Figure 1.3. *Representation of DNA double helix (B-formed) with characteristic inter-helical and inter-helical turn distances [4].*

As anticipated before, the interaction between two strands of DNA forms by the hydrogen bonds between the nucleotides. To form base pairs, while Adenine and Thymine make two hydrogen bonds, Guanine and Cytosine form three hydrogen bonds. This variety of hydrogen bonds between nucleotides affects the melting point of the DNA. The backbones of the DNA molecules are made of deoxyribose molecules connected to each other through phosphates. While the base pairs point to the center of the helix, the backbones stay in the outer part of the helix (see figure 1.3.A). Additionally, the strands of DNA molecules run antiparallel, meaning that the 5' end of one strand is next to 3' end of the other. Therefore, the double helix forms the secondary structure of the DNA [4].

The double helix does a full turn every ten pairs of bases, and the period length is around 3.4 nm. It thus turns +36 degrees per base pair and the distance between two strands is about 2 nm (see figure 1.4.B) [7]. Additionally, since Adenine-Thymine and Guanine-Cytosine interaction creates the minor and major groove of double helix, relying on the groove, double helix turns either closer to each other or further apart [8].

The length of the genome depends on the organism, and it ranges between a few thousands to several millions of base pairs. In the human genome, the length of DNA is about two meters, and it is estimated to be composed of around $3x10^9$ base pairs. However, the 2 meters of DNA in the human genome is packed tightly with histones proteins. As the length and packaging of the genome relies on the organism, this phenomenon is different in bacteria. The bacterial genome can range in size from 130 kbp to 14 Mbp and the packing of genomic DNA in bacteria is mediated by the DNA binding proteins [9].

1.1.3. Other nucleotide structures

The structure of oligonucleotides is not restricted to only double helix formation but also other structures with the numerous nucleotides associations might be found both in nature and synthetic systems. As an example, hairpin formation occurs when the single strand DNA has a part of sequence that are complementary to each other. Hence, the complementary part of the sequence matches, the molecule folds up on itself. The parts of the sequence that completes each other are called stems; the rest of the sequence is defined as a loop. These structures are mostly found on single

stranded DNA molecules and occur in RNA molecules (Ribonucleic acids which have a similar structure to DNA and crucial role on gene expression). The hairpin patterns on RNA molecules can act as recognition sites for RNA binding to proteins for example and therefore they are of crucial interest in living organisms. These kinds of structures are mostly appeared cellular processes such as: replication on the template for lagging-strand synthesis, DNA repair, rolling-circle replication (RCR), therefore they have crucial roles in living organisms [10].

1.2. Evolution of DNA nanotechnology

Nanotechnology is one of the most promising technologies capable of converting the nanoscience theory to useful applications by measuring, manipulating, assembling, controlling, and manufacturing matter at nanometric scale [11]. Nanotechnological construction can be designed in two different ways as top-down and bottom-up systems. In the top-down systems, the structures require some manipulations at microscopic scale. On the other hand, in bottom-up constructions, building nanostructures is done by utilizing to advantage material properties. Among the materials which can be usable for this purpose, DNA has attractive features; its diameter around 2 nm and its short structural repetitions of about 3.4 nm-3.6 nm which are quite convenient to work at nanometric scale. As a chemical property, DNA is one of the best candidates to be part of bottom-up construction in nanotechnology [12]. This is especially true since it is structurally well defined on nanometric scale and because of the specificity of the base pairing, which makes the material programmable on the nanometric scale. Additionally, it can be rapidly synthesized and modified by various reactive groups depending on the purpose and has a high chemical and structural stability. All these properties make the DNA an ideal material for DNA nanofabrication [13].

The creation of the programmable molecules started in early 1970 by using the advantage of sticky ends. Sticky ends are sequences which can exist at the end of the dsDNA. Therefore, when the sticky ends are designed as complementary to each other, they can hybridize with each other. By this way, the connection between two double strand DNA would be possible. Thus, sticky ends allow to design

programmable molecules by DNA [12]. However, tacking together double strand DNAs limits the design in the sense of dimension and offers just longer lines.
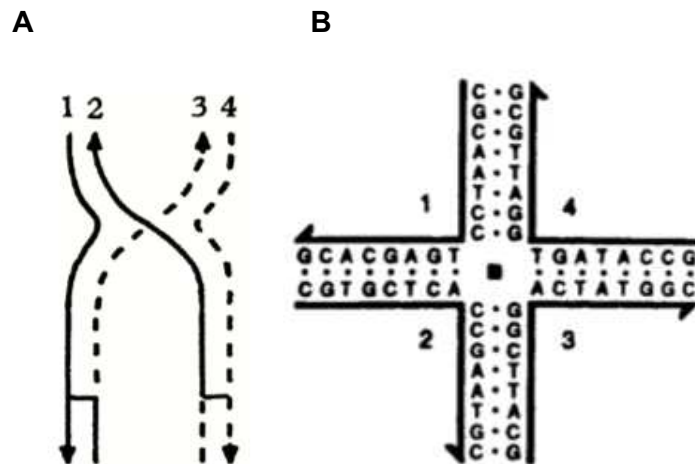
Figure 1.4. *Holliday Junctions **A** Representation of DNA strands of chromatids [14]. **B** Representation of four arm Holliday junction [15].*

Besides from assembly of dsDNA with sticky ends, there is a way to create a 2D Structure, which is composed of multiple dsDNA, by the aid of Holliday junctions. It can consist of four strands complementary by two or four strands partially complementary to two other strands [14] (See Figure 1.4).

In the 80's, the perspective about DNA was changed in a more comprehensive way by the first publication of a DNA nanotechnology paper which is Nucleic Acid Junctions and Lattices [16]. Nadrian Seeman and Alexander Rich predicted that DNA could be used as a building material for nanoscale fabrication. With the contribution of Marvin Caruthers [17], who developed DNA synthesis methodology to make the DNA synthesis easily available, Seeman introduced the possible formation of complex DNA nanostructures. The complexity of DNA nanofabrication increased with the contribution of N. Seeman over the past thirty years. To show how much DNA nanotechnology was developed by the time, an overview is presented in Figure 1.5.

A   **B**

C
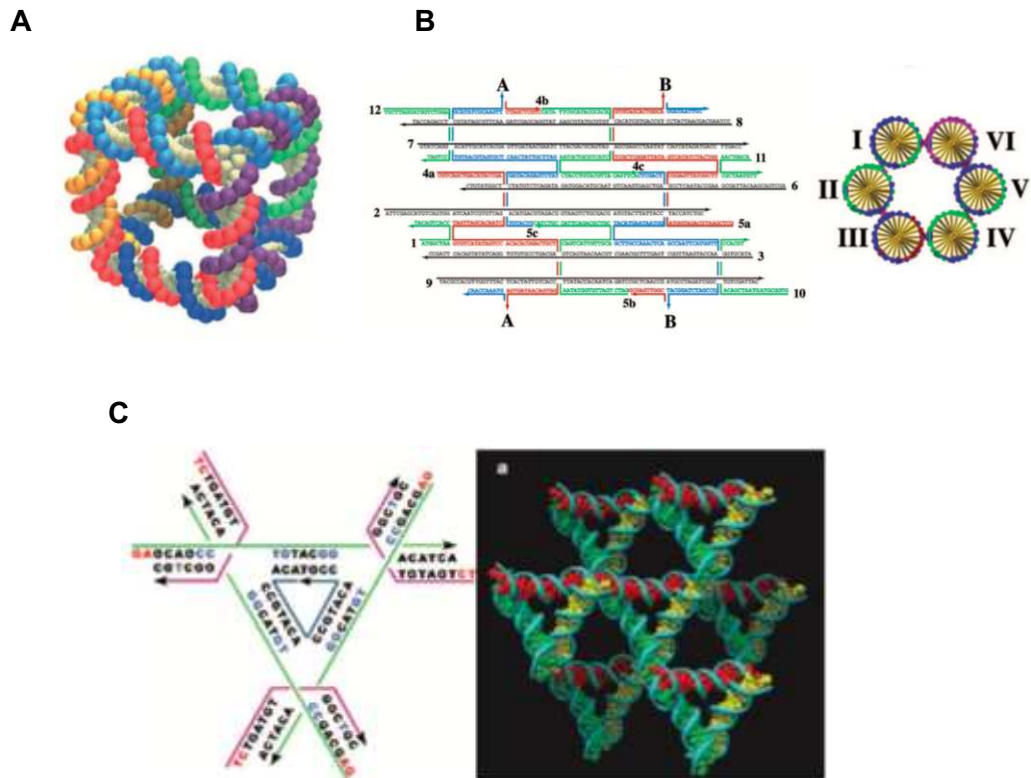
Figure 1.5. Summary of N.Seeman et coll. achievements in DNA nanotechnology. *A* DNA cube constructed from connected three-arm junctions Adapted from [13]. *B* Six helix bundle constructed from DNA. The strand sequences are shown for a version of this motif with 14 nucleotide pairs between crossovers. Points A and B are the places where the cyclic bundle closes. Strand numbering is indicated. There are six helical strands, shown in black, and the other strands perform the crossovers (left). A cross-sectional geometrical view of a six-helix bundle in which the helices are indicated by Roman numerals (right). Adapted from ref [18]. *C* 3D DNA crystals consist of triangular DNA motifs. The three unique strands are shown in magenta (strands restricted to a single junction), green (strands that extend over each edge of the tensegrity triangle) and dark blue (one unique nicked strand at the center passing through all three junctions). Arrowheads indicate the 39 ends of strands. Nucleotides with A-DNA-like characteristics are written in bright blue. Cohesive ends are shown in red letters (left). This stereoscopic image distinguishes three independent directions by base-pair color. The central triangle is flanked by six other triangles (right) [19].

The 3D DNA cube, which is composed of DNA strands in three-armed junctions, was reported at the beginning of the development of DNA nanotechnology (figure 1.5.A). Even though the structure formed in a low yield because of the technology of the time (the gel based structural characterization used could detect the strand in closed cycles only), it demonstrated that the possible formation of 3D DNA structures was incredibly promising for the future of DNA nanotechnology [13]. As another significant example,

a cyclic DNA motif was designed, composed of six DNA double helices that are linked to each other at two crossover sites (see figure 1.5.B). Thanks to Watson-Crick hydrogen bonding, small sticky ends can direct the associations of larger DNA molecules in a sequence-specific fashion such as crossover molecules (DX, TX, PX) [18]. This study demonstrated that sticky-ended cohesion pioneers not only to programmable affinities between molecules, but also to predictable local product structures. A first example of a self-assembled 3D DNA crystal was published in 2009 (figure 1.5.C). This 3D DNA crystal relied on the association of branched, rigid DNA 3D triangular motifs tailed through the short sticky ends [19]. This report demonstrated that the association of these sticky ends in directions that span 3D space results in a crystal-line arrangement which might be used as template for protein crystallization and as porous host for nano-materials.

Even though the development of DNA nanofabrication was started with N. Seeman, developments were not limited to just his works. In 2006, formation of 2D DNA nanostructures (DNA origami) was reported by Paul Rothemund, and Yin developed single stranded tile technology by the assembly of a high number of single stranded DNA into a predefined shape. Both technologies will be the subject of the following paragraph.

1.2.1. Techniques to create DNA nanostructures as 2D and 3D

Even though the work of Seeman made an enormous contribution to the nanofabrication field, the limitation of 3D DNA nanostructure was that they were not suitable to form in large dimensions. Thus, it was crucial to raise the size and complexity of the structures for the DNA nanofabrication field. Therefore, this evolution path was improved through the two technologies as DNA tiles and DNA origami.

1.2.1.1. DNA tiles

Yin and his colleagues presented a solution to the challenge of 2D nanostructure, limited in size and complexity, by introducing the technology of DNA Tiles. These DNA tiles are single strand DNA molecules that are designed to form the desired shape through the sticky ends complementary with each other, relying on the Watson-Crick base pair rule [20]. Sticky ends are the base pairing of single-stranded overhangs [13]. One sticky-ended molecule self-assembles to form one-dimensional arrays 1-15 microns [18].
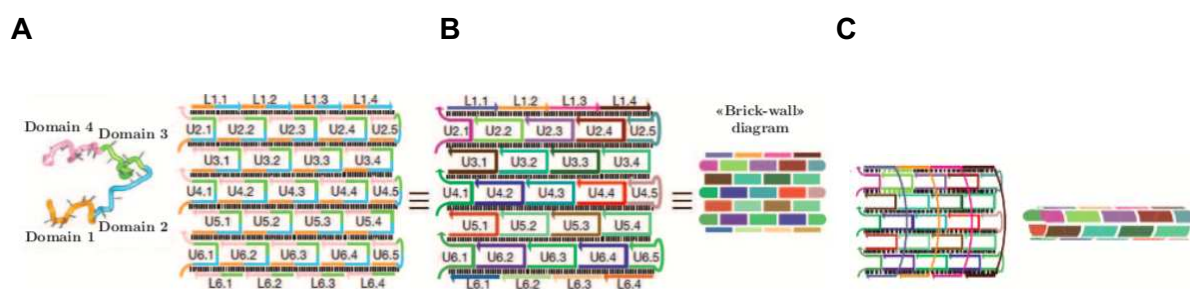
Figure 1.6. *Formation of Single strand Tiles from a mix of unique ssDNA **A** Each ssDNA is separated into four domains. **B** Formation of "Brick wall "diagram **C** Formation of DNA nanotube. Adapted from [20].*

In the design of Yin and his colleagues', single stranded tiles are made of a 42-base strand of DNA, composed of concatenated sticky ends, and they bind to four local neighbors during self-assembly. In figure 1.6.A, each brick (labeled as U) is composed of four domains and each brick has its own unique sequence. Each domain can bind the other domain of the other brick according to their sequence; hence the bricks can connect intermolecularly with each other to become the desired shape as shown in figure 1.6.B. It is prevented the formation of unpaired sequences by addition of shorter DNA strands (labeled as L). Long L "side" strands are also used to force the wall to fold on itself and create a tube-like shape as it is shown in figure 1.6.C. Since the complexity and the size of the molecule is controlled by the number and design of the DNA tiles, other shapes could be obtained by selectively removing some of the bricks [20]. Thus, thanks to this technique, the size and complexity is not limited with a few geometric structures. Additionally, Yin and his colleagues designed bricks from the single strand tile and hence obtained 2D and 3D structures. In this design, all bricks form a unique shape, when they are integrated into the target structure.
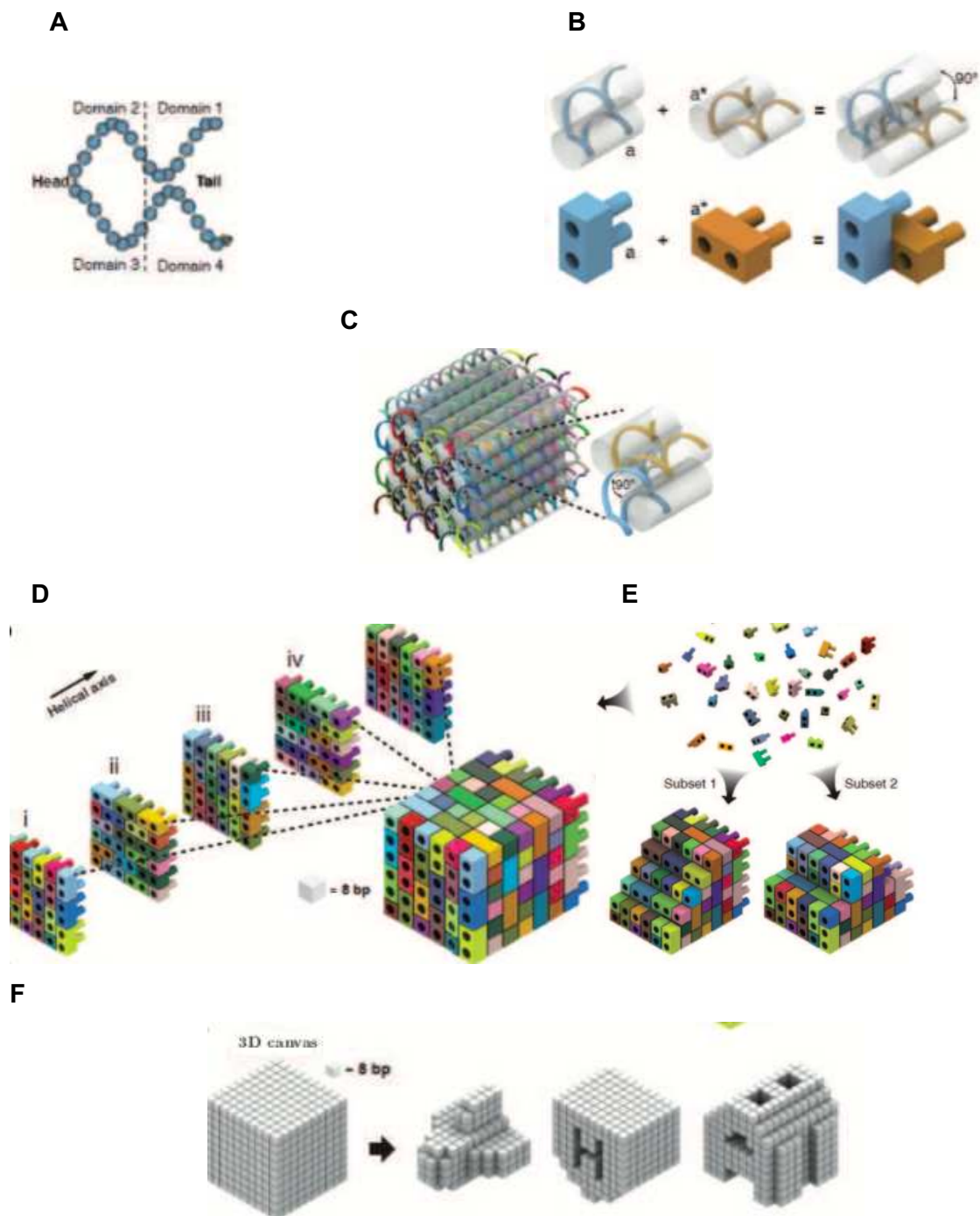
Figure 1.7. *3D Single strand tiles.* ***A*** *3 D structures consist of ss-DNA of 32 bases separated into four domains.* ***B*** *Assembly of two strands.* ***C*** *Molecular model of 3 D assembly composed of all strands.* ***D*** *LEGO model of 3 D cube.* ***E*** *Through selection of strands it is possible to obtain any desired 3 D shape.* ***F*** *Representation of 3 D cubes as a 3D canvas composed of vowels. Adapted from* [21].

A single DNA brick is composed of 32 nt and has four domains and each domain (sticky ends) consists of 8 nt (figure 1.7.A). Among the four domains, the ones that are linked to each other by phosphodiester bonds are called head domain, the other two are called tail domains as presented. An orange DNA brick with a tail domain with a sequence represented as ''a'' makes a linkage with a complementary sequence as

''a*'' head domain of the blue brick. All brick pairs form three parallel helices with 90 dihedral angles (figure 1.7.B). The difference here is that the strands bind with other strands by forming a 90° angle, allowing for the construction of a 3D DNA object. The structural design of this 3D object was shown in 6 helix by 6 helices by 48B (bp) cuboid structure. (figure1.7.C and D). Since each brick has a particular sequence that guides it to fit only to its pre-designed position, the cuboid is self-assembled from DNA bricks in a one-step reaction (figure 1.7.E). It is convenient to create a more complex structure like a 3D cuboid canvas which includes thousands of voxels that are incorporated with each other via the distinct nucleotide sequence for each. The authors reported the formation of a hundred different 3D shapes ranging from the "simple" cube to more complex sculpted shapes [21].

In 2017, the group of Yin managed to produce a 20 times larger 3D DNA Canvas than the previous one. This demonstrated that SST technology can create highly complex structures [22].

The technology of SST has reached a level that it is possible to control the shape and size both in 2D and 3D. Also, since there is a high specificity of DNA base pairing, folding reaction conditions and sequence complementary between the SST, the presence of non-complementary DNA strands does not create a critical issue during the experimental procedure. Nevertheless, there is still a possibility that the efficiency of formation might be low. The appearance of these cases triggered the development of new technologies to create DNA nanostructures as DNA Origami.

1.2.1.2. DNA origami

Differently from the single strand tiles technique, in DNA origami, a nanostructure is assembled from a very long single-stranded scaffold molecule held in place by many short single-stranded staple oligonucleotides instead of the usage of DNA sticky ends to connect different tiles. In this technique, DNA nanostructure forms through hybridization of staple strands and scaffold due to each staple strand being targeted to the specific location in the scaffold in the sequence design of the structure [23].

As I described in section 1.2.1.1, there are attractive features of single stranded tiles technique. Every single component in the DNA nanostructures can be included, excluded, or replaced without changing the remaining structure and 3 D structures can be built through the short oligonucleotides without requiring long or fragile single stranded DNA. However, the assembly in solution of a high number of DNA strands to form a large DNA nanostructure is limited by the encounter probability of complementary strands. The formation of the "brick wall" is a sequential process and the integration of each brick into the structure leads to entropy loss for the system. It is probable that the entropic cost for the formation of complex DNA nanostructures with the SST technique is one of the reasons why the observed yields are low. Thus, this challenge triggered the birth of another way of technology to design DNA nanostructure called scaffold DNA origami in 2006 by Paul Rothemund. The DNA origami technique has potential to decrease the entropy loss because staple strands are targeted to the specific location in the scaffold, and it makes the system organize before the final assembly. Pre-organized systems have fewer degrees of freedom and the entropic loss over the assembly process is therefore less significant.

As I mentioned before, there are two elements of this reaction. The first one is circular or linear ssDNA which is composed of 7249 nucleotides from the M13mp18 virus, called the scaffold. The second one is small oligonucleotides that are named as staples (30 nucleotides). The principle of the technique is that each staple can bind different domains of the scaffold and form a desired shape. The folding is achieved by staple strands and the complexity of the structure relies on composition of the staples. In general, the size of 2D origami molecules is around 100 nm [24].

The standard experimental procedure consists of heating up to 90°C-95°C that prevents the misfolding between staples and gradually cooling down to around 20°C

that ensures that the folding procedure has been done among staples and scaffold M13. Here, the buffer composition, which is TAEMg (Tris, Acetate, EDTA and Mg^{+2}) and large excess of staples, such as 100 X, is important to ensure to obtain the desired shape.

There are a few advantages of DNA origami technology compared to the SST technology. Firstly, since the number of staples in the solution was more than the scaffold, the folding process is maintained by staples. These excessive staples provide an advantage to the process by shifting the equilibrium through the formation of origami since a large excess of staples increases the possibility that the staples might find the right position easily on the scaffold. Secondly, the incubation time of DNA origami molecules is less than SST technology [25].

In 2009, the DNA origami technology application field was expanded to 3D structures by Shih and his colleagues. This design was done through stacking parallel helices in a honeycomb lattice.
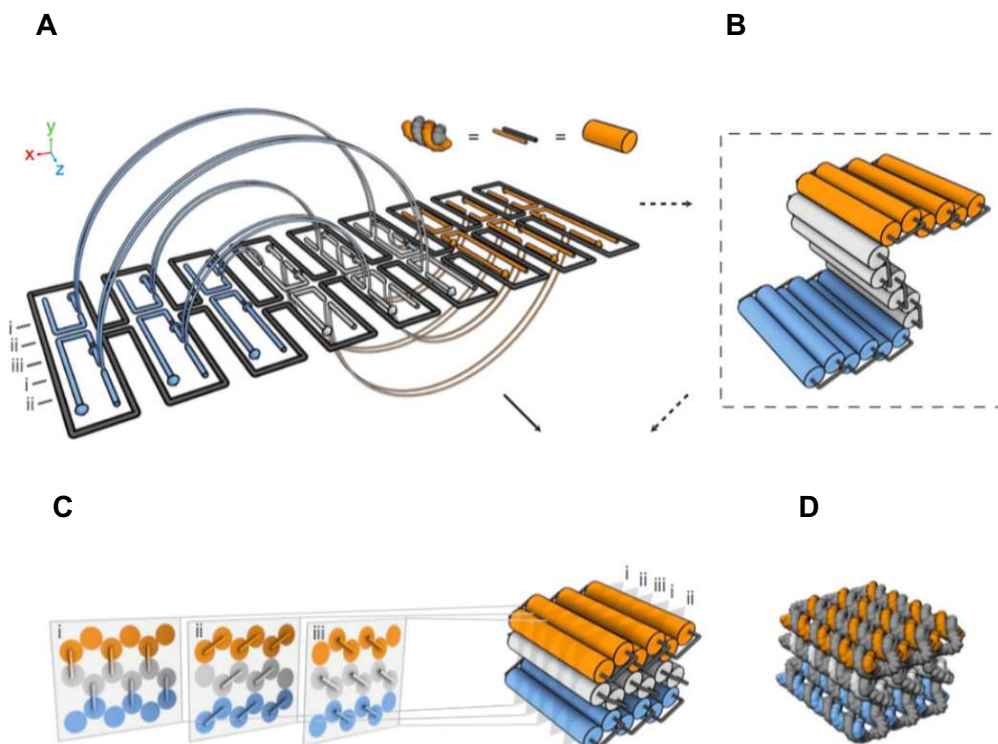
*Figure 1.8. Formation of 3D Origamis through stacking sheets of helices on top of each other. **A** Formation of honeycomb lattices. **B-D** Example of shapes achieved with this method. **C** Principle of the technique for the assembly of square lattices. Adapted from [26].*

The purpose of this study was assembling a target three-dimensional shape using the honeycomb-pleat-based strategy. An unrolled two-dimensional schematic of the target shape is presented in figure 1.8.A. In this scheme, double helices are composed of scaffold (gray) and staple strands (orange, white, blue) and phosphate linkages form crossovers between adjacent helices, with staple crossovers bridging different layers. The cylinders in the model represent the double helices, with loops of unpaired scaffold strand linking the ends of adjacent helices in figure.1.8. B. Complementary staple strands wind the scaffold strands in an antiparallel direction to assemble B-form double helices. The geometrical configuration is characterized by a diameter of 2 nm, a base-pair pitch of 0.34 nm and a rotation of 34.3 degrees base-pair average twist (or 21 base pairs / 2 turns). Crossovers between adjacent scaffold helices occurred at positions displaced upstream or downstream of the corresponding staple-crossover points by 5 base pairs or a half-turn, thanks to the features of B-form DNA. Each helix in the honeycomb lattice was connected to three others, making angles of 0°, 120°, and 240° (figure 1.8.C). A more explicit DNA model, showing individual strands, of the obtained honeycomb is shown in figure 1.8.D [26].

Thus, Shih achieved to design curved 3D DNA origami structures and later the method of Shih was furtherly improved by Yan and his colleagues for the formation of various 3D nanostructures [27].

Even after those kinds of revolutions in the DNA Origami technology, there were still some limitations in the field, since the total number of folded objects was quite low because of the usage of low M13 concentrations (1nM). However, this limitation was overcome by Dietz et coll., thanks to his method for biotechnological mass production of DNA origamis. The principle of the method is to build the structure on generated single-stranded oligonucleotides via excision from circular ssDNA with type IIS restriction endonucleases [28]. However, instead of using auxiliary proteins for the isolation of single oligonucleotides from a precursor ssDNA, they defined the staple list during the design process and used it to generate a pseudogene in which the target staple sequences are concatenated and interleaved with self-excising DNAzyme cassettes. First, self-cutting DNA cassettes generate the desired target strands. Afterwards, the pseudogene is assembled and cloned into phagemid backbones. ssDNA production is performed in *Escherichia coli* cultures using either a helper-

plasmid or a helper-phage approach. In both approaches, the expression of phage gene stimulates the intracellular rolling circle amplification of the phagemid DNA in a single-stranded form and then, the packaging into extracellular phagemid particles. By this way, both ssDNA and staple strands can be isolated from phagemid particles [29].

Another limitation of DNA origami structure was the size of the object that is guided by the size of scaffold M13, which is why the largest size of the object was around 100 nm. Qian and Dietz reported methods for the assembly of large DNA origami structures. The study of Qian et coll. was the fractal assembly of multiple DNA origami platforms in a pre-arranged manner to form a large 2D squared DNA origami assembly of 1 μm [30]. The strategy of Dietz was the stacking of multiple 3D origami nanostructures for the formation of a larger composite around 300 nm [31].

The emergence of DNA nanotechnologies led to a new promising method for the formation of complex nanostructures of highly controlled shape and size through the perfect specificity of DNA base-pairing interactions. Single Stranded Tile and DNA origami techniques represented revolutions in the field since they allowed the formation of always more complex and elaborate 2D and 3D DNA nanostructures. Both techniques are similar on some aspects, but the SST technology has a limitation as a low yield production of objects, partially due to the stoichiometric proportions between the constituting DNA strands. On the other hand, DNA origami depends on the folding of a circular single-stranded template strand which, in the presence of a large excess of small DNA strands (staple strands), folds quantitatively into the final nanostructure.

The recent developments tend to bring this field of nanotechnology toward the microtechnology, with increasing complexity but still high control of the bottom-up process.

1.2.1.3. Designing and formation of DNA nanostructure

<u>Design</u>

The design phase is the first and most important step for the formation of DNA nanostructure and the principle relies on the Watson-Crick base pairs. That is the reason why several software packages have been developed to minimize the error rate of the design. The software to design DNA nanostructures can be categorized in two main categories: manual and automated tools. CaDNAno is a well-known computer-aided design (CAD) tool for the manual design of DNA nanostructures using the bottom-up approach. This software allows users to determine the optimal crossover positions, but it is limited by the expertise of the designer. This limitation makes it difficult to carry out and compare a range of possible designs. To overcome this challenge, the scientist developed automated tools which accelerate the design process and expand the design space. DAEDALUS, TALOS, PERDIX are some of the popular automated tools. However, these automated tools also are limited to top-down design since they require a designer to fully conceptualize the structure. Therefore, while CaDNAno and its newer versions remain more suitable for bottom-up designs despite their limitations, the automated tools likely seem more convenient for the top-down designs [32].

1.2.1.3.1. Thermal and isothermal Assembly techniques for DNA nanostructures

There are basically two techniques for the formation of DNA nanostructures at nanometric scales. Even though both allow us to build nanostructures which are made by DNA and improve the complexity of the objects in 2D and 3D, they differ from each other in the experimental procedure. These are categorized as thermal and isothermal assembly. The main difference between them as a principle is that while in thermal assembly, annealing temperature is carried out for the formation, only a fixed temperature is used in isothermal reactions. Additionally, their approach of avoiding mismatches among the DNAs differs in thermal and isothermal assembly.

<u>Thermal assembly DNA tiles and DNA origami</u>

Both techniques have several similarities, they allow one to build highly controlled nano objects out of DNA, and one can thus obtain a large variety of possible 2D or 3D shapes in a very simple manner, while the experimental procedures are different for the two technologies.

<u>Thermal assembly of DNA tiles</u>

The first methodology to create DNA nanostructure through the single strand tiles was reported by Yin in 2012. This method was based on the self-assembly of many single strand DNA strands into desired shape. The detail of this protocol is presented below as:

Formation buffer	TE 0,5x buffer with 12,5 mM $MgCl_2$
Concentration of each ssDNA tile	100 nM

Table 1.1. Thermal assembly conditions through DNA Tiles.

The mixture of ssDNA in the formation buffer is thermally annealed through two steps. A first annealing goes from 90°C to 61°C (-0.2 C/min), the second annealing goes more slowly from 60°C to 25°C and total annealing time was between 5.4h and 72.5h. During thermal annealing it is crucial to keep the staples apart at the beginning of the process to prevent the mismatches between them. The formation time of DNA nano objects is long, and the efficiency of the method is around 20% for DNA Tiles. The reason for the low efficiency is that ssDNA molecules are required in stoichiometric proportions in DNA tiles. Since the usage of small oligonucleotides are limited in DNA tiles techniques, the possibility of hybridization between DNAs is dramatically reduced. Therefore, the folding time is increased, and the yield of the reaction is decreased [20].

<u>Thermal assembly DNA origami</u>

The first protocol of thermal DNA origami folding as demonstrated by Paul Rothemund and according to his method, the desired shape can be obtained through folding long single strand DNA with short oligonucleotides, named as staples under the conditions listed in Table 1.2.

Formation Buffer	TAE 1 X buffer with 12,5 mM $MgCl_2$
Template Strand (Scaffold)	M13mp18
Ratio between template and staples	From 1:40 to 1:100

Table 1.2. Thermal assembly conditions through DNA Origami.

The mixture of M13 template and staple strands is heated up to 95°C and kept at 95°C for 10 min and gradually cooled down to 20°C at the rate of -1°C /min [33]. This protocol was shown to be very robust, with the yields of origami formation at around 70% and even at 90% for the rectangle origami. Since M13 is the limiting reactant of the reaction and no unbound M13 is left in solution at the end of the annealing process. The system can furthermore no longer evolve at room temperature. After the thermal annealing, purification of the objects is required and this can be performed by agarose gel electrophoresis [34], size exclusion chromatography [35], ultra-centrifugation [36] or PEG precipitation [37]. Among these methods, agarose gel electrophoresis is the most usual one, but it can have low recovery yield because of the contamination with the residual agarose. However, through the optimization of running buffer and pH etc., this method keeps its popularity since it is easy to perform and provides bulk estimation for the nano objects. Chromatography methods have great potential for the purification of DNA nano objects. While other chromatography methods, such as ion exchange and affinity purification, often require high ionic strength buffer to elute the molecules of interest, size exclusion chromatography can be performed in a variety of biologically compatible buffers. This feature of size exclusion columns is an advantage compared to the other purification methods because it allows easy optimization of parameter such as Mg^{+2} and Na^+ ions concentration and a high yield recovery of DNA nano

objects. Ultracentrifugation and PEG-precipitation methods that involve pelleting or concentration of samples have a higher tendency to aggregate the samples during the purification process [35].

The concentration of magnesium, annealing temperature and ratio between scaffold and staples all have an important role in the formation of DNA nano objects. The presence of magnesium is known to be crucial as it reduces the electrostatic repulsion between DNA strands. Thus, optimization of magnesium concentration plays an important role in the folding of nano objects [38]. The role of magnesium on the interaction between strands of DNA was known as ion bridging [39]. However, in case of DNA origami formation, the situation is different, and it is demonstrated that magnesium concentration affect the Debye screening. Low magnesium concentration causes the longer Debye length and longer Debye length means a larger inter helical distance [40]. The presence of multivalent ions such as Mg^{+2} affects electrostatic interaction potential more than monovalent ions such as Na^+. The reason is that in the lack of ionic strength, the efficiency of charge screening of DNA phosphate groups decreases, and this causes the induction of electrostatic repulsion between DNA nano objects.

Another factor is the thermal annealing temperature; it is significant that the temperature should be higher than the melting temperature of the mixture since the thermal annealing prevents the mismatch between template and staple strands. The other factor is the ratio between template and staple. High excess of staples increases the possibility of encounter between template and staple strands [25]. This ratio might differ between 1:1 to 1:100 according to the protocol and the specific aim of the experimental procedure.

The two main differences between the DNA origami and DNA tile techniques formation procedures is that the DNA tile method requires longer incubation time (over 5 hours for DNA tiles and 1 hour for DNA Origami) and the presence of all strands in stoichiometric proportions (whereas the origami staples are used in overwhelming excess compared to the M13 template). Because of the stoichiometric proportions needed for the formation of SST, the thermal annealing process has a low yield. Since the system almost does not evolve any more at 25 °C, and therefore often largely constituted of partially formed SST [20,21,41,42]. Folding times are significantly higher

for SST than for DNA origami and yields significantly lower. The difference in formation time might have originated from the usage of the staples excess because this creates a kinetic advantage in DNA origami compared to the stoichiometric conditions for SST.

Isothermal assembly of DNA origami and SST

The formation of DNA nanostructures has become a process which is controllable and promising. However, even though the formation protocol of nano objects is precise, it still has some weak points like in every recent technology. Especially since this process can create complex nano objects which require a very precise approach, the probability of having defects should be as low as possible. Since this process requires thermal annealing, it is convenient to discuss the thermodynamic effects on the stability of the nano objects.

The folding process includes gradually decreasing temperature and means if the temperature change is faster than the time needed to reach the equilibrium state of DNAs at each temperature, there would be lots of mismatches between the oligonucleotides and it causes creation of unstable complex nano objects.

To overcome these limitations, the larger objects need a longer incubation time to reach thermodynamic equilibrium. Therefore, the need of isothermal folding came out at this point to allow the system to reach equilibrium without limitation of time at constant temperature. These systems provide a thermodynamic base for folding rather than a kinetic one. The isothermal folding not only has many contributions (as explained above) for the formation of complex structures but also it offers a favorable path for a wide range of applications including proteins and enzymes [43] since they are thermosensitive biomolecules, and they might lose their properties and functions under high temperatures. In the isothermal process, not only the certain temperature but also the assembly time needs to be optimized since they are the base factors which control the system thermodynamically. Since there were several logical reasons for the need of isothermal folding, some researchers have developed several methods for the formation of DNA nanostructures at constant temperatures. These methods can be categorized as two groups as isothermal assembly with and without denaturing agents.

Isothermal protocol without denaturing agents

The history of isothermal folding has been started by Simmel et coll. in a negative way since the TAEMg buffer and room temperature were not the suitable conditions for this process even with a long incubation time [44].

Dietz et coll. thus investigated the folding and melting processes for a variety of DNA nanostructures by fluorescence spectroscopy to determine the optimal temperature for the isothermal process to occur (see figure 1.9.A). They found out that the formation of a rectangular structure was possible even in a TAMg buffer; however, it required the pre-denaturation step at 65°C (see Figure 1.9.A). The optimal temperature was 53 °C and below this temperature, the folding rate decreases. (See figure.9. A 1). The folding process was completed after 5 minutes (figure1.9.2) and resulted in flawless rectangular DNA origami (figure1.9 3) [45]. However, even though isothermal folding is partially possible with those research, pre-denaturation step remained problematic for the design, involving thermosensitive biomolecules. Thus, the requirement of pre-denaturation step was removed by the report of the Fritzsche. They showed that the assembly of DNA nanostructure is feasible in the temperature range between 43°C-57°C without denaturing step and 54°C was the optimal temperature for the highest yield. Outside of this temperature range, it was not possible to observe well folded structures (Figure 1.9.C). Since those temperature trials were performed maximum for 60 minutes, it was not reported how the yield changes by also optimizing the time versus the temperature [44].

To understand comprehensively the folding kinetics of DNA nanostructures under isothermal conditions, Dong, Besenbacher and his colleagues came out with a different method where they used partially folded (obtained by thermal annealing process but incomplete mix of the staples) rectangle DNA origami instead of starting with the mixture of scaffold and staples. They immobilized the partially folded DNA origami on mica surface and gradually added missing staples at 60°C and 25°C separately. The measures were continuously taken by scanning the sample through AFM. The pore-filling at 60°C for an adsorbed origami was resulting in well-folded objects in less than 24 minutes, as shown in Figure 1.9.B., however, the same process at 25°C showed an incomplete pore-filling, even after 120 minutes. Thus, these results

pointed out that the pre-folded origami has an entropic advantage since the template was already half organized [46].
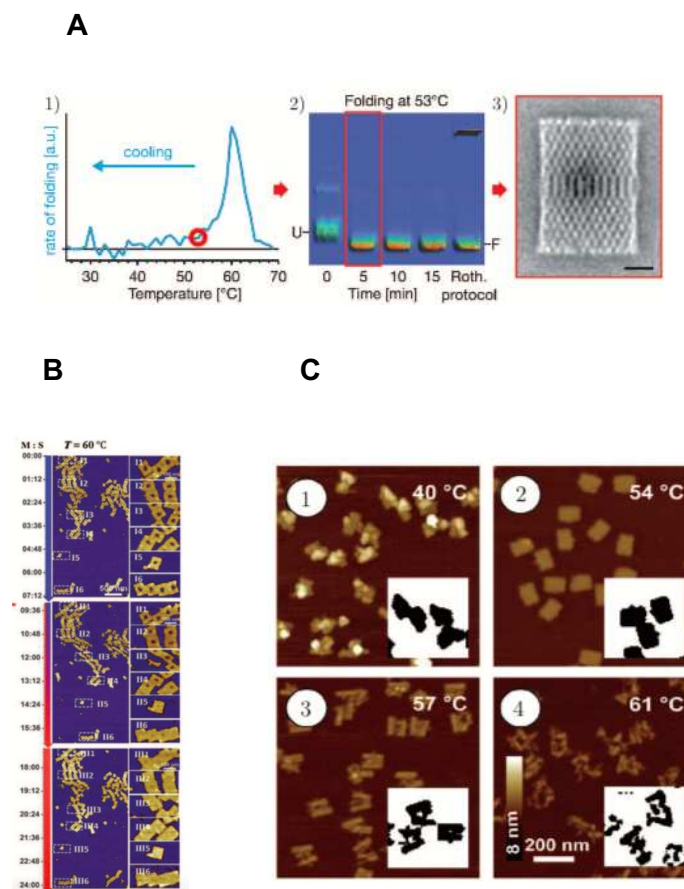
Figure 1.9. *Isothermal formation of DNA nanostructures without chemical modification.* ***A*** *Isothermal formation of the tall-rectangle origami. 1) Folding rate. Determination of the optimal temperature for isothermal folding (53°C). 2) Isothermal folding's kinetics at T = 53°C."U" represents the unfolded,"F" represents the folded DNA origami. 3) TEM image of the isothermally folded tall rectangle at 53°C. Scale bar: 20 nm [45].* ***B*** *Isothermal pore-filling at T = 60°C. Red arrow = addition of staples [46].* ***C*** *Isothermal formation of the tall-rectangle origami at different temperatures [47].*

Isothermal assembly for DNA tiles

Additionally, isothermal assembly can be profitable not only for the DNA Origami, but also for the DNA tiles method. Yin and his colleagues published a study showing that the isothermal assembly can be applicable without any denaturing step. The key point of this study is that assembly of the object is controlled by GC content and the length of the binding domain of the bricks.

It was demonstrated that length of the domain and the needed temperature of the system are related to each other. While longer binding domains of the bricks require higher temperatures, shorter ones need lower temperature for the formation [48]. In the other study of Yin, the authors reported that isothermal assembly of 3-D crystal bricks was achieved through isothermal assembly. They found out the optimal temperature 33°C for this process without any need of denaturation step [42].

As a brief conclusion of those studies above, there are some methods which have been developed with denaturation steps and on the other hand a few ones not including the denaturation step, with an average temperature around 53°C for the isothermal assembly through DNA origami technique. However, 53°C still can be counted as high for the thermosensitive biomolecules. Therefore, it was crucial to develop new methods such as isothermal assembly to overcome the challenge of the previous methods and to increase range usage of DNA nano objects. Thus, some group of researchers demonstrated the protocol of isothermal assembly through the chemical reagents which can decrease the melting temperature of DNA nanostructures.

<u>Isothermal protocol with denaturing agent</u>

From the previous section, the denaturing step remained as a challenge for the isotherm assembly process. Even though some studies showed that isothermal assembly is possible at certain temperatures around 50°C - 55°C, this challenge led the scientist to work on decreasing the melting temperature through denaturing agents [49]. These agents can disrupt the hydrogen bonds between DNA strands, thus the stability of DNA is reduced. But at the same time, there would be a trap for DNA if the reagent reacts with a DNA base with an interaction between two bases. In this case, denaturing reagents might work in the opposite direction from the main purpose by building multiple hydrogen bonds. Simmel and his colleagues reported isothermal assembly protocol through the formamide and urea [44] which can reduce the melting temperature of DNA by 0,60-0,72°C per 1% of formamide [50]. In their protocol, templates, and staples in a 1X TAEMg with 85% of the agent were incubated at room temperature. Afterwards, to lower the concentration of agents, dialysis was carried out to lower the formamide progressively over 24h as presented in Figure 1.10.A 1).

The purified samples were measured by AFM and resulted in flawless structures (see Figure 1.10. A 2 and 3 respectively). Even though the formation of DNA nanostructure was possible with the aid of denaturing agents, dialysis remained as the limitation of the process. Then, the modification of protocol was done by Dong, Gothelf, and his group and thus, they incubated the template and staples at 25°C with 30 to 40% concentration of formamide instead of dramatically changing the concentration of the agent [51].
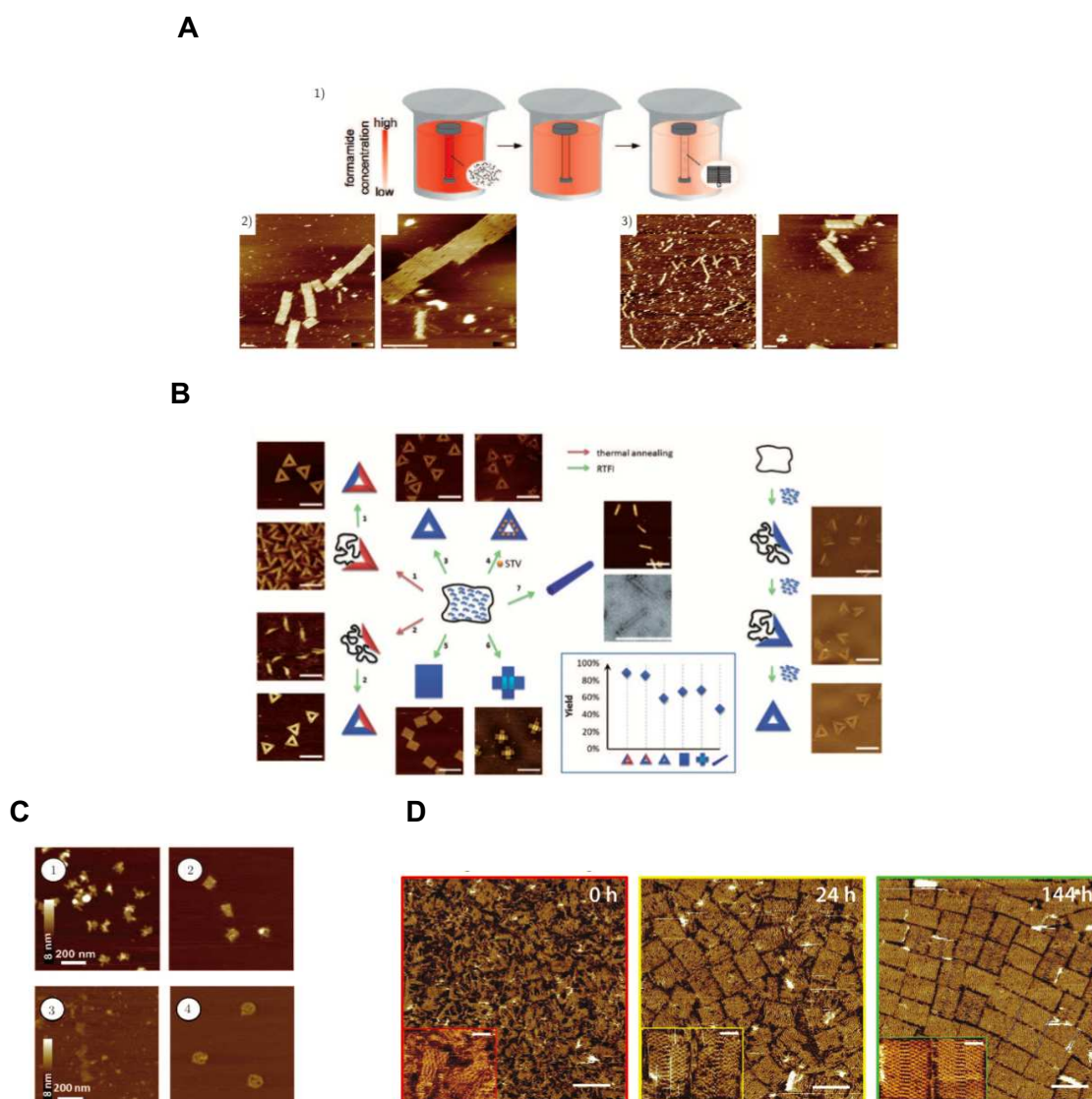
Figure 1.10. Isothermal formation at room temperature using denaturing agents. A Isothermal formation of the tall-rectangle origami using a formamide and urea at room-temperature [44]. B Isothermal formation of several DNA nanostructures through a room-temperature formamide incubation process [51]. C Isothermal formation at 33°C of rectangles (1 and 2) and smiley origamis (3 and 4) through incubation with betaine [47]. D Isothermal folding of the tall-rectangle origami in dehydrated glycholine as solvent at 20°C [49].

At the end of the incubation, 2D and 3D DNA origami were successfully folded (see Figure 1.10.B). Another advantage in using denaturing agents is that they prevent aggregation of DNA structures in the bulk. Thus, they also help the isothermal folding process through increasing the possibility of having non-aggregated DNA nanostructures in the solution. Thus the usage of denaturing agents aids the process to avoid non-specific hydrogen bonds between template and staples strands and that is why isothermal folding can be more suitable for the complex nano object.

Additionally, isothermal assembly of DNA nanostructures including DNA origami was carried out with betaine at 37°C by Fritzsche and betaine is known as an effective reagent to disrupt the binding between base pairs. The experimental protocol was the incubation of template and staples in the presence of 1M betaine for 24 hours to form a rectangle origami and smiley origami. Sufficiently folded structures were obtained at 37°C (see Figure 1.10.C.). But the same efficiency was not observed at room temperature. The advantage of this process was that several biomolecules like some enzymes and proteins can still protect their function. Therefore, this method was promising for several applications [47]. Moreover, Hud and his colleagues developed a method allows to fold DNA origami at room temperature in the presence of 1:4 mixture of glycerol and choline chloride named glycholine in 6 days, after a thermal denaturation step at 70°C [49]. The structures are perfectly folded and folding of the structures were improved with the longer incubation (see Figure 1.10.D). This method is convenient except for deep-eutectic solvent which are known to be very unstable with time and especially to dramatically change their properties through water absorption from the air humidity. Also, non-standard conditions limit the use of this method for biological applications.

Methods	Advantages	Disadvantage	Best Practices
Gel electrophoresis	Provide bulk estimation about the structure Allows extraction of desired sample	Cannot provide deep structural inside at the single molecule level	Gel should run at 0 -4 C to prevent sample degradation Running buffer should not contain more 100 mM monovalent ions
Fluorescence Spectroscopy	Highly sensitive Provide distance dependent information when applied to FRET Can be used to monitor dynamic events in real time	Requires fluorescence modified standards Structural information is related to the local dye's environment	Fluorescently labeled oligomers should be checked for purity Accurately prepare control donor-only samples for reliable quantification of the FRET effect
AFM	High fidelity Lateral resolution up to 1-2 nm Imaging can be done in fluid or in air	Unsuitable for 3D or multilayer DNA structures Time-consuming	Purified samples give best result Nickel acetate can be used to fix the DNA sample on the mica surface
Transmission Electron Microscopy	Best for 3 D structures Highest Resolution Suitable for metal nanoparticle modified structures	Negative staining is cumbersome and time consuming Difficult to obtain sample height information Structural deformation due to the drying	TEM grid should be discharged to enhance hydrophilicity Before staining, add NaOH solution to adjust the pH of the uranyl solution
Single Molecule Force Measurement	Detects molecular forces with piconewton resolution	Highly specialized technique	A dual optical tweezer is the preferred instrumental

Table 1.3. *Quality control of DNA origami with different techniques adapted from [25].*

There are plenty of methods to confirm the structure of DNA nano objects after performing the assembly reaction. As a general aspect, one method which gives a bulk estimation is necessary to predict the best conditions depending on the purpose. Then, another method is essential to visualize exact geometry of the structure and efficiency of the assembly. According to the objective, selection of visualization

methods is as critical as selection of formation and design. Every method has its limitation for different structures such as 2 D and 3 D. Also, while some of the methods can show the deep internal structure, some of them cannot. These methods were described in table 1.3 including their advantages and disadvantages.

1.2.2. Applications of DNA nanotechnologies
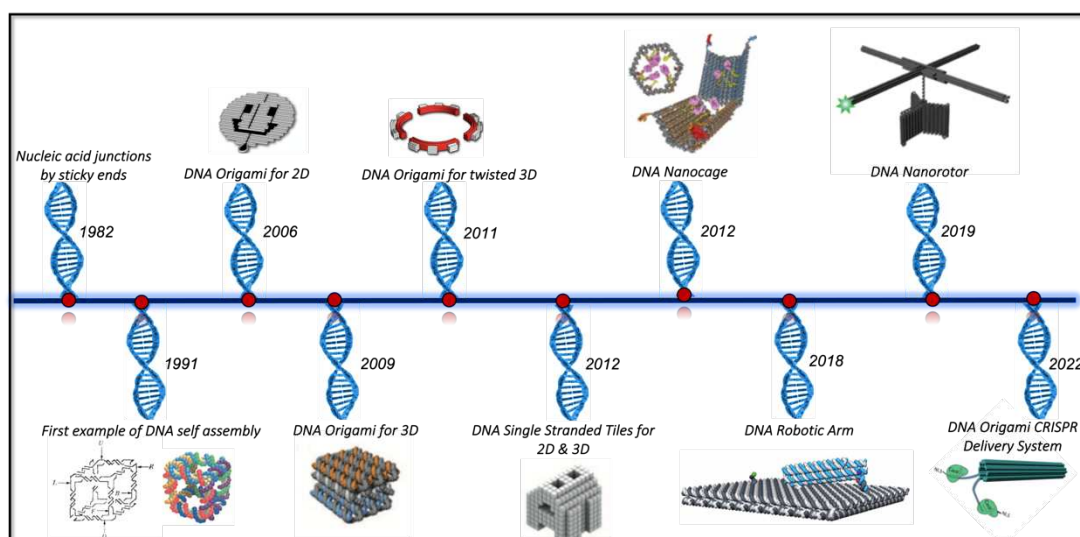
Figure 1.11. Evolution of DNA Nanostructures.

The main reason which makes DNA technology quite interesting is that the design and production are not complicated, and the usage range is incredibly wide for site-specific functionalization, biophysical and biomedical purposes. The main steps that allowed the evolution of DNA nanostructures were shown in figure 11. In 1982 the DNA started to be used as a material by the guide of a publication called as nucleic acid junction by sticky ends. Then, the first example of DNA self-assembly was reported in 1991 and 15 years later the first comprehensive technique, DNA origami, was developed for formation of DNA nanostructure by Paul Rothemund. Three years later, DNA origami technique was improved for designing 3D structures and, in 2011 this technique was advanced for twisted 3D nanostructures. A year later, the DNA single stranded tiles technique was developed to construct DNA nanostructures by Yin. The demonstration of these techniques opened the path to many useful designs.

1.2.2.1. The utilization of DNA nanotechnology for a site-specific functionalization

Thanks to the Watson-Crick base pairing and the high chemical and structural stability of the DNA, we can control the design of staples and thus the shape of structure according to the sequence. In general, site specific functionalization of proteins has been done through attaching a protein of interest to the one of the DNA strands of the nanostructures via chemical modifications.
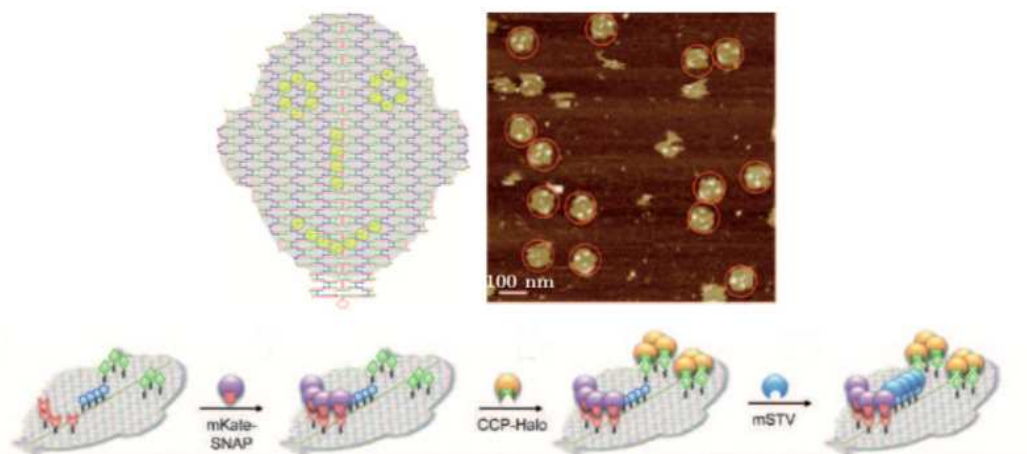
Figure 1.12. *Positioning of proteins on DNA nanostructures. **A** Examples of monofunctionalized DNA–origami conjugates are shown for CH-, biotin-, and BG-bearing origami with mKate-Snap (indicated by violet spheres), CCP-Halo (yellow spheres), and mSTV (blue spheres) Adapted from [52].*

In the study of Christof M. Niemeyer et coll. (illustrated in figure 1.12) the authors modified selected strands of DNA origami by attaching biotin, BG- and CH-. The BG-, CH-, and biotin-tagged origami was initially labeled with mKate-Snap, thus giving rise to the mouth feature. This origami was then allowed to bind CCP-Halo, which resulted in an appearance of the eye features. Finally, treatment with mSTV led to formation of the fully decorated smiley face structure (figure 1.12.A bottom). The smiley face origami sample was observed through AFM, since the coupling between the chemically modified DNA strands and mSTV, mKate- Snap, CCP-Halo, and EYFP-Halo occurred in a highly specific manner and the smiley face origami with nose, mouth and eyes was flawless (figure1.12. A up). In this study, protein-DNA coupling, and DNA origami techniques are mainly used [52]. This report is interesting because; the sequential functionalization with the various proteins proved an orthogonal specificity

of the binding process, thus paving the way for synthesize and use of more complex patterns of proteins.

1.2.2.2. Use of DNA nanostructure for biophysical and biomedical applications

There are numerous studies of use of DNA nanostructures for biomedical and biophysical applications. For instance, the application of DNA nanostructures is popular in drug delivery field, while for biophysical applications DNA nanostructures are used to organize and reposition the proteins and enzymes in complex systems on DNA scaffolds.
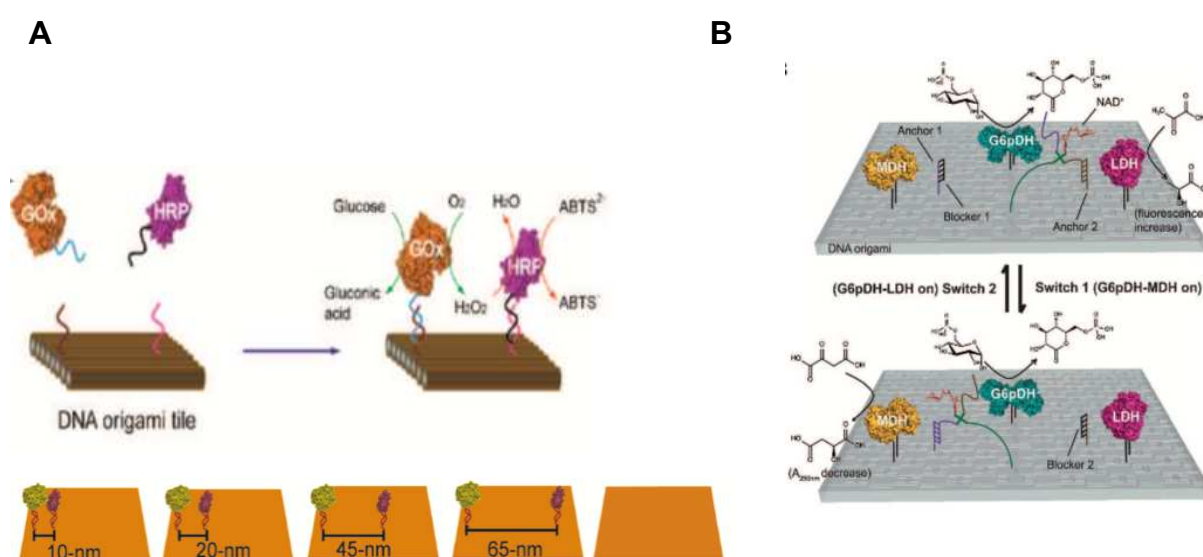
Figure 1.13. *DNA nanostructure-directed assembly of GOx and HRP enzymes with control over inter enzyme distances. **A** The assembly strategy and details of the GOx/HRP enzyme cascade (top). Rectangular DNA origami tiles with assembled Gox/HRP pairs spacing from 10 to 65 nm (bottom). **B** The nanostructured complex consisting of G6pDH, MDH, LDH, and NAD+-anchored HJ on a rectangular DNA origami platform. In the presence of switch 1, the NAD+ substrate channeling binds to anchor 1 and is located between G6pDH and MDH, thus activating the G6pDH–MDH pathway. In the presence of switch 2, the substrate channeling is located between G6pDH and LDH, activating the G6pDH–LDH pathway.*

In the studies of enzymatic cascades, it is quite challenging to fully understand their specific mechanisms. In this area, the DNA origami technology offers a certain advantage: it allows to position enzymes of various types at well-defined positions on a 100 nm scaffold, to bring mechanically two enzymes at a pre-defined distance. As

an example, Yann and his colleagues studied the impact of the distance between two enzymes involved in an enzymatic cascade. To do this, they coupled protein with a ssDNA and attached the complementary of the strand, which is attached to the protein, to rectangle DNA nano object. Then, they placed two enzymes glucose oxidase (GOx) and horseradish peroxidase (HRP) with the controlled inter enzyme spacing on specific positions the rectangle DNA nanostructure (Figure1.13.A). They studied the efficiency of enzyme activity with different distances (from 10 nm to 65 nm) and proved the optimal distance (45 nm) between the enzymes for the maximum efficiency of enzymatic reaction [43]. The proximity between enzymes was here shown to play an important role in their activity, therefore other systems with DNA nanostructures have been developed.

The metabolism of living systems is a complex structure which involves various delicate multi-enzyme systems and the precise and dynamic control of enzymes and substrates [53]. Because of this, the construction of smart artificial multi-enzyme systems that directionally activates different enzyme pathways in response to different environmental conditions (e.g., different stimulants, different substrate levels) is critical for enzyme regulatory circuitry and signal transduction pathways [54]. As another important example in this domain, Yann showed the possibility of regulation two cascade enzymes (G6pDH-MDH and G6pDH-LDH) by being able to control the NAD^+ substrate (Figure 1.13.B). Because the proximity of this substrate induces the enzyme activity for both systems (bottom and up of figure.13. B), the continuous changing of position of NAD^+ allows the location of the enzyme activity on a DNA origami structure [55].

DNA origami technique has also been used for medical applications (see Figure 1.14). It is crucial for the drug delivery systems since the carrier molecule is made by DNA and this increases the possibility of success due to the biocompatibility. Thus, it is more probable that the drug can be accepted by cells during *in vivo* studies. Another advantage of this technique for medical purposes is that a small receptor can be inserted to the surface of the DNA nanostructure with the purpose targeting the structure and facilitating cell entry. In *in vivo* studies, this technique has been used for the cancer treatment as illustrated in Figure 1.14 where a triangular DNA nanostructure is loaded with doxorubicin for cancer treatment [56].
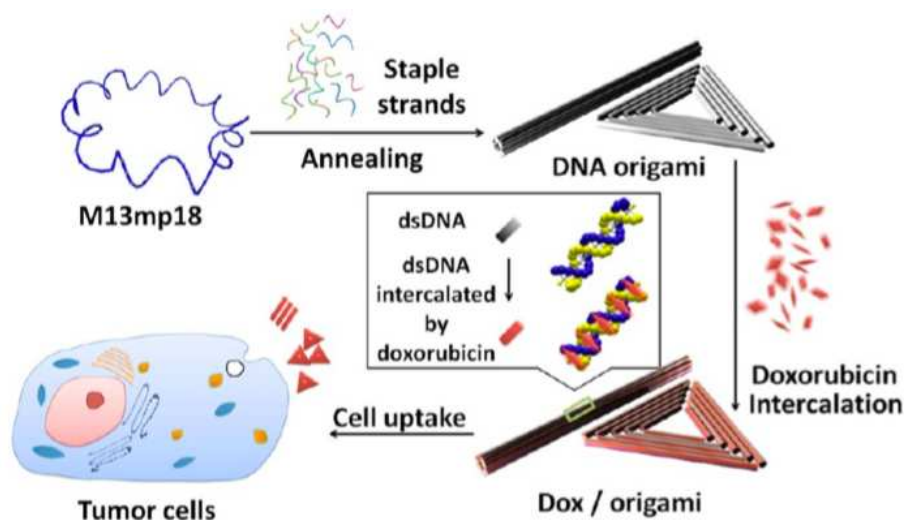
Figure 1.14. *DNA origami and doxorubicin origami delivery system assembly. The long single-strand M13mp18 genomic DNA scaffold strand (blue) is folded into the triangle and tube structures through the hybridization of rationally designed staple strands. Watson−Crick base pairs in the double helices serve as docking sites for doxorubicin intercalation. After incubation with doxorubicin, the drug-loaded DNA nanostructure were delivered to MCF-7 cells* [56]*.*

The DNA nanotechnologies has a wide range of applications as mentioned above. However, since these nano objects are static instead of being dynamic the system needs to be induced to make them dynamic. Therefore, it is significant to design and produce those nano machines in more reversible, dynamic and a controllable way.

Conclusion

The DNA molecule appears to be an ideal building material for supramolecular constructions. The double-stranded DNA itself is a supramolecular object self-assembled through the exquisite specificity of the base-pairing interactions between the two strands composing the DNA double helix.

Over the years, complex architectures including many interacting molecules were developed by chemists. Nadrian Seeman led to improvement of two technologies as the DNA origami and Single-Stranded DNA Tiles. Both methods were feasible for the formation of 2D and 3D DNA nanostructures. Thanks to their constitutive material,

DNA nanostructures were functionalized with several entities and with nanometric resolution.

For the self-assembly process, two assembly protocols were developed so far as thermal and isothermal assembly. Thermal assembly protocol was optimized through numerous studies, whereas there are few studies for isothermal self-assembly of DNA nanostructures at room temperature, at high temperature and using the denaturing agent or not. Particularly, isothermal self-assembly at room temperature was just achieved in the presence of denaturing agents which could decrease the melting and folding temperature of the DNA objects. However, the inconvenience of this process is the need for removal of the denaturing agent after folding.

Since DNA nanostructures are formed through self-assembly of custom-designed DNA strands, it is possible to know for each architecture the precise position of every strand, every sequence and even every base pair. Thus, this programmability feature gave birth to the usage of DNA technology for biomedical and biophysical applications.

Therefore, when DNA nanostructures are functionalized with biomolecules, DNA they enable molecular transportation, sensitive molecular detection or to control amplification of molecular events. Additionally, they pave the way of engineering and mimicking biological complex systems.

Chapter 2 - Protein-DNA coupling methods and challenges

2.1. Principles of chemical modifications of proteins

In biological organisms, post translational modifications occur after the translation of the proteins. These Post Translational Modifications (PTMs) contribute to the variety of protein structure and functionality and post-transcriptionally modified protein molecules are included in many complex biological processes. Inspired by natural modifications of proteins, synthetic modifications have been made both chemically and genetically [57]. The broad range of modification methods allows us to construct and assemble various structures by modifying residues [58].

In early studies, the group of Heinz Fraenkel-Conrat has performed studies to identify amino acid residues which are crucial for the biological activity of the protein [59]. By the improvement of biotechnology, chemical and genetic modification methods have been used both *in vivo* and *in vitro* studies. The tracking of protein-fluorophore probes [60], probing the mechanism of pathological enzymes [61] are some of the studies that have been done *in vitro*. In general, chemical modification of proteins includes post-translational modifications and chemical engineering of proteins. The PTM is based on the chemical reactions with amino acids residues, so it is affected by the chemical reaction types and the reactivity of residues. Chemical engineering of proteins has the potential to introduce non-natural amino acids into peptides and synthetic peptides that scientists design without the limitations imposed by the genetic code. Despite the important progress in the field of modification of proteins, scientists still face many challenges, not only about synthesis but also from processing, manufacturing, and stability [62].

2.2. Chemo diversity of proteins in nature

The post transcriptional modification increases the variety of proteins by covalent modifications, hydrolytic cleavage of peptide bonds and reductive cleavage of disulfide bonds. Half-life and the turnover of proteins are maintained by the hydrolytic cleavage of peptide bonds. Since proteolytic cleavages are irreversible, proteases are firmly regulated. The features coming from the PTMs determine the working principle of proteases. For instance, proteases are not able to reach their substrates until required and zymogen forms of proteases are activated by other proteases when they are needed [63]. Another form of PTMs is reductive cleavage of disulfide bonds and this maturation is maintained in Golgi, ER and post Golgi complex [64]. Most of the disulfide bonds are in function in the extracellular environment of Endoplasmic reticulum, Golgi and endosome. Also, plasma and organelle membrane proteins consist of a high proportion of disulfide bonds [65]. Therefore, PTMs has crucial roles in many aspects as, protein functions, activation of proteases, maturation of proteins.

The many basic types of post translational modifications contain additions to the protein such as; acylation, methylation, phosphorylation, sulfation, famesylation, ubiquitination, glycosylation [66]. While all of these have their own significant roles in cellular processes, phosphorylation and glycosylation are known as the most common modifications.

2.2.1. Phosphorylation

Phosphorylation is described as addition of phosphate group to the serine, tyrosine, threonine in eukaryotes and histidine, aspartate for prokaryotes. It is achieved by various by protein kinases [67]. Addition of a negative charge to the protein (-2) results not only in conformational changes, alteration of enzymatic activity and protein-protein interaction [68,69] but also, creates new docking sites for interactions [70]. Phosphorylation has a major role in cell-cell communication and coordination. For instance, cell cycle progression and cellular growth control are maintained by phosphorylation of serine and threonine residues [69]. Protein phosphorylation is used also for biotechnological applications. As an example, chemical reaction was utilized to add the sodium tripolyphosphate to the soybean protein and resulted in increasing

the particle size distribution of phosphorylated soybean protein concentrate and improvement of the emulsifying property of SPC through phosphorylation [71].

2.2.2. Glycosylation

Protein glycosylation is controlled by over 200 glycosyl transferase enzymes. In this process, these proteins can be glycoproteins, positioned glycans on those proteins or glycans [72]. The proteins can be glycosylated in two different ways as enzymatic and non-enzyme glycosylation. In non-enzymatic glycosylation, lysine and arginine residues of proteins react with glucose and then go through further changes and then become ready to be involved in pathways which are important specifically for aging and diabetes. Enzymatic glycosylation occurs through the enzymes and formation of a glycosidic bond using a sugar nucleotide donor during synthesis of glycoproteins which is an example of enzymatic glycosylation. Glycopeptide bonds are separated into three groups as N-, O - and C- linked glycosylation, glypiation and glycosylation [73]. N- linked glycosylation occurs through attachment of sugar molecules to the nitrogen atom. There is the intrinsic and extrinsic role of N-glycans in maintaining glycoproteostasis. N-glycans have an impact on protein structure as maintaining folding of the proteins [74]. For instance, proteins that are involved in the eukaryotic secretory pathway are commonly modified with N-linked carbohydrates. These bulky amphipathic modifications at asparagine intrinsically enhance solubility and change folding energetics through carbohydrate-protein interactions. N-linked glycans can also extrinsically enhance glycoprotein folding by utilizing the glycoprotein homeostasis or "glycoproteostasis" network [75].

Another important type of glycosylation is O- linked glycosylation which plays an important role in the determination of secondary, tertiary, and quaternary structures of fully folded protein [76].

N- and O- linked glycosylation are known as the most common types in cellular processes [77]. The enzymatic process of glycosylation occurs and mature in endoplasmic reticulum and Golgi complex and mostly secreted proteins are glycosylated [78].

2.3. Engineering of proteins

Besides from natural PTM, proteins can be chemically modified to create a wide range of chemo diversity. Protein engineering is a field that refers to the development of new proteins through modification of amino acid sequences that exist in nature or the invention of new sequences of the amino acids naturally found in proteins [57].

There are basically two ways to manipulate proteins as genetic modification and site-specific modification. The first one modifying the protein features by implementing the modification directly to the genes. Site directed mutagenesis of the gene of interest can be an example. In that regard, modified genes programs that specific feature of the protein according to modification [79]. This was proposed by Schultz and Yarus in 1996 [80]. Even though the consequences of this technique are promising, there is a limitation of probes and the number of cell lines which are usable [81]. The second category, chemical modification of proteins, was started by the discovery of crystallization of protein structure in 1920. By the development of organic chemistry, X-ray crystallography, and molecular biology, understanding of established reagents and procedures was improved. These developments also led to development of many new, and more sophisticated, reagents and procedures [82]. In 1945, the scientists started to study on amino acid residues necessary for the biological activities of proteins. This interest led to the publication of two important reviews of protein modification in 1947 [83].

By the time, methods for the chemical derivatization of proteins continued to expand and develop. Various studies were published about site specific modifications, side chain modifications, modifications by fluorescence label, crosslinking and immobilization. The purposes of these chemical modification of proteins are categorized in three; labeling with probes [84] and observing the interaction between proteins and other biomolecules [85].

Chemical modification of proteins with special probes allows their functional characterization, including dynamics, localization, and protein-protein interactions [86]. As example, Toboya and his colleagues reported a labelled probe, TLSHalo, to Halo-Tag proteins which was expressed by cells on the plasma membrane for the investigation of concentration changes in K+ caused by transfer channel [87]. As an

example of modification of protein functions, the group of Mohsen chemically modified the lysine residues of *Bacillus licheniformis* α-amylase (BLA) by citraconic anhydride and succinic anhydride and they observed that the specificity of enzymes was changed and caused dramatic increasing in maltosidase and a reduction in amylase activity [88]. As an example of observing the interaction between proteins with other biomolecules, chemical crosslinking method was used by Yeo, to obtain DNA-green fluorescent protein complex, DNA-23 mer peptide complex and DNA-10 mer peptide complex. Investigation of these DNA-protein lesions indicated the polymerases cannot read through the larger DNA-protein complexes [89].

Furthermore, better understanding of fundamental biochemical structure and function also contributed to the pharmacological or medical diagnostic applications.

For instance, the modification of protein therapeutics with polyethylene glycol (PEG) improves their stability and half-life in the pharmaceutical industry [90]. As a result of these studies, a wide variety of protein modification methodologies have been studied to obtain an accurate control of these macromolecules [91].

In the protein engineering field by chemical modifications, the main purpose is to engineer the proteins without damaging the structure and folding properties of proteins. That is why, it is critical to consider the critical factors and provide careful control about reaction conditions such as pH, solvents, and temperature. By following this approach, it is also possible to overcome the limitations of chemical modification of proteins as degree of site-selectivity and regioselectivity. So far, various techniques, that allows engineering of proteins without damaging the structure and function, are demonstrated for the chemical modifications of proteins.

2.3.1. Modification of proteins through reactive group and modifying reagents

Chemical modification of protein methods relies on the roles of amino acid residues in the manner of biological, chemical, and functional properties of proteins [92].

Four basic chemical modifications, which are commonly used, are listed below as:

- Primary Amines (-NH2): This group is located at the N-terminus of each polypeptide chain and at the side chain of Lysine amino acid(K).
- Sulfhydryl (-SH): It is found at the side chain of cysteine amino acids. Cysteines forms disulfide bridge (-S-S-) in the protein structures.
- Carboxyl (-COOH): It is located at C terminus of each polypeptide chain and side chain of Aspartic and (D) and Glutamic acid (E).
- Phenol (-Ph-OH): This group is found at the side chain of Tyrosine (Y).

-NH2 groups are located at the N terminus of each polypeptide change and at the side chain of lysine amino acids. The chemical reaction with an amine group is required to perform in the alkaline conditions since the amino acids are supposed to be deprotonated [93]. Since there are several reactions which can be done with the primary amines group, usage of lysine remains popular for chemical modifications. Also, lysine is an amino acid that is found on the surface of proteins. This feature makes lysine accessible and facilitates the chemical modification reactions [94].

N-hydroxy succinimide ester (NHS) reaction
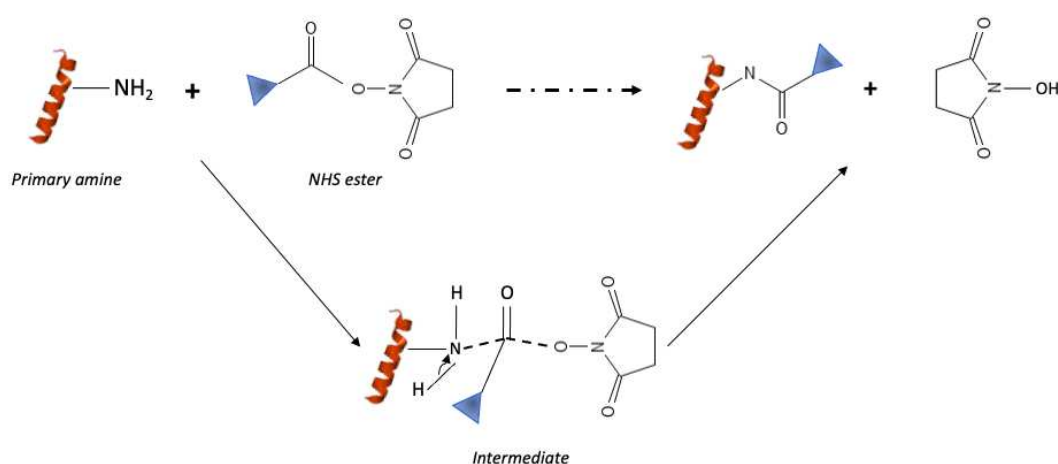
Figure 2.1. The mechanism of chemical reaction of NHS ester with primary amine.

The primary amines are mostly able to be protonated in physiological conditions, but they can still function as nucleophiles in the neutral state. The N-hydroxy succinimide groups are commonly used for the modification of amines due to the proportion of amines, and they are transiently deprotonated even at pH values below the pK_a. The electron distribution, which occurs in C=O bound, of the NHS ester molecule to get closer to O atom due to its electronegativity and the carbon atom has a partial positive charge. The reaction occurs through the attack of the N atom of the NHS to the C atom to form an intermediate. Thus, the C=O bound breaks and the primary amine loses H+(see figure 2.1).

The preparation of stable NHS esters should be in non-aqueous solutions such as DMSO. NHS ester cross linking reagents react with primary and secondary amines and form amide and imide bonds. Therefore, they can couple with the α-amines at the polypeptide N-terminals and the ε-amines of lysine side chains. Additionally, NHS esters can react with the tyrosine, serine, and threonine –OH groups [95].

NHS reagents react with nucleophiles and at the end of the reaction, they release the NHS, to form an acylated product. The reaction between NHS ester and primary amines is pH dependent. When the pH is quite low, the amino group is protonated. However, when the pH is high, NHS esters can be hydrolyzed. Therefore, the optimal pH is around 8.3-8.5. This reaction to work efficiently, requires 2 hours incubation at 25°C. To maximize the coupling primary amines with NHS ester groups and minimize the rate of hydrolysis, high concentration of protein usage and molar ratio between protein and NHS ester groups are crucial factors. The control of over sites of protonation of peptides is maintained through the difference between the pK_a values of N terminus and amino group of lysines [96].
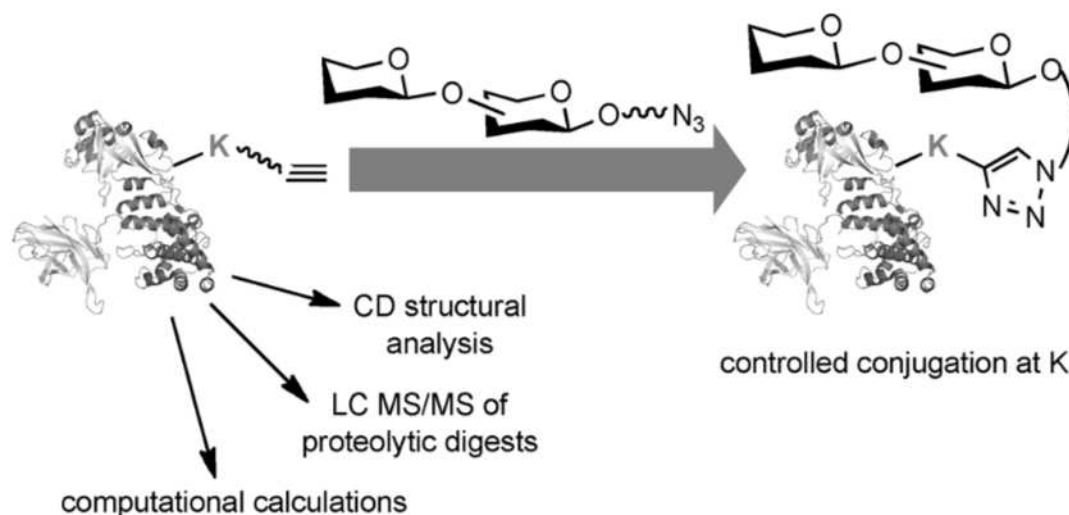
Figure 2.2. Two-step conjugation enabled identification of the more exposed lysines of CRM197.

As an example, the reactivity of the lysines were characterized from nontoxic cross-reactive material of diphtheria toxin (CRM197) through NHS ester linkers bearing alkyne or azide by Stefano and his colleagues in 2014 (see figure2.2). This characterized method enabled the preparation of lysine-mediated glycoconjugates with developed batch-to-batch reproducibility, thereby producing neo-glycoconjugates with more-consistent biological activity. This study demonstrated and mapped the reactivity of lysine residues via LC-MS/MS analysis [97].
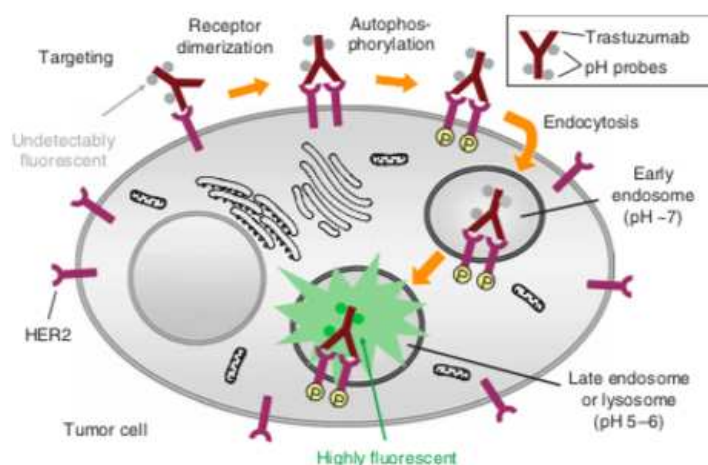
Figure 2.3. A schematic representation of highly selective tumor imaging with an activatable fluorescence probe-antibody conjugate. The probe is nonfluorescent when outside the tumor cells. After internalization by endocytosis, the probe is accumulated in late endosomes or lysosomes, where the acidic pH activates the probe, making it highly fluorescent.

Particularly, conjugation of proteins with fluorescently labeled NHS esters can be a useful tool to construct *in vivo* probes [98]. As an example, NHS ester was used to label the cancer cells with fluorescent probes in the study of Yasuteru and his colleagues for developing the specific and sensitive tumor-imaging techniques (see figure 2.3). They conjugated this fluorescent probe to a cancer-targeting monoclonal antibody. This conjugation was conducted through fluorescence probes to trastuzumab (monoclonal antibody) by using their mono NHS ester derivatives to form an amide bond with lysine residues of the antibody. The feature of this fluorescent probe is being acidic pH- activatable fluorescent probe. It is activated after cellular internalization by sensing the pH change in the lysosome [99].

In addition, due to the selectivity and specificity of the method, it can be a useful technique to manipulate complex proteins in biological systems such as oligomeric proteins in cell membranes. Therefore, I can conclude that the main advantage of this technique is being able to be selective and specific for several purposes in numerous fields and applications.

Other reactive groups that react with primary amines

The other groups that can react with primary amines are isothiocyanates, aldehydes, and imidoester. Isothiocyanates can react with the primary amines but just in aqueous solution and the disadvantage of this molecule is that it can be hydrolyzed to ammonia, CO_2 and H_2S. The potential of being hydrolyzed leads to a risk of deterioration during the storage [100].

Aldehydes can react with the amines to compose Schiff base intermediates. The reaction between aldehydes and amines group is pH dependent and it forms Schiff base at low pH [101]. The limitation of aldehydes is that the indecisive form of Schiff.

It can be stabilized through reduction with the addition of sodium borohydride. However, it requires an additional step [96].

The imidoester is a functional group that is involved in the group of acylating, reacting specifically with primary amines. Like the other groups, there is a certain pH 7-10. The reaction between imidoester and primary amines results in amidine which is positively charged in physiological conditions due to it being protonated [102]. However, the

stability of the bond relies on the pH; while at low pH the amidine bond is stable, at high pH it becomes sensitive for hydrolysis [103].

Besides the functional groups which are mentioned above, there are sulfonyl chloride, acyl azide, anhydride, epoxide, fluorophenyl ester. Sulfonyl chloride reacts with primary amines. However, since it includes a tetrahedral configuration of substituents, there is possibility that formation of an unstable pentavalent intermediate in case of the reaction between nucleophile and sulfonyl chloride [104].

Acyl azides are known as activated carboxylate groups and can react with primary amines. In the presence of hydrazine with sodium nitrate, the reactive acyl azide forms. Acid anhydrides composed of dehydration of two carboxylic acid groups. Even though they can react with cysteine, tyrosine, and histidine, side chains form unstable complexes. However, amine functional groups of proteins are stable to acylation with anhydride reagents to form the amide bonds [105].

Epoxide groups couples with nucleophile molecules in a ring-opening process. They react with primary amines, sulfhydryl, or hydroxyl. The pH of the reaction depends on the groups that epoxide reacts with. However, for the coupling of amines with epoxide, optimal pH is about pH=9 [96].

Fluorophenyl ester is another functional group which couples with amines. There are several types of fluorophenyl esters such as a pentafluoro phenyl (PFP) ester, a tetra fluorophenyl (TFP) ester, and a sulfo-tetrafluoro- phenyl (STP) ester. They all react with amines, but their electronegativity changes in the aqueous solutions as uncharged ones are hydrophobic and the sulfonated one is negatively charged. The reaction occurs at alkaline conditions around pH=7-9 and forms the amide bond like NHS ester. Additionally, they are more resistant to hydrolysis compared to the NHS ester molecules [96].

Chemical modification of sulfhydryl (–SH) groups

Cysteine labeling is known as more specific compared to the amine labeling. However, generally cysteines are not located on the surface of protein molecules and it is a versatile amino acid for selective chemical modification of proteins. This selective chemical modification is conducted through site-directed mutagenesis [106]. Site-

directed mutagenesis methods are used for generating DNA sequences with mutated codons, insertions, or deletions. These mutations are generated by PCR using a pair of oligonucleotide primers designed with mismatching nucleotides at the center of the primers. Thus, this method allows to produce site selectively modified protein through the DNA manipulating [107]. By this approach, it is feasible to mutate the amino acid residues to other residues on the surface of the protein. This allows the usage of accessibility advantages of the amino acid residues on the protein surface.
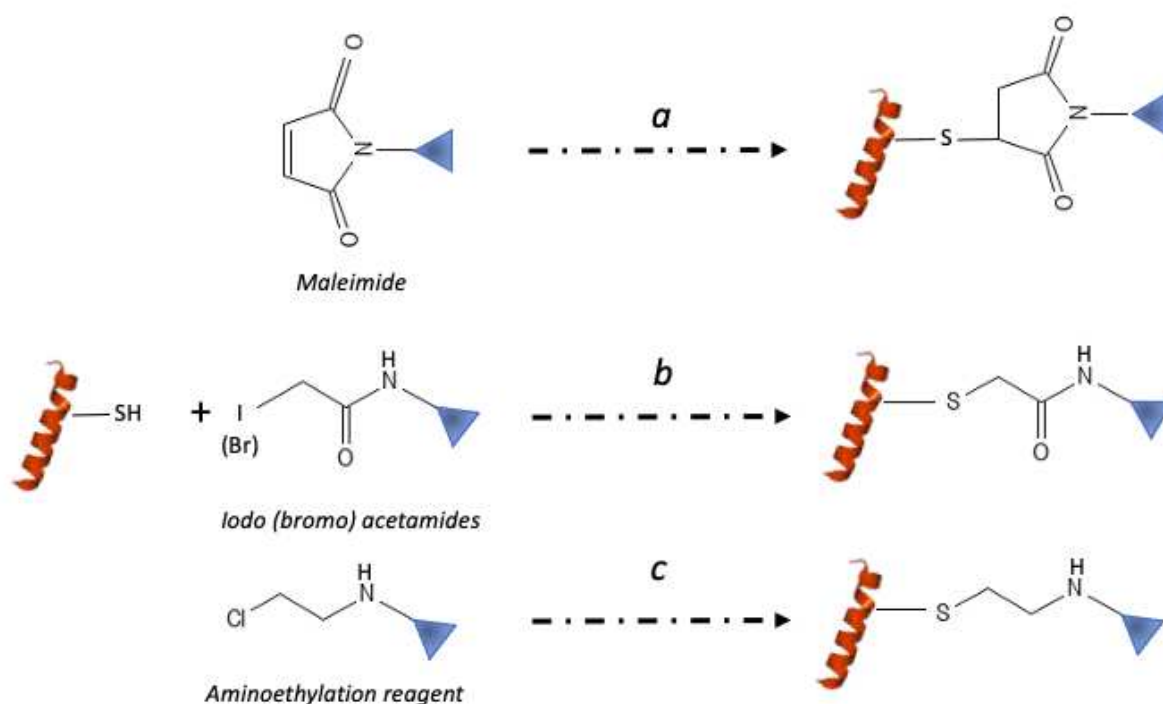
Figure 2.4. *Classic methods for the modification of sulfhydryl (-SH)* *a* *maleimide reaction* *b* *iodo (or bromo) acetamides reaction* *c* *aminoethylation reaction.*

Maleimide derivatives have been commonly used for the modification of the cysteine residues [108]. The reaction is called the Michael reaction and it occurs with the addition of nucleophile to an α, β-unsaturated carbonyl compound (see figure 2.4). In the Michael reaction, while the cysteine residue of the protein molecule is the donor, maleimide derivatives are the acceptor. In the manner of Michael reaction conditions, pH plays an important role to control the reaction since maleimide can also react with other groups too. In pH of higher than 8.5, it can lose the specificity to cysteines and can react with amines. Another scenario is that the maleimide can be hydrolyzed and become an unreactive maleimide acid. The optimal pH is around 6.5 -7.5 [106].
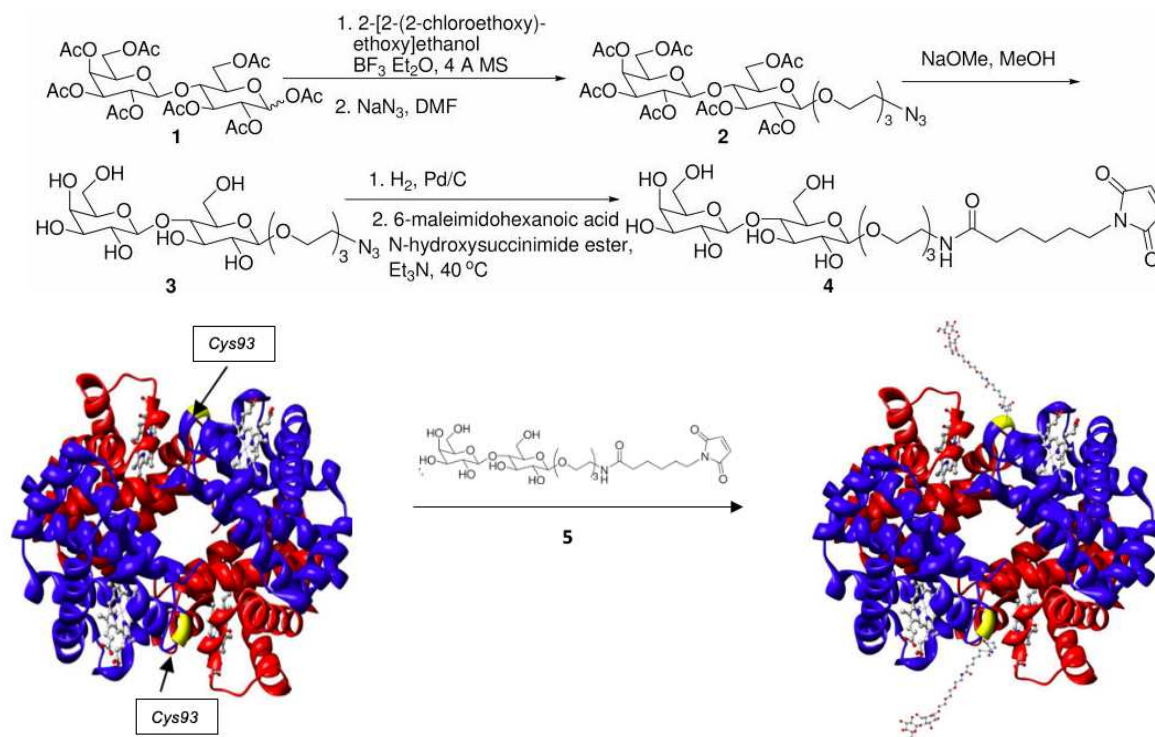
Figure 2.5. Site-selective glycosylation of bovine Hemoglobin at Cys93 by the lactose derivative.

An example of usage of chemical modification with maleimide is demonstrated that maleimide- lactoside was coupled with cysteine for the glycosylation of myoglobin to provide higher oxygen affinity to the protein by Yalong (see figure 2.5). The process is composed of two phases. As a first phase, maleimide-lactoside was prepared. The peracetyl protected lactoside **1** was treated with 2-[2-(2-chloroethoxy) ethoxy] ethanol catalyzed by boron trifluoride and then reacted with sodium azide to yield 1-azido-4,6-dioxa-9-octyl-peracetyl-lactoside **2**. The acetyl groups were removed by treating with sodium methoxide to yield compound**.** The azido group of **3** was hydrogenated to an amine and reacted with 6-maleimidohexanoic acid N-hydroxysuccinimide ester to yield maleimide-lactoside **4**. As a second phase, the maleimide lactoside was introduced to the hemoglobin on Cys 93 **5** [109]. Halo carbonyl electrophiles are the other reagents to couple the cysteine residues. In early studies, halo carbonyls were used to modify the cysteines of the Keratin [110]. Even though coupling the cysteines with Halo carbonyls offers stable conjugation, the reaction rate and chemo selectivity of this reagent reduces the convenience of usage of the halo carbonyls [111].

Modification methods of disulfide bonds

The modification of disulfide bonds by bis-alkylation has become attractive in recent years since it is crucial for designing and improving the properties of protein-based pharmaceuticals. For the protein function, the disulfide bonds have a role to stabilize tertiary and quaternary of the protein structure. Even though the modification of disulfide bonds is required, breakage might alter the stability of the proteins [112].

There are several methods to reduce disulfide bonds. As first α, β-unsaturated-β'-mono-sulfone functionalized PEG reagents were conjugated, and it resulted in a stable thioether bridge. Another method is to introduce aryloxy maleimide and bridging of disulfides permits the incorporation of PEG chains. This approach has been demonstrated by Baker and co- workers in 2014 [113]. The other method, which is mostly commonly used in recent years, is reducing the disulfide bonds by reducing reagents. There are three most used to reduce reagents as Tris (2- carboxyethyl) phosphine (TCEP), dithiothreitol (DTT) and 2-mercaptoethanol (2-ME) [114].

TCEP is a very stable reducing agent and soluble in aqueous solutions. Since it does not include any thiol molecule, there is no need to remove the reagent from the solution after the reaction and it is resistant to the oxidation rather that the other reducing reagents. This reduction also does not require special pH (the pH range is 1.5-9) and as a reducing time is quite short. It is commonly used in the coupling reaction of cysteines by the maleimides [115]. Another well-known reducing reagent is DDT (dithiothreitol). Due to the oxidation of DTT by the air, it is not stable compared to the TCEP. Once it is oxidized, it becomes a molecule with a six membered ring containing a internal disulfide bond. Additionally, its half-life is limited to about 40 hours, and it requires quite specific pH = 6.5 for the protection of stability of the reagent [116]. 2-mercaptoethanol (2-ME) is another reducing reagent which can reduce the disulfide bonds in proteins. It is commonly used since it is more stable than DTT and soluble in water. However, 2-ME is volatile and considered a toxic reagent [117].

Many reagents have been developed for cysteine-specific protein modification of conjugation of cysteine residues through the disulfide bound. However, a few of them allow for multi-functionalization of a single cysteine residue and disulfide bridging bio-conjugation. The study of Zhu and his colleagues is an example for this conjugation

and this conjugation was conducted through reducing the disulfide bonds by reducing reagent (see figure 2.6).
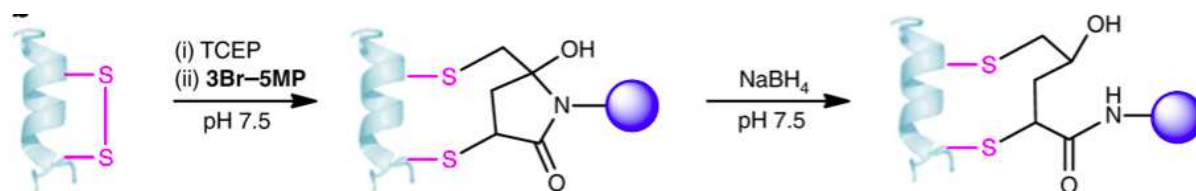
Figure 2.6. *Disulfide bridging bioconjugation of protein using 3Br-5MP.*

They reported a disulfide bridging conjugation through the addition of cysteine to 3Br-5MP, which is a robust reagent for cysteine-specific protein modification. They demonstrated the disulfide bridging bioconjugation with 3Br-5MPs by using the endogenous hormone somatostatin (SST), which contains a single disulfide bridge. The disulfide bond of SST was reduced with (TCEP) in HEPES buffer (pH 7.5) and addition of 3Br-5MP reagent allowed to disulfide re-bridging products. The products were converted to the reduced product by the addition of 2.5 mM NaBH$_4$. Purified products were examined with Ellman's test, which showed that they do not have any free thiol groups. Thereby, they confirmed that the 3Br-5MPs were attached to SST via disulfide bridging [118].

Chemical modification of Carboxyl (–COOH)

Since carboxylate groups are less nucleophilic in aqueous solution, reactions with carboxyl groups are not widespread. Nevertheless, some methods have been developed for the modification of proteins via carboxyl groups. Diazo acetic compounds have ability to couple with carboxyl groups naturally. The reaction occurs through the attacking of carboxylate oxygen atoms to a protonated diazo alkyl group. It results in covalent linkage. The reaction needs a specific pH range and the optimal one was considered as 5. However, at pH=5, the diazo acetic compounds lose their specificity to the carboxyl group and tend to react with the sulfhydryl groups at the same pH. In case of higher pH, the specificity of the reaction stays low since the reaction with other nucleophiles [119].

Chemical modification of Phenol (ph-OH)

Hydroxylation of phenylalanine forms the tyrosine amino acid. Since tyrosine has an electron rich side chain, it can be modified through chemical reagents by different methods such as non-transition metal-catalyzed reaction and transition metal-catalyzed reaction. The first modification of the tyrosine method was demonstrated by Francis and his colleagues. The method was named as a three component Mannich reaction, and the reaction occurs with the aldehydes and anilines. Another method is to use preformed imines for tyrosine modification. This method has a usage in DNA-protein conjugations [120].

Chemical modification with non-natural amino acids

In the last 20 years, site-specific mutagenesis of DNA has been one of the most significant advances in biology. The site-specific replacement of an amino acid in a protein enabled the examinations of the role of individual amino acids in folding, structure, stability, and function of a protein by any of the other nineteen amino acids. Despite of these advantages, the limiting factor of this technique is that can be used to replace an amino acid only with one of the other nineteen naturally occurring amino acids. The ability of unnatural amino acids (amino acid analogues) with novel chemical, physical and biological properties either at multiple sites or at specific sites into proteins brings a new dimension to studies of protein structure and function. The novel properties of these analogues include being photoactivatable or fluorescent amino acids, those that carry heavy atoms such as iodine, reactive side chains such as keto groups, spectroscopic probes, and those that mimic natural phospho-amino acids. This method not only offers the design of proteins with novel chemical and biological properties, but also proteins bearing such unnatural amino acids can be used for in vivo and in vitro studies on protein folding, structure, function, protein-protein interactions, and protein localization [121].

These non-natural amino acids can be synthesized as both biologically and chemically. In biological synthesis, they can be synthesized through metabolic engineering approach. Zhang and co-workers synthesized L-homoalanine directly from glucose using engineered E. coli. This process carried out as threonine synthesized after glycolysis and then it was converted to 2-ketobutyrate by threonine

dehydratase and thus obtained 2-ketobutyrate was diverted to synthesize L-homoalanine. By the time, this approach expanded by the other scientist, they also managed to produce L-2-aminobutyric acid (L-ABA), L-proline, L-azidohomoalanine, β-phenylalanine (from phenylalanine), nor-leucine, and several others, have been successfully biosynthesized. This non-natural amino acid can be synthesized through biocatalytic methods such as asymmetric methods. The limitations of biocatalytic methods were overcome by protein engineering. Chemical routes are more profound than biocatalytic routes, possibly due to versatile synthetic ways. In the group of chemical routes, there are asymmetric synthesis, diastereoselective synthesis, enantioselective synthesis, catalytic synthesis, chiral pool synthesis. Among them, it is proved that the most efficient way is the asymmetric synthesis [122].
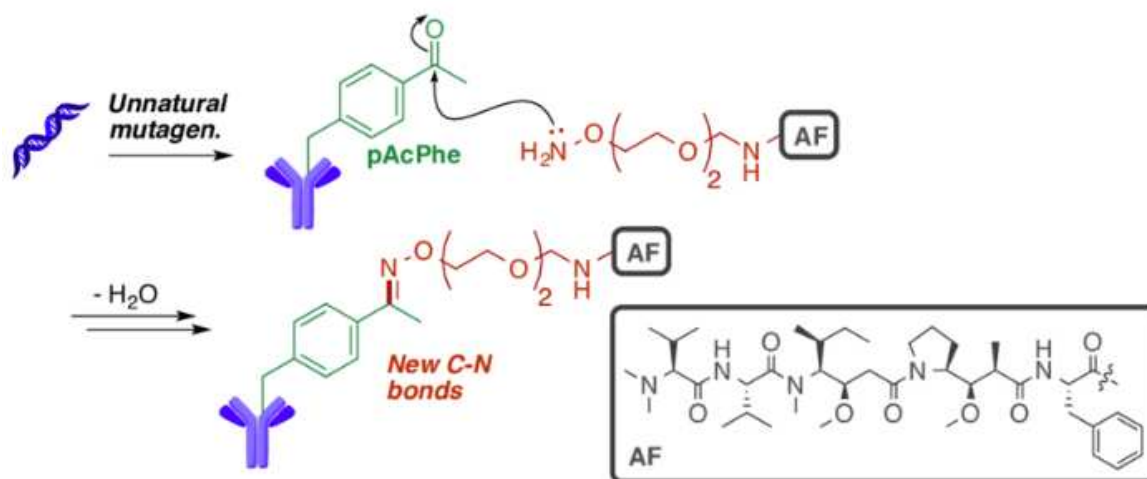
Figure 2.7. *Production of trastuzumab (blue)-AF conjugate. Reaction of genetically encoded pAcPhe residue (green) with alkoxyamine-polyethylene glycol-derivatized AF in an oxime (brown) coupling.*

An example of the usage of non-natural amino acids were shown in figure 2.7. They coupled Auristatin F(AF), which is a highly potent tubulin inhibitor, to an anti-Her2 antibody (trastuzumab). p-Acetyl phenylalanine (non-natural amino acid) was site-specifically incorporated into an anti-Her2 antibody Fab fragment in *Escherichia coli* and mammalian cells. The mutant protein was efficiently and selectively conjugated to an auristatin derivative through a stable oxime linkage, which contains two ethylene glycol moieties modified with a terminal alkoxy-amine group. By this approach, they overcame the limitation of non-selectively cysteine or lysine labeling in the antibody that leading to make optimization of the biological, physical, and pharmacological properties of an ADC difficult through heterogeneous products [123].

2.3.2. Comparison of chemical modification methods

The main purpose of chemical modifications is to create bio conjugated molecules while protecting the function and structure of proteins. To reach the high modification efficiency without an overly complicated experimental protocol, the requirements for these reactions are aqueous reaction buffer, or near neutral pH and room temperature (up to 40°C), high reaction rates, low reactant concentrations, and non-toxic reagents.

Among the modification groups, the modification of phenol and carboxyl groups have several limitations. For instance, carboxylate groups become low nucleophilicity in aqueous solutions, they lose their reactivity during the nucleophilic addition. Furthermore, some modification reactions targeting these amino acids require two steps. This makes the modification more complex and reduce the yield. The addition of primary amines in the system can also make the chemistry difficult to control. The reaction of Pd-catalyzed allylic O- alkylation of tyrosine has solved this problem to some extent. However, metal-catalyzed reactions are not always compatible with proteins. Each time it is necessary to find a suitable concentration of metal to allow rapid reaction while preserving the function of target protein.

The pH is another crucial factor for the future of the reaction because almost all the chemical modification processes are pH dependent. The challenge of carboxyl group modification is the requirement of pH5. However, pH5 is not always compatible for the proteins, they mostly maintain their function around neutral pH which is 7. Therefore, phenol and carboxyl groups are not first choices for the chemical modification of proteins compared to the lysine and cysteines.

Modification of lysine and cysteine residues have been commonly used since they have more advantages compared to the other chemical modification groups. Thanks to the nucleophilicity of the lysines and cysteines, as an average modification of these residues is efficient and simple. For instance, lysine is an amino acid residue which is located mostly on the surface of the proteins and so, it makes conjugation with lysine residues quite convenient because they are mostly accessible. [124]. However, this accessibility can be both an advantage and disadvantage. Since lots of lysines are accessible on the protein, it leads to a lack of selectivity especially in protein-DNA coupling. When the DNA attaches with the aid of lysine conjugation to the critical part

of the proteins such as the binding domain, that can preclude its function [58]. In contrast, cysteine does not usually expose on the surface of proteins, this results in an accessibility problem. At the same time, it prevents the lack of selectivity. Even though the accessibility problem was overcome by introducing cysteines to the surface of proteins, this might lead to protein dimerization or aggregations, through disulfide bond formation. These new disulfide bonds might cause critical changes in integrity and stability of the proteins [125]. Therefore, modification of lysine residues is more favorable than the cysteine residues.

For the modification of amine groups of lysine residues, reagents contain NHS ester are more favorable and thus highly popular compared to the others since isothiocyanates, aldehydes, and imidoester have some limitations as stability, hydrolyzation and pH.

NHS esters are reagents that have half time depending on pH. However, it is quite easy to prevent its hydrolysis by just preparing the reagent just before the reaction. Also, it does not require extreme temperatures and pH for the bioconjugation reaction. The rate of the reaction is high since the NHS group is an excellent leaving group.

2.4. Protein-DNA coupling reactions

Development of the protein-DNA coupling methods offers a new path for the immobilization of the proteins to DNA nanostructures. These DNA nanostructures might be the one ssDNA, dsDNA, or more complex DNA structures.
DNA is chemically inert, and nanostructures based on unmodified DNA mostly lack function. However, functionality can be obtained through chemical modification of DNA nanostructures and the opportunities are endless. This chemically modified DNAs is being used to create functional nanodevices and make DNA nanostructures more applicable. DNA nanostructures provide a powerful method for spatial organization of objects on the nanometric scale. Such nanometric organization of different materials opens the way to the formation of diverse meta-materials with properties emanating from the spatial organization of their components. In this context, it is interesting to improve methods for coupling proteins to DNA, to increase the possibilities for functionalizing DNA nanostructures with proteins. Indeed, the wide range of protein functions provides the possibility of constructing many interesting and useful materials.

2.4.1. Coupling strategies

There are different ways to attach oligonucleotides to proteins: by using covalent links, which is known as conjugation; or based on non-covalent interactions, such as absorption and encapsulation. Antibody-Hapten interactions, aptamer binding, biotin-streptavidin interactions are examples for non-covalent interactions [126].

2.4.1.1. Non-covalent coupling strategies

<u>Antibody</u>

Target-specific therapy was discovered by the inventor of chemotherapy, Paul Ehrlich [127]. The ensuing development of hybridoma technology for generating monoclonal antibodies by Ehrlich along with the therapeutic success of antibodies led to the development of several novel antibody drugs conjugates (ADCs) [128]. In this novel antibody drugs, both polyclonal and monoclonal antibodies can be used. Polyclonal antibodies are mixtures of serum immunoglobulins and collectively are likely to bind to multiple epitopes on the antigen. Monoclonal antibodies contain only a single antibody clone and have binding specificity for one epitope. Haptens are many small molecules of M_r < 1000 such as toxins, drugs and hormones do not have an ability to invoke immune response by direct injection. Thus, they are not immunogenic by themselves. However, it is discovered that these small molecules can triggers an immune response when attached to a large carrier such as a protein thanks to the revealing of the first X-ray structure of an antibody [129] and the structure of the F_{ab} fragment of antibody with the cognate hapten [130]. Furthermore, the knowledge of structural composition of antibody binding regions which are constant and hypervariable regions led the many antibody engineering approaches that were emerging in subsequent years. They are still widely applied as essential capturing, cross-linking, or detection reagents in many diagnostic assays. Hapten-binding antibodies can serve as versatile bridging elements between small chemically synthesized entities and large molecules that are produced recombinantly in bacterial or eukaryotic expression systems. Many scientific reports published that shows designed recombinant hapten-binding antibodies, hapten-binding IgGs for modulation of pharmacokinetic properties of small molecules, hapten-binding antibodies for targeted and pre-targeted payload delivery, bispecific hapten-binding antibodies as cell biology research tools through engineering approach [131].

Additionally, they engineered a technology that connects haptenylated payloads by a covalent disulfide bond to the antibody through by spontaneous redox shuffling reaction. However, this strategy requires the wise use of informatic tools and various optimization, thus it could not compete with other covalent conjugations. ADCs has the limitation of linkage and conjugation chemistry to link an optimized number of the payloads to the antibody in predefined location homogeneously [132].

Streptavidin-Biotin

The interaction between Streptavidin and biotin is one of the strongest non-covalent interactions known so the use of this interaction is popular. Streptavidin is a homo-tetramer from the bacterium Streptomyces avidin that binds up to four biotin molecules with Kd ~ 10^{-14} M. The protein also has high thermostability (Tm of 73°C for apo-SA and 112°C for biotin-SA) and is resistant against extreme pH, denaturing agents, and enzymatic degradation, which are important traits for use under a wide range of experimental conditions. There is also engineered variants of streptavidin as monovalent and monomeric. Due to it being the strongest non-covalent interaction and streptavidin being a very stable protein, they are used in pre-targeted cancer therapy, enzyme engineering, cell biology studies development of biosensors, and designing DNA nanostructures (see figure 1.11). Among these applications, there are some limitations of use of streptavidin-biotin interactions. For drug targeting systems, there is a challenge that prevent its long-term use, which originates from the immunogenicity of streptavidin. Another challenge exists in cell biology studies that coupling of biotinylated cell surface receptors with streptavidin can causes receptor crosslinking, which perturbs cellular signaling, induces receptor internalization and changes receptor dynamics [133].

Aptamers

Aptamers are defined as short, single-stranded DNA, RNA, or synthetic molecules that can be developed with high affinity and specificity to interact with any desired targets. Aptamers has several advantages that make them amenable to play significant roles on the improvement of clinical diagnostics, and therapeutics, such as cancer and pathogen recognition. They are cheap to synthesize in a short time and can easily be

modified without affinity loss. Their specificity can be provided through single point mutations. They are resistant to temperatures. While their potential targets vary widely from small molecules to cells. Their cellular uptake is fast, however, target specific delivery is a limitation for these molecules particularly for *in vivo* studies [134].

Aptamers have more advantages compared to the monoclonal antibodies and monoclonal or polyclonal antibody-hapten interactions [135]. Aptamers are produced synthetically, modified by chemical substitutions and these substitutions raise the stability and half-life of the molecule [136]. Secondly, they have an ability to be stored for a long time, have low immunogenicity and the production of aptamer is low cost compared to antibodies especially for targeted medical therapies [137]. In contrast, antibodies can be easily denatured and are temperature sensitive, furthermore incorporation of modifications to antibodies often leads to reduced activity. Thus, this increases the possibility of binding different antibodies to the same antigenic molecules.

Among the antibodies, the use of aptamers is more favorable for conjugations with the other molecules and particularly with DNA. In the context of coupling with DNA, aptamer immobilization onto the DNA nanostructures is more carried out concept than protein immobilization. Therefore, there are more published reports about the coupling of aptamers than coupling of the proteins with DNA. Nevertheless, these non-covalent interactions are not as stable as covalent conjugation because of the limitations explained above. Particularly, if they are RNA or peptide aptamers, they are prone to quick degradation in the biological environment because of the biomolecule's interactions [138].

2.4.1.2. Covalent coupling strategies

In DNA nanostructures, the individual DNA strands are the building blocks, and their nature eventually decides the structure and properties of the desired design. Adapting the chemical functionality of the DNA building blocks is the first step toward obtaining nanostructures with desired function. In many cases, a chemical handle is incorporated to the DNA strand in-synthesis. The desired modification is subsequently coupled to the DNA strand in a post synthesis conjugation reaction between the chemical handle on the DNA and a complementary chemical functionality on the

modification. Amino groups, thiols, alkynes, and azide are the most common oligonucleotide modifications [112].

<u>Amino Handles</u>

Amino modifications enable selective reaction with activated carboxylates in amide bond forming reactions. Since there is a wide range of commercially available probes containing activated carboxylates, such as N-hydroxy-succinimide (NHS) esters, amino-modified oligonucleotides offer access to a broad range of functionality.

Despite the coupling of NHS esters with amino-modified oligonucleotides is highly popular, this method has some disadvantages. While at low pH to neutral pH, amines are protonated and thus, hardly react with NHS esters, at high pH NHS esters undergo hydrolysis, which also complicates the conjugation. Thus, appropriate buffer composition and high concentrations of reactants are required for the high yield from conjugation of NHS ester [139].

<u>Thiol</u>

Thiol-modified DNA also provides access to functionalization with a wide range of probes. Their efficient reaction occurs with maleimides, α-halo carbonyls, and vinyl sulfones [140]. Particularly, the reaction of thiols with maleimides is commonly used due to its high reaction rate at pH 6.5−7.5. The reaction is also very selective even in the presence of reactive primary amines. A limitation of the thiol-maleimide conjugation is the risk of thiol-exchange in complex medium, but this can be compensated by hydrolytic ring-opening of the thioether product [141]. Another attractive property of thiol-modified DNA is linkage using disulfide formation. Thiol modifiers are commonly conserved as disulfides, and reduction of the resulting disulfide modified oligonucleotides with reagents such as Dithiothreitol or Tris (2-carboxyethyl)-phosphine is essential before conjugation. Also, DNA can be modified with free thiols using benzoyl or MMT protection. However, the obtained thiols are prone to intermolecular disulfide formation and reduction can still be required before further reactions.

Alkynes

Terminal alkynes are favorable to incorporated in oligonucleotides during solid phase synthesis both internally as well as in the 5′ or 3′ end. They are crucial for conjugation with azides in copper(I)- catalyzed alkyne−azide cycloaddition (CuAAC). Copper-catalyzed "click" reaction is a popular method since large number of available azide-probes make alkyne-modified oligonucleotides reliable in bioconjugation. The invention of strain- promoted alkyne−azide cycloaddition (SPAAC) offers another way for efficient reactions with azide probes without the need for copper(I), which tends to complicate certain applications of oligonucleotides due to its cytotoxicity and tendency to generate reactive oxygen species that may degrade DNA. SPAAC is popular for its ease of use and the mild reaction conditions required, but since it does not have a catalyzer as copper, it has lower reaction rate compared to CuAAC [142].

Azide Handles

Reaction of oligonucleotides with alkynes in CuAAC or SPAAC needs introduction of azides although they can be incompatible with phosphoramidite formation due to Staudinger reduction of the azide in the presence of phosphoramidite reagents [143]

Thus, azides are generally incorporated in a separate step after oligonucleotide synthesis either by reaction of sodium azide with haloalkyl modifications on the solid support, diazo transfer reactions to alkyl amines, or reaction of amines with bifunctional NHS-azide molecules [144].

Site-specific attachment of DNA to proteins

Proteins generally contain numerous reactive handles at their surface thereby linking molecular Cargo such as a DNA strand to specific site of protein have some struggles. However, protein engineering approach has a high potential to allow site-specific functionalization of proteins. In this context, when we consider nucleophilic residues of proteins, lysine residues are abundant on the proteins surface, but this is not the case for cysteine residues. To conduct a site-specific attachment DNA to proteins, genetic engineering is a straightforward approach for incorporation of cysteine residue to the protein during protein expression. However, cysteine-incorporation for specific

conjugation did not become a generalized tool due to many proteins have several reactive cysteine residues. Thereby, to incorporate handles for site-specific protein conjugation led to development of expanded genetic codes which allow incorporation of unnatural amino acids. Numerous reports about the incorporation of unnatural amino acids highlighted the importance of the attachment site to protein activity and illustrate the potential for tuning the function of protein−DNA nano assemblies by rational design of the protein−DNA attachment site [145].

Peptide tags for formation of Protein−DNA conjugates

Integration of peptide tags allows sequence-specific enzymatic ligation. It is also effective for site-specific conjugation of DNA to proteins [146]. This strategy requires the modification of the polypeptide sequence with a peptide tag recognizable to a specific enzyme such as Sortase A, Formyl generating enzyme, Phosphopantetheinyl transferase, Farnesyl transferase, Transglutaminase [147]. As an example, Khatwani and his colleagues incorporate CVIA-tags to the C-terminus of GFP and mCherry. This tag is recognized by Saccharomyces cerevisiae Farnesyl transferase (ScFTase), and by adding azidofarnesyl as substrate, an azide modification was attached enzymatically to the C-terminal of the proteins. The azide handles then allowed site-specific attachment of ssDNA via CuAAC or SPAAC [148].

Self-Labeling Polypeptides for Formation of Protein−DNA Conjugates

Despite peptide tags are very useful for formation of protein−DNA conjugates, it requires the use of ligases, and expression of the tagged proteins. The use of self-labeling polypeptides (SLPs) is another strategy for attaching DNA to proteins. SLPs are proteins or peptides that react substrate-specifically to form covalent bonds between the SLP and the substrate. In contrast to the peptide tags, SLPs do not require a ligase to mediate the ligation and they can be attached to any protein of interest. The most famous examples of self-labeling proteins are the Halo-tag and the SNAP-tag [149].

For instance, in the group of Niemeyer 's study, they have labelled Gre2 and BMR oxidoreductase enzymes with SNAP and Halo tag instead of linker (see figure 1.11), to attach them to the more complex DNA structure. This DNA nanostructure, which is

composed of several dsDNA, has been designed as a rectangular shape by DNA origami technique. While this study and our study has common advantage point which provides the stability of the enzymes for the more efficient interaction between them by immobilizing the proteins, this tool's design provides a convenient platform that offers a wide range of places to position them. Nevertheless, in the study of Christof M. Niemeyer, the bridge between proteins and DNA nanostructure, is introduced as a SNAP and Halo tag which are not trackable without any addition of fluorophore molecules [150].

Conjugation using heterobifunctional linkers

Both ways of covalent conjugations, which are by using amino acid residue offer high selectivity, specificity, and yield under simple conditions. In the bioconjugation domain, click chemistry is known as a specific and controllable biorthogonal chemistry. Since click chemistry is biorthogonal, the reactions never occur with functional groups of biological reactants or products. This feature offers specificity and selectivity for the reaction. To obtain the high yield from the reaction, reaction rate should be high and reactive groups should be stable. But since click chemistry is a highly feasible method to perform under simple conditions, it is not difficult to optimize the reaction conditions such as molar ratio, pH for raising the yield of the reaction [89]. Discovery of biorthogonal reactions has started by inactivated azide and alkyne cycloaddition. Even though (3+2) azide cycloaddition was commonly used and seemed suitable [151], there was a crucial challenge as high temperature and high pressure were needed for the reaction. Since high temperature and high pressure were not suitable conditions for several biomedical and biomaterial applications, copper as a catalyst was introduced to allow reactions to be performed at room temperature and atmospheric pressure [152]. Copper-azide- alkyne (3+2) cycloaddition (CuAAC) click reaction became to be more convenient due to the performing the reaction became simpler and usage of this bioconjugation had more wide range of applications [153,154].

For example, Cu- catalyzed Azide- alkyne (3+2) cycloaddition was used for immobilizing modified proteins to the solid surface [155] and ODNs [89]. Although biorthogonal click reactions become more attractive by these optimizations, there are

still a few challenges that can drastically impact on efficiency of the reaction. Cu can cause oxidation of amino acid chains [156] and degradation of DNA in *in vitro* studies, furthermore for *in vivo* studies, copper can be problematic in living systems because of its toxicity [157,158]. To overcome these obstacles, researchers introduced a cyclooctene ring, which reacts with azide [142,151]. Although these reactions were rapid and able to reduce the activation energy, they were not providing the rise of stability because nitrones can be hydrolyzed or dimerized. Hence, there was a development which overcomes this obstacle about instability. A catalyst-free, strain-promoted alkyne azide cycloaddition (SPAAC) was introduced to bioconjugation chemistry (see figure 2.8). For strained alkynes, Difluorinated cyclooctyne (DIFO), was the molecule which works for the click reaction without catalysis. It was able to reduce the activation energy, making them form regioisomeric mixtures of triazoles [159]. So far, several strained cycloalkynes were developed for the click chemistry.
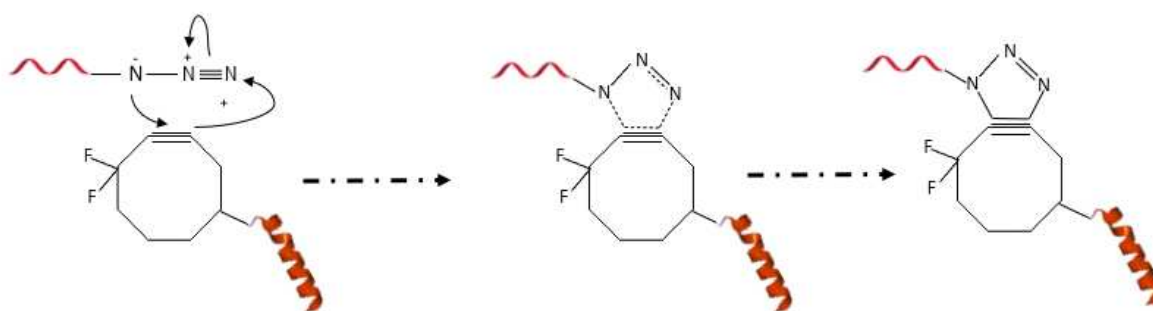
Figure 2.8. *The chemical reaction process of SPAAC.*

Many of the methods used for the construction of covalent protein-DNA adducts require the manipulation of the protein sequence to allow specific and quantitative modification of the protein. Thus, for example engineering unique cysteine residues, the incorporation of amber codons in specific locations in the sequence or the addition of peptide tags. The notable exception to this is the attachment of DNA through the modification of native amino-groups in the protein (the N-terminus and lysine residues). This approach does not require genetic manipulation of the gene encoding the protein of interest. However, this comes at the cost of lack of specificity due to the presence, typically, of multiple reactive amines on the protein surface.

Chapter 3 – My Objectives

DNA nanostructures provide a powerful method for spatial organization of objects on the nanometric scale. Such nanometric organization of different materials opens the way to the formation of diverse meta-materials with properties reflecting spatial organization of their components. In this context, it is interesting to develop methods for coupling proteins to DNA nanostructures, in particular because of the very diverse functional and structural properties observed in proteins.

Due to protein-DNA coupling specificity and addressability, the biochemical conjugation method has been used for a wide range of biological applications [126]. Manipulation of proteins by modifying nucleotides is commonly used for immunoassays [160], biomedical diagnostic and nanofabrication. In the research studies of nanofabrication, protein-DNA coupling can be used for engineering and mimicking complex biological systems. Those model systems are not only used to enable demonstration of systems such as signaling enzyme cascade reactions and light harvesting systems but also, they have high potential to allow designing bioinspired systems such as cellulose degrading and photosynthetic systems [126]. Also, Protein-DNA coupling methods have an important potential as a tool to organize proteins, enzymes, and different macromolecules in multi complexes [161–164].
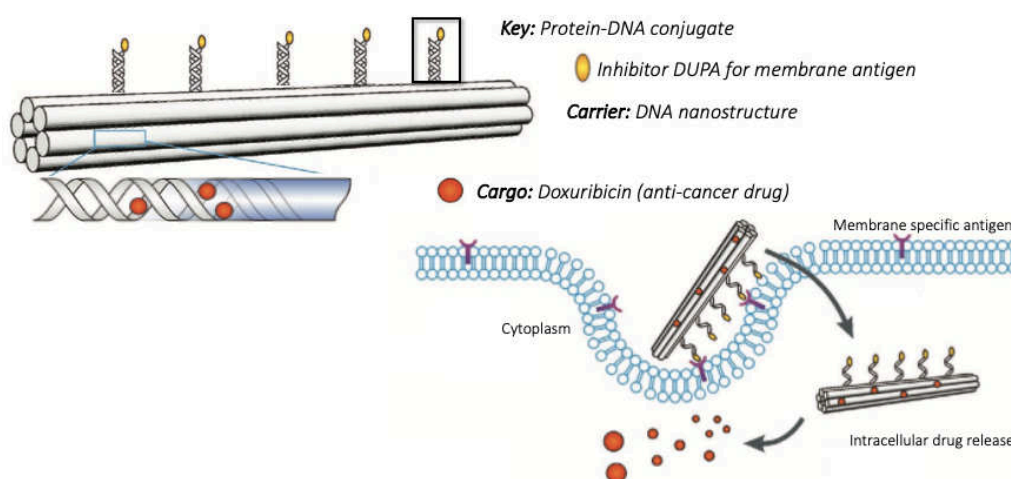
Figure 3.1. Six-helical-bundle DNA origami nanostructure (6HB-DONs) as its working principle [165].

As an example of molecular transportation, specific drug delivery system was shown in figure 3.1. In this example, the target was the prostate cancer cells which has a specific antigen on its membrane. Li and his colleagues built the DNA nanostructure as a carrier to deliver the anti-cancer Doxorubicin to the target cells. They loaded Doxorubicin inside of the nanostructure since the drug is a DNA binding molecule.

To make the delivery specific, they attached protein-DNA conjugate on the nanostructure. This molecule was composed of one double helix DNA and specific inhibitor called DUPA with affinity for the membrane antigen. At the end of the process when administered, the protein-DNA conjugate was binds to the membrane antigen. This leads to the DNA nanostructure being internalized and reaching the cytoplasm where the Doxorubicin is released directly in the targeted cancer cells.
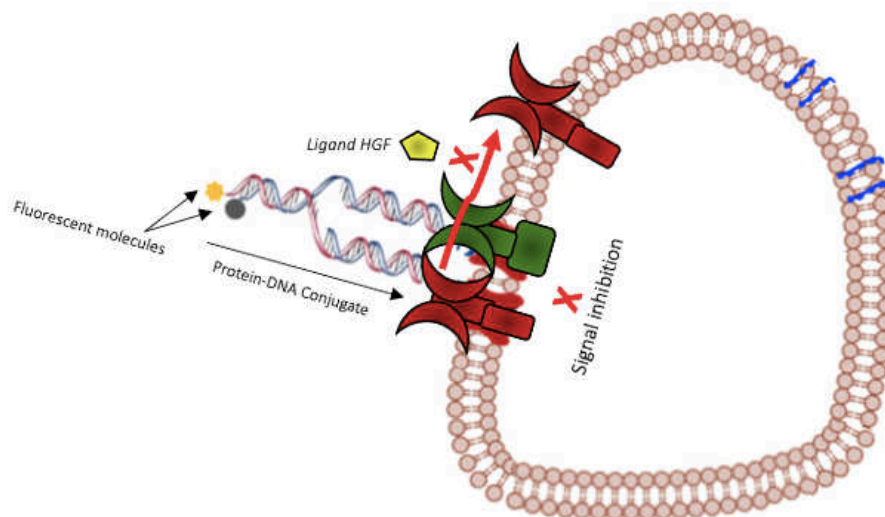
Figure 3.2. Bispecific aptamer induced protein dimerization to specifically regulate Met receptor function and downstream signaling pathway [166].

An example of regulation of signalling pathways was shown in figure 3.2. There are three receptors on the cell membrane. The MET receptors were shown in red from the met signalling pathway, while TFR was shown as green and HGF ligand was shown in yellow. In case of dysfunction of the MET receptors in the signalling system, the dimerization of MET proteins can give also abnormal signal by interacting with their ligand HGF and this can trigger the metastasis or proliferation. To block dimerization of MET proteins, Wang and his colleagues modified MET and TFR which is highly express in cancer cell by DNA aptamers. Then they connected them with double helix

DNA to prevent the dimerization of the MET proteins. Then, they labelled DNA aptamers with fluorescent dyes. At the end, they demonstrated that the interaction of ligand and MET receptor was disrupted. Therefore, protein-DNA conjugate was used as useful tool to regulate membrane receptors in case of dysfunctional signalling pathway.
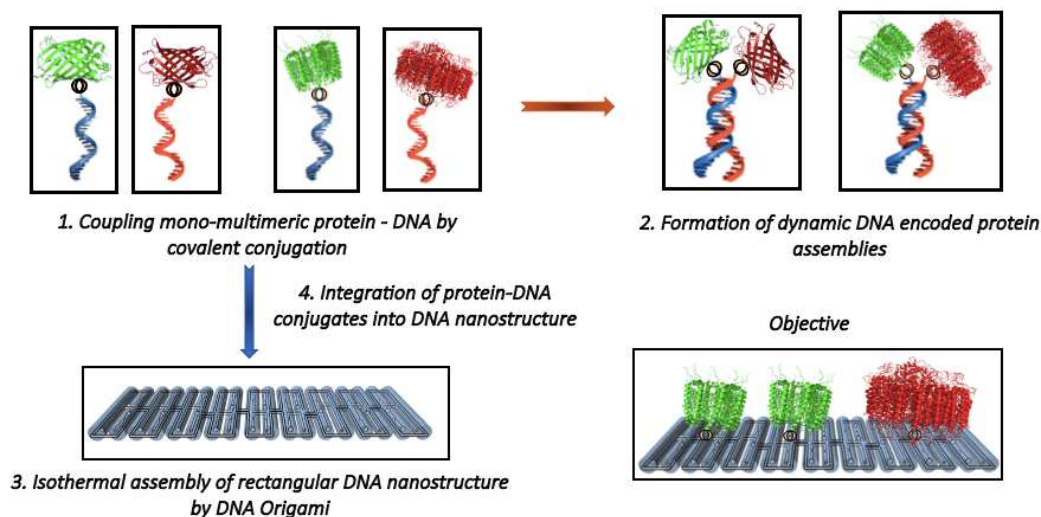
Figure 3.3. 1) Coupling mono-multimeric protein-DNA by covalent conjugation. 2) Formation of dynamic DNA encoded protein assemblies. 3) Isothermal assembly of rectangular DNA nanostructure by DNA origami. 4) Integration of protein-DNA conjugates into DNA nanostructure.

In this context, my objective in the work presented here was to develop a methodology for protein-DNA coupling applicable to first homo-multimeric proteins for producing a 1:1 adduct that is one protein with multiple reactive groups, attached to one DNA molecule. Allowing, as far as possible, the accurate positioning of the protein in DNA nanostructures based on freely available chemicals. This method allows the controlled formation of 1:1 protein-DNA adducts, and the DNA in these adducts is accessible for the formation of DNA scaffolded nanostructures. The path of this procedure is composed of four steps; chemistry for attaching protein to DNA, assembly of protein-DNA chimeras, DNA guided membrane protein assemblies, measuring the effect of the positioning of the functional properties of proteins and integration of DNA guided protein assemblies to the DNA nanostructure (see figure 3.3).

In the context of homo-multimeric proteins, genetic manipulation is unable to introduce unique labeling sites, therefore I decided to opt for a controlled labeling of the surface amine residues using a bi-functional reagent. The notable exception to this is the

attachment of DNA through the modification of native amino-groups in the protein (the N-terminus and lysine residues). This approach does not require genetic manipulation of the gene encoding the protein of interest. As proteins I have chosen to study two fluorescent proteins sfGFP and mRuby3. The reason why is that, since they are fluorescent proteins, it is easy to follow them during an experimental procedure. Secondly, they both have lots of lysines, which makes them possible to label with linker. As a bridge between DNA and proteins, I have chosen the linker DBCO-NHS-ester. Lysines of proteins can bind to the DBCO-NHS ester and azide modified DNA are able to bind cyclooctene of the linker. This chemistry has the advantages of being widely applicable, easy to follow to adjust reaction conditions to optimize the yield (thanks to the UV absorption of DBCO) and give readily separable reaction products, thanks to the charge on the DNA molecules and the charge modifications resulting from the loss of primary amines. Nevertheless, the reason behind choosing DNA as material for this construction is that it is easy to synthesize the DNA, design the geometric DNA structures and it has high chemical and structural stability compared to the RNA. Furthermore, it is well known that thermodynamically duplex stability might be calculated accurately and the Watson-Crick base pairing between strands is a strong interaction.

I believe that DNA based structures provide an experimental route to controlling organization and assembly of proteins also in complex environments such as cell membranes. It is crucial to understand functioning of biological membranes and organization of the integral membrane proteins because they are responsible for a wide variety of biological processes. Over the years, genomic studies have showed that 30% of genes are expected to code for membrane proteins [165]. Furthermore, 50–60% of drugs currently on the market target membrane proteins, and account for 25 of the best-selling drugs [166]. Despite membrane protein organization is of major biological and medical interest, they remain poorly studied. However, integration of DNA nanotechnology to this biological challenge can be promising since these DNA based nanostructures can enable us to test different hypotheses on the complex biological environments like cell membranes and thereby understanding non-elucidated part of these complex systems comprehensively.

Because of the reasons mentioned above, I decided to bring membrane proteins and systems together with the advantages of DNA nanotechnology. Among the numerous of membrane systems, I chose a light harvesting system in photosynthetic membrane from purple photosynthetic bacteria since it has an ability to convert the energy from the light to the chemical energy with around 99% of efficiency [167].

For this reason, to mimic a complex energy transfer system of the photosynthetic unit through DNA based nanostructures offers a route to comprehend the importance of critical factors for organization of biomolecules on the photosynthetic membrane. Understanding the secrets of this systems and the reason why it is that much efficient can be promising for the design of new efficient systems which utilizes the solar energy through the principle of this mechanism.
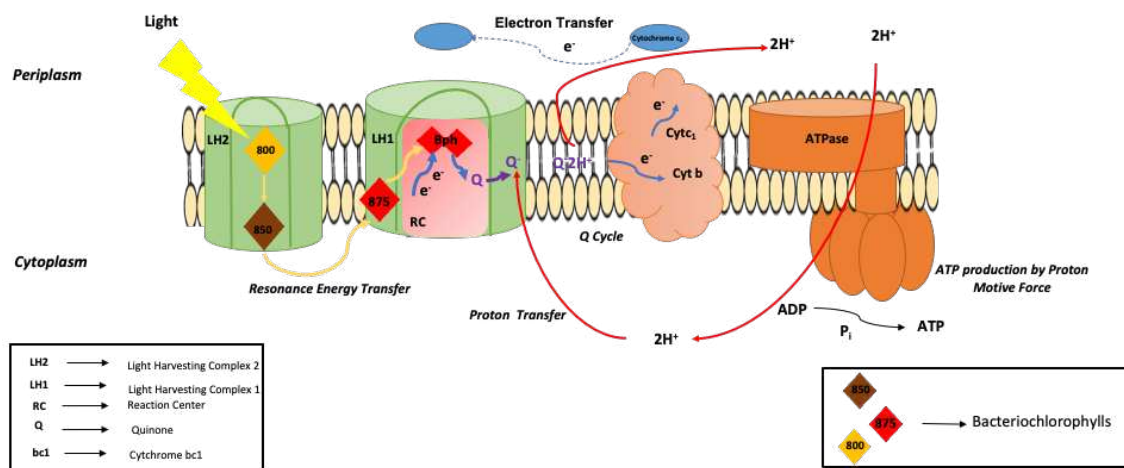
Figure 3.4. *Organization of Photosynthetic Unit in Purple Photosynthetic Bacteria.*

The photosynthetic unit from purple photosynthetic bacteria is responsible for harvesting light and converting the electromagnetic energy into chemical energy (see figure 3.4). The system involves four integral membrane proteins: LH2, the core complex-LH1, cytochrome bc1 and ATP synthase. In the system, light is initially absorbed by the carotenoids and Bacteriochlorophyll which are the pigments of light-harvesting complex 2, and then the absorbed energy is transferred to the core-complex via LH1. Here, this resonance energy transfer requires very close association of these two proteins thus the organization of proteins is crucial for energy transfer [168]. The energy from the light excites the electrons of CC and leads them into electron carrier Bacteriochlorophyll. After the mobile electron carrier Q gains electrons

and is reduced, quinol takes two protons from the cytoplasm and then diffuses to the cytochrome bc1 complex. In cytochrome bc1, when quinol gives the electrons to cytochrome bc1 complex, it releases the protons into the periplasm and the electrons are carried back to the reaction center by cytochrome c2. Therefore, this electron and proton transfer changes the density gradient of protons between cytoplasm and periplasm. ATP-synthase then uses the proton motive force to make ATP for the cell. Thus, the efficiency of this process depends on the organization of those proteins. The absorption, energy transport, and dissipation properties are managed by the balance of pigment-pigment and pigment-protein couplings [169]. However, the molecular machinery responsible for this balance, and for the couplings themselves, is still poorly understood. The reason is mainly related to the fact that the couplings are highly sensitive to intermolecular distances and microenvironments. As a result, the light-harvesting properties differ drastically among individual photosystems, because of small differences in protein conformation, and vary drastically even within a single photosystem over time, because of protein fluctuations [170]. This biological system is therefore particularly sensitive to the organization of the components in the membrane and that is why it is an interesting and challenging model system.

Having developed a tool with sfGFP and mRuby3, I will transfer it to the proteins of the photosynthetic unit but LH2 from *Rhodobacter sphaeroides* makes the procedure difficult because of some additional considerations. The first is solubility of membrane protein. To make the membrane soluble, I require to use a certain percent of detergent. But still, it is not easy to provide the exact favorable environment because of hydrophobicity. This hydrophobic environment of the reaction can prevent the reaction between the elements since it surrounds only hydrophobic reaction elements. The second is the heterogeneity of LH2. Which originates from the lipids which are in the center of the barrel shape of it, preventing their removal during the purification process.
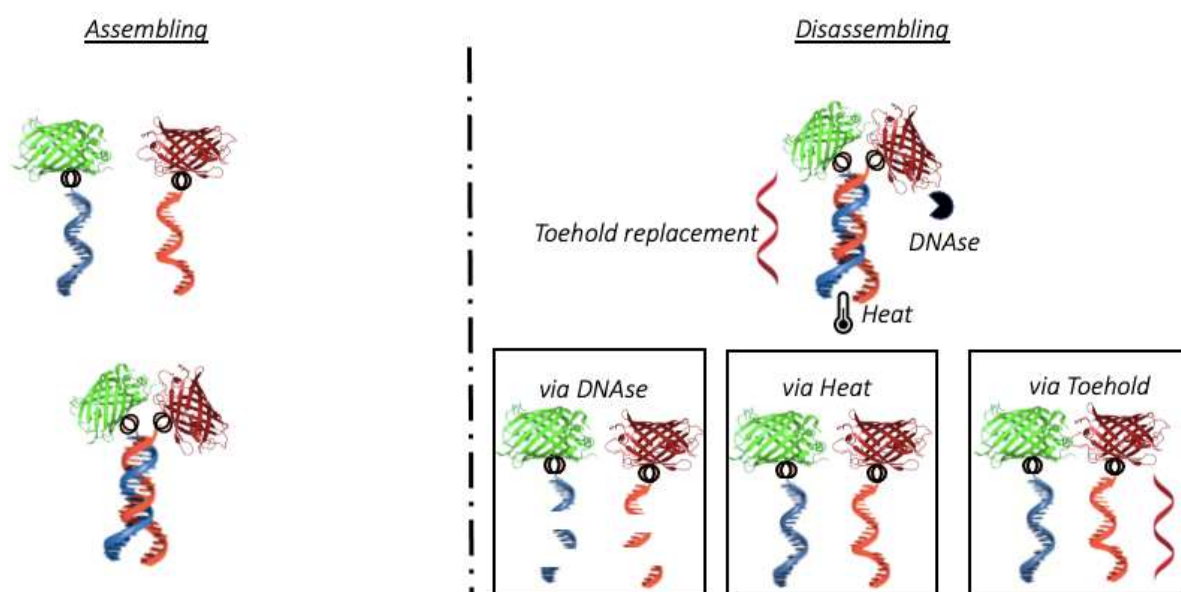
Figure 3.5*. protein-DNA conjugates can be formed in dynamic structures**. They can be assembled at room temperature and can be disassembled by different method as DNAse, heat or toehold strand replacement.*

One of the interesting aspects of DNA nanostructures is that they can be dynamic and be induced to change structure. This depends on the availability of the DNA elements to different agents. Having formed the sfGFP-DNA and mRuby3-DNA 1:1 adducts, I wanted to assemble them in a DNA directed manner to form dimers. To verify that the DNA elements in the dimeric structure, I wanted to use several methods to disassemble the dimer (Figure 3.5). Then, since these two DNA were part of sequence design of the rectangular origami structure, I will integrate these dimers into the rectangular DNA nanoobject by the isothermal DNA origami assembly.

There are basically two approaches to design DNA based nanostructures as ssDNA tiles and DNA origami in the literature. I chose the DNA origami method instead of ssDNA tiles since it is proven that one of the biggest limitations of ssDNA tiles is the low formation efficiency compared to the DNA origami. The techniques for the formation through DNA origami are categorized in two so far in the literature: thermal assembly and isothermal assembly with and without denaturing reagents. In my context, the structure includes proteins and hence I cannot use thermal assembly which requires denaturation step at high temperature. Therefore, the isothermal

assembly is preferred since it is feasible to conduct formation of rectangular nanoobject at favorable temperatures for stability of proteins. [51]. Among these favorable temperature for DNA nanostructures with proteins, there are a few optimization studies which were explained in chapter 1, for the isothermal assembly of DNA origami. It is shown that isothermal assembly at 25°C, 37°C and 50°C. While the formation efficiency at 37°C and 50 °C was around 50%, the yield of the assemblies at 25°C were not improved much and stayed around 20% [46]. Thus, my purpose was to improve this isothermal assembly efficiency at 25°C and obtain DNA nanostructure directed protein assemblies.

In this thesis, I carried out experiments of the Protein-DNA coupling with sfGFP, mRuby3 and membrane proteins. While I succeeded in obtaining DNA encoded protein assemblies with mRuby3 and sfGFP, I had struggles with LH2 molecules via methods which were explained in chapter 4.

Chapter 4- Materials and Methods

4.1. Methods for homomeric proteins

### 4.1.1. Protein expression and purification

The mRuby3 gene from mRuby3 C1 Cloning vector (plasmid #127808 from Addgene) was PCR-amplified using Q5 DNA polymerase (New England Biolabs) and oligonucleotides fp_mRuby3 and rp_mRuby3 from table 4.1. Amplified PCR products were inserted into pET-6His-TEV vector using the EcoRI and XhoII restriction sites. Similarly, the sfGFP gene was PCR-amplified using Q5 DNA polymerase (New England Biolabs) [171] using the oligonucleotides fp_sfGFP and rp_sfGFP in table 4.1 and inserted into pET-6His-TEV. PCR products were transformed into *Escherichia coli* DH5α and plasmid constructs verified by sequencing. Proteins were expressed in *E. coli* BL21-DE3 freshly transformed with the appropriate plasmid.

Name of Plasmid	Plasmid Number	Bacterial Strains
pET-6His-TEV vector	plasmid #84863 from Addgene	*Escherichia coli* BL21-DE3
mRuby3 C1 Cloning vector	plasmid #127808 from Addgene	*Escherichia coli* DH5α
pET-6His-TEV sfGFP	engineered plasmid #25630 from novagen	*Escherichia coli* BL21-DE3
pIND4-RC1	engineered plasmid	*Rhodobacter sphaeroides GA*

Table 4.1. Plasmids and strains used in this work.

2 ng of pET -6 His-TEV (100 ug/ml amp) was transformed to 40 ul of competent *E. coli* BL21 cells by first incubating on ice for 10 minutes, then, the mixture of plasmid and competent cell were put in 42°C for heat shock for 1 minute, incubated back on the ice 4 minutes. 500 ul of sterilized LB was added to the mixture and incubated on the shaker at 37°C for 1 hour. The mixture was centrifuged at 1000 rpm for 4 min. 400 ul of supernatant was discarded. The pellet was resuspended in 100 ul of LB and it was spread out to LB Plate with 100 ug/ml ampicillin. It was placed into an incubator at 37°C for overnight.

For protein production, a single isolated colony was taken from the plate and grown with shaking overnight at 37°C in 10 ml 2YT medium supplemented with 25 µg/ml of Ampicillin. This culture was used to inoculate 500 ml LB media and cells were grown at 37°C with shaking (180 rpm) to an O.D at 600nm of 0.8, about 3 hours. Protein production was then induced by addition of 1mM IPTG and the temperature reduced to 30°C. Cells were harvested when the OD at 600 nm reached about 2.0 (3 hours). For mRuby3 a better yield was achieved if cell growth was overnight at 25°C, prior to induction with 0.1 M IPTG for 3 hours at 16°C.
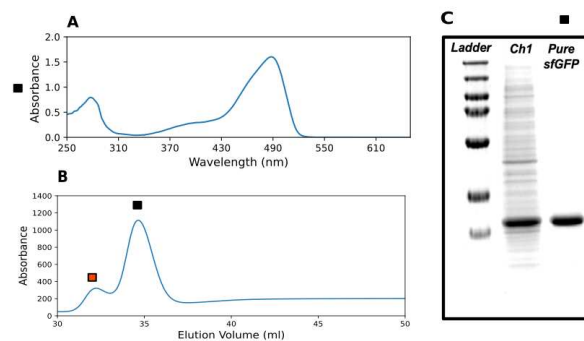
Figure 4.1. *Separation of sfGFP from unwanted proteins. **A** Purity of sfGFP was assessed through visible UV spectrophotometer. **B** Purification of sfGFP using His-trap column. The initial peak contains the unwanted proteins (red) and the later peak contains sfGFP. **C** Process of the purification was shown through the SDS-PAGE method. Charge 1 represents (Ch1) the protein mixture before purification of sfGFP and it contains lots of unwanted proteins in the first well. The second well is the product of the later peak in AB (black).*

Cells were harvested by centrifugation, 4500 rpm for 20 minutes, then cell pellets were washed phosphate buffered saline (PBS). Cell pellets (about 5g) were either used immediately or flash frozen and kept at -80°C until use. Cells were lysed by resuspending the cell pellet in 50 ml of lysis buffer (50 mM Tris pH=8,100 mM NaCl, 1mM EDTA, 10 mM $MgCl_2$,0.1 mg/ml DNAse, 0.1mg /ml Lysozyme, 1mM PMSF) for 1 hour at 4°C with shaking prior to lysis using 3 passages of an Emulsiflex C5 high

pressure homogenizer (Avestin, Ottawa, Canada) at between 10000 -15000 psi. The cell lysate was cleared by centrifugation (45K rpm for 35 minutes).

The cleared cell lysate was loaded onto a 1ml Histrap (Cytiva) column equilibrated with buffer A (50 mM K_2HPO_4 ,100 mM NaCl ,10 mM, Imidazole pH 7.5), the column was washed with a further 5ml of buffer A before elution with a 15 ml gradient to buffer B (50 mM K_2HPO_4,100 mM NaCl, 500 mM, Imidazole pH 7.5). Proteins were purified on Ni-NTA columns thanks to the oligo-histidine attached tag. The Ni-NTA column has metal ions chelated on the surface of beads and the oligo histidine tag has a higher affinity for these beads than the other proteins. After binding the histidine tagged proteins, the gradient of imidazole was used to elute them from the column since the imidazole competes with His-tag for binding to the metal-charged resin [172].

During elution 0.5 ml fractions were collected. sfGFP were eluted at 450 mM imidazole and the initial peak containing the unwanted proteins (red square) peak and the later peak (black square) (see figure 4.1.B.) contained the purified sfGFP. The fractions corresponding to the later peak (black square) were collected and loaded on the SDS-PAGE with the initial protein solution as charge (Ch). The difference between charge and sfGFP in figure 4.1.C, showed that sfGFP is pure enough since there are no other bands corresponding to the unwanted proteins as in the clarified lysate. Fractions were pooled and dialyzed against 100 mM $NaHCO_3$ pH 8.0, and then concentrated to 200 µM using a 10 kDa concentrator (Amicon Ultra 4 Centrifugal Units Merck). Purified protein was used immediately or flash frozen and stored at -80°C until required. The ratio of absorbance of protein at 280 nm to the absorbance of the sfGFP specific peak at 488 nm is a convenient measure of the purity of protein. To ensure the purity, this ratio must be less than 0.5. In this purification, the ratio A280/ A488 was calculated as 0.23 by UV-visible spectroscopy in Figure 4.1.A.
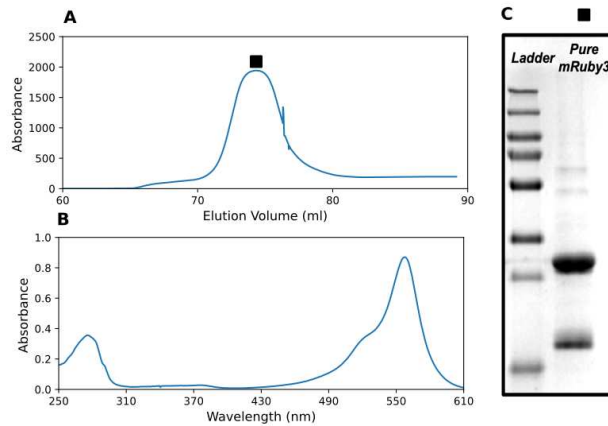
Figure 4.2. *Separation of mRuby3 from unwanted proteins with affinity chromatography.*
A Purification of mRuby3 using his-trap column. The peak contains the pure mRuby3 (black square).
B The first well is the product of the peak in A as pure mRuby3 (black square). C Purity of mRuby3
was assessed through visible UV spectrophotometer.

mRuby3 was purified from clarified cell lysate through His-trap column. mRuby3 was eluted at 500 mM Imidazole as a single peak (black square) (see figure 4.2.A). The fractions from the peak (black square) were assessed by UV-visible spectroscopy (see figure 4.2.B). The purity was confirmed through UV-visible spectrophotometer and SDS-PAGE gel analysis. The ratio A280/A488 which was calculated through UV-visible spectroscopy, and it was 0.3 in Figure 4.2.A. Then, the fractions corresponding to the peak (black square) were loaded on the 12% SDS-PAGE to examine the purity and mRuby3 eluted as two bands (see figure 4.2.C).
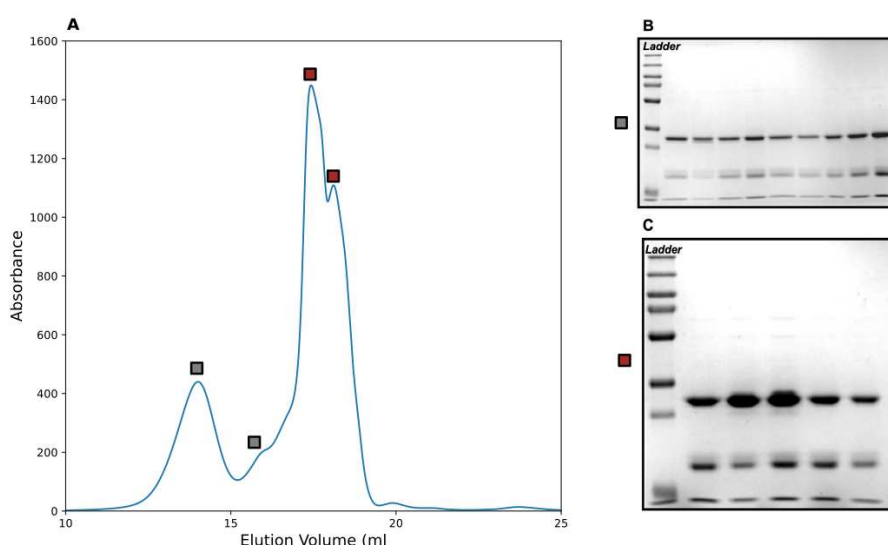
Figure 4.3. *Separation of mRuby3 from unwanted proteins with affinity chromatography.*
A *Purification of mRuby3 using his-trap column. The peak contains the pure mRuby3 (black square).*
B *The first well is the product of the peak in A as pure mRuby3 (black square).* **C** *Purity of mRuby3*
was assessed through visible UV spectrophotometer.

To understand whether the second band was originated from the dimerization of
mRuby3, the sample was analyzed with the size exclusion chromatography. The
fractions corresponding to these peaks (gray and red) shown in figure 4.3.A They were
loaded on SDS-PAGE, and all showed the same profile (see figure 4.3.B-C). Thus, the
origin of the double bands was not clear but dimerization of mRuby3 does not seem
to be the cause. All the fractions loaded on SDS-PAGE were collected into the dialysis
membrane for an exchange buffer with 100 mM $NaHCO_3$ pH:8 for the further steps.

One Gel	H₂O	Acrylamide 30%	3M Tris pH8.8	SDS 10%	APS 10%	TEMED	Final Volume
Separating	2.2 ml	2 ml	0.65 ml	0.05 ml	0.05 ml	0.005 ml	5 ml
Stacking	1,4 ml	0,33 ml	0.25 ml	0.02 ml	0.02 ml	0.03 ml	2 ml

Table 4.2. *Preparation of 12% of SDS-PAGE.*

4.1.2. Formation of DNA-Protein assemblies

4.1.2.1. DNA

The oligonucleotides as shown in table 4.3 were purchased from Biomers (GmBH, Germany) and dissolved in water at a concentration of 100µM. Oligonucleotides were frozen and kept at -20°C prior to use.

Use	Sequence of Oligonucleotides
FP _sfGFP	5'GCG-GCA-GAG-AAA-ACT-TGT-ATT-TCC-AGG-GCC-AAT-TGT-CTA-AAG-GTG-AAG-AAC-TGT-TCA-CCG-GTG'3
RP_sfGFP	5'GTG-CGG-CCG-CAA-GCT-TGT-CGA-CGG-AGC-TCT-TAT-TTG-TAG-AGC-TCA-TCC-ATG-CCG-TGC-G 3'
FP_ mRuby3	5'CCA-CTC-GAG-TTA-CTT-GTA-CAG-CTC-GTC-CAT-GCC 3'
RP_ mRuby3	5'CCA-CTC-GAG-TTA-CTT-GTA-CAG-CTC-GTC-CAT-GCC 3'
Attached to mRuby3	AZIDE- 5'TCA-ACC-TTC-TAC-AAG-ATG-GTC-GTT-TTT-T 3'
Attached to sfGFP	5'TTT TTT CGA CCA TCT TGT AGA AGG TTG A 3'-AZIDE
Toelhold Displacement	5'TCA-ACC-TTC-TAC-AAG-ATG-GTC-GAA-AAA 3'

Table 4.3. *Sequences of synthetic oligonucleotides used in this work.*

4.1.2.2. Coupling reactions

To couple protein and DNA various protocols were used, see Results. The final protocol on which I settled was as follows: to 60 nmol of protein (at 200uM in 100 mM NaHCO$_3$ pH 8.0) was added 240 nmol of DBCO NHS ester (Jena Biosciences, GmBH) in dry DMSO and the reaction allowed to proceed for 2 hours at room temperature. After incubation, the reaction was stopped by addition of 100 mM Tris pH 8.0. Excess DBCO NHS ester was removed by repeated (3 cycles) 10-fold dilution and concentration with 100 mM NaHCO$_3$ pH 8.0 using a 10 kDa concentrator, as above. DBCO coupling to the protein was assessed by a UV-visible spectrophotometry to calculate the labeling yield. Then, 24 nmol of azide modified oligonucleotide was added to the mixture and incubated overnight at room temperature. Reaction products were separated on a 1 ml Resource Q anion exchange column (Cytiva). The column was equilibrated and washed with 50 mM Tris, 10 mM NaCl pH 8.0 and then products eluted with a gradient between 500 mM and 700mM NaCl. The various products were collected in fractions (0.5 ml) and characterized by UV-visible spectrophotometry. If necessary purified and pooled fractions were concentrated using a 10kDa concentrator. Several variants of this protocol will be discussed.

4.1.2.3. DNA directed dimer formation

For DNA directed dimer formation 1:1 Protein-DNA adducts (sfGFP or mRuby3 with the appropriate oligonucleotide) were added to a final concentration of 0.5 µM in the hybridization buffer (5 mM Tris pH 8.0, 1 mM EDTA ,15 mM MgCl$_2$) and incubated at room temperature.

4.1.2.4. Analytical methods

Fluorescence images of SDS-PAGE gels containing fluorescent proteins were obtained with a Typhoon FLA 9500 scanner (GE Healthcare Life Sciences, Uppsala, Sweden). Images measured at 488 nm and 555 nm were combined as the red and green, respectively, channels in the color image.

To measure melting curves 100 µl of hybridized molecules were placed in a microcuvette. Absorption at 260 nm was measured using a spectrophotometer fitted with a Peltier temperature controller (Agilent Cary 3500 UV-Vis Compact Peltier, USA).

The sample was heated from 25°C to 70°C at the rate 1°C/ min and the absorption changes at 260 nm measured.

To measure hybridization kinetics 0.5 µM of mRuby3-DNA adduct in hybridization buffer was put in a microcuvette thermostated at 25°C. Absorption changes at 260 nm were measured after sfGFP-DNA adduct was added to a final concentration of 0.5 µM and the absorption changes at 260nm were followed.

FRET measurements were made of donor sensitized acceptor emission using a TECAN M200 PRO (Tecan Life Science, Switzerland) microplate reader equipped with excitation and emission monochromators. Samples (100µl with 0.5 µM proteins) were prepared in Greiner 96 Flat Bottom Black Polystyrene plates in 5 mM Tris pH 8.0, 1mM EDTA, 15 mM $MgCl_2$. Emission spectra were measured between 500 and 700 nm with 470 nm excitation. DNAse treatment was performed by adding 0.1 mg/ml DNAse 1 (Roche) to the hybridized molecules (0.5 µM) in the presence of 5 mM EDTA and digestion allowed to proceed at room temperature for 15 minutes.

4.2. Methods for multimeric proteins

I have chosen membrane proteins from purple photosynthetic bacteria to work in this thesis project. There are five steps which are growth of bacteria, breaking bacterial cells, purification of intracellular membrane, purification of LH2 and quality control of the protein.

4.2.1. Cell growth

Among several species, I have just chosen *Rhodobacter (Rb.) sphaeroides GA* because they have lots of lysines and they are all located at the N-terminal of the protein. *Rb. sphaeroides GA* cell was taken from the glycerol stock -80°C and grown in sterile Hutner medium in anaerobic bottles (2ml) at 28°C for four days. Then, these starter cultures were inoculated 1 L sterilized Hutner medium from the bottle cap by the aid of sterile shringe (4ml per 1L) and left for two days to be grown at 28°C. Cells were collected at O.D 650 nm of 2.0 at late log phase [173].

Ingredients	Mass / Volume
Nitrilotriacetic	0.1 g
K_2HPO_4 / KH_2PO_4 (pH=7, 1mol)	20 ml
$NH_4(C_2H_4O(COO)_2)$ (pH=7, 1mol)	20 ml
$Mg_2SO_47H_2O$	0.29 g
$CaCl_2$	25 mg
$(NH_4)_6Mo_7O$ $4H_2O$	0.092 mg
$FeSO_4$ $7H_2O$	1 mg
EDTA	62.5 mg
$ZnSO_4$ $7H_2O$	275 mg
$MnSO_4$ H_2O	38.5 mg
$CuSO_4$ $5H_2O$	9.75 mg
$Na_2B_4O_7$ $10H_2O$	4.5 mg
$Co(NO_3)_2$ $6H_2O$	6.25 mg
H_2SO_4(6N)	37.5 ul
Biotine	0.04 mg
$NaHCO_3$	1 mg
Vitamine B_3	2 mg
Thiamine hydrochloride	1 mg
4-aminobenzoic acid	2 mg
Yeast Extract	1 mg
Sodium Glutamate	1.2 mg
Sodium Acetate	1.4 mg
H_2O	Final volume = 1 L

Table 4.4. *The components of Hutner's Media [173].*

Lysis Buffer
10 mM Tris-HCl pH=8
50 mM NaCl
0.6 mg/ml Lysozyme (Sigma, USA)
1 mM EDTA
400 ug/ml DNAse (Roche, USA)
20 ug/ml RNAse (Roche, USA)
1mM PMSF
10 mM MgCl$_2$

Table 4.5. *Lysis Buffer.*

4.2.2. Harvesting bacteria

The collected bacterial cells were centrifuged at 4500 rpm with F10S rotor for 20 minutes and the supernatant was removed. Then, pellets were washed 40 ml of PBS 1X pH=7 to remove media and metabolites. The washed pellets were resuspended in an equal volume (5 ml/g) of lysis buffer (Table 4.5) without enzymes. After the cells were homogenized in the lysis buffer for 20 minutes on ice, enzymes from lysis buffer recipe were added and cells were incubated at 4°C to allow the enzymes to work for 40 minutes. Then, the cells were broken by the French press (Thermo Electron Corporation, USA) in three passages at 10000 psi. The cells were always in the ice during harvesting of bacterial cells. After passage through the French press, unbroken cells were removed by centrifugation at 4000 rpm for 10 min. The supernatants were collected in a clean tube [174].

4.2.3. Isolation of membrane by continuous sucrose gradient

To separate intra cellular membrane, 10% and 70% of sucrose solution were prepared in 10 mM Tris-HCl pH=8, 50 mM NaCl. A continuous sucrose gradient was prepared by mixing low- and high-density sucrose solutions with Watson Marlow mixer (Hoefer SG100 Gradient Marker, Amersham Biosciences, Buckinghamshire, UK). 33 ml of 70% and 15% of sucrose solutions were added to separate tubes of Watson Marlow mixer. Firstly, 3 ml of 70% sucrose solution was added to the SW28 tube from the outlet. Secondly, the lighter solution was added to the tube nearer the outlet and let a little of this into the outlet tube. Then, a small magnet was placed into the 70% sucrose solution, Watson Marlow mixers were placed onto the stirrer and turned it on to create continuous sucrose gradient. To transfer the sucrose solution to the plastic SW28 tube, peristaltic pumps were used (Watson Marlow 205 U, UK). The valves of the mixer were opened, and the solution was pumped out into a plastic SW28 tube in an ice bucket. The tube is placed at an angle, so the solution does not drip but runs into the tube. The lysate from *Rhodobacter (Rb.) sphaeroides* were gently layered on top of the gradient by plastic Pasteur pipette. Each tube was filled with 20 ml of gradient solution and 4 ml of cleared cell lysate. Loaded gradients were balanced accurately with the buffer of 10 mM Tris-HCl pH=8, 50 mM NaCl and placed into the SW28 buckets and then SW28 rotor (Beckman Coulter, USA. They were centrifuged at 24000 rpm for 4 hours.

At the end of 4 hours, the intracellular membrane band was carefully collected with a glass pipet and put into Ti45 tubes (see Figure 4.1). Since the ICM was in the sucrose solution, it was diluted 4 times with a buffer (10 mM Tris-HCl pH 8, 50 mM NaCl) and pelted by the centrifugation at 40000 rpm for 90 minutes. The supernatant was removed, and the pellet was collected by the help of a spatula into a Dounce homogenizer. It is resuspended with the 7 ml buffer of 10 mM Tris-HCl pH=8, 50 mM NaCl and transferred to the dark 15 ml falcon tube [174].

Figure 4.4. *The Intra Cellular Membrane (ICM) of Rhodobacter sphaeroides.*

4.2.4. Protein purification

The protein purification is crucial for the characterization of the function, structure The protein purification is crucial for the characterization of the function, structure, and the biomolecular interactions of the protein of interest. Since LH2 and core complex are both embedded in the membrane, after the membrane isolation, it requires good separation of LH2 from core complex and other proteins.

Solubilization

To solubilize protein, 4% of dodecyl-maltoside (Anatrace, U.S.A) was added to the membrane suspension and the tube was placed on a wheel mixer and solubilized overnight at 4°C. After solubilization, insolubilized material was separated by centrifugation at 40000 with Ti90 rotor (Beckman Coulter, USA) for 90 minutes. Then, the supernatant was immediately used for protein isolation [174].

Protein complex purification

The separation LH2 and reaction core complex, which are the two main proteins of purple photosynthetic bacterial membrane, was complicated by using anion exchange and gel filtration chromatography and therefore discontinuous sucrose gradients have been found to offer better separation of these two proteins [175].

The sucrose solutions of discontinuous gradient were prepared as (from top to bottom) layers of 0.2 M (4ml), 0.4 M (5 ml) ,0,8 M (7 ml) and 1 M (4 ml) sucrose in 10 mM Tris-HCl pH=8, 50 mM NaCl, 0,1% DDM [176–178]. Then, solubilized proteins were carefully layered on top of the 0.2 M sucrose solution in the tube Ti70 (Beckman Coulter, USA). Gradients were centrifuged in a Ti 70 rotor (Beckman Coulter, USA) for 16 hours at 30000 rpm. The principle of the discontinuous density separation relies on the density of proteins. After the separation, LH2 is found as a dark band in the top of the fractions of the tube, while the core complex is found in the second dark band below the LH2 nearer the bottom of the tube. (See Figure 4.2). Those bands fractionated as 500 ul to be sure not load core complex to the purification column. These fractions were analyzed by a UV-visible spectroscopy since the core complex absorbs the UV light at between 875 nm-900nm, while the LH2 has strong absorption bands at 800 and 850nm.
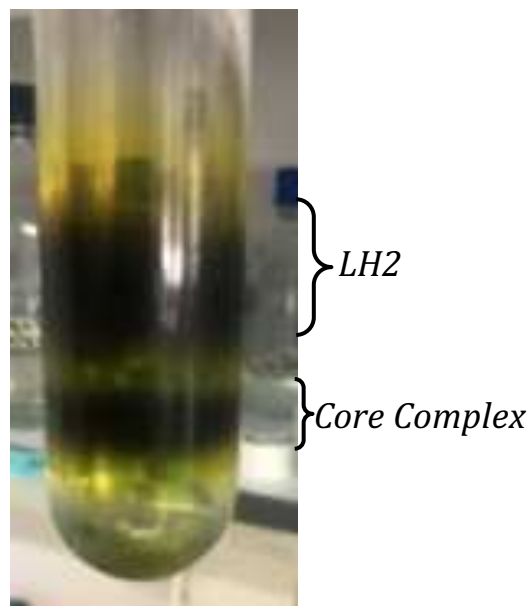
*Figure 4.5. The separation is based on gradient solution ultracentrifugation. **A** The extraction was layered carefully on the top of the gradient and centrifuged 16 hours at 4°C **B** The results of sucrose gradient centrifugation.*

There are several simple methods to confirm the quality and quantity of the proteins such as SDS-PAGE, Native-PAGE, HPLC and UV spectrophotometer [179]. I have used HPLC, a UV-visible spectrophotometer.

4.2.5. Quality control of protein

<u>Chromatography</u>

After sucrose density gradient purification LH2 proteins were required for further purification as they were contaminated with other proteins and ribosomes which led to the shift of the peak at 280 nm to 260nm - 270 nm. It was convenient to purify LH2 with a Resource Q (1 ml) anion exchange column (Cytiva) [177]. The principle of the Resource Q column relies on charge-charge interaction. The elution of the protein was achieved by the competition between Cl- ions of NaCl. When the protein molecules were introduced to the column, Cl- molecules were replaced by the protein molecules. To elute the protein from the column, high concentration of NaCl was introduced and Cl- replaced the protein molecules.

<u>UV visible spectrophotometer</u>

The purified LH2 were diluted 200 times and assessed by UV visible spectrophotometer (UV-1800 Shimadzu, USA). The ratio A280 / A850, was used to measure the level of purification and it was 0.33. In this ratio, while the amino acids which are tryptophan and tyrosine absorb the UV light at 280 nm, bacteriochlorophyll (Bchl) absorbs the UV light strongly at 800 nm at 850 nm. The Visible UV absorption between 400 nm and 532 nm represents the carotenoid molecules. Bchl and carotenoids are known as a fingerprint of LH2 [177].
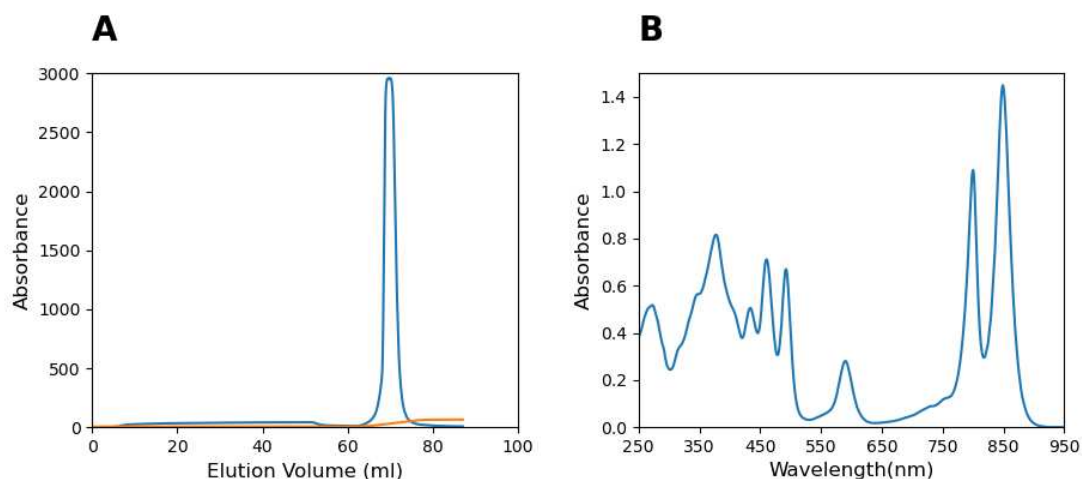
Figure 4.6. *Purification of LH2 from Rhodobacter sphaeroides. **A** The purification of LH2 from Rhodobacter sphaeroides by Resource Q anion exchange chromatography. **B** Absorbance spectra of pure LH2 from Rhodobacter sphaeroides.*

As the last part of LH2 from *Rhodobacter sphaeroides* purification, I have been using the Resource Q anion exchange column since I separated LH2 firstly from the other component of the cells and then from the core complex in the intra cellular membrane complex. In the purification process, LH2 was eluted at 275 mM NaCl and the peak was not sharp because it had a high signal due to the quantity of protein that saturated the detector. I have obtained one peak which has signal around 3000 at 280 nm. This peak showed that the purification process for LH2 was successful, and I obtained enough quantity of pure LH2 (Figure 4.6.A).

To confirm that LH2 is separated from core complex and other proteins at 280 nm, I have assessed LH2 with a visible UV spectrophotometer. I observed that the bacteriochlorophyll peaks on the right at 850 nm and not at 850 nm-900 nm since the core complex is absorbed between those wavelengths. This showed that LH2 was separated well from the core complex. For the separation of LH2 from other proteins at 280 nm, the ratio of A280/A850 was less than 0.5 (Figure 4.6.B). Thereby, pure LH2 was obtained for further experiments.

4.2.6. Coupling reaction

All fractions from the purification of LH2 from *Rhodobacter sphaeroides GA* by Resource Q were pooled down and dialysis against coupling reaction buffer 100 mM NaHCO$_3$ for 3 hours at 4°C through the 20kDa (-A-Lyzer MINI Dialysis Device, 0.5 ml, Thermo Scientific, USA) dialyzer. 10 nmol of LH2 from *Rhodobacter sphaeroides* was in 100 mM NaHCO$_3$ pH=8 was taken and concentrated to 200 uM using a 10 kDa concentrator (Amicon Ultra 4 Centrifugal Units Merck Milliporebrand, Germany). 80 nmol of DBCO NHS ester (Jena Biosciences, Germany) in DMSO was added to the protein and incubated for 2 hours at 25°C. After incubation, the reaction was stopped by addition of 100 mM Tris pH:8 and the mixture were incubated a further 15 minutes at room temperature. Excess DBCO NHS ester was removed by dialysis with a 20 kDa concentrator (Slide-A-Lyzer MINI Dialysis Device, 0.5 ml, Thermo Scientific, USA). The solution DBCO coupled protein was assessed by a UV-visible spectrophotometer to establish the labeling yield. Then, 5 nmol of the desired azide modified oligonucleotide was added to the mixture and incubated over night at room temperature. Reaction products were separated on a 1 ml Resource Q anion exchange column (Cytiva). The column was equilibrated and washed with 50 mM Tris,10 mM NaCl pH 8.0 and then products eluted with a certain gradient. The various products were collected in fractions (0.5 ml) and characterized by UV-visible spectrophotometry. If necessary purified and pooled fractions were concentrated using a 10kDa concentrator. Same procedure has been done with the linkers which are DBCO PEG$_4$ NHS ester and Maleimide PEG$_2$ NHS ester as described in chapter 5.

4.3. Methods for isothermal assembly of DNA origami

4.3.1. Preparation of rectangular DNA origami

<u>Isothermal Folding Reaction of scaffold and staples</u>

2 pmol of p7249 scaffold was incubated with 10 pmol of staples (staples) (purchased from Tilibit nanosystem GMBH, Germany) in TANa 1 X Buffer during 96 h at Eppendorf shaker. At different times (1 min, 5 min, 10 min, 30 min, 1h, 2h, 5h, 10h, 24h, 48h, 72h and 96h),15 ul of the rectangular origami was taken and frozen in liquid nitrogen to be loaded on 1.5% agarose gel.

4.3.2. Agarose gel preparation

For 1,5% agarose gel in TBEMg 1X, 2.25 g agarose (Euromedex, France) was dissolved in TBE 1X with 11 mM $MgCl_2$. Then it was heated for 3 minutes in the microwave until it was dissolved. During heating by microwave, the agarose solution should be mixed by stopping the microwave. The time for heating up is crucial because it might change the concentration of the agarose gel and it should not pass more than 3 minutes. The gel was poured into a rectangular 10 x 8 cm mold with 17 combs placed on the top of it and left to cool down for 2 hours in the cold room. The gel was inserted into the electrophoresis cell 30 minutes before the beginning of the experiment. The electrophoresis cell was placed 1 hour before the run into the ice bath. Then 500 ml of TBE1X were poured to the electrophoresis cell and the gel placed within cassettes and the migration was carried out at 70 V during two and a half hours in the cold room [25]. Then, the gel was stained in a gel red solution for 20 minutes. The gel image was captured by Baby Imager.

4.3.3. Negative staining

5 µl drops of the sample were placed directly on glow discharged carbon coated grids (EMS) for 3 minutes. The grids were then washed with two drops of 2% aqueous uranyl acetate and stained with a third drop for 1 min. Grids were dried on filter paper and the samples were analyzed using a Tecnai 200 KV electron microscope (FEI), and digital acquisitions were made with a numeric camera (Oneview, Gatan).

Chapter 5 - Coupling Protein-DNA and assembling proteins

DNA directed assembly of proteins represents an exciting prospect for the construction of new active meta-materials. In this context, my objective was to characterize a method for coupling protein to DNA which is applicable to homo-multimeric proteins for producing a 1:1 adduct and then assembling formed Protein-DNA adducts. To achieve this objective, a two-step procedure will be required.

For the first phase, coupling protein-DNA, I decided to use bi-functional reagents, which contains amine reactive groups to create links between surface amine residues on the protein and a reactive group attached to the DNA. The method I chose allows full control over this formation and can be favorable for both monomeric and multimeric proteins. Furthermore, it has few limitations among the methods for manipulation of proteins. While the other methods such as site-directed single side chain modifications are limited to monomeric proteins, this chemistry reaction is not limited by complexity of the proteins, and it can be applicable either for multimeric proteins. Besides, it also provides controllable junction of complex molecules in complex environments.

For the second phase, I aimed to assemble the previously formed protein-DNA adducts and to characterize this assembly by absorption spectroscopy, fluorescence measurements and gel electrophoresis. Furthermore, I wanted to demonstrate that the DNA strands in such structures can be still available for disassembly, either through enzyme digestion or toe-hold strand displacement. The above mentioned results in a procedure that allows controlled formation of 1:1 protein-DNA adducts, which are accessible for the formation of DNA scaffolded nanostructures.

However, to maximize the yield of the process, first for coupling protein-DNA and afterwards for assembling of protein-DNA adducts, I started by selecting the most suitable reactants and conditions, as it will be described and discussed in section 5.1.

5.1. Coupling Protein-DNA: Design of reactants and conditions for the chemistry

I designed and performed various experimental studies trying to investigate the most favorable reactants and conditions for the reaction, involving labeling proteins with a bifunctional reagent, to provide the best results possible. The chemistry reaction is composed of two steps. In the first step, a bifunctional reagent reacts with primary amines on the protein in a second step, the DBCO labeled protein is attached to the azide-labeled DNA molecules, through the Cu-free cyclo-addition of DBCO and the azide groups on the DNA.

After this brief introduction, in the following paragraphs, I will describe more in detail the selection process highlighting the main reasons behind the choices.

5.1.1. Selection of reactants

<u>Selection of proteins</u>

Selection of proteins is one of the main factors that can affect the results since the reaction between bifunctional reagents and proteins are influenced by the number and location of the lysines. In this context, I decided to use fluorescent proteins because it allows to easily follow reaction results in all the steps of the procedure.

Among the numerous fluorescent proteins, I wanted to choose two proteins, which can act as donor and acceptor to study the formation of complexes via FRET. Criteria for the selection of proteins included the number of lysines and the brightness of proteins. Thus, mCherry, sfGFP and mRuby3 were the best options.

Additionally, after developing this chemistry with model proteins, I wanted to examine the behavior of the chemistry with multimeric proteins as Light Harvesting Complex 2 (LH2) and Core Complex (LH1-Reaction Center) from purple photosynthetic bacteria since they are the proteins of an interesting system which is included in purple photosynthetic bacteria.

mCherry
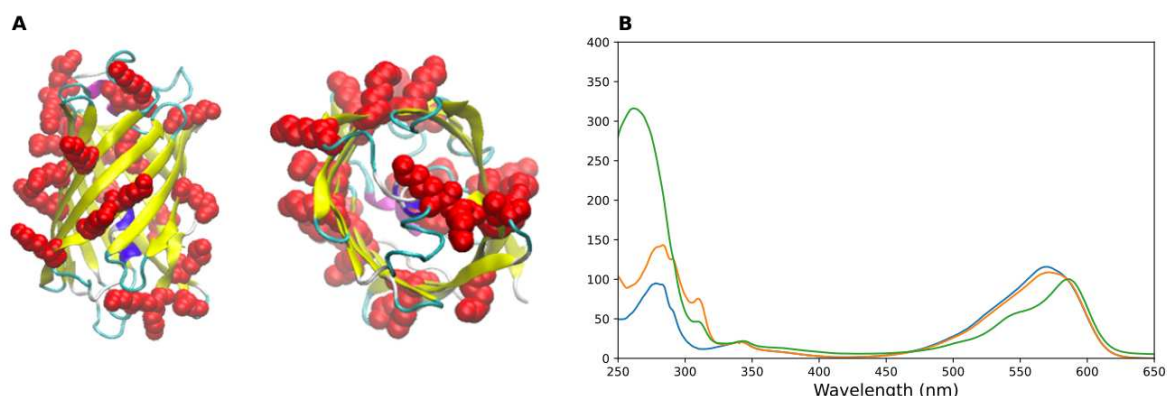
Figure 5.1. *Structure of mCherry was shown and a spectroscopic following of the reactions performed with mCherry, DBCO NHS ester and azide modified DNA.* ***A*** *Accessible lysines of mCherry are shown in red (both side and top view).* ***B*** *Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing mCherry protein (blue), mCherry reacted with DBCO NHS ester (orange); and both non-attached, attached DNA (green).*

mCherry comes from the red fluorescent protein family and contains 3 alpha helices and 13 beta sheets which make up the beta barrel shape and it has 24 lysines which are accessible as shown in Figure 5.1.A. Since it is known that the position of charged groups like lysines is likely cause the significant but limited spectral perturbations for mCherry. Thus, I decided to check whether any spectral perturbation would be observed. Hence, the first reaction of the chemistry was performed with amine reactive reagent DBCO NHS ester (summarized in selection of bifunctional reagent) under the conditions recommended (provided by Jenabiosciences). In the normalized absorption spectra, figure 5.1.B, the absorption spectrum of pure mCherry showed a peak at 575 nm (blue line), and a peak at 280 nm as expected. After reaction with DBCO NHS ester, and removal of excess reagent, the absorption spectrum (orange line) showed an additional peak at 310 nm. I observed a red-shift in the peak of mCherry-DBCO NHS ester (orange line) adduct from 575 nm to 580 nm. After the second reaction of the chemistry with azide-labeled DNA through Cu-free cyclo-addition and the absorption peak was at 260 nm as seen in the spectrum of the protein-DNA adduct (green line). The red-shift was also observed in the peak of mCherry-DNA (green line) adduct from 575 nm to 595 nm and change in shape. Due to the movement of charged lysines are proposed to modify the chromophore electron-density distribution, inducing

the red shift caused a spectral perturbation [180]. There is a possibility that this phenomenon affects the yield of chemistry reaction since it is complicated to predict properties of the mCherry and how it will behave in the mixture for the chemistry.
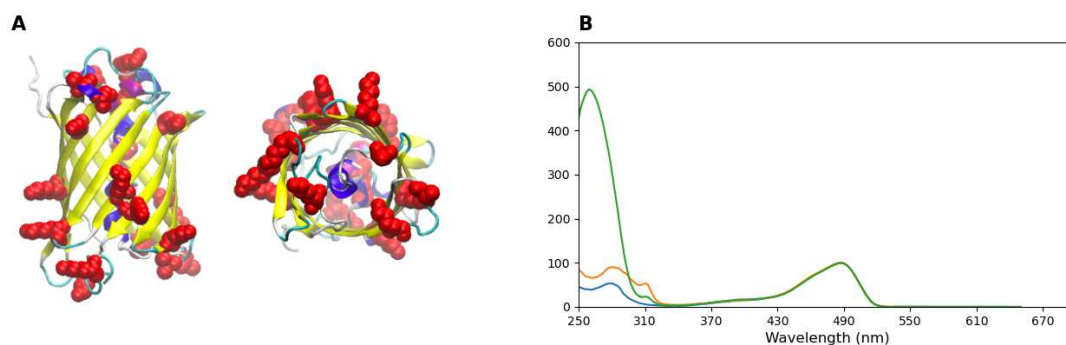
mRuby3
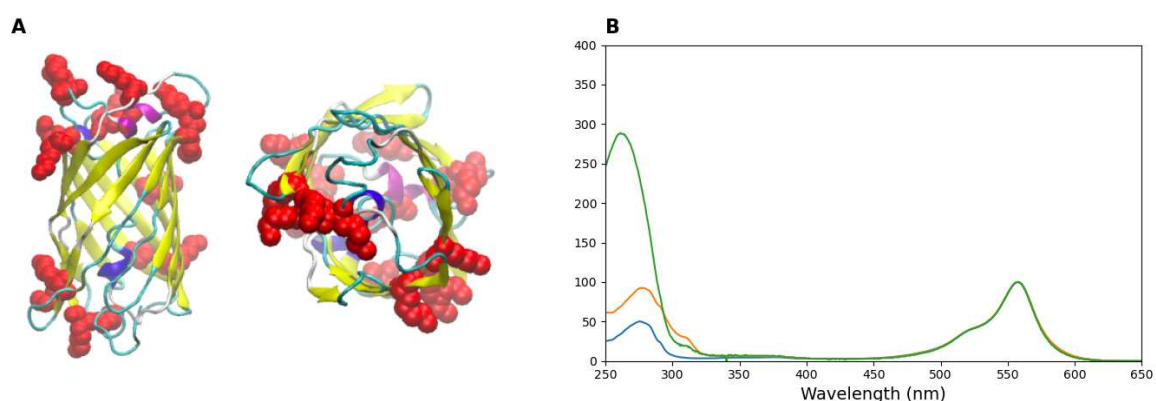
Figure 5.2. *Structure and accessible lysines of mRuby3 were shown. **A** Accessible lysines of mRuby3 were shown in red (both side and top view). **B** Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing mRuby3 protein (blue), mRuby3 reacted with DBCO NHS ester (orange); and DNA (green).*

mRuby3 is shaped as a barrel and it has 17 lysines which are accessible as shown in figure 5.2. mRuby3 is a variant protein originated from Ruby2 by the aid of crystal structure of mRuby. Ruby2 is known as a favorable acceptor fluorophore in GFP-RFP FRET due to its high quantum yield, large stroke shift, fast maturation, and high photostability. Jun Chu and his colleagues designed mRuby3 with higher brightness, to improve its performance as a FRET acceptor by engineering of mRuby2 through performing mutagenesis in a combinatorial and simultaneous manner at several locations. They concluded that mRuby3 improves photostability by 200% over the previous generation of fluorophores. Notably, mRuby3 is also 35% brighter than mRuby2, making it both the brightest and most photostable monomeric red fluorescent protein yet characterized. Compared to mCherry, lysines of mRuby3 are only distributed on N-terminal and C-terminal not on center of the Beta barrel. This feature makes mRuby3 more suitable than the mCherry since it provides controlled orientation for the protein-DNA coupling. As it is presented in figure 5.2.B, I did not observe any change in shape in absorption spectra of mRuby3. Despite it being reported that there

can be slight (3nm to 6nm) shift in absorption of mRuby3 in cells, this shift did not observe in purified systems due to lysine reaction [181].

sfGFP

*Figure 5.3. Structure and accessible lysines of sfGFP were shown in red (both side and top view) **A** Accessible lysines of sfGFP were shown in red (both side and top view) **B** Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing sfGFP protein (blue), sfGFP reacted with DBCO NHS ester (orange); and DNA (green).*

sfGFP is a robustly folded version of GFP which was discovered in the jellyfish *Aequorea victoria*, engineered by Geoffrey S Waldo and his colleagues since the existing variants of green fluorescent protein (GFP) often misfold when expressed as fusions with other proteins. Superfolder GFP folds well even when fused to poorly folded polypeptides. Another property of this variant is that sfGFP fusions were considerably brighter than folding reporter GFP fusions [171]. The structure of sfGFP and accessible lysines of sfGFP were shown in figure 5.3.A. sfGFP is composed of 11 β-strands from a cylindrical β-barrel with the α-helix harboring the chromophore spanning straight through the barrel center [182]. Despite engineering of sfGFP led some conformational changes of amino acid residues, it did not cause a shift in the absorption as it is shown in figure 5.3.B. However, it is reported that red shift can be observed depending on ionic agent [183]. Since this parameter is controllable, sfGFP is a good candidate for the chemistry.
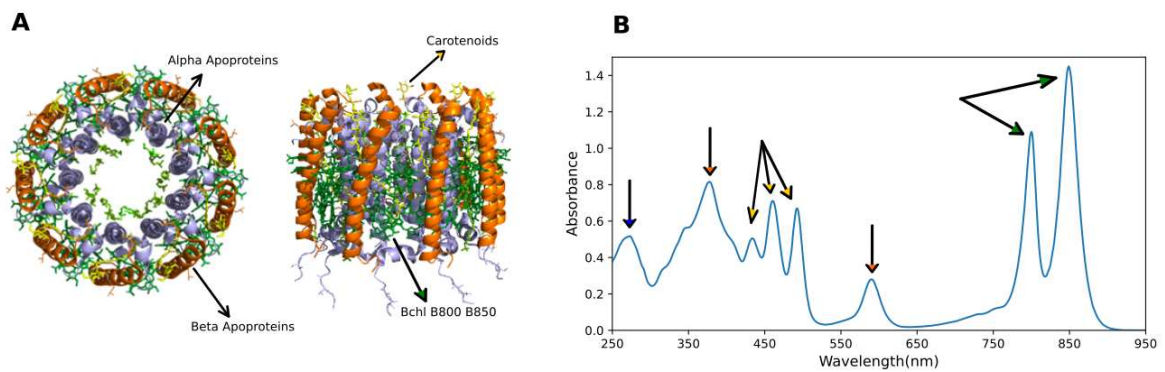
Figure 5.4. *Structure of LH2 and its subunits. **A** Structure of LH2. **B** Absorbance spectra of pure LH2 from Rhodobacter Sphaeroides. (280 nm-Other Proteins-LH2 Bx By, 376 nm-Soret Transition of Bacteriochlorophyll, 400 nm-532 nm-Carotenoids, 590 nm-Qx Transition of BChl, 800-850 nm BChl -Qy)*

Multimeric structure of LH2 was shown in Figure 5.4.A. LH2 (130 kDa) is shaped like a barrel and composed of two rings as inner and outer rings. While the inner ring is composed of alpha apoproteins, beta apoproteins are the subunits of the outer ring. It is a nonamer of dimers with a total of 18 polypeptides. Each subunit of LH2 is formed of 1 apoproteins alpha and beta apoproteins and their associated 3 BChl and 1 Carotenoids. Bchl and carotenoids are known as a fingerprint of LH2, and they make LH2 pigmented. It has 27 of BChl and while 18 of them are tightly coupled rings, 9 of them are larger looser rings. 2 rings absorb the light at different wavelengths 800 nm and 850 nm [184]. Also, absorbance spectra of pure LH2 from *Rhodobacter sphaeroides* was shown in figure 5.4.B. Bx and By of LH2 (blue), soret transition of bacteriochlorophyll (orange), carotenoids (yellow), Qx Transition of BChl, BChl-Qy (green) were shown with the colored arrows same as in the legend.
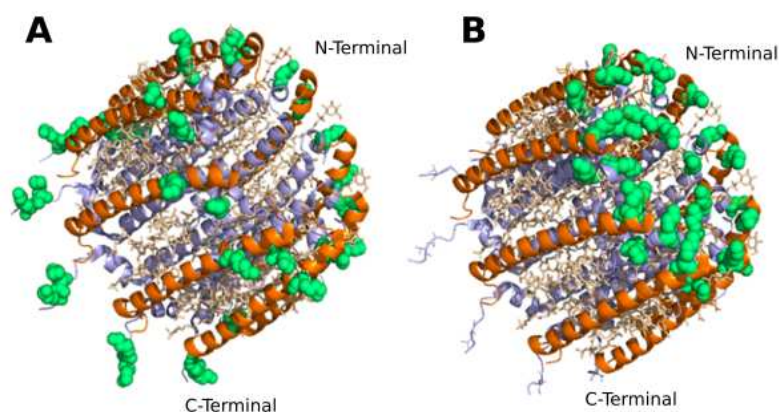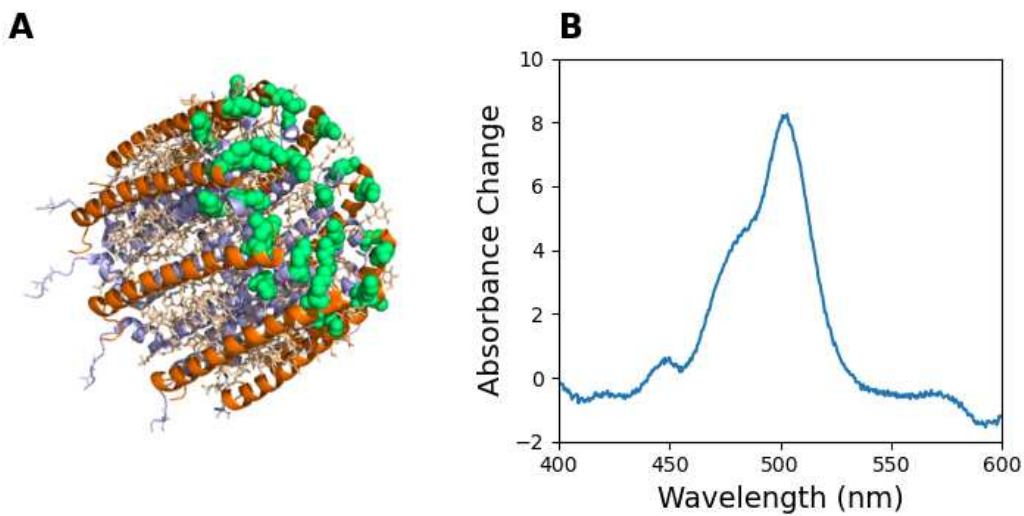
*Figure 5.5. Accessible lysines are shown in green. **A** Accessible lysines of LH2 from Rhodopseudomonas acidophila at N and C-Terminal. **B** Accessible lysines of LH2 from Rhodobacter sphaeroides at N-Terminal.*

The first reaction between protein and linker relies on the interaction between amines of the side chain of the lysine amino acid of the LH2. That is why, accessibility of lysine of LH2 is crucial for the reasonability of this reaction. Since the purple photosynthetic bacteria have different species, I had a chance to select the most reasonable one for the accessibility of the lysines. Different species of purple photosynthetic bacteria have different numbers of lysines and different distributions.

As examples, *Rhodopseudomonas acidophila* have both lysines located on N-terminal and C-terminal (Figure 5.5.A). Nevertheless, the lysines of Rhodobacter spaheroides are located only in the N-terminal of the protein (Figure 5.5.B)*. Therefore, LH2 from *Rhodobacter spaheroides* seemed more suitable than from *Rhodopseudomonas acidophila* for reacting lysine of LH2 with DBCO NHS ester.

However, the multimeric structure of proteins makes the labeling more complicated compared to the homomeric ones. That is why, I wanted to figure out how many of them are accessible among 27 lysines from *Rhodobacter sphaeroides* located on alpha and beta apoproteins, even though it was shown that lysines of LH2 from *Rhodobacter spaheroides* in Figure 5.5. For this purpose, I decided to label lysine residues of LH2 since NHS-Fluorescein can be followed by the visible-UV spectrophotometer and NHS ester is part of the chemistry.

Figure 5.6. *Lysines of LH2 (green) from Rhodobacter sphaeroides are labeled through NHS-Fluorescein.* **A** *Accessible lysine of LH2 from Rhodobacter sphaeroides (green).* **B** *Absorption of attached NHS Fluorescein to LH2 was shown. It was obtained through subtraction of pure LH2 from LH2 reacted with NHS fluorescein.*

Labeling lysine residues of LH2 by the NHS-Fluorescein molecules was done to figure out how much lysine is accessible at the N-terminal of LH2 from *Rhodobacter sphaeroides* (Figure 5.6.A). Attached NHS-Fluorescein showed a peak at between 494 -518 nm in figure 5.6.B. This absorption was obtained through subtraction of pure LH2 from LH2 reacted with NHS fluorescein LH2 to show the attached NHS fluorescein better since not only NHS fluorescein but also LH2 absorbs the visible UV light at 450-500 nm. After the incubation of 15 nmol of NHS-Fluorescein with 3 nmol pure LH2, then the excess NHS-Fluorescein was removed. At the end of the reaction, 9 nmol of NHS Fluorescein attached to lysine of LH2.

$$\frac{A\,(493) - A\,(850)\,/\,2.17}{70{,}000\ \text{M}^{-1}\text{cm}^{-1}}$$

At the end of the reaction, the number of NHS-Fluorescein which are attached to the LH2 protein was calculated by using absorbance and extinction coefficients of LH2 and NHS-Fluorescein. There are 81 of lysines in the nonameric LH2 structure and only 9 lysines were labeled by NHS Fluorescein, under these conditions.

Core Complex
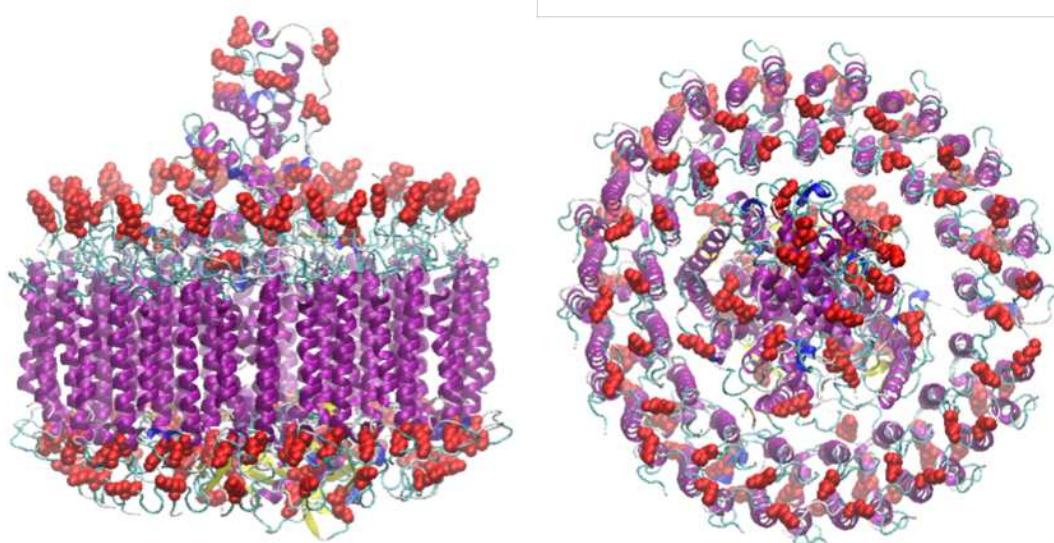
Figure 5.7. *Accessible lysines of LH2 from Blastochloris viridis were shown in red **A** Side view. **B** Top view.*

Core complex is a multimeric protein from the photosynthetic unit of purple photosynthetic bacteria. In the photosynthetic unit, there is resonance energy transfer between LH2 and core complex. Core complex is composed of both LH1 and the reaction center. The LH1 complex surrounds the RC to form a closed elliptical ring. The triple-ring LH1 complex comprises a circular array of 17 β-polypeptides sandwiched between 17 α- and 16 γ-polypeptides [185]. Since the reaction center is the partner of LH2 to perform resonance energy transfer I thought that it can be favorable to consider core complex as an option. Thus, I looked at the accessibility of lysines and its distribution. Lysine distribution varies from species to species as in LH2. While its lysines are accessible, as orientation they are located both at N- terminal and C terminal in *Blastochloris viridis* (see figure 5.7) since the structure has both reaction center and LH1 in all the species of core complex. Like *Blastochloris viridis*, accessible lysines are located on both N and C terminal, in *Phsp. molichenium*. However, the location of accessible lysines is different in *Roseobacter denitrificans* as only located on N-terminal of alpha-apoproteins.

Therefore, core complex is favorable as another multimeric protein for this chemistry since it has accessible lysines. Among species, *Roseobacter denitrificans* seems most suitable one because majority of accessible lysine are located at terminal N-terminal.

In conclusion, as a first pair of proteins, I selected mRuby3 and sfGFP, they have 20 and 17 accessible lysines respectively. Brightness and stability of mRuby3 was sufficient for the further fluorescent measurements that I aim. Even though sfGFP does not differ much about brightness from the other variant of GFP proteins, it overcame the folding limitation of the GFP family. Additionally, no change in the spectrum was observed during the reaction for both sfGFP and mRuby3. As multimeric proteins, I decided to choose LH2 in photosynthetic unit of purple photosynthetic bacteria. It is not only because its system is interesting but also the number (27) and location of lysines was suitable. Additionally, I predict that this chemistry may be applicable later for core complex since it is better to try other multimeric proteins after observing the behavior of this chemistry on LH2.

Bifunctional reagent and modification of DNA

Besides selection of most suitable proteins for this chemistry, conditions of the reaction are one of the important factors that strongly affect the yield of the reaction. In this selection, my priorities were having an amine reactive group, specificity for attaching modified DNA and not extreme conditions since I wanted to attach a bifunctional reagent to the proteins by amine of lysines and the specificity of linker will help to provide a control over the reaction. In this manner, I selected di-benzyl cyclo-octyne (DBCO)-N-hydroxysuccinimide ester, di-benzyl cyclo-octyne (DBCO) PEG$_4$ N-hydroxysuccinimide ester, Maleimide PEG2 N-hydroxysuccinimide ester as options. While the common point of them was having NHS ester to attach through the amines groups of proteins, the other part of the bifunctional reagent was the part which attach DNA via reactive group. The selection of reactive group of DNA is guided by the bifunctional reagents. For example, while this reactive group is azide for DBCO NHS ester and DBCO PEG$_4$ NHS ester, for Maleimide PEG$_2$ NHS ester is thiol. Each of these three linkers were commercially available hence I chose these three linkers as options. However, stability of DBCO is higher than maleimide in aqueous solutions and UV absorption of DBCO allows following the reaction by UV-visible absorption spectroscopy unlike the maleimide. Additionally, di-benzyl cyclo-octyne (DBCO) reacts only with azide, even in presence of -NH2, -SH, -COOH and this DBCO-azide conjugation has long term stability. All the reasons mentioned above led us to choose DBCO NHS ester as first.

5.1.2. Selection of conditions

In this section, I discussed the effect of the reaction conditions on the reaction efficiency to select the most favorable reaction conditions. It is important to choose most favorable conditions because this chemistry strongly pH dependent and I can keep the control on the reaction through this way. The most important parameters in determining the yield in the reaction, involving coupling amines via NHS ester, are protein concentration, protein-NHS ester ratio and to a lesser degree buffer conditions and temperature.

Protein Concentration
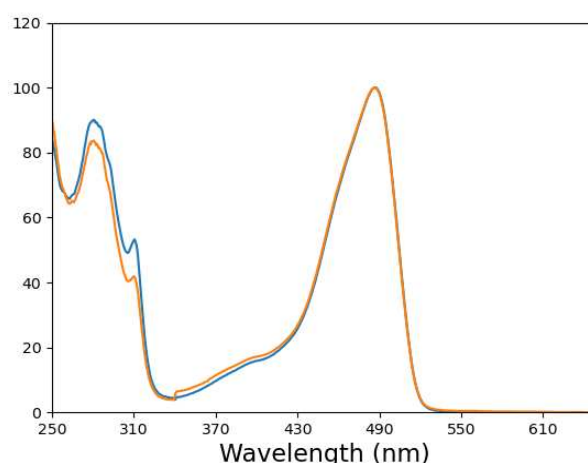
Figure 5.8. *Absorption spectra of different conditions shown are the normalized absorption spectra of samples containing 200 uM sfGFP protein reacted with DBCO NHS ester (blue),60 uM sfGFP protein reacted with DBCO NHS ester (orange) at room temperature for 2 hours.*

Protein concentration is critical for the reaction efficiency since DBCO NHS ester also can attach water molecules. This creates a competition between proteins and water molecules to attach DBCO NHS ester. The best way to avoid this competition is to use concentrated protein as much as possible. However, while increasing protein concentration increases the efficiency of labeling it is hard to increase protein concentration above a certain value. In the experiment, I used sfGFP concentration of 200 uM (blue line) and 60 uM (orange line) and kept constant the quantity of sfGFP, the ratio between sfGFP and DBCO NHS ester (1:4), temperature (25°C) and the incubation time (2 hours) to clearly observe the effect of sfGFP concentration. The

spectrum shows the comparison of sfGFP molecules reacting with DBCO NHS ester in the two different concentrations after normalization in figure 5.8. The percentage of DBCO NHS ester which are attached to sfGFP was calculated by using absorbance and extinction coefficients of sfGFP and DBCO NHS ester. While sfGFP concentration was 200 uM, 60% of DBCO NHS ester was attached to the sfGFP. However, while the sfGFP concentration was 60 uM, 38.3% of DBCO NHS ester linked to the sfGFP.

This result suggests that the yield of the reaction, involving protein-DBCO NHS ester, is affected by the concentration of sfGFP. For a high a yield, it is important to keep sfGFP concentrated as much as possible, otherwise the competition between water and DBCO NHS ester occurs during the reaction.

Temperature

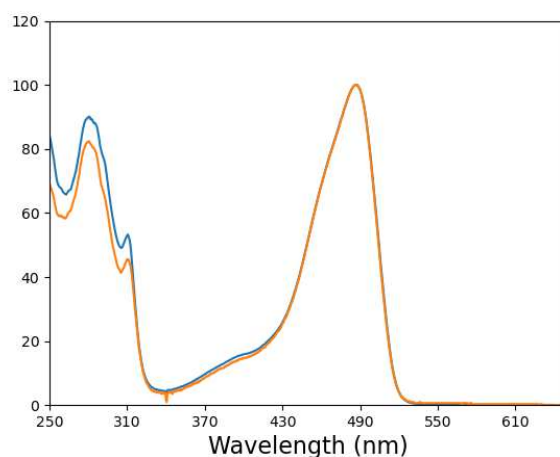
Figure 5.9. Absorption spectra of different conditions shown are the normalized absorption spectra of samples containing sfGFP protein reacted with DBCO NHS ester at room temperature protein (blue), sfGFP protein reacted with DBCO NHS ester on ice (orange) for 2 hours.

Temperature is another important factor that can impact the efficiency of the reaction and incubation time. I studied 25°C and 4°C to understand the impact of temperature on the reaction and kept constant the other parameters (described in figure 5.9). sfGFP reacted with DBCO NHS ester at both 25°C (blue) and 4°C (orange) and the results were normalized to compare the effect on the absorption of DBCO NHS ester at 310 nm in figure 5.9. The efficiency of the reaction was decreased by around 7%, when the incubation was on ice compared to the incubation of molecules at 25°C.

Temperature thus had relatively little effect on the efficiency of this reaction. However, using low temperature as < 4°C cause the increasing of incubation time, thus makes the reaction more time consuming. Also, I observed that room temperature does not lead protein degradation. Therefore, 25°C is favorable for the reaction and does not result in any observable protein degradation.

Molar ratio among Protein-DBCO NHS ester
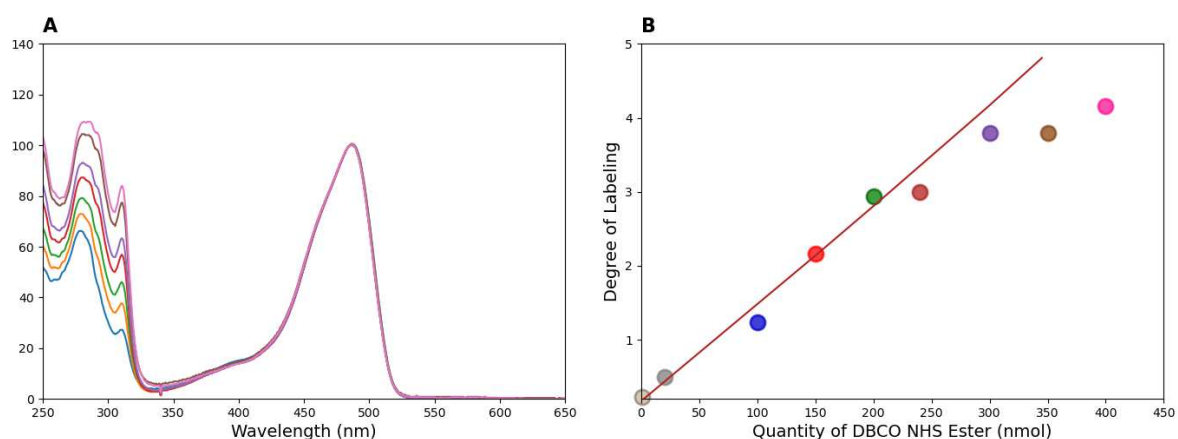
Figure 5.10. *Absorption spectra of sfGFP protein reacted with different quantities of DBCO NHS ester and degree of labeling are shown from 100 nmol to 400 nmol. (Colors in the table correspond to the colors in the graph).*

The molar ratio between protein and bifunctional reagents was significant factor that I can control to determine the number of DNA linked per protein since this can be partially directed by the number of bifunctional reagents. Hence, to determine the optimal quantity of DBCO NHS ester, I wanted to observe the correlation between the quantity of DBCO NHS ester and the degree of labeling of sfGFP. 60 nmol of sfGFP were reacted with different quantities of DBCO NHS ester, the results were presented as normalized spectra in figure 5.10.A. At 310 nm was raised by the increase in the quantity of DBCO NHS ester. Change in protein: DBCO-NHS ester ratio alters the degree of labeling in a predictable way, with a reaction efficiency approximately constant at 76% +/- 3% (N=8), though there is a slight decline above 250 nmol.

I show the correlation between the degree of labeling and the amount of DBCO NHS ester in figure 5.10.B. I can see that 20 nmol of DBCO NHS ester was not sufficient to

encounter with sfGFP molecules and react with them all, however in the presence of 100 nmol or 250 nmol of DBCO NHS ester, degree of labeling increased proportionally. This suggest that at least around 100 nmol of linker is required for the labeling reaction.

In the presence of 300 nmol, 350 and 400 nmol of DBCO NHS ester, the degree of labeling of sfGFP slightly reduced. This result pointed out that reaction started to slow down and number of available lysines decreased and thus they become less reactive after the breaking point which is 250 nmol of DBCO NHS ester.

Therefore, while 100 nmol is a starting point of labeling, 250 nmol is suggested here as a saturation point. In the presence of 250 nmol, 3 lysines reacted per protein, this degree of labeling is suitable for this reaction. About the quantity of sfGFP, I decided to set it as 60 nmol considering that there will be non-reacted sfGFP molecules. Therefore, I decided on a 1:4 (60 nmol: 240 nmol) ratio for the reaction of labeling via DBCO NHS ester.

<u>Reaction buffer</u>

Other crucial factors are the reaction buffer and pH to get a high yield for protein-DNA coupling since the reactions of molecules in this chemistry are highly pH dependent.
In the literature, among the reaction buffers, KH_2PO_4, PBS and $NaHCO_3$ are suggested as options for performing the coupling between amine groups of proteins and NHS ester molecules, while the pH should be in the range 7.5-9.0. Thus, for coupling amine groups of proteins via NHS ester molecules, three different trials were performed in $NaHCO_3$ pH8.0 and two phosphate buffers with different pH as KH_2PO_4 pH8.0 and PBS pH7.5. I obtained lower yields in KH_2PO_4, pH8 (50%) compared to $NaHCO_3$ pH8 (75%) and reducing the pH to 7.5 reduced the yield further (20% in phosphate buffer). It is possible that this reduction of yield in phosphate buffers might be caused by the higher ionic strength. Thus, $NaHCO_3$ pH8 was the most favorable buffer of those tested.

Reaction conditions for the chemistry

The DBCO-azide-DNA and lysine of protein-NHS ester reactions are strictly related, and the outcome of the chemistry reaction is strongly affected by the yield of the both DBCO and azide-DNA and lysine of protein-NHS ester reactions.

Selection of protein is indirectly important for the reaction, involving coupling of the DBCO labeled protein with azide-DNA, since to obtain Protein-DNA adduct, bifunctional reagent should attach both as protein and DNA. As bifunctional reagent, DBCO NHS ester was favorable for this chemistry since DBCO reacts only with azide even in the presence of -NH2, -SH, -COOH and the DBCO- azide conjugation has a long-term stability.

The molar ratio between labeled protein and azide-DNA is one of the factors that allows partially control of the average number of DNA molecules per protein. I observed from the experiments that 20 nmol of azide-DNA results in protein-DNA couples as ratio from 1 to 4 DNA per protein. From this result, I decided that 20 nmol is favorable for the reaction.

In the context of the reaction conditions, 4°C was time consuming since the binding between DBCO and azide is specific enough and therefore I decided that room temperature is favorable. As buffer, after I conducted the chemistry reaction in KH_2PO_4 and $NaHCO_3$ around 8.0, I decided that both buffers are favorable.

Reaction sequence

The last factor to analyze in this section is the decision of the reaction sequence. The main question to answer was which reaction to perform first between the two:

- Attaching Protein to DBCO NHS ester and then adding azide-DNA to the mixture.

- Attaching DNA to DBCO NHS ester and then adding protein to the mixture.

DBCO is more stable than NHS ester in aqueous solutions. In the second reaction sequence, the reaction efficiency was low since azide-DNA got dissolved in water and DBCO NHS ester was in dry DMSO. In this case, it is still possible to adjust the pH of the buffer by the addition of ions, however, since the pH8 is provided by the buffer of protein, it was easier to perform both reaction in the same buffer starting with the first

reaction protein-NHS ester and then following with the azide labeled DNA-DBCO reaction.

Hence, I can summarize by saying that the pH of the reaction, the composition of buffer, the protein concentration and the protein-NHS ester ratio are all factors that play an important role on the yield of the overall reaction for both proteins-DBCO NHS ester and the DBCO NHS ester-azide DNA steps. It is better to keep protein concentrated as high as possible, the buffer pH around 8.0-8.5 and avoid higher ionic strength in buffer composition. For the protein-DBCO NHS ester ratio, the competition between NHS ester and water molecules should not be ignored.

5.1.3. Selection of separation method for reaction products

Selection of the separation method was crucial since it allows us to separate all the reaction products. I decided to use an anion exchange column to take advantage of the negative charge of the DNA. The main parameters for the selection of the column were good resolution and efficiency. I needed to be sure that the molecules which contain different numbers of DNA are separating with a good resolution since I need only the 1:1 protein DNA adducts. In this direction, I decided that DNA affinity and ion exchange column were best options since both allow to selectively bind DNA.

<u>DNA affinity column</u>

DNA-affinity chromatography is potentially a powerful method with broad applicability and the principle of the DNA affinity chromatography is that it separates the biomolecules depending on interactions between biomolecules. This chromatography has been used for the purifications DNA binding proteins. There have been developments in coupling methods and hence applicability of this chromatography was improved [186]. I used a Hitrap NHS-activated column prepare a single stranded oligo-A column. This column was designed to separate non-reacted proteins and coupled protein-DNA. The NHS-activated support is composed of an N-hydroxy-succinimide (NHS) ester attached by epichlorohydrin to Sepharose High Performance via a 6-atom spacer arm. I introduced into the NHS-activated column the amino labeled oligo A DNA (composed of 12 Adenine nucleotides) and did the coupling reaction according

to the manufacturer's recommendations. By using a large excess of amine labeled DNA and a long incubation time, the efficiency of the first step reached up to 70%.

However, the separation of non-coupled and coupled protein-DNA molecules using this oligo-A column proved difficult. It was difficult to find conditions for the hybridization between coupled protein-DNA and amino labeled DNA. The important factors for this reaction were the concentration of Mg^{+2}, the temperature and optimization of these two conditions had relatively little effect on the efficiency of this reaction. Even though I carried out the reaction at both 4°C and 25°C in the presence of the optimal $Mg+^2$ concentration (12 mM), the efficiency of attaching coupled protein-DNA to the column remained as 27%. It is likely that hybridization between Thymine and Adenine tails did not occur sufficiently. Thus, while the hybridization reaction adds the possibility of sequence specific interactions, the low efficiency of DNA retention by the column under chromatographic conditions prevented by using this method of separation.

Resource Q column is based on quaternary ammonium cation to retain negatively charged molecules by coulombic interaction. After some initial trials testing the important parameters of the separation, I discovered that the separation of protein-DNA adducts with good resolution can be guided by selecting a compatible buffer, flow rate and elution gradient. About the buffer, as described in section 5.1.2, both 50 mM KH_2PO_4 pH=8 and 100 mM $NaHCO_3$ pH=8 was suitable. However, after performing the separation of protein-DNA adducts for both buffers, 100 mM $NaHCO_3$ proved to be more compatible with the Resource Q column to achieve a good resolution. Also, I observed that step elution was also suitable for the separation of unlabeled protein from the labeled ones while the gradient elution could separate the protein DNA adducts which contain different number of DNA molecules.

Lastly, the flow rate I chose of between 0.5 ml/min and 1 ml/min was better than faster flow rates as it allows the protein-DNA adducts more time for the accurate separation. Additionally, I found out a way to improve the resolution of the separation, by using only the monomeric proteins. Thus, it means that I am not completely sure that these factors might not improve the resolution for the separation adducts of multimeric proteins. As I mentioned at the beginning of section 5.1, characterizing protein-DNA

coupling with multimeric proteins is more difficult compared to the monomeric ones for all steps of the entire procedure, including separation.

<u>Separation reference for multimeric proteins</u>

It is more difficult to perform chemistry with the multimeric membrane proteins due to their structure and the fact that they are not water soluble. For this reason, detergent is necessary for each step of the reaction and for separation of products. Another challenge comes from the fact that some, possibly charged, lipid molecules co-purify with the protein in the middle of the inner ring. These lipids can have different charges on each purification which can reflect in a change in the charge of LH2 itself. Since the separation method relies on the negativity of LH2-DNA adducts, the variable charge of LH2 might create a challenge for the separation. Therefore, I decided to perform some experiments on pure LH2 and pure DNA to have some references for the separation products and avoid the charge confusion which comes from LH2 in the analysis.
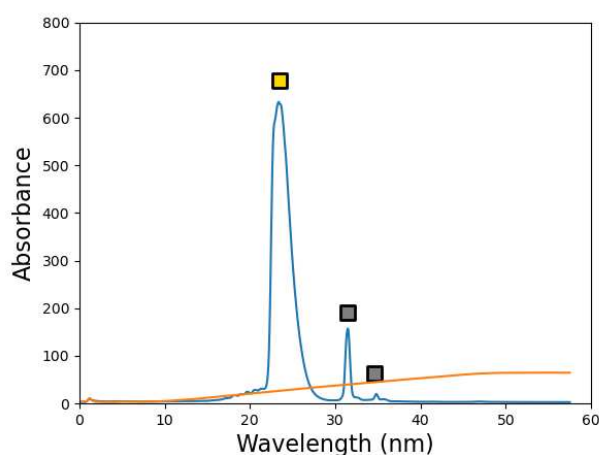
Figure 5.11. *Elution of Pure LH2 and LH2-DNAs without a linker using Resource Q anion exchange column. LH2 (yellow square) + DNA (gray square).*

In figure 5.11, I wanted to observe the chromatographic profile of LH2 and DNA on the Resource Q column according to their charge as a reference. I believed that using these chromatography profiles, it would be easier to detect the excess of DNA and LH2-DNA adducts in the analysis of the separation products of chemistry with LH2. While pure LH2 and DNA are loaded to the column without any coupling, pure LH2

(yellow) which was eluted at 275 mM NaCl, was represented by a single peak. I could see that DNAs (gray) eluted as two peaks from the column at 400 mM NACI and 450 mM NACI. The reason of DNA elution as two peaks is probably that the DNA may form several secondary structures. Thanks to these references in the separation, it is easier to observe the elution of reaction products.

Even though the conditions for NHS activated Hitrap column could be still optimized, in case of insufficient optimization, it likely can add other struggles to the chemistry. As a separation method, among NHS activated Hitrap column and Resource Q since as working principle, Resource Q has less complications than the NHS activated Hitrap column. Therefore, I believe that using advantage of DNA negativity through Resource Q, is more favorable for the separation of chemistry products.

5.1.4. Summary

My purpose was to select reactants and reaction conditions to provide an easy monitoring for each step, better control and improving yield of the reaction. Following will be presented a summary of the selections done for the reactants and conditions as described in the precedent sections.

Proteins

- sfGFP
- mRuby3
- LH2

Bi-functional reagent and modification of DNA

- DBCO NHS ester and azide labeled DNA

Conditions

- Quantity: 60 nmol of protein, 240 nmol of DBCO NHS ester (1:4) and 20 nmol of azide labeled DNA
- Temperature: 25°C
- Buffer: 100 mM $NaHCO_3$ pH8.0

Reaction Sequence

- Attaching Protein to DBCO NHS ester and then adding azide-DNA to the mixture

Separation Method

- Resource Q

I used these selected reactants and conditions as starting point for this experimental procedure. Obviously, whenever challenges arose in the chemistry, I investigated possible changes and optimization to improve the results.

5.2. Coupling proteins to DNA

In this section, I used three proteins, sfGFP, mRuby3 and LH2 that can easily be distinguished (by their color). I presented an analysis of chemistry between sfGFP, mRuby3, LH2 and azide labeled DNA via DBCO NHS ester.

5.2.1. Attaching Azide-DNA to mRuby3, sfGFP and LH2 via DBCO NHS ester

5.2.1.1. mRuby3
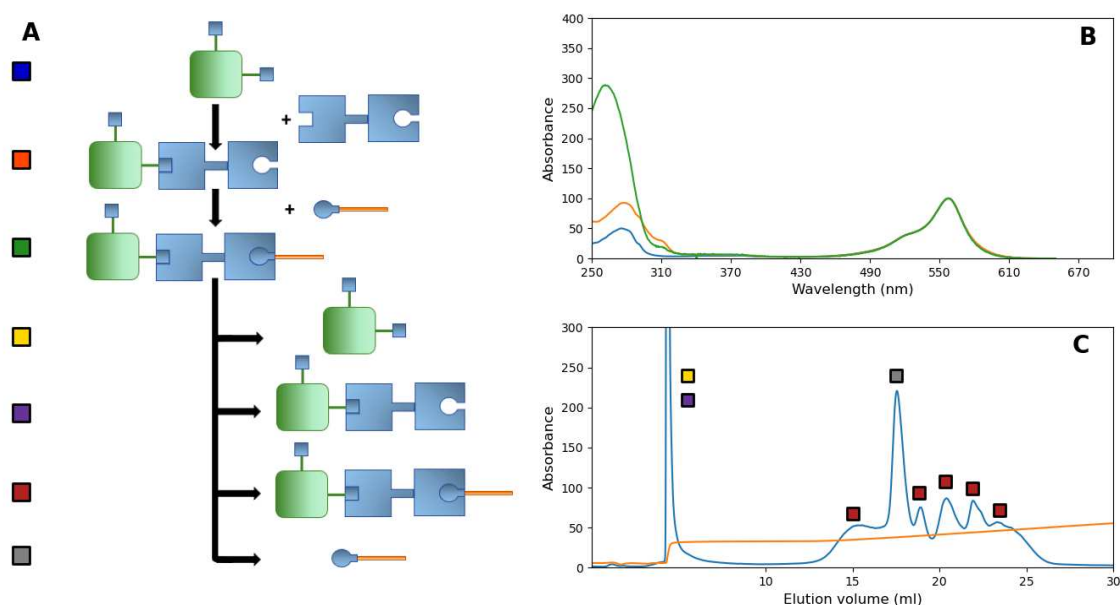
Figure 5.12. *Attaching DNA to protein. **A** The general scheme. Protein (green square) containing many reactive groups reacts partially with the bifunctional reagent (shown as a blue dumbell with specific binding sites), then in a second step this first adduct reacts with azide labeled DNA. Finally, the different reaction products are separated. The small squares on the left provide keys to parts B and C. **B** Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing mRuby3 protein (blue), protein reacted with DBCO NHS ester (orange); and DNA (green). **C** Separation of the reaction products using anion exchange chromatography. The initial peak contains both unreacted protein and protein DBCO adducts, the later well resolved peaks contain pure DNA, and protein linked to various numbers of DNA molecules.*

In figure.5.12.A, the outline of the procedure is shown: first, the protein with many reactive lysine residues is coupled to the bi-functional reagent DBCO-NHS ester, by the reaction of some of the primary amine groups with the NHS ester; in a second step, after the removal of excess linker, the DBCO labeled protein is attached to the azide-labeled DNA molecules, through the Cu-free cyclo-addition of DBCO and the azide groups on the DNA; finally the various reaction products are separated by Resource-Q anion exchange chromatography. As possible reaction products from separation by Resource Q, there are non-labelled via DBCO NHS ester protein (yellow square), non-reacted with DNA but labeled protein via DBCO NHS ester (purple square), coupled protein-DNA (red square) and non-reacted DNA (gray square).

One of the reasons for choosing this chemistry was that the progress of the reaction can be conveniently monitored, and the reaction products can be analyzed using absorption spectroscopy. In the figure 5.12.B, I show the absorption spectrum of mRuby3, with peaks at 558 nm and 280 nm, the spectrum of the protein-DBCO adduct (orange) with an additional peak at 310 nm due to the DBCO moiety, and the absorption spectrum of the 1:1 protein-DNA adduct after purification with a strong absorption at 260 nm due to the DNA. The intensity of the DBCO peak at 310 nm allows the accurate determination of coupling efficiency and the degree of labeling, using the extinction coefficient of 12,000 $M^{-1}cm^{-1}$ [187] at 310 nm. In the sample presented the degree of labeling is 2.07. Of course, this degree of labeling is an average, and so the sample includes proteins with 0, 1, 2, 3, 4 and more reacted lysines, I expect these different forms to follow a Poisson distribution. The relative intensity of the absorption at 260 nm allows calculation of the DNA-protein ratio. While the calculations of labeling efficiencies can be done using the protein absorption at 280 nm these calculations are greatly facilitated here using colored proteins. The result was a good separation of the absorption maxima for the protein at 558 nm, 488 at nm and at 850 nm for mRuby3, sfGFP and LH2 respectively, at 310 for DBCO nm and at 260 nm for DNA.

All these reactions modify the charge of the protein, since the reaction of the NHS ester with primary amines removes the positive charge of the primary amine, while the second reaction attaches a DNA poly-anion to the protein. Thus, the potential multitude of different reaction products can be conveniently separated by ion-

exchange chromatography. I used a strong anion-exchange column (resource-Q) and a typical elution profile (measured at 280 nm) as shown in Figure 5.12.C. The chromatographic profile shows seven well resolved peaks. The absorption spectra of the different peaks were analyzed to determine the composition of each peak in terms of protein, linker, and DNA. The first peak contains protein with 0, 1, 2 or more DBCO attached. This peak elutes during an initial step in the salt gradient and a more gentle gradient allows separation of proteins containing different numbers of reacted lysines. In the chromatogram, this is followed by a group of peaks in which I found protein molecules containing an increasing amount of DNA: 1, 2, 3, 4 and more molecules (red square), and a peak corresponding to unreacted DNA (gray square). It is worth noting here that the form of the chromatogram is a little misleading as each additional DNA molecule greatly increases the UV absorption, thus while the first 2 peaks are of similar height this means that the maximum concentration of simple protein DNA adduct is about twice that of the doubly coupled protein (Protein + 2 DNA molecules).

In the example used for the figures 5.12.B and 5.12.C, 60 nmoles of mRuby3 were reacted with 240 nmoles of DBCO NHS ester, then the labeled proteins were reacted with 24 nmoles of azide-labeled DNA. This resulted in about 6 nmoles of purified labeled mRuby3-DNA adducts, a yield on protein terms of 10% and on DNA terms of 25%. 2 nmol of purified labeled mRuby3-DNA adducts was singly labeled mRuby3-DNA adduct and this makes a yield on labeled protein terms of 33.3%. Overall, only 12 nmol of DNA attached to the 6 nmol of mRuby3 and the rest of DNA eluted as excess. Thus, the reaction between DBCO and azide labeled DNA had a yield of about 50%.

5.2.1.2. sfGFP
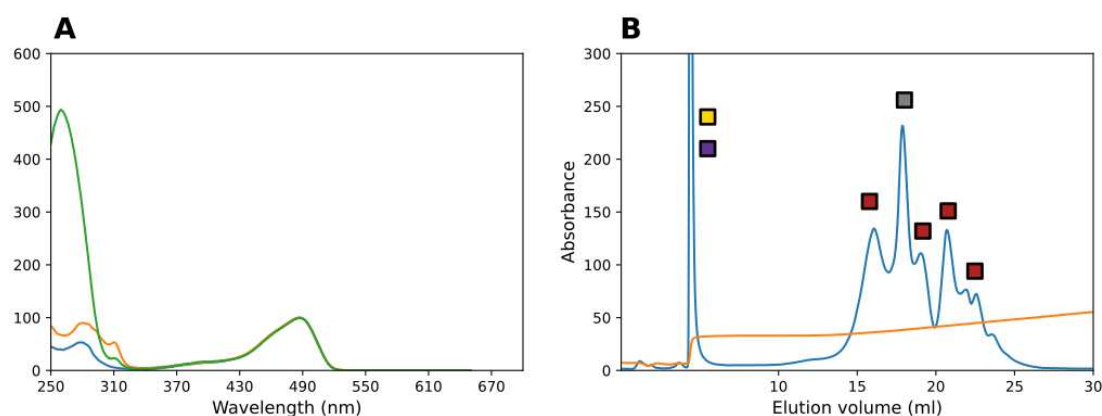
Figure 5.13. *Attaching DNA to sfGFP.* **A** *Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing sfGFP protein (blue), protein reacted with DBCO NHS ester (orange); and DNA (green).* **B** *Separation of the reaction products using anion exchange chromatography. The initial peak contains both unreacted protein and protein DBCO adducts, the later well resolved peaks contain pure DNA, and protein linked to various numbers of DNA molecules.*

Figure 5.13.A. shows the absorption spectra of pure sfGFP, protein after reaction with the linker, and a 1:1 protein DNA adduct. The absorption spectrum of pure sfGFP shows a peak at 488 nm (blue line), and a peak at 280 nm as expected. After the reaction with DBCO-NHS ester, and separation of excess reagent, the absorption spectrum (orange line) shows an additional peak at 310 nm. The relative size of this peak allows calculation of the reaction efficiency based on the extinction coefficients of the protein and DBCO. 60% of DBCO NHS ester reacted with sfGFP. In the spectrum shown, each protein has on average 4.46 linkers attached, while the absorption spectrum of the 1:1 protein-DNA adduct shows a strong absorption at 260 nm due to the DNA.

Separation of protein-DNA adducts was done again for sfGFP. Typical elution profile from a strong anion-exchange column (resource-Q) and (measured at 280 nm) was shown in figure.5.13.B. The chromatographic profile shows six well resolved peaks. Analysis of the different peaks were done to understand the composition of each peak, in terms of protein, linker and DNA. The first peak contains proteins with 0, 1, 2 or more DBCO attached. In the chromatogram this is followed by a group of peaks in

which I found protein molecules containing an increasing amount of DNA (i.e., 1, 2, 3, 4 and more molecules) and a peak corresponding to unreacted DNA.

In figures 5.13.A and 5.13.B, 60 nmoles of sfGFP were reacted with 240 nmoles of DBCO NHS ester, then the labeled proteins were reacted with 24 nmoles of azide-labeled DNA. This resulted in about 7.25 nmoles of purified labeled protein-DNA adduct, a yield on protein terms of 12% and on DNA terms of 30%. 3.75 nmol of purified labeled sfGFP-DNA adducts was singly labeled sfGFP-DNA adduct and this makes a yield on labeled protein terms of 51.7%. The reaction, involving coupling of the DBCO labeled protein with azide-DNA, had a yield of about 66% in my hands since 16 nmol of DNA reacted with sfGFP.

The yield of labelling proteins via DBCO NHS ester and DNA for sfGFP was around 15% higher than mRuby3. For the labeling protein via DBCO NHS ester, this reason is likely that sfGFP has more lysine than mRuby3 and the lysine distribution difference between them. In sfGFP, lysine is distributed not only N and C-terminal but also on beta sheet contrary to mRuby3. This distribution possibly increased the chance of labeling. I believe that higher degree of labeling raised the probability of reacting labeled protein with DNA. The phenomena occurred for both sfGFP and mRuby3 is that quantity of reacted protein is low and as an average 6.6 nmoles. Clearly this yield is expected to be strongly dependent on the reagents and the reaction conditions. It is not clear why this reaction is not more efficient, though it is possibly due to interactions of the DBCO with the protein surface or degradation of the DBCO.

5.2.1.3 LH2

One of the reasons for choosing this chemistry is that it is not limited by complexity of the protein, and it can be applicable for multimeric proteins such as LH2. Besides being multimeric, LH2 behaves differently from sfGFP and mRuby3 since it is a membrane protein. Hence, 0.1% -Dodecyl-B-D-Maltoside which is a maltoside based non-ionic detergent with an alkyl tail, was used to ensure solubility. The use of detergent led to some modification of the procedure as dialysis membrane usage to prevent precipitation of LH2 by centrifugation in the critical step and removing excess linker to prevent the competition between free DBCO NHS ester molecules and LH2-DBCO NHS ester molecules to bind azide modified DNAs. Nevertheless, some of the

factors had to be adjusted. The ratio between LH2 and DBCO NHS ester was decided as 1:8 instead of the 1:4 used for monomeric proteins since detergent makes the labelling reaction more complicated and LH2 itself has an absorption at 310 nm. Also, the quantity was decided as 10 nmol instead of 60 nmol since there is a limitation to increase of protein concentration. The reason behind this limitation is that size of LH2 (76 Å in diameter) is bigger compared to sfGFP (24 Å in diameter) and mRuby3.
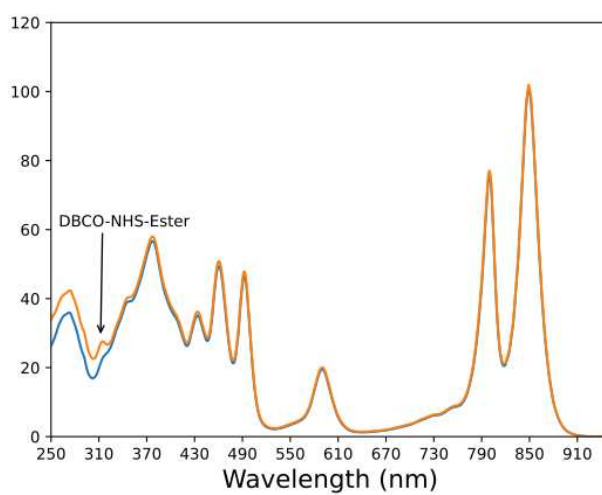
Figure 5.14. *Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing LH2 protein (blue), protein reacted with DBCO NHS ester (orange).*

Figure 5.14 shows the absorption spectrum of LH2, with peaks at 376 nm, 400 nm-532 nm, 590 nm, 800-850 nm, and 280 nm (blue) and the spectrum of the protein-DBCO NHS ester (orange) with an additional peak at 310 nm due to the DBCO moiety. Thanks to the UV absorption of DBCO NHS ester at 310 nm, I determined the coupling efficiency and degree of labeling, using the extinction coefficient as before. 33% of DBCO NHS ester attached to LH2. In the spectrum shown each protein has on average 3.6 linkers attached.
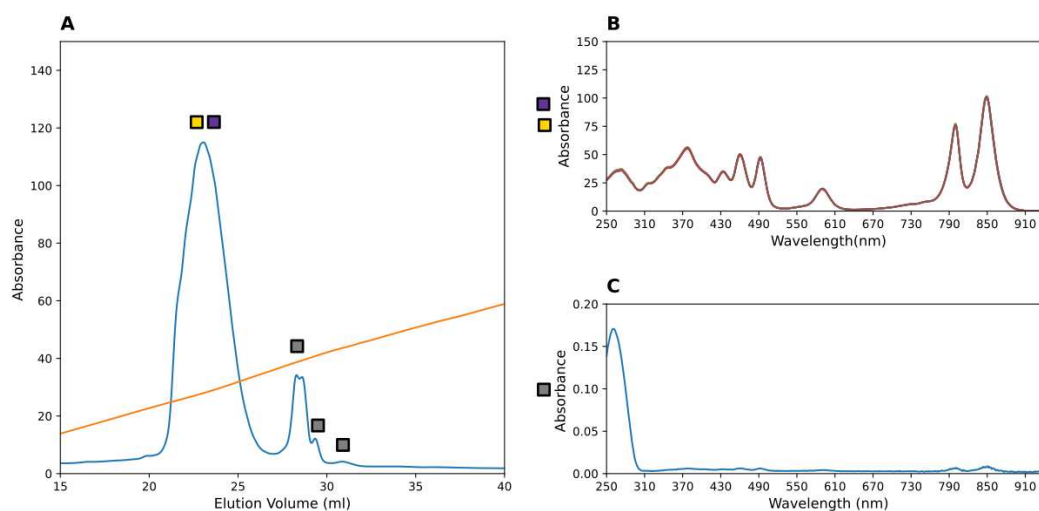
Figure 5.15. *Separation of reaction products for the chemistry reaction with LH2-DBCO NHS ester - DNA. **A** Separation of the reaction products using anion exchange chromatography. **B** Absorption spectra of the initial peak contains both unreacted protein and protein DBCO adducts (yellow and purple squares respectively). **C** Absorption spectra of the later peaks contain pure DNA with a small amount of LH2 (gray).*

Even with the non-stable charge of LH2, I predicted that it can be still favorable to use anion-exchange column to separate the reaction products of the reaction since the second reaction attaches a DNA poly-anion to the protein. Typical elution profile (measured at 280 nm) is performed by Resource-Q column. The chromatographic profile shows three poorly resolved peaks and these peaks are shown with colored squares (yellow, purple, gray) in figure 5.15.A. The first peak contains non-reacted LH2 as yellow and purple squares (see figure 5.15.B). This peak elutes during an initial step in the 280 mM salt gradient. Normally, gradient descent is supposed to allow separation of LH2 containing different numbers of reacted lysines with a well resolved peak. However, the first peak represented by gray square, were not separated as expected in figure 5.15.C. They were eluted in a way of splitting at the 400 mM salt gradient. I determined the reason for this splitting was related to the presence of 0.001 uM LH2 and 0.4 uM of DNA (gray), thanks to the UV absorption and extinction coefficient of LH2. The other following peaks (Gray) had more DNA but not more LH2. Thus, LH2 and DNA rate was not constant in these two peaks. This suggested that there was no LH2 - DNA Conjugation but coelution of LH2 - DNA.

In this chemistry, 10 nmoles of LH2 were reacted with 80 nmoles of DBCO NHS ester, then the labeled proteins were reacted with 5 nmoles of azide-DNA. This resulted in labeling 70% of LH2 with of DBCO NHS ester. The absorption spectrum of the reaction adducts (see figure 5.16.C), containing LH2 and DNA and it is noted that the quantity of LH2 here was negligibly low. Thus, I concluded that the elution of azide-DNA was contaminated with a tiny quantity of LH2 rather than coupling with each other. This result suggests that while the labeling LH2 via DBCO NHS ester reaction works efficiently, the azide-DNA did not react with labeled LH2.

So far, I detected two main challenges as efficiency of the azide-DNA with DBCO reaction and poor resolution which probably originates from the charge of LH2 molecules. As first, I focused on the efficiency of the azide-DNA with DBCO reaction.

It is not clear why the reaction between azide-DNA and labeled LH2 reaction is not efficient, but it is possibly either due to somehow DDM preventing coupling of azide-DNA with labeled LH2 or because of the hydrophobic micelle environment incorporating both the LH2 and reactive DBCO. In this case, it prevents coupling of hydrophilic azide-DNA and DBCO NHS ester labeled LH2.

Chemistry reaction in a reverse sequence

To test the first hypothesis about the DDM effects on the azide-DNA and labeled LH2 reaction, I decided to change reaction sequence. Instead of treating LH2 and DBCO NHS ester as first, I reacted azide-DNA and DBCO NHS ester under the same conditions as incubation and temperature and again separated reaction adducts with the anion exchange Resource Q column. Thus, I carried out the reaction between azide-DNA and DBCO NHS ester without DDM around.

Even in the reverse sequence, while NHS ester still binds LH2 with the degree of labeling 2.15, azide-DNA did not react with LH2. Since performing the reaction in the reverse sequence did not improve the reaction efficiency, I suggest that the possible reason can be pH of the buffer for DNA-DBCO NHS ester reaction. The pH of the reaction became acidic because of the DBCO NHS ester and DNA's solvent (DMSO and water respectively). Thus, I adjusted the pH of the first reaction as DNA-DBCO NHS ester to 8, and then reacted DNA-DBCO NHS ester with LH2. While the degree of labeling was 2.6, in the separation procedure, I obtained only one LH2-DNA adduct.

However, this adduct contains DNA (3.6 uM) and the negligible amount of LH2 (0.003 uM). This result pointed out again that non-reacted azide-DNA molecule coeluted with a small amount LH2.

5.2.1.3.1. LH2 with DBCO PEG$_4$ NHS ester

For the efficiency of the DNA-DBCO reaction, I suspected that failure of the azide labeled DNA-DBCO reaction can be related to the hydrophobic environment which comes from the nature of membrane protein. Hence, I wanted to decrease the hydrophobicity of the coupling reagent. This can be achieved by adding Polyethylene Glycol moieties to the linker. The reaction was performed with this modification of the coupling reagent. This Polyethylene Glycol addition not only provide decrease the hydrophobicity of the environment but also makes the linker longer and contributes to the reaction between DBCO and azide-DNA.
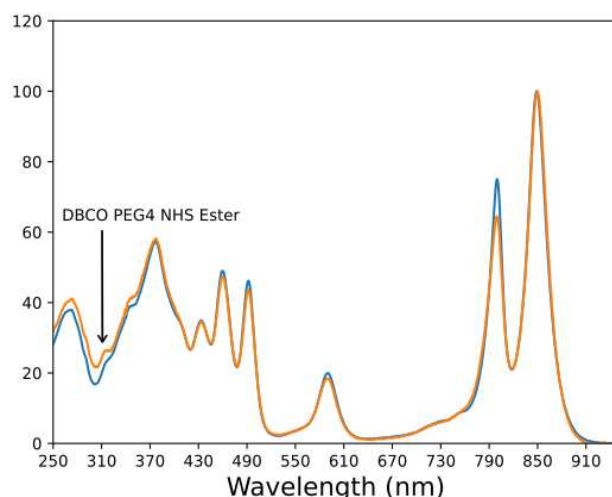
Figure 5.16. *Absorption spectra of different adducts shown are the normalized absorption spectra of samples containing LH2 protein (blue), protein reacted with DBCO PEG$_4$ NHS ester (orange).*

Figure 5.16 shows the absorption spectrum of LH2 (blue), the spectrum of the LH2-DBCO PEG$_4$ NHS ester with an additional peak (orange) at 310 nm due to the DBCO PEG$_4$ NHS ester. With the addition of PEG molecules, the degree of labeling increased, each protein has on average 5.8 linkers attached.
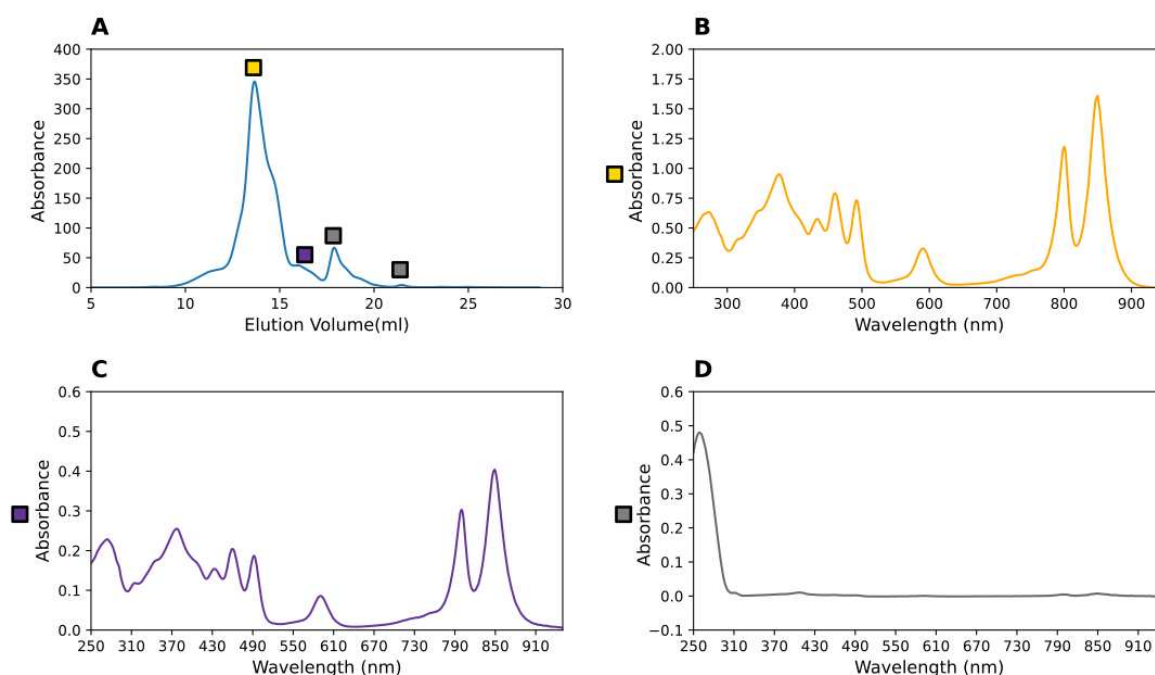
Figure 5.17. *Separation of reaction products for the chemistry reaction with LH2-DBCO NHS ester - DNA **A** Separation of excessive DNAs by Superose 6 increase 10/300 G/L size exclusion column **B** Absorbance spectra of a fraction corresponding to pure LH2 (yellow square). **C** Absorbance spectra of fractions corresponding to LH2-DBCO-PEG$_4$-NHS ester (purple square). **D** Absorbance spectra of four fractions corresponding to excess DNA (gray square).*

For the poor separation process as a second challenge, if I remove the excess DNA somehow, it would be the best way to eliminate the possibility of co-elution of excess DNA and LH2 molecules. Before going to separation adducts step with Resource Q,

I decided to modify the protocol to prevent the elution of DNA with LH2 and used size exclusion chromatography (Superose 6 increase 10/300 GL column) thanks to the size difference between DNA and LH2.

The size exclusion chromatography profile shows 3 peaks, which have color code squares. (Figure 5.17.A). Analysis of the peaks was done and pure LH2 was eluted, first two peaks (yellow square), according to the principle of the size exclusion chromatography as it is shown in figure 5.17.B. Since attaching DBCO NHS ester does not create a big change in size, LH2 and DBCO PEG$_4$ NHS ester (purple square) were eluted at just near to the second peak, figure 5.17.C. Lastly, as I expected, excess DNAs arrived as two small peaks (gray squares) because of their size (figure 5.17.D).
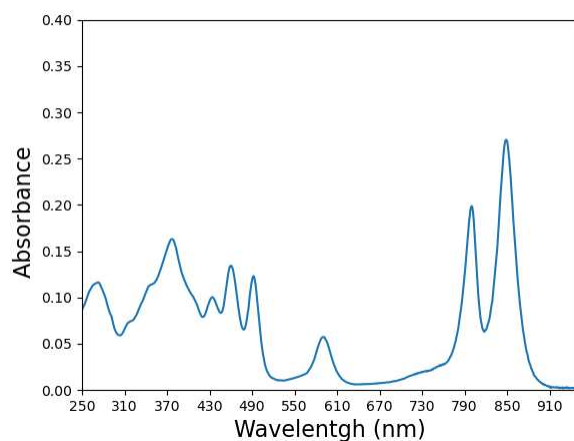
Figure 5.18. Separated sample from size exclusion chromatography contains both unreacted and reacted LH2 with DBCO PEG₄ NHS ester to load to the Resource Q anion exchange column.
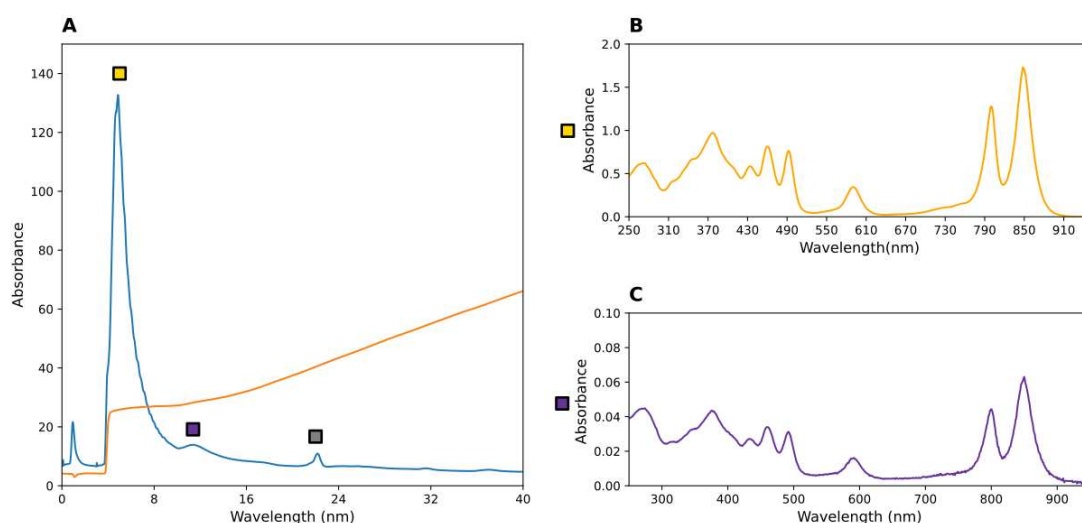
*Figure 5.19. Separation of reaction products with LH2 DBCO PEG₄ NHS ester DNA **A** Separation of LH2 reacted with DBCO PEG₄ NHS ester and LH2 reacted with DBCO PEG₄ NHS ester DNA by the Resource Q anion column. **B** Absorbance spectra of fractions corresponding to unreacted LH2 (yellow square). **C** Absorbance spectra of LH2 reacted with DBCO PEG₄ NHS ester (purple square).*

Except for excess DNA, which was obtained by the size exclusion chromatography, I collected and concentrated all the adducts of the reaction, represented in figure 5.18, to separate them by Resource Q anion exchange column.

To provide better separation between non-reacted and LH2-DNA adducts, I decided to use different gradients since I already know the salt gradient for non-reacted LH2

elution (Figure 5.19.A). In the end, adducts of the reaction were separated as three peaks (yellow, purple, and gray squares). As usual, non-reacted LH2 appeared in the chromatographic profile as first with the 260 mM salt gradient (Figure 5.19.B). Second peak was eluted at 280 mM salt gradient. This salt gradient was still in the range of salt gradient for the non-reacted LH2. This adduct was composed of a 0.01 nmol LH2 with 0.05 nmol DBCO molecules (purple) and shown in 5.19.C. The third peak on the chromatography represents excess DNAs since it was eluted from the column at 400 mM NaCl (gray).

I kept temperature, incubation time and ratio as same as previous experiment with LH2 and observed that addition of PEG_4 improved the yield of coupling LH2 via NHS ester since it increased from the previous 40% up to 75% and the degree of labeling increased from 4 to 6. However, there was no indication of the formation of LH2-DNA adducts. Hence, addition of PEG molecules to DBCO NHS ester did not contribute to the efficiency of the second reaction, involving azide labeled DNA-DBCO.

The increase of degree of labeling with the addition of PEG_4 shows partially improved reaction efficiency since it only increases the efficiency of the LH2-NHS ester not DBCO-DNA. This may be an indication that the hydrophobicity of DBCO is indeed resulting in it being sequestered by micelles. Preventing the simultaneous coelution of LH2 with excess DNA via size exclusion chromatography provided us better vision to clearly understand that as reaction adducts there is no diversity in the sense of having different amounts of DNA attached to the LH2. This can be counted as another proof that the reaction of DNA-DBCO is not efficient.

5.2.1.3.2. Summary

To summarize what I have done so far for the chemistry with LH2 and how it has turned out, I encountered two difficulties. One of these was the disadvantage that LH2 presented in the separation process, LH2 exhibits variable negativity due to the lipids of the inner ring. Secondly, LH2 reacted with the NHS ester, while the DBCO did not react with azide-DNA despite many attempts such as changing the reaction order, pH adjustment, PEG_4 addition and modification conditions. While this was the case for LH2, I also conducted the chemistry with the core complex, which comes from the

same system, although the core complex is more complex in terms of structure. All this attempts for both LH2 and core complex were presented in the table below.

Protein	DBCO NHS ester		Reaction Order		Incubation Time & Conditions		Chromatography Type		Protein not coupled but coeluted with DNA	Coupled Protein with DNA
	With PEG$_4$	W/O PEG$_4$	1)Protein 2)Linker 3)DNA	1)Linker 2)DNA 3)Protein	O.N & 2 h R. T	30min & 2h R. T	SEC	Anion		
LH2		✓	✓		✓			✓	✓	Not observed
LH2	✓		✓			✓	✓	✓	✓	Not observed
CC	✓			✓		✓		✓	Not observed	✓

Table 5.1. *Summary of the chemistry reaction with LH2 and CC.*

Looking at the table above, in the first line the chemistry with LH2-DBCO NHS ester was summarized and it resulted in coelution of LH2 and excess DNA. While addition of PEG$_4$ to decrease the hydrophobicity did not solve the binding difficulty between azide-DNA and DBCO, separation of excess DNA via SEC did not result in LH2-DNA coupling. This supported my hypothesis that LH2 and DNA coelutes rather than reacting with each other. In the third line, when I examined coupling core complex-DNA via DBCO PEG$_4$ NHS ester, I obtained 3 core complex molecules with different amounts of DNA even though the excessive DNA was not removed by size exclusion chromatography. However, all of products of the reaction eluted under the same peak and did not show separation in the chromatography same as LH2.

Comparing the LH2 and CC results, I first looked at the main differences between them, CC has a higher number of lysine than LH2, and then the lysines of the core complex were presented in both the N-terminal and the C-terminal of the CC. As a result, two phenomena have occurred as:

● These features of the core complex, which are different from LH2, increase the possibility of coupling with DNA, hence LH2 requires these properties.

● There is a parameter that disrupts the azide and DBCO interaction in case of both LH2 and CC, like being multimeric or non-soluble. However, since the core complex has more lysines located on both N- and C-terminal, differently from LH2 (has just lysine on N-terminal), these advantages compensate the impact on interaction of DNA-DBCO and the reaction that takes place.
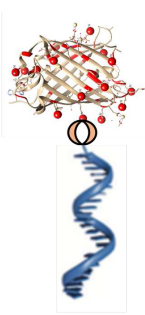
In conclusion, LH2 and DBCO NHS ester reacted with a degree of labeling as 3.6 and this degree of labeling increased up to 5.8 with the addition of PEG_4 to DBCO NHS ester. However, neither DBCO NHS ester nor DBCO PEG_4 NHS ester did not attach DNA.

5.2.2. Testing conditions for coupling LH2-DNA and Maleimide PEG_2 NHS ester using sfGFP

In the previous section, even though I achieved to label LH2 via DBCO NHS ester and DBCO PEG_4 NHS ester with a sufficient degree of labeling, it could not contribute to the efficiency reaction of DNA-DBCO, so I considered changing the linker (see 5.1.1.) and decided to test Maleimide PEG_2 NHS ester and thiol modified DNA. The reason for choosing this linker is not only it is commercially available but also Maleimide is less hydrophobic than DBCO. Maleimide thiol reaction is one of the most used reagents although Maleimide PEG_2 NHS ester cannot be followed by a visible UV spectrophotometer.

As previously, I need to determine the optimal conditions, buffer-pH, reaction order and molar ratio of the reactants to obtain high yield from the reaction. Since coupling LH2 is a more complicated process as I experienced from the previous results, I wanted to optimize the important parameters for the reaction through sfGFP. The efficiency, which is presented in the table, was calculated in terms of DNA since the quantity of DNA limiting the coupling efficiency of LH2-azide labeled DNA.

Protein	Reaction Buffer			Reaction Order		Buffer pH		Protein Quantity		DDM	Attached DNA to sfGFP/ Total DNA used for reaction
	100mM NaHCO$_3$ pH8	50 mM KH$_2$PO$_4$ pH7.5	PBS pH7.5	1)Linker 2)DNA 3)Protein	1)Linker 2)DNA 3)Protein	pH7 then pH8	pH8	10 nmol	60 nmol	0,1%	
sfGFP	✓				✓		✓	✓			1.25%
sfGFP	✓				✓		✓		✓	✓	5.5%
sfGFP			✓		✓		✓		✓		11.25%
sfGFP		✓		✓			✓		✓		30.25%
sfGFP	✓			✓			✓		✓		33.75%
sfGFP		✓		✓		✓			✓		37.5%

Table 5.2. *Optimization of sfGFP-DNA coupling via Maleimide PEG$_2$ NHS ester under different conditions.*

In **line 1,** since Maleimide PEG$_2$ NHS ester cannot be followed by spectrophotometer, I could not arrive at the accurate determination of the degree of labeling but coupling efficiency. The chemistry under the condition presented in the table resulted in only 1.25% of DNA labeled sfGFP. Also, not only the quantity of attached DNA but also the quantity of sfGFP-DNA couple was low after the separation of reaction adducts by Resource Q anion exchange column. I believed that the possible reason could be the quantity of sfGFP as 10 nmol. The quantity of sfGFP raised to 60 nmol in the presence of the DDM in **line 2** to examine the effect of the quantity of LH2 and the impact of detergent chemistry efficiency. The efficiency of coupling reaction was increased around 5 times. This raise possibly provided by the increase of sfGFP quantity rather

than detergent. To ensure that in *line 3*, only the chemistry performed without detergent. The coupling efficiency was raised around 2 times by the absence of detergent and at pH7.5. Here, I interpret that the reason for this increase likely be both absence of the detergent and pH. Since pH7.5 seemed convenient, to discover the effect of the reaction sequence and the composition of the reaction buffer I modified them as it is shown in *line 4.* The efficiency was around 2-3 times higher than trial, explained in line 3. Here, the main reason for this raises possibly originated from avoiding the usage of excess salt ions like PBS and the effect of reaction sequence. I also examined **in *line 5*** that whether the chemistry is producible even in the different composition of the buffer. I observed that avoiding high ionic strength and keeping pH around 7.5-8.0 make the chemistry producible. Since DNA-Maleimide PEG_2 NHS ester and protein-Maleimide PEG_2 NHS ester requires different pH, *in line 6,* I altered the pH during the reaction. While pH was 7 for the reaction of DNA-linker, for the reaction of linker-protein pH was 8.0 and the efficiency of the chemistry was recorded as 37.5%.

In conclusion, I found that the detergent disrupted the interaction between reactants. However, it is essential to use detergent for LH2 despite challenges. Also, the quantity of protein must be decided by considering the complexity and the steps of the procedure since it is likely to protein in long process. However, in my case LH2 can be concentrated until a certain value, 10 nmol of LH2 is enough for the chemistry. Another critical factor was here that altering the pH during the chemistry increased the efficiency up to 37.5%. The other parameters of this chemistry with Maleimide PEG_2 NHS ester, were two optimal pH, reaction sequence and avoiding excess salt ions in the buffer compositions.

5.2.2.1. LH2- Maleimide PEG$_2$ NHS ester

I optimized the conditions for the chemistry with Maleimide PEG$_2$ NHS ester and here I tried optimized conditions with LH2, however there is one difference which is the presence of 0.1% of DDM.
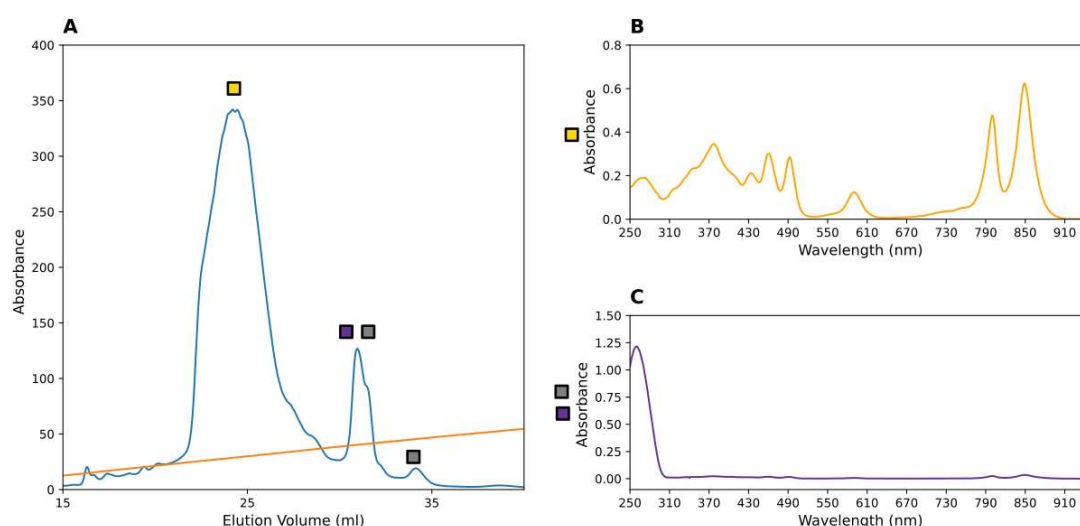
Figure 5.20. 10 nmol pure LH2 was reacted first with 50 nmol Maleimide PEG$_2$ NHS ester and then 5 nmol of azide-DNA. Reaction products was separated by Resource Q anion. *A* Separation of LH2 reacted with Maleimide PEG$_2$ NHS ester and LH2 reacted with Maleimide PEG$_2$ NHS ester-DNA by the Resource Q anion column. *B* Absorbance spectra of fractions corresponding to unreacted LH2 with Maleimide PEG$_2$ NHS ester (yellow square). *C* Normalized Absorbance spectra of excess DNA with a small amount of LH2 (gray and purple squares respectively).

In the separation of adducts step, the chromatography profiles show three main peaks with poor resolution in figure 5.20.A. As separation of the peaks, I obtained a profile like the LH2-DNA coupling via DBCO NHS ester which was explained in 5.2.1.3. The analysis of peaks showed that in figure 20.B, the peak which has yellow square represents just the non-reacted LH2 and I cannot say that either it is only pure LH2 or maleimide attached LH2 since maleimide cannot be tracked by the spectrophotometer. However, I am sure that it contained non-reacted LH2 with the DNA since it was eluted 280 mM NaCl. A second peak which is in a way of splitting, was eluted from the column at 420 mM NaCl which corresponds to excess DNA. (Purple and gray color). In figure 20.C, it was shown that these adducts under the

second peak contain just excess DNA with negligible amount of LH2. Here again, a small quantity of LH2 was coeluted with DNA molecules instead of the way they reacted to each other. The last small peak was the part of the oligonucleotide since it eluted a bit further than usual at around 460 mM NaCl (gray). The reason is that the dimerization of thiol since the DNA is thiol labeled. Because of this dimerized DNA had twice as much negative charge and eluted further than expected.

In conclusion, despite the replacement of linker via Maleimide PEG$_2$ NHS ester and optimization of conditions by considering this linker, the reaction between thiol labeled DNA and the linker did not work. Since I did not obtain any coupling of LH2 and DNA, it remained unknown whether the maleimide PEG$_2$ NHS ester labeled LH2 or not. However, lysine of LH2 reacted with the linker NHS ester group, I believe that while LH2 likely labeled via Maleimide PEG$_2$ NHS ester, thiol-DNA could not arrive to react with maleimide.

5.2.2.2. Comparison of coupling reaction with sfGFP and LH2

In the reaction for coupling DNA with LH2, I have been encountered with two challenges. Firstly, while the NHS ester group of linkers were reacting to the amines of LH2, the reaction between DNA and various groups remained problematic since the linker did not react with DNA. To introduce a solution to this case, I examined the reaction with all the linker options which were decided in 5.1.1 as DBCO NHS ester, DBCO PEG$_4$ NHS ester and Maleimide PEG$_2$ NHS ester.

While DBCO NHS ester and Maleimide NHS ester reacted with sfGFP and thus I obtained protein-DNA adducts, I did not observe the adducts in the case of LH2.
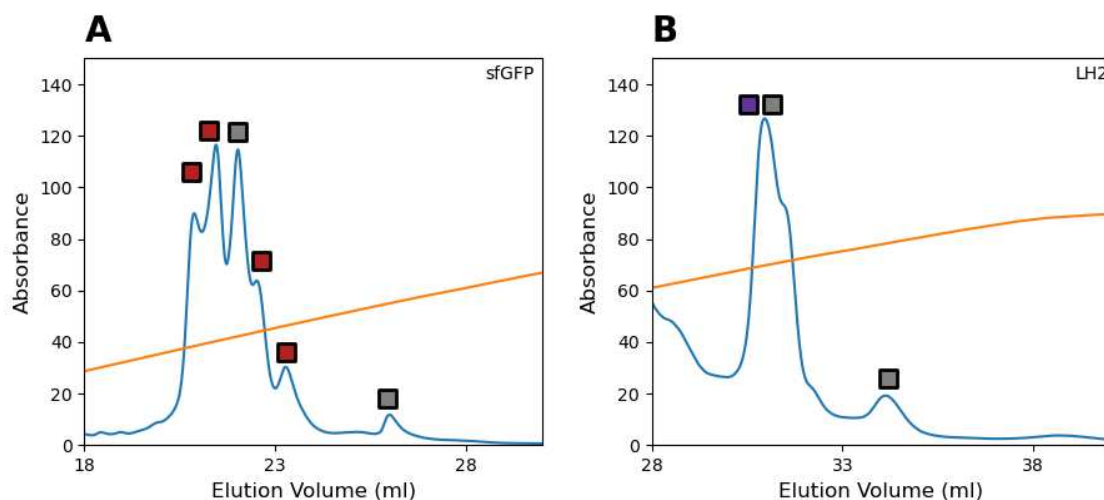
Figure 5.21. *After the chemistry reaction separation of LH2-Maleimide PEG$_2$ NHS ester-DNA and sfGFP-Maleimide PEG$_2$ NHS ester-DNA reaction products by Resource Q anion column. **A** Separation of the reaction products using anion exchange chromatography. The well resolved peaks contain pure DNA (gray squares), and sfGFP linked to various numbers of DNA molecules (red squares). **B** There is no separation of reaction adducts. The initial peaks contain LH2-Maleimide PEG$_2$ NHS ester (purple square), small amounts of excess DNAs (gray square) and the second peak contains excess DNA (gray square).*

The second challenge was about the separation step via anion exchange Resource Q. Even though LH2 and DNA did not react with each other, they coeluted from the Resource Q column. I showed this fact by comparing chromatography profile of sfGFP and LH2 in figure 5.21. Here, the peaks containing non-reacted LH2 and sfGFP were not shown in the figure to show clearly while sfGFP adducts were separated well, LH2 coelutes with DNA rather than separating.

5.2.3. Summary

My aim was coupling DNA with sfGFP, mRuby3 then LH2 through the chemistry reaction. I examined several conditions such as reaction buffer-pH, molar ratio of reactants, temperature, protein concentration, linkers, and separation methods to characterize an efficient method. Then, under selected conditions and with separation method, I carried out the chemistry with sfGFP, mRuby3 and LH2 with DBCO NHS ester which is the first choice as a linker. While I obtained enough sfGFP-DNA and

mRuby3-DNA adducts 1:1, the chemistry worked only at the first step which is labeling LH2 with DBCO NHS ester and it did not go further for the reaction between DNA and linker labeled LH2. The possible reason was that the environment was too hydrophobic for the reaction. Hence, PEG$_4$ was introduced to the DBCO NHS ester. However, the presence of PEG$_4$ did not overcome the challenge of reaction attaching DNA to the linker. Therefore, I decided to replace the DBCO PEG$_4$ NHS ester with Maleimide PEG$_2$ NHS ester which is less hydrophobic than DBCO. The result of the chemistry with Maleimide PEG$_2$ NHS ester showed that replacement of linker did not improve the efficiency of the reaction. Here the question was while the first reaction works between NHS ester and amines of LH2, why the reaction between DBCO-azide -DNA and Maleimide-thiol-DNA does not. It is not clear why this reaction is not efficient, though it is possibly due to interactions of the DBCO / Maleimide with charged azide/thiol groups in the presence of detergent. The reaction between NHS ester and the amines of LH2 works in the presence of detergent due to the fact that micelles naturally surround the hydrophobic molecules, however, the reactive groups of DNA are strongly charged and thus they cannot arrive to interact with the linker.

Additionally, it is worth noting here that there was a phenomenon, which was observed coupling LH2-DNA via all these three linkers, non-reacted LH2 always coeluted with DNA. This suggested that this challenge originated from the LH2. Since the principle of the column relies on the negativity of the molecules, I believed that this complexity has originated from negativity of the LH2 in the reaction and the literature supported my hyphothesis. *In Rhodobacter capsulatus* and *Rhodospirillum rubrum* and *Rhodobacter sphaeroides*, it is known that LH2 which associated phosphatidylglycerol (PG) can be detected after purification [188] and these lipids can alter the negativity of LH2. Thus, the apparent poor resolution likely be due to the presence of multiple unresolved forms with local charge characteristics.

Overall, I could not obtain LH2-DNA adducts but obtained enough sfGFP-DNA and mRuby3-DNA 1:1. This allowed to go through further process which is assembling formed DNA-protein adducts in a DNA directed manner.

5.3. Assembling proteins

As I explained in the beginning of chapter 5, my aim in this work is composed of two phases. First phase was coupling protein-DNA 1:1 adducts and from the first phase I obtained sfGFP-DNA and mRuby3-DNA 1:1 adducts. Here in the second phase, I assembled the formed sfGFP-DNA and mRuby3-DNA 1:1 adducts in a DNA directed manner to form dimers and characterized this assembly by absorption spectroscopy, fluorescence measurements and gel electrophoresis. Furthermore, I demonstrated that the DNA strands in such structures can be still available for disassembly, either through enzyme digestion or toe-hold strand displacement.
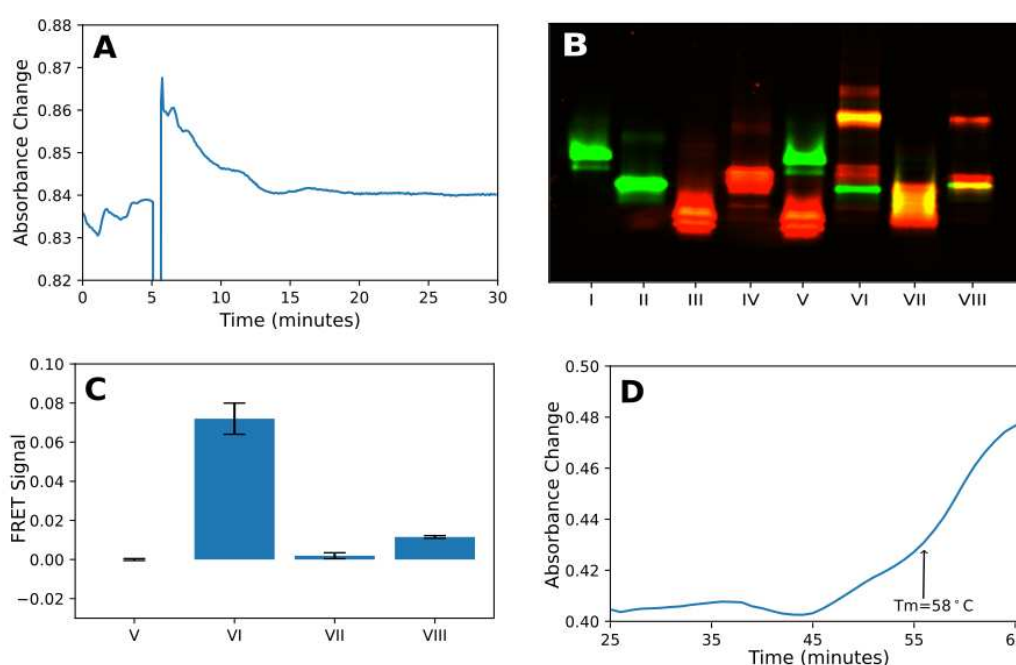
Figure 5.22. Assembling and dis-assembling DNA assemblies. (A) Absorbance changes measured at 260 nm following the mixing of sfGFP-DNA and mRuby3-DNA pair at 0.5 uM in hybridization buffer at 25°C. (B) SDS-PAGE analysis of proteins, adducts and assemblies. Samples containing 18 pmoles protein: pure sfGFP (I) , sfGFP-DNA (II), pure mRuby3 (III), mRuby3-DNA (IV) , mixture of sfGFP and mRuby3 (V) dimer of protein-DNA adducts (VI), DNAse treatment dimer (VII) and dimer with toehold replacement strand (VIII) were separated on 12% SDS-PAGE Gel (without sample heating). The gel was then imaged with 488 and 555nm excitation using a fluorescence imager. (C) Förster resonance energy transfer between sfGFP and mRuby3 was measured by donor (sfGFP) sensitized acceptor (mRuby3) emission for the dimer (V), DNAse treated dimer (VII) and dimer with the toehold replacement strand (VIII). (D) The hybridized dimer was heated from 25 C to 65 C and absorbance followed at 260 nm to observe the thermal denaturation of the double helix.

As an initial approach to show that these protein-DNA adducts can be used for making DNA directed assemblies I have followed the assembly and disassembly using absorption spectroscopy (following dsDNA hypochromism at 260 nm), fluorescence energy transfer between different fluorescent proteins, and SDS-PAGE of assemblies.

Here, I used sfGFP and mRuby3 coupled respectively to the pair of complementary oligonucleotides. These oligonucleotides are designed to hybridize, forming a double helix with the 2 coupled proteins at the same end of the DNA duplex, and a short unhybridized oligo T region on each DNA strand. This unhybridized region allows disassembly by toehold strand displacement, see below. As the protein part of the DNA-protein chimeras is temperature labile (above 60°C) I have used an isothermal assembly method. Indeed, the necessity of establishing isothermal annealing methods for the assembly of covalent DNA-protein adducts into nanostructures is what pushed us to initially investigate assembly using this simple construct.

In figure 5.22, I demonstrate the formation of a heterodimeric protein complex, driven by DNA hybridization, and then disassemble this dimer using several different methods. After mixing protein-DNA chimeras containing sfGFP and mRuby3 coupled to complementary DNA strands (final concentration 0.5 uM) I can observe the formation of DNA duplex by following the decrease in absorption at 260 nm (Figure 5.22.A). That shows that annealing in these conditions takes about 15 minutes.

To verify that the DNA-protein chimeras assemble correctly I have used SDS-PAGE revealed by fluorescent imaging (Figure 5.22.B). This allows us to follow the hybridization reaction easily. In the first wells I have loaded sfGFP, sfGFP-DNA, mRuby3 and mRuby3-DNA. As is expected the addition of the negatively charged DNA molecule to the sfGFP results in an increase in mobility of the green fluorescent protein (compare lanes I and II) surprisingly addition of the DNA molecule to the mRuby3 protein results in a decrease in the mobility of the red fluorescent protein (lanes III and IV). In this case apparently the change in molecular mass has more effect than the change in charge. Possibly due to interactions between the DNA and the protein surface. Gratifyingly there is no sign of doubly labeled protein containing 2 DNA molecules and more anomalous migration. Mixing the sfGFP and mRuby3 proteins, without any attached DNA, (lane V) results in bands at the appropriate size with no sign of higher molecular weight dimers being formed. However, when the

hybridized dimer is examined on the gel (lane VI) most of the fluorescence, both green and red, is associated with a band migrating further up the gel corresponding to a dimer. This band contains almost all the red fluorescence, however there is some green fluorescence that remains at the same position as observed in lane II suggesting that there is probably a slight excess of sfGFP-DNA in the sample. Lanes VII and VIII will be discussed below.

As a second method to study the formation of complexes I used FRET, (Figure 5.22.C), where I report the donor (sfGFP) sensitized acceptor (mRuby3) emission. The mixture of purified proteins gave a consistent zero background reading (V) while I obtained a reliable and reproducible measurement from the hybridized protein-DNA adducts (VI) confirming that the hybridization of the DNA molecules brings the proteins into proximity and can transfer energy between them. The characteristic distance, R0, for this pair is about 6 nm so given the levels of FRET observed, about 10% transfer efficiency, the center of mass of the 2 proteins must be about 8 nm apart. This suggests that though the DNA molecules bring the 2 proteins the length of the flexible linkers between the DNA and the protein, about 0.8 nm, together the electrostatic repulsion between these 2 basic proteins is sufficient to reduce the transfer efficiency.

Dis-assembling DNA structures

One of the interesting aspects of DNA nanostructures is that they can be dynamic and be induced to change structure. This depends on the availability of the DNA elements to different agents. To verify that the DNA elements in the dimeric structure I formed were available I used several methods to disassemble the dimer.

The easiest and most used method to disassemble DNA nanostructures is by heating. This is also possible with the protein dimers that I formed, as is shown in Figure 5.10.D where I follow the absorption at 260 nm as the solution of dimeric proteins is heated. As expected, there is a marked increase in absorption and the DNA duplex melts. In my simple case I observe a melting centered at about 58°C. Unfortunately, the fluorescent proteins used in this study are not very stable at 60°C and above so this DNA melting also resulted in protein denaturation and changes in the visible absorption spectrum.

Another method to modify or destroy DNA nanostructures is to treat them with enzymes. To verify that the DNA in my system is available for enzymatic digestion I treated dimers with DNAse I, Figure 5.22.B, and C VII. In the fluorescent gel image, this treatment removes the high molecular weight dimer band, and the fluorescence is all found in the monomer region of the gel. However, I do not revolver the same distribution as with proteins containing no DNA (lane V). Thus, while the DNA'ase treatment has clearly destroyed the DNA directed dimer formation there is almost certainly a small amount of DNA still attached to the proteins, but not sufficient to drive assembly of a stable structure. To confirm that enzyme treatment destroys the dimers I also measured FRET (Figure 5.22.C) in agreement with the SDS-PAGE measurements I see no energy transfer indicating that the two proteins have been separated.

As a final method to separate the 2 proteins and verify accessibility of the DNA I used toehold strand replacement. Here I added to the preformed dimers an oligonucleotide fully complementary to the DNA attached to sfGFP, this single stranded DNA can displace the oligonucleotide attached to the mRuby3 protein since it is complementary over a longer sequence (see Table 1). After toehold replacement, I observed that the dimers had been disrupted both by SDS-PAGE measurements (Figure 5.22.C) and by FRET measurements (Figure 5.22.D).

These results show that I can construct and purify 1:1 protein-DNA adducts obtained at a reasonable yield using the bifunctional reagent DBCO NHS ester. These adducts can be formed under mild conditions, of pH, temperature, redox potential that do not compromise protein or chromophore structure or activity. I demonstrate that the different reaction products can be readily separated using ion-exchange chromatography, relying on the modification to the surface charge distribution caused by reaction of the protein amine groups and attachment of a polyanionic DNA strand. I observe that the DNA-protein adducts that are so formed can be assembled in a DNA directed manner by absorption spectroscopy, fluorescence measurement and gel electrophoresis. Furthermore, I demonstrate that the DNA strands in such structures are still available for disassembly, either through enzyme digestion or toe-hold strand displacement.

<u>Conclusion</u>

DNA directed assembly of proteins represents an exciting prospect for the construction of new active meta-materials. However, building such structures is currently hampered by several problems. Among these is the ability to obtain pure and stable DNA-protein adducts with reasonable yield. I showed that using DBCO NHS ester I can obtain monomeric proteins coupled to a single DNA molecule at reasonable yield. This chemistry has the advantages of being widely applicable, as most proteins contain surface amino groups; easy to follow to adjust reaction conditions to optimize the yield, thanks to the UV absorption of DBCO, and give readily separable reaction products thanks to the charge on the DNA molecules. I believe this method is particularly useful in situations for attaching DNA to proteins where the introduction of single reactive residues into the protein is difficult, for example in oligomeric proteins or proteins from organisms that are difficult to genetically manipulate. However, the other specific properties of oligomeric proteins might add new difficulties to this chemistry as it happened in my case with LH2. Since LH2 protein is membrane protein, I needed to use detergent at each step of the reaction and the presence of detergent prevented the reaction of DBCO NHS ester labeled LH2 with azide labeled DNA. Thus, this chemistry is promising to work with complex proteins, even though some modification can be needed depending on the structural and physical features of protein for a high efficiency of the reaction.

One of the main advantages of this procedure was that obtained protein-DNA adducts are accessible for the formation of DNA scaffolded nanostructures. For that reason, in the following section it will be explained the design of DNA nanostructure through DNA origami technique to integrate these sfGFP- DNA and mRuby3- DNA 1:1 adducts into this DNA nanostructure.

Chapter 6 – Formation of Rectangular DNA Nano Object

It is interesting to develop methods for coupling proteins to DNA nanostructures, in particular because of the very diverse functional and structural properties observed in proteins. One of the most prominent features of DNA nanostructures is the possibility to place and control guest molecules and particles with high precision. These suggest that molecules can be attached and site-specifically arranged on the structures by the selection and elongation of a certain set of oligonucleotides pointing outward from the structure. This allows for positioning of any guest molecules that can be attached to any DNA strand in DNA nanostructure. Because of all the reasons mentioned above, I aimed to obtain a DNA nanostructure and then integrate sfGFP-DNA and mRuby3-DNA 1:1 adducts into rectangular DNA nanostructure because it allows to create a tool which has a control on positioning of proteins thanks to the chemistry reaction and DNA origami technique.

In this chapter, I have looked at the formation of rectangular DNA origami. The techniques for the formation through DNA origami were categorized in two as thermal assembly and isothermal assembly with and without denaturing reagents so far in the literature. In this context, the isothermal assembly without denaturation agent is preferred as I aim to include proteins in this structure. Isothermal assembly process differs from thermal assembly process as the way of temperature and the prevention of mismatches between scaffold and staples. The principle of isothermal assembly is that the formation of scaffold and staples occurs at a constant temperature. Among the temperatures of isothermal assembly process, the formation of scaffold and staples was conducted at 25°C, 37°C and 50°C in the literature. While the formation at 37°C and 50°C resulted in a reasonable yield, the formation at 25°C remained challenging. The prevention of mismatches between DNA strands is provided through controlling the salt ions instead of high temperature in thermal assembly. This formation of DNA nanostructure through isothermal assembly is controlled by two factors in isothermal assembly as concentration of NaCl and the ratio between scaffold and staples. It is critical to find the best salt concentration because while the low salt prevents the mismatches between DNA molecules, it will also disrupt the correct assembly of scaffold and staples. The significance of the ratio between scaffold and staples is that increases the possibility

of encountering of DNA molecules to meet in the right formation. Therefore, my purpose was to improve this assembly efficiency through optimizing these two factors since they are critical to obtain high efficiency of the structure for the isothermal assembly of rectangular DNA nanostructure 25°C.

6.1. Optimization of NaCl concentration for isothermal assembly

Since the concentration of salt ions is the main factor that prevents mismatches between small oligonucleotides in isothermal assembly, I tried to find the optimal salt concentration for the formation of rectangular DNA object. To optimize the salt concentration, I kept scaffold:staples ratio, buffer composition as constant and different concentrations. Then, I analyzed the DNA nanoobjects samples, which were incubated at different concentrations of NaCl, through the agarose gel which allows bulk estimation of the yield of structure formation, the purity of the structure and the extraction of the desired sample.

A
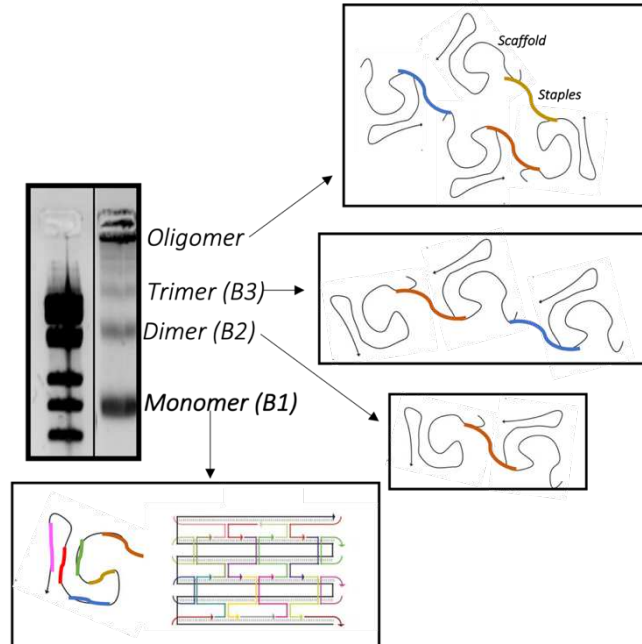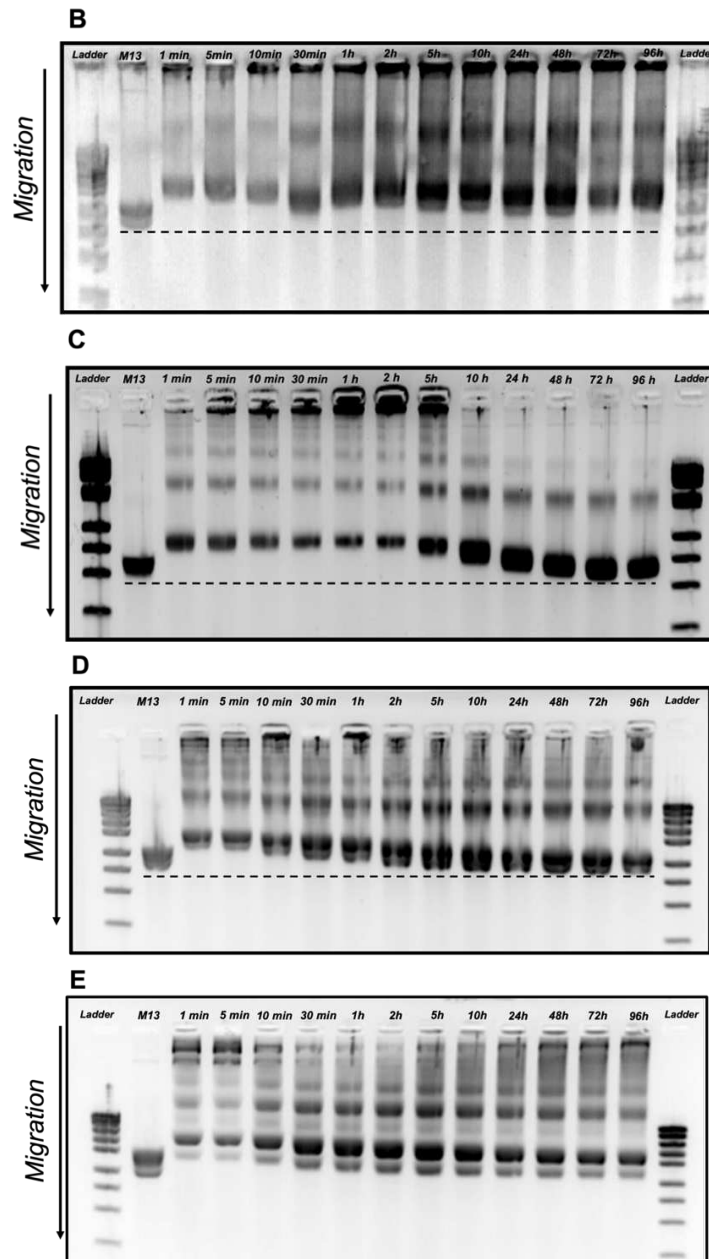
Figure 6.1. *Separation of oligomers and monomers of DNA nano object was shown.* **A** *There were many different bands as oligomers, dimers and monomers. The isothermal formation reaction for rectangular DNA nano object was performed in the presence of* **B** *20 mM NaCl* **C** *50 mM NaCl* **D** *100 mM NaCl* **E** *200 mM NaCl with the scaffold: staples ratio 1:5 for 96 hours at 25 ℃.* On each agarose gel image ladder (first well) and p7249 (M13mp18) scaffold (second well) were shown as reference. *The dash line is added to show the migration comparison between scaffold M13mp18 and the rectangular DNA origami samples, taken at certain times.* At each desired time, 15 μl of DNA nano objects was taken and frozen by the liquid nitrogen to stop the reaction to examine the impact of incubation time for the formation of DNA origami. *The rectangular DNA nano objects were obtained at certain times (minutes) and loaded on 1.5% agarose gel.*

As general description of gel images, the different bands were observed in figure 6.1.A. A big band aggregates on the right at the top are oligomers (many scaffolds with staples). There are bands in the middle that correspond to trimers and dimers (three and two scaffold respectively with staples) when it is partially folded as rectangular DNA. Also, there is band which is monomers (only one scaffold with a certain number of staples) at the bottom it changes its position according to how well folded it is. Because the principle of the migration relies on the quality of folding since as much as origami molecules fold, they become more compact and start to progress faster. Lastly, since the quantity of staples was 5 times more than scaffold, the excess staples were released at the end of reaction.

In figure 6.1.B, here because of lack of salt, the aggregation of oligomers of rectangular DNA at the top and the presence of dimers in the middle were increased by the time. Despite from obtaining rectangular DNA nano objects at the bottom of the gel, it is clear image that the lack of salt led these monomers do not fold properly in the gel. In Figure 6.1.C, the quantity of oligomers of rectangular DNA nano objects were dramatically decreased after 5 hours and this led the formation of the bands at the bottom of the gel image which represents monomers even though this formation took a while. The folding process was about to complete after 48 hours. At the same time, the oligomers, and dimers at 24 hours, 48 hours, 72 hours, 96 hours are less intense than the oligomers and dimers of the DNA nanoobject which are incubated less than 24 hours. It points out that misfolding is decreased by raising the incubation time. These results suggested that it is feasible to obtain monomer form of DNA objects in the presence of right amount of salt.

In the figure 6.1.D, the scaffold and staples encountered with each other after the 10 minutes incubation instead of 5 hours incubation as it was in figure 6.B. However, this formation is complicated because the formation of monomers progressed only slowly. In figure 6.E, the oligomerization of nano objects was not strongly observed here on the top of the gel image. However, usage of high amount of salt caused that the scaffolded and staples formed to a shape in a short time and the DNAs were stacked and hence progression of right formation became extremely slow. There is also another phenomenon which was observed in only figure 6.E. I observed two bands instead of one not only as a monomer form of DNA nano object but also as scaffold in

the first well. This suggests that the existence of the second band was not caused by the folding reaction.
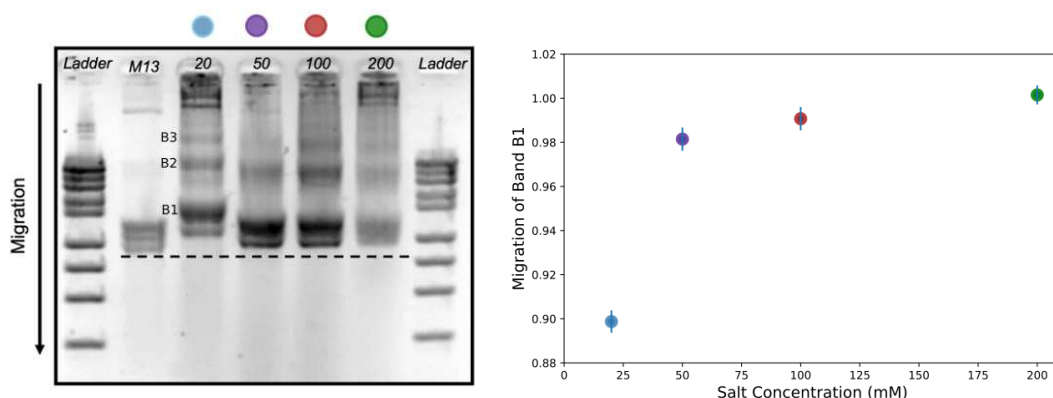
Figure 6.2. The migration difference increased in direct proportion to the NaCl concentration, and the greatest difference was observed between samples with 20 mM and 50 mM NaCl. **A** The rectangular DNA nano object with different concentrations of NaCl (mM) incubated during 96 hours at 25 ℃ were loaded on 1.5% agarose gel to observe the impact of concentration of NaCl on the formation of rectangular DNA nano objects. **B** The migration distance of band B1 for the samples, which were incubated with 20 mM (blue), 50mM (purple) ,100 mM (pink) and 200 mM NaCl (green) in the presence of 1:5 scaffold: staples ratio, were measured through Image J. Error bars were shown on scatter plot (vertical blue lines).

In Figure 6.2.A, the DNA nanoobjects, which were formed with different concentrations of NaCl, were analyzed through agarose gel to compare well the impact of NaCl concentrations on folding of DNA nano objects. Oligomers, dimers, and monomers were shown as B3, B2 and B1. The progression of assembled DNA nano objects in the presence of 100 and 200 mM NaCl, did not differ much. It possibly suggests that high amount of salt did not improve the folding efficiency after a certain amount of salt which is 50 mM.

Furthermore, I analyzed the migration distance of the monomer (B1) for all samples which were assembled in the presence of 20 mM ,50 mM, 100 mM and 200 mM NaCl through Image J software, to clearly show the correlation between NaCl concentration and formation of DNA origami samples in figure 6.2.B. The migration distance increased with the NaCl concentration, and the greatest difference was observed between samples with 20 mM and 50 mM NaCl. Between 50 mM-100 mM and 100 mM-200 mM, the migration was very similar.

Therefore, low salt concentration (< 50mM), results in high electrostatic repulsion and, thus reduced the formation of DNA nanoobject. In the presence of excess salt (> 50 mM), the formation of DNA nano objects falls into a kinetic trap, and this can be seen as formation is complicated and progressed slowly. However, when the salt concentration is around 50 mM, the results show that the monomers formation progressed very well. Therefore, I concluded that 50 mM NaCl is the optimal concentration of salt to provide the balance between preventing the mismatches and supporting the folding of rectangular DNA Origami.

6.2. Optimization of staples scaffold ratio for isothermal assembly

Another important parameter was the ratio between scaffold and staples to provide the isothermal formation of rectangular DNA nano objects. The increase of ratio between scaffold and staples decreases the possibility of mismatches in the sequence of the rectangular DNA origami. In this section, I examined the effect of scaffold: staples ratio on the folding of the DNA origami and wanted to observe correlation between the scaffold: staples ratio versus formation of the DNA nano object and find out the optimal ratio in my case. To achieve this, I kept the salt concentration, buffer composition as constant and altered the ratio of scaffold: staples as 1:2, 1:5, 1:10, 1:20 and 1:40. The reaction depends strongly on the number of the staples. Staples ratio and reaction speed are directly proportional. However, this proportion can depend on other factors. For example, if the staples have a protein attached, the number of staples should probably be low. Excess staples can also reduce formation of completely folded structures. Other phenomena which is controlled by number staples is oligomer formation.
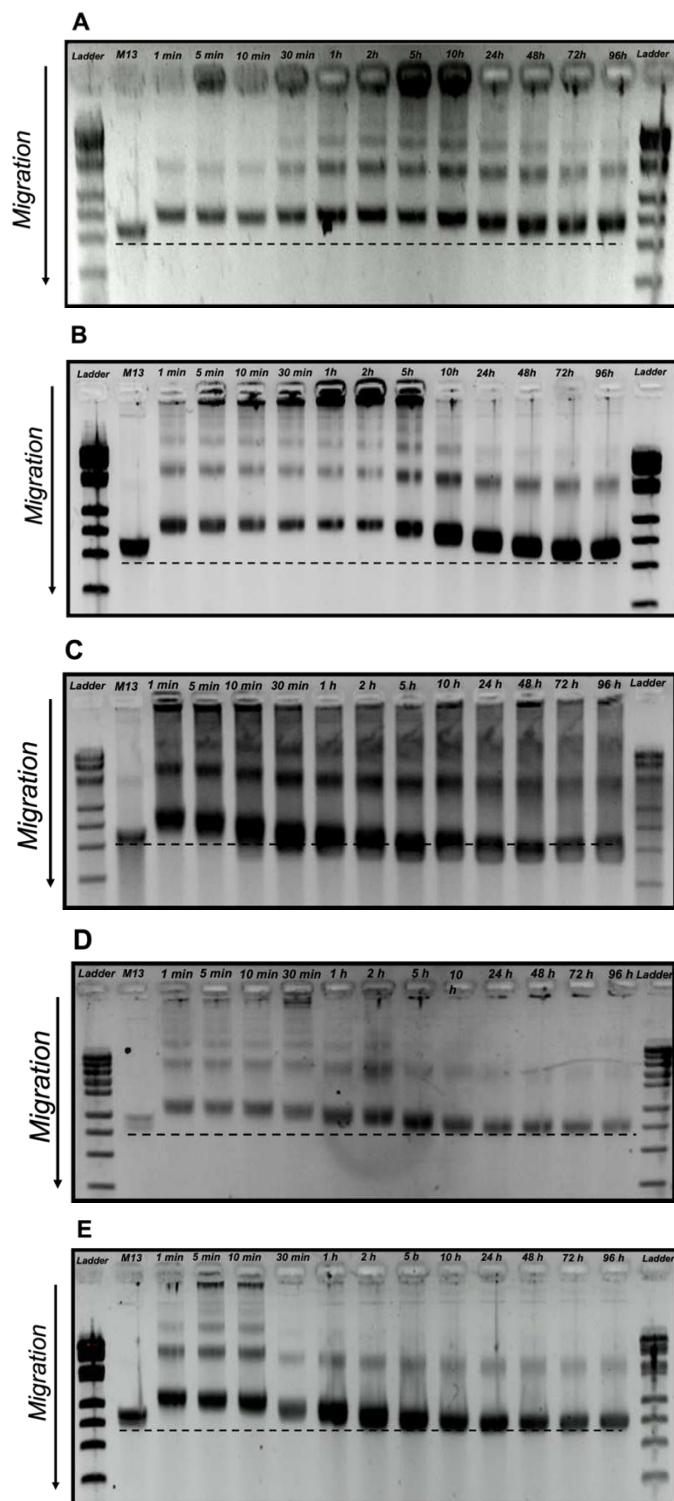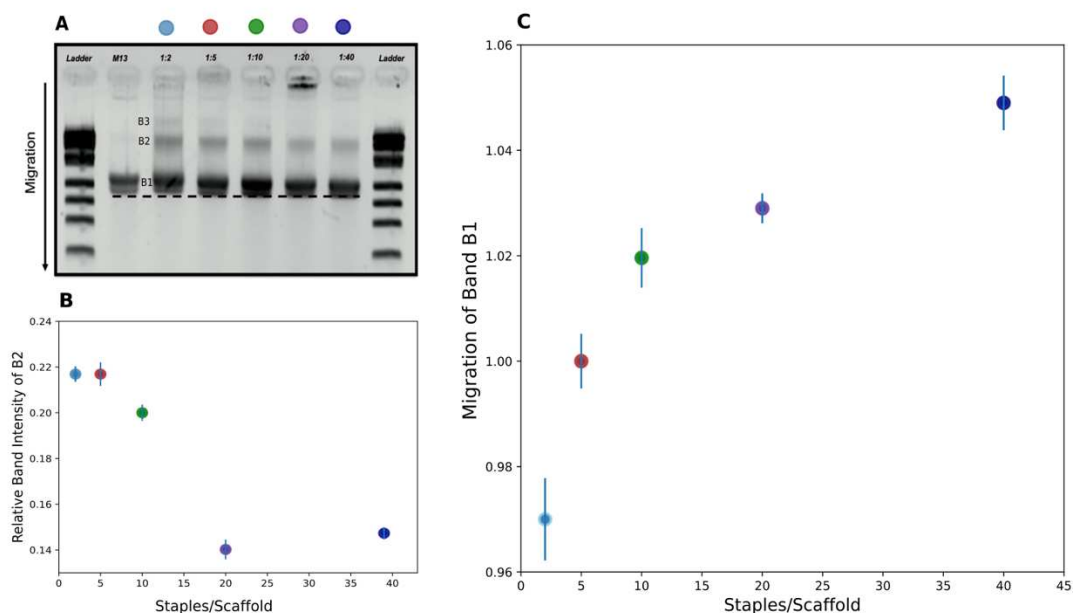
Figure 6.3 *Increase of scaffold: staples ratio up to 1:40 prevented oligomerization of rectangular DNA nano objects on the 1.5% agarose gel. The isothermal formation reaction for rectangular DNA nano objects was performed in the presence of 50 mM NaCl with the scaffold: staples ratio **A** 1:2 **B** 1:5 **C** 1:10 **D** 1:20 **E** 1:40 for 96 hours at 25℃.*

As explained in figure 6.1, there are different bands in here figure 6.3. In the wells of the gel image, there is aggregation of oligomers and oligomers is followed by trimer, dimer, and monomer respectively from top to the bottom of the gel. Each DNA origami sample at each different time points was shown in Figure 6.3. On each agarose gel image ladder (first well) and p7249 (M13mp18) scaffold (second well) were shown as reference.

In figure 6.3.A, the folding reaction completed after 48 hours incubation. The possible reason of that having low number of staples caused the low speed of formation reaction. In the presence of scaffold: staples (1:5) ratio, after 5 hours incubation, the migration is faster due to folding of rectangular nano object getting improved. Compared to the 1:2 ratio in figure 6.3.A, the samples here in figure 6.3.B not only migrated faster in a short incubation time but also the folding efficiency was drastically improved at the end of 96 hours incubation compared to the beginning of incubation with the effect of increasing the ratio as I expected. In the figure 6.3.C, the folding path of the DNA nano object improved even after 30 minutes incubation and the speed of migration was stabilized at 5 hours incubation. This suggests that the folding reaction became faster since the ratio between scaffold and staples was 1:10.

In figure 6.3.D, the starting point of folding time was short as 30 minutes. At the end of 96h incubation, the monomer DNA nano object migrated more compared to the one with 1:10 ratio. This points out that increase of number of staples improves the formation of the nano object. Nevertheless, after 10 hours incubation the trimers and dimers of the nano object which existed in the presence of 1:2, 1:5 and 1:10 were disappeared. It is likely that the formation of dimers and trimers was prevented through the combination of longer incubation time as 10 hours and the presence of sufficient scaffold: staples ratio. In figure 6.3.E, the improvement of the formation of monomer nano objects started around 1 hour as same as the samples with 1:20 ratio. The band which represents trimer and dimer formations become to lose its intensity at 1 hour incubation which is earlier than the sample with 1:20 ratio. This shows that increasing the ratio of scaffold:staples likely reduced dimer and trimer formations of the nano object. In the image gel of 1:10, 1:20 and 1:40 ratio, the aggregates in the well of the gels and intensity of the band which represents dimer forms dramatically decreased comparing to the one's ratio with 1:2 and 1:5. In the presence of 1:20 and 1:40 ratio,

the band of trimer forms which exist at 1:2, 1:5 and 1:10 were completely disappeared after 96 hours incubation.

Figure 6.4*. The intensity of band B2 decreased in direct proportion to the increase of scaffold: staples ratio and the greatest difference was observed between samples with the ratio of 1:10 (green) and 1:20 (purple).* ***A*** *The rectangular origami samples with different scaffold: staples ratio incubated in the presence of 50 mM NaCl for 96 hours were loaded on 1.5% agarose gel, to observe the impact of scaffold: staples ratio on the formation of rectangular DNA nano objects.* ***B*** *The intensity of band B2 for the samples, which were incubated with the scaffold: staples ratio of 1:2 (blue), 1:5 (pink) 1:10 (green), 1:20 (purple), 1:40 (blue) were measured through Image J.* ***C*** *The migration distance of band B1 for the samples, which were incubated with 1:2 (blue),1:5(pink) 1:10 (green),1:20 (purple),1:40 (blue) in the presence of 50 mM NaCl, were measured through Image J. Error bars were shown on scatter plot (vertical blue lines) for both B and C.*

In figure 6.4.A, I examined all the samples which are incubated for 96 hours with different ratios of scaffold: staples through agarose gel to analyze the difference between them. It is obvious that while the band which represent trimer form of DNA nano object exist in the presence of 1:2 ratio, it becomes to disappear by the increase of the ratio. The fastest migration of monomer form of rectangular DNA nanoobject was observed in the sample with ratio 1:40 (see Figure 6.4.C). The correlation between the scaffold: staples ratio and the intensity of B2, which represents the dimerization of DNA nano object, was analyzed by Image J and shown in figure 6.4.B.

The biggest difference of the intensity of B2 was observed when the ratio changed from 1:10 to 1:20. It points out that 1:10 ratio is required to avoid having trimerized form of DNA nano objects. Increase of scaffold: staples ratio improved both the formation of assembly and prevented from oligomerization of assembly. There is the possibility at high ratios of not getting all assembled properly. Therefore, the ratio 1:10 was possibly optimal conditions for an efficient isothermal assembly reaction.

In conclusion, I optimized the isothermal assembly protocol for rectangular origami

without any denaturing agents. The ratio > 1:10 and 50 mM NACI was possibly optimal conditions for an efficient isothermal assembly reaction. All explained in detail above showed us the concentration of salt and the scaffold: staples ratio was mainly the driver of the isothermal reaction. It is also important to keep the salt concentration under control since in case of excess NaCI, the formation process of the origami likely be disrupted. As scaffold: staples ratio, I observed increasement of the quantity of the staples led to the prevention of oligomerization of DNA nano object. Since DNA agarose gel technique provide a bulk estimation, investigating the sample through negative staining and then visualizing with electron microscopy would inform the shape of the structure as 2 D.

Bibliography

[1]   L.A. Pray, Discovery of DNA Double Helix : Watson and Crick | Learn Science at Scitable Discovery of DNA Structure and Function : Watson and Crick The landmark ideas of Watson and Crick relied heavily on the work of other scientists . What did the duo Aa The First P, Nat. Educ. 1 (2008) 100.

[2]   P.A. Levene, W.A. Jacobs, Über die Pentose in den Nucleinsäuren, Berichte Der Dtsch. Chem. Gesellschaft. 42 (1909) 2102–2106. https://doi.org/10.1002/cber.19090420295.

[3]   A.U. SMITH, C. Polge, © 1950 Nature Publishing Group, Nat. Publ. Gr. (1950) 2.

[4]   J. Lewis, M. Raff, K. Roberts, Molecular Biology of the Cell (4th Ed), J. Biol. Educ. 37 (2002) 45–47. https://doi.org/10.1080/00219266.2002.9655847.

[5]   A. Travers, G. Muskhelishvili, DNA structure and function, FEBS J. 282 (2015) 2279–2295. https://doi.org/10.1111/febs.13307.

[6]   D. Cleden, The Anatomy of a Bid, Bid Writ. Proj. Manag. 0 (2019) 15–28. https://doi.org/10.4324/9781315569345-2.

[7]   J.C. Wang, Helical repeat of DNA in solution., Proc. Natl. Acad. Sci. U. S. A. 76 (1979) 200–203. https://doi.org/10.1073/pnas.76.1.200.

[8]   J.B. Chaires, N. Dattagupta, D.M. Crothers, Studies on Interaction of Anthracycline Antibiotics and Deoxyribonucleic Acid: Equilibrium Binding Studies on Interaction of Daunomycin with Deoxyribonucleic Acid, Biochemistry. 21 (1982) 3933–3940. https://doi.org/10.1021/bi00260a005.

[9]   S.Y. Lee, C.J. Lim, P. Dröge, J. Yan, Regulation of Bacterial DNA Packaging in Early Stationary Phase by Competitive DNA Binding of Dps and IHF, Sci. Rep. 5 (2015) 1–10. https://doi.org/10.1038/srep18146.

[10]   D. Bikard, C. Loot, Z. Baharoglu, D. Mazel, Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes, Microbiol. Mol. Biol. Rev. 74 (2010) 570–588. https://doi.org/10.1128/mmbr.00026-10.

[11]   S. Bayda, M. Adeel, T. Tuccinardi, M. Cordani, F. Rizzolio, The history of nanoscience and nanotechnology: From chemical-physical applications to nanomedicine, Molecules. 25 (2020) 1–15. https://doi.org/10.3390/molecules25010112.

[12]   N.C. Seeman, DNA in a material world, Nature. 421 (2003) 427–431. https://doi.org/10.1038/nature01406.

[13]   N.C. Seeman, H.F. Sleiman, DNA nanotechnology, Nat. Rev. Mater. 3 (2017). https://doi.org/10.1038/natrevmats.2017.68.

[14]   R. Holliday, A mechanism for gene conversion in fungi, Genet. Res. (Camb). 89 (2008) 285–307. https://doi.org/10.1017/S0016672308009476.

[15]   M.L. Petrillo, C.J. Newton, R.P. Cunningham, R. -I Ma, N.R. Kallenbach, N.C. Seeman, The ligation and flexibility of four-arm DNA junctions, Biopolymers. 27 (1988) 1337–1352. https://doi.org/10.1002/bip.360270902.

[16]   N.C. Seeman, Nucleic acid junctions and lattices, J. Theor. Biol. 99 (1982) 237–247. https://doi.org/10.1016/0022-5193(82)90002-9.

[17]   M.H. Caruthers, Gene synthesis machines: DNA chemistry and its uses, Science (80-.). 230 (1985) 281–285. https://doi.org/10.1126/science.3863253.

[18]   F. Mathieu, S. Liao, J. Kopatsch, T. Wang, C. Mao, N.C. Seeman, Six-helix bundles designed from DNA, Nano Lett. 5 (2005) 661–665. https://doi.org/10.1021/nl050084f.

[19]   J. Zheng, J.J. Birktoft, Y. Chen, T. Wang, R. Sha, P.E. Constantinou, S.L. Ginell, C. Mao, N.C. Seeman, From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal, Nature. 461 (2009) 74–77. https://doi.org/10.1038/nature08274.

[20]   B. Wei, M. Dai, P. Yin, Complex shapes self-assembled from single-stranded DNA tiles, Nature. 485 (2012) 623–626. https://doi.org/10.1038/nature11075.

[21]   Y. Ke, L.L. Ong, W.M. Shih, P. Yin, Three-dimensional structures self-assembled from DNA bricks, Science (80-.). 338 (2012) 1177–1183. https://doi.org/10.1126/science.1227268.

[22]   L.L. Ong, N. Hanikel, O.K. Yaghi, C. Grun, M.T. Strauss, P. Bron, J. Lai-Kee-Him, F. Schueder, B. Wang, P. Wang, J.Y. Kishi, C. Myhrvold, A. Zhu, R. Jungmann, G. Bellot, Y. Ke, P. Yin, Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components, Nature. 552 (2017) 72–77. https://doi.org/10.1038/nature24648.

[23]   H.S. Ong, M.S. Rahim, M. Firdaus-Raih, E.I. Ramlan, DNA tetrominoes: The construction of DNA nanostructures using self-organised heterogeneous deoxyribonucleic acids shapes, PLoS One. 10 (2015). https://doi.org/10.1371/journal.pone.0134520.

[24]   P.W.K. Rothemund, Folding DNA to create nanoscale shapes and patterns, Nature. 440 (2006) 297–302. https://doi.org/10.1038/nature04586.

[25]   C.E. Castro, F. Kilchherr, D.N. Kim, E.L. Shiao, T. Wauer, P. Wortmann, M. Bathe, H. Dietz, A primer to scaffolded DNA origami, Nat. Methods. 8 (2011) 221–229. https://doi.org/10.1038/nmeth.1570.

[26]   S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W.M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes, Nature. 459 (2009) 414–418. https://doi.org/10.1038/nature08016.

[27]   H. Dietz, S.M. Douglas, W.M. Shih, Folding DNA into twisted and curved nanoscale shapes, Science (80-.). 325 (2009) 725–730. https://doi.org/10.1126/science.1174251.

[28]   C. Ducani, C. Kaul, M. Moche, W.M. Shih, B. Högberg, Enzymatic production of "monoclonal stoichiometric" single-stranded DNA oligonucleotides, Nat. Methods. 10 (2013) 647–652. https://doi.org/10.1038/nmeth.2503.

[29]   F. Praetorius, B. Kick, K.L. Behler, M.N. Honemann, D. Weuster-Botz, H. Dietz, Biotechnological mass production of DNA origami, Nature. 552 (2017) 84–87. https://doi.org/10.1038/nature24650.

[30]   G. Tikhomirov, P. Petersen, L. Qian, Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns, Nature. 552 (2017) 67–71. https://doi.org/10.1038/nature24655.

[31]   K.F. Wagenbauer, C. Sigl, H. Dietz, Gigadalton-scale shape-programmable DNA assemblies,

Nature. 552 (2017) 78–83. https://doi.org/10.1038/nature24651.

[32]   B. Babatunde, D.S. Arias, J. Cagan, R.E. Taylor, Generating dna origami nanostructures
        through shape annealing, Appl. Sci. 11 (2021). https://doi.org/10.3390/app11072950.

[33]   P.W.K. Rothemund, Folding DNA to create nanoscale shapes and patterns, Nature. 440
        (2006) 297–302. https://doi.org/10.1038/nature04586.

[34]   G. Bellot, M.A. McClintock, C. Lin, W.M. Shih, Recovery of intact DNA nanostructures after
        agarose gel-based separation, Nat. Methods. 8 (2011) 192–194.
        https://doi.org/10.1038/nmeth0311-192.

[35]   A. Shaw, E. Benson, B. Högberg, Purification of functionalized DNA origami nanostructures,
        ACS Nano. 9 (2015) 4968–4975. https://doi.org/10.1021/nn507035g.

[36]   K.F. Wagenbauer, F.A.S. Engelhardt, E. Stahl, V.K. Hechtl, P. Stömmer, F. Seebacher, L.
        Meregalli, P. Ketterer, T. Gerling, H. Dietz, How We Make DNA Origami, ChemBioChem. 18
        (2017) 1873–1885. https://doi.org/10.1002/cbic.201700377.

[37]   E. Stahl, T.G. Martin, F. Praetorius, H. Dietz, Facile and scalable preparation of pure and
        dense DNA origami solutions, Angew. Chemie - Int. Ed. 53 (2014) 12735–12740.
        https://doi.org/10.1002/anie.201405991.

[38]   C. Kielar, Y. Xin, B. Shen, M.A. Kostiainen, G. Grundmeier, V. Linko, A. Keller, On the Stability
        of DNA Origami Nanostructures in Low-Magnesium Buffers, Angew. Chemie - Int. Ed. 57
        (2018) 9470–9474. https://doi.org/10.1002/anie.201802890.

[39]   X. Qiu, K. Andresen, L.W. Kwok, J.S. Lamb, H.Y. Park, L. Pollack, Inter-DNA attraction
        mediated by divalent counterions, Phys. Rev. Lett. 99 (2007) 1–4.
        https://doi.org/10.1103/PhysRevLett.99.038104.

[40]   S. Fischer, C. Hartl, K. Frank, J.O. Rädler, T. Liedl, B. Nickel, Shape and interhelical spacing
        of DNA origami nanostructures studied by small-angle X-ray scattering, Nano Lett. 16 (2016)
        4282–4287. https://doi.org/10.1021/acs.nanolett.6b01335.

[41]   B. Wei, L.L. Ong, J. Chen, A.S. Jaffe, P. Yin, Complex reconfiguration of DNA nanostructures,
        Angew. Chemie - Int. Ed. 53 (2014) 7475–7479. https://doi.org/10.1002/anie.201402437.

[42]   Y. Ke, L.L. Ong, W. Sun, J. Song, M. Dong, W.M. Shih, P. Yin, DNA brick crystals with
        prescribed depths, Nat. Chem. 6 (2014) 994–1002. https://doi.org/10.1038/nchem.2083.

[43]   J. Fu, M. Liu, Y. Liu, N.W. Woodbury, H. Yan, Interenzyme substrate diffusion for an enzyme
        cascade organized on spatially addressable DNA nanostructures, J. Am. Chem. Soc. 134
        (2012) 5516–5519. https://doi.org/10.1021/ja300897h.

[44]   R. Jungmann, T. Liedl, T.L. Sobey, W. Shih, F.C. Simmel, Supporting Information for
        Isothermal assembly of DNA origami structures using denaturing agents, Structure. i (2008) 1–
        15.

[45]   J.P.J. Sobczak, T.G. Martin, T. Gerling, H. Dietz, Rapid folding of DNA into nanoscale shapes
        at constant temperature, Science (80-.). 338 (2012) 1458–1461.
        https://doi.org/10.1126/science.1229919.

[46]   J. Song, Z. Zhang, S. Zhang, L. Liu, Q. Li, E. Xie, K.V. Gothelf, F. Besenbacher, M. Dong,

Isothermal hybridization kinetics of DNA assembly of two-dimensional DNA origami, Small. 9 (2013) 2954–2959. https://doi.org/10.1002/smll.201202861.

[47]   A. Kopielski, A. Schneider, A. Csáki, W. Fritzsche, Isothermal DNA origami folding: Avoiding denaturing conditions for one-pot, hybrid-component annealing, Nanoscale. 7 (2015) 2102–2106. https://doi.org/10.1039/c4nr04176c.

[48]   P. Yin, C. Myhrvold, M. Dai, P.A. Silver, Isothermal Self-Assembly of Complex DNA Structures under Diverse and Biocompatible Conditions, Nano Lett. (2013) 130826090418005. http://pubs.acs.org/doi/abs/10.1021/nl4019512%0Apapers3://publication/doi/10.1021/nl401951 2.

[49]   I. Gállego, M.A. Grover, N. V. Hud, Folding and imaging of DNA nanostructures in anhydrous and hydrated deep-eutectic solvents, Angew. Chemie - Int. Ed. 54 (2015) 6765–6769. https://doi.org/10.1002/anie.201412354.

[50]   J.R. Hutton, Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea, Nucleic Acids Res. 4 (1977) 3537–3555. https://doi.org/10.1093/nar/4.10.3537.

[51]   Z. Zhang, J. Song, F. Besenbacher, M. Dong, K. V. Gothelf, Self-assembly of DNA origami and single-stranded tile structures at room temperature, Angew. Chemie - Int. Ed. 52 (2013) 9219–9223. https://doi.org/10.1002/anie.201303611.

[52]   B. Saccà, R. Meyer, M. Erkelenz, K. Kiko, A. Arndt, H. Schroeder, K.S. Rabe, C.M. Niemeyer, Orthogonal protein decoration of DNA origami, Angew. Chemie - Int. Ed. 49 (2010) 9378–9383. https://doi.org/10.1002/anie.201005931.

[53]   D.F. Savage, B. Afonso, A.H. Chen, P. a Silver, Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery David F. Savage,* Bruno Afonso,* Anna H. Chen, Pamela A. Silver† Cyanobacterial, Science (80-.). 1641 (2010) 1–4.

[54]   L.H. Tan, H. Xing, Y. Lu, DNA as a powerful tool for morphology control, spatial positioning, and dynamic assembly of nanoparticles, Acc. Chem. Res. 47 (2014) 1881–1890. https://doi.org/10.1021/ar500081k.

[55]   G. Ke, M. Liu, S. Jiang, X. Qi, Y.R. Yang, S. Wootten, F. Zhang, Z. Zhu, Y. Liu, C.J. Yang, H. Yan, Directional Regulation of Enzyme Pathways through the Control of Substrate Channeling on a DNA Origami Scaffold, Angew. Chemie - Int. Ed. 55 (2016) 7483–7486. https://doi.org/10.1002/anie.201603183.

[56]   Q. Jiang, C. Song, J. Nangreave, X. Liu, L. Lin, D. Qiu, Z.G. Wang, G. Zou, X. Liang, H. Yan, B. Ding, DNA origami as a carrier for circumvention of drug resistance, J. Am. Chem. Soc. 134 (2012) 13396–13403. https://doi.org/10.1021/ja304263n.

[57]   O. Boutureira, G.J.L. Bernardes, Advances in chemical protein modification, Chem. Rev. 115 (2015) 2174–2195. https://doi.org/10.1021/cr500399p.

[58]   N. Stephanopoulos, M.B. Francis, Choosing an effective protein bioconjugation strategy, Nat. Chem. Biol. 7 (2011) 876–884. https://doi.org/10.1038/nchembio.720.

[59]   H. Fraenkel-Conrat, H.S. Olcott, Reaction of Formaldehyde With Proteins, J. Biol. Chem. 174

(1948) 827–843. https://doi.org/10.1016/s0021-9258(18)57292-6.

[60]   M.J. Hangauer, C.R. Bertozzi, A FRET-based fluorogenic phosphine for live-cell imaging with the Staudinger ligation, Angew. Chemie - Int. Ed. 47 (2008) 2394–2397. https://doi.org/10.1002/anie.200704847.

[61]   P.E. Dawson, M.C. Fitzgerald, T.W. Muir, S.B.H. Kent, Methods for the chemical synthesis and readout of self-encoded arrays of polypeptide analogues, J. Am. Chem. Soc. 119 (1997) 7917–7927. https://doi.org/10.1021/ja963122t.

[62]   S.B. Sun, P.G. Schultz, C.H. Kim, Therapeutic applications of an expanded genetic code, ChemBioChem. 15 (2014) 1721–1729. https://doi.org/10.1002/cbic.201402154.

[63]   K.M. Cook, P.J. Hogg, Post-translational control of protein function by disulfide bond cleavage, Antioxidants Redox Signal. 18 (2013) 1987–2015. https://doi.org/10.1089/ars.2012.4807.

[64]   I. Braakman, N.J. Bulleid, Protein folding and modification in the mammalian endoplasmic reticulum, Annu. Rev. Biochem. 80 (2011) 71–99. https://doi.org/10.1146/annurev-biochem-062209-093836.

[65]   I. Azimi, L.J. Matthias, R.J. Center, J.W.H. Wong, P.J. Hogg, Disulfide bond that constrains the HIV-1 gp120 V3 domain is cleaved by thioredoxin, J. Biol. Chem. 285 (2010) 40072–40080. https://doi.org/10.1074/jbc.M110.185371.

[66]   A.G. Beck-Sickinger, K. Mörl, Posttranslational Modification of Proteins. Expanding Nature's Inventory. By Christopher T. Walsh., Angew. Chemie Int. Ed. 45 (2006) 1020–1020. https://doi.org/10.1002/anie.200585363.

[67]   F. Ardito, M. Giuliani, D. Perrone, G. Troiano, L. Lo Muzio, The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review), Int. J. Mol. Med. 40 (2017) 271–280. https://doi.org/10.3892/ijmm.2017.3036.

[68]   D.J. Mandell, I. Chorny, E.S. Groban, S.E. Wong, E. Levine, C.S. Rapp, M.P. Jacobson, Strengths of hydrogen bonds involving phosphorylated amino acid side chains, J. Am. Chem. Soc. 129 (2007) 820–827. https://doi.org/10.1021/ja063019w.

[69]   L.N. Johnson, R.J. Lewis, ChemInform Abstract: Structural Basis for Control by Phosphorylation, ChemInform. 32 (2010) no-no. https://doi.org/10.1002/chin.200140284.

[70]   T. Pawson, Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems, Cell. 116 (2004) 191–203. https://doi.org/10.1016/S0092-8674(03)01077-8.

[71]   J. Guo, X.Q. Yang, X.T. He, N.N. Wu, J.M. Wang, W. Gu, Y.Y. Zhang, Limited aggregation behavior of β-conglycinin and its terminating effect on glycinin aggregation during heating at pH 7.0, J. Agric. Food Chem. 60 (2012) 3782–3791. https://doi.org/10.1021/jf300409y.

[72]   R. Aebersold, J.N. Agar, I.J. Amster, M.S. Baker, C.R. Bertozzi, E.S. Boja, C.E. Costello, B.F. Cravatt, C. Fenselau, B.A. Garcia, Y. Ge, J. Gunawardena, R.C. Hendrickson, P.J. Hergenrother, C.G. Huber, A.R. Ivanov, O.N. Jensen, M.C. Jewett, N.L. Kelleher, L.L. Kiessling, N.J. Krogan, M.R. Larsen, J.A. Loo, R.R. Ogorzalek Loo, E. Lundberg, M.J. Maccoss, P. Mallick, V.K. Mootha, M. Mrksich, T.W. Muir, S.M. Patrie, J.J. Pesavento, S.J.

Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S.A. Slavoff, L.M. Smith, M.P. Snyder, P.M. Thomas, M. Uhlén, J.E. Van Eyk, M. Vidal, D.R. Walt, F.M. White, E.R. Williams, T. Wohlschlager, V.H. Wysocki, N.A. Yates, N.L. Young, B. Zhang, How many human proteoforms are there?, Nat. Chem. Biol. 14 (2018) 206–214. https://doi.org/10.1038/nchembio.2576.

[73]   R.G. Spiro, Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, Glycobiology. 12 (2002) 43–56. https://doi.org/10.1093/glycob/12.4.43R.

[74]   M. Aebi, N-linked protein glycosylation in the ER, Biochim. Biophys. Acta - Mol. Cell Res. 1833 (2013) 2430–2437. https://doi.org/10.1016/j.bbamcr.2013.04.001.

[75]   D.N. Hebert, L. Lamriben, E.T. Powers, J.W. Kelly, The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis, Nat. Chem. Biol. 10 (2014) 902–910. https://doi.org/10.1038/nchembio.1651.

[76]   P. Van Den Steen, P.M. Rudd, R.A. Dwek, G. Opdenakker, Concepts and Principles of O-Linked Glycosylation, 33 (1998) 151–208.

[77]   H. LIS, N. SHARON, Protein glycosylation: Structural and functional aspects, Eur. J. Biochem. 218 (1993) 1–27. https://doi.org/10.1111/j.1432-1033.1993.tb18347.x.

[78]   H. Xiao, F. Sun, S. Suttapitugsakul, R. Wu, Global and site-specific analysis of protein glycosylation in complex biological systems with Mass Spectrometry, Mass Spectrom. Rev. 38 (2019) 356–379. https://doi.org/10.1002/mas.21586.

[79]   L. Wang, J. Xie, P.G. Schultz, Expanding the genetic code, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 225–249. https://doi.org/10.1146/annurev.biophys.35.101105.121507.

[80]   D.W. Schultz, M. Yarus, On Malleability in the Genetic Code, J. Mol. Evol. 42 (1996) 597–601. https://doi.org/10.1007/BF02352290.

[81]   Y. Takaoka, A. Ojida, I. Hamachi, Protein organic chemistry and applications for labeling and engineering in live-cell systems, Angew. Chemie - Int. Ed. 52 (2013) 4088–4106. https://doi.org/10.1002/anie.201207089.

[82]   G.E. Means, R.E. Feeney, Chemical Modifications of Proteins: History and Applications, Bioconjug. Chem. 1 (1990) 2–12. https://doi.org/10.1021/bc00001a001.

[83]   M.E. Bar, M.P. Damborsky, E.B. Oscherov, C. Wisnivesky-colli, A. Ferrari, F.S. Weill, M.L. Paz, J. Leoni, E. Coghlan, L.B. Quero, M. Schwab, D. Pellegrini, H. Trimarchi, D. Cocozzella, C. Méndez, J. Malacalza, M.G. Cersósimo, S. Juri, S.S. De Chandler, R. Clerici, F.E. Micheli, P. Foresto, M.D. Arrigo, F. Filippini, M.C. Buys, L.N. Guerra, B. Martín, C.E. Miranda, I. Torrejón, T. Garrot, E. Malvino, M.C. Bruno, J.P. Gallo, J.C. Medrano, D. Ferrante, L. Schwartz, J.A. Barcat, L. Schwartz, 170 173 175, 65 (2005).

[84]   M.Z. Lin, L. Wang, Selective labeling of proteins with chemical probes in living cells, Physiology. 23 (2008) 131–141. https://doi.org/10.1152/physiol.00007.2008.

[85]   C. Study, L. Milroy, L. Brunsveld, C. Ottmann, Stabilization and Inhibition of Protein − Protein

Interactions : The 14-, (2012).

[86] J.A. Shadish, C.A. DeForest, Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials, Matter. 2 (2020) 50–77. https://doi.org/10.1016/j.matt.2019.11.011.

[87] T. Hirata, T. Terai, H. Yamamura, M. Shimonishi, T. Komatsu, K. Hanaoka, T. Ueno, Y. Imaizumi, T. Nagano, Y. Urano, Protein-Coupled Fluorescent Probe to Visualize Potassium Ion Transition on Cellular Membranes, Anal. Chem. 88 (2016) 2693–2700. https://doi.org/10.1021/acs.analchem.5b03970.

[88] A.E. Habibi, K. Khajeh, M. Nemat-Gorgani, Chemical modification of lysine residues in Bacillus licheniformis α-amylase: Conversion of an endo- to an exo-type enzyme, J. Biochem. Mol. Biol. 37 (2004) 642–647. https://doi.org/10.5483/bmbrep.2004.37.6.642.

[89] J.E. Yeo, S. Wickramaratne, S. Khatwani, Y.C. Wang, J. Vervacke, M.D. Distefano, N.Y. Tretyakova, Synthesis of site-specific DNA-protein conjugates and their effects on DNA replication, ACS Chem. Biol. 9 (2014) 1860–1868. https://doi.org/10.1021/cb5001795.

[90] N. Nischan, C.P.R. Hackenberger, Site-specific PEGylation of proteins: Recent developments, J. Org. Chem. 79 (2014) 10727–10733. https://doi.org/10.1021/jo502136n.

[91] T. Rutherford, M. Kibrick, M. Burchinal, L. Richland, K. Osborne, S. Schneider, L. Duran, A. Coulson, 1 of 26, Mind. (2010) 1–26.

[92] Hemaprabha E, Chemical Crosslinking of Proteins: a Review, J. Pharm. Sci. Innov. 1 (2012) 22–26. www.jpsionline.com.

[93] M.T. Skoog, W.P. Jencks, Reactions of Pyridines and Primary Amines with N-Phosphorylated Pyridines, J. Am. Chem. Soc. 106 (1984) 7597–7606. https://doi.org/10.1021/ja00336a047.

[94] E.M. Sletten, C.R. Bertozzi, Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality, Angew. Chemie - Int. Ed. 48 (2009) 6974–6998. https://doi.org/10.1002/anie.200900942.

[95] S. Mädler, C. Bich, D. Touboul, R. Zenobi, Chemical cross-linking with NHS esters: A systematic study on amino acid reactivities, J. Mass Spectrom. 44 (2009) 694–706. https://doi.org/10.1002/jms.1544.

[96] G.T. Hermanson, Functional Targets for Bioconjugation, 2013. https://doi.org/10.1016/b978-0-12-382239-0.00002-9.

[97] S. Crotti, H. Zhai, J. Zhou, M. Allan, D. Proietti, W. Pansegrau, Q.Y. Hu, F. Berti, R. Adamo, Defined conjugation of glycans to the lysines of CRM197 guided by their reactivity mapping, ChemBioChem. 15 (2014) 836–843. https://doi.org/10.1002/cbic.201300785.

[98] L.M. Wysocki, L.D. Lavis, Advances in the chemistry of small molecule fluorescent probes, Curr. Opin. Chem. Biol. 15 (2011) 752–759. https://doi.org/10.1016/j.cbpa.2011.10.013.

[99] Y. Urano, D. Asanuma, Y. Hama, Y. Koyama, T. Barrett, M. Kamiya, T. Nagano, T. Watanabe, A. Hasegawa, P.L. Choyke, H. Kobayashi, Selective molecular imaging of viable cancer cells with pH-activatable fluorescence probes, Nat. Med. 15 (2009) 104–109. https://doi.org/10.1038/nm.1854.

[100] V. Dufour, M. Stahl, C. Baysse, The antibacterial properties of isothiocyanates, Microbiol.

(United Kingdom). 161 (2015) 229–243. https://doi.org/10.1099/mic.0.082362-0.

[101]   L. Robert, F.S. Penaranda, Studies on aldehyde-protein interactions. I. Reaction of amino acids with lower aldehydes, J. Polym. Sci. 12 (1954) 337–350. https://doi.org/10.1002/pol.1954.120120128.

[102]   A. Sinz, Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes, J. Mass Spectrom. 38 (2003) 1225–1237. https://doi.org/10.1002/jms.559.

[103]   R.B. Pepinsky, D. Cappiello, C. Wilkowski, V.M. Vogt, Chemical crosslinking of proteins in avian sarcoma and leukemia viruses, Virology. 102 (1980) 205–210. https://doi.org/10.1016/0042-6822(80)90081-1.

[104]   A. Narayanan, L.H. Jones, Sulfonyl fluorides as privileged warheads in chemical biology, Chem. Sci. 6 (2015) 2650–2659. https://doi.org/10.1039/c5sc00408j.

[105]   H. Fraenkel-Conrat, L.K. Ramachandran, Structural Aspects of Tobacco Mosaic Virus, Adv. Protein Chem. 14 (1959) 175–229. https://doi.org/10.1016/S0065-3233(08)60611-7.

[106]   J.M. Chalker, G.J.L. Bernardes, Y.A. Lin, B.G. Davis, Chemical modification of proteins at cysteine: Opportunities in chemistry and biology, Chem. - An Asian J. 4 (2009) 630–640. https://doi.org/10.1002/asia.200800427.

[107]   J.R. Deng, S.F. Chung, A.S.L. Leung, W.M. Yip, B. Yang, M.C. Choi, J.F. Cui, K.K.Y. Kung, Z. Zhang, K.W. Lo, Y.C. Leung, M.K. Wong, Chemoselective and photocleavable cysteine modification of peptides and proteins using isoxazoliniums, Commun. Chem. 2 (2019) 1–10. https://doi.org/10.1038/s42004-019-0193-5.

[108]   Y. Kim, S.O. Ho, N.R. Gassman, Y. Korlann, E. V. Landorf, F.R. Collart, S. Weiss, Efficient site-specific labeling of proteins via cysteines, Bioconjug. Chem. 19 (2008) 786–791. https://doi.org/10.1021/bc7002499.

[109]   Y. Zhang, V.S. Bhatt, G. Sun, P.G. Wang, A.F. Palmer, Site-selective glycosylation of hemoglobin on Cys β93, Bioconjug. Chem. 19 (2008) 2221–2230. https://doi.org/10.1021/bc8003205.

[110]   A.W. Erian, S.M. Sherif, H.M. Gaber, The chemistry of α-haloketones and their utility in heterocyclic synthesis, Molecules. 8 (2003) 793–865. https://doi.org/10.3390/81100793.

[111]   K.J. Nelson, D. Parsonage, A. Hall, P.A. Karplus, L.B. Poole, Cysteine pKa values for the bacterial peroxiredoxin AhpC, Biochemistry. 47 (2008) 12860–12868. https://doi.org/10.1021/bi801718d.

[112]   M. Madsen, K. V. Gothelf, Chemistries for DNA Nanotechnology, Chem. Rev. 119 (2019) 6384–6458. https://doi.org/10.1021/acs.chemrev.8b00570.

[113]   G. Badescu, P. Bryant, M. Bird, K. Henseleit, J. Swierkosz, V. Parekh, R. Tommasi, E. Pawlisz, K. Jurlewicz, M. Farys, N. Camper, X. Sheng, M. Fisher, R. Grygorash, A. Kyle, A. Abhilash, M. Frigerio, J. Edwards, A. Godwin, Bridging disulfides for stable and defined antibody drug conjugates, Bioconjug. Chem. 25 (2014) 1124–1136. https://doi.org/10.1021/bc500148x.

[114] S.F. Betz, Disulfide bonds and the stability of globular proteins, Protein Sci. 2 (1993) 1551–1558. https://doi.org/10.1002/pro.5560021002.

[115] D.E. Shafer, J.K. Inman, A. Lees, Reaction of tris(2-carboxyethyl)phosphine (TCEP) with maleimide and α- haloacyl groups: Anomalous elution of TCEP by gel filtration, Anal. Biochem. 282 (2000) 161–164. https://doi.org/10.1006/abio.2000.4609.

[116] W.W. Cleland, Dithiothreitol, a New Protective Reagent for SH Groups, Biochemistry. 3 (1964) 480–482. https://doi.org/10.1021/bi00892a002.

[117] M. Tri-gas, Material safety data sheet 1., (2008) 1–7.

[118] Y. Zhang, C. Zang, G. An, M. Shang, Z. Cui, G. Chen, Z. Xi, C. Zhou, Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones, Nat. Commun. 11 (2020) 1–10. https://doi.org/10.1038/s41467-020-14757-4.

[119] R.L. Lundblad, W.H. Stein, On the reaction of diazoacetyl compounds with pepsin., J. Biol. Chem. 244 (1969) 154–160. https://doi.org/10.1016/s0021-9258(19)78205-2.

[120] N.S. Joshi, L.R. Whitaker, M.B. Francis, A three-component Mannich-type reaction for selective tyrosine bioconjugation, J. Am. Chem. Soc. 126 (2004) 15942–15943. https://doi.org/10.1021/ja0439017.

[121] C.J. Noren, S.J. Anthony-cahill, M.C. Griffith, P.G. Schultz, I. April, Method for Site-Specific, (n.d.).

[122] T. Hohsaka, M. Sisido, Incorporation of non-natural amino acids into proteins, Curr. Opin. Chem. Biol. 6 (2002) 809–815. https://doi.org/10.1016/S1367-5931(02)00376-9.

[123] J.Y. Axup, K.M. Bajjuri, M. Ritland, B.M. Hutchins, C.H. Kim, S.A. Kazane, R. Halder, J.S. Forsyth, A.F. Santidrian, K. Stafin, Y. Lu, H. Tran, A.J. Seller, S.L. Biroc, A. Szydlik, J.K. Pinkstaff, F. Tian, S.C. Sinha, B. Felding-Habermann, V. V. Smider, P.G. Schultz, Synthesis of site-specific antibody-drug conjugates using unnatural amino acids, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 16101–16106. https://doi.org/10.1073/pnas.1211023109.

[124] M. Haque, N. Forte, J.R. Baker, Site-selective lysine conjugation methods and applications towards antibody-drug conjugates, Chem. Commun. 57 (2021) 10689–10702. https://doi.org/10.1039/d1cc03976h.

[125] L. Wang, P. Quan, S.H. Chen, W. Bu, Y.F. Li, X. Wu, J. Wu, L. Zhang, Y. Zhao, X. Jiang, B. Lin, R. Zhou, C. Chen, Stability of Ligands on Nanoparticles Regulating the Integrity of Biological Membranes at the Nano-Lipid Interface, ACS Nano. 13 (2019) 8680–8693. https://doi.org/10.1021/acsnano.9b00114.

[126] C.M. Niemeyer, Semisynthetic DNA-protein conjugates for biosensing and nanofabrication, Angew. Chemie - Int. Ed. 49 (2010) 1200–1216. https://doi.org/10.1002/anie.200904930.

[127] K. Strebhardt, A. Ullrich, Paul Ehrlich's magic bullet concept: 100 Years of progress, Nat. Rev. Cancer. 8 (2008) 473–480. https://doi.org/10.1038/nrc2394.

[128] © 1975 Nature Publishing Group, (1975).

[129] R.J. Poljak, L.M. Amzel, H.P. Avey, B.L. Chen, R.P. Phizackerley, F. Saul, Three dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8 Å resolution, Proc. Natl.

Acad. Sci. U. S. A. 70 (1973) 3305–3310. https://doi.org/10.1073/pnas.70.12.3305.

[130] D.M. Segal, E.A. Padlan, G.H. Cohen, S. Rudikoff, M. Potter, D.R. Davies, The three dimensional structure of a phosphorylcholine binding mouse immunoglobulin Fab and the nature of the antigen binding site, Proc. Natl. Acad. Sci. U. S. A. 71 (1974) 4298–4302. https://doi.org/10.1073/pnas.71.11.4298.

[131] S. Dengl, C. Sustmann, U. Brinkmann, Engineered hapten-binding antibody derivatives for modulation of pharmacokinetic properties of small molecules and targeted payload delivery, Immunol. Rev. 270 (2016) 165–177. https://doi.org/10.1111/imr.12386.

[132] S. Dengl, E. Hoffmann, M. Grote, C. Wagner, O. Mundigl, G. Georges, I. Thorey, K.G. Stubenrauch, A. Bujotzek, H.P. Josel, S. Dziadek, J. Benz, U. Brinkmann, Hapten-directed spontaneous disulfide shuffling: A universal technology for site-directed covalent coupling of payloads to antibodies, FASEB J. 29 (2015) 1763–1779. https://doi.org/10.1096/fj.14-263665.

[133] C.M. Dundas, D. Demonte, S. Park, Streptavidin-biotin technology: Improvements and innovations in chemical and biological applications, Appl. Microbiol. Biotechnol. 97 (2013) 9343–9353. https://doi.org/10.1007/s00253-013-5232-z.

[134] T. Adachi, Y. Nakamura, Aptamers : A Review of Their Chemical Properties, MDPI Mol. 24 (2019). https://doi.org/10.3390/molecules24234229.

[135] Q. Zhu, G. Liu, M. Kai, A.O.A. Miller, DNA aptamers in the diagnosis and treatment of human diseases, Molecules. 20 (2015) 20979–90997. https://doi.org/10.3390/molecules201219739.

[136] W.A. Pieken, D.B. Olsen, F. Benseler, H. Aurup, F. Eckstein, Kinetic characterization of ribonuclease-resistant 2′-modified hammerhead ribozymes, Science (80-.). 253 (1991) 314–317. https://doi.org/10.1126/science.1857967.

[137] Y. Morita, M. Leslie, H. Kameyama, D.E. Volk, T. Tanaka, Aptamer therapeutics in cancer: Current and future, Cancers (Basel). 10 (2018). https://doi.org/10.3390/cancers10030080.

[138] G. Mayer, The chemical biology of aptamers, Angew. Chemie - Int. Ed. 48 (2009) 2672–2689. https://doi.org/10.1002/anie.200804643.

[139] M.G. Murray, W.F. Thompson, Nucleic Acids Research Nucleic Acids Research, Methods. 12 (1980) 8235–8251.

[140] M.H. Stenzel, Bioconjugation using thiols: Old chemistry rediscovered to connect polymers with nature's building blocks, ACS Macro Lett. 2 (2013) 14–18. https://doi.org/10.1021/mz3005814.

[141] S.D. Fontaine, R. Reid, L. Robinson, G.W. Ashley, D. V. Santi, Long-term stabilization of maleimide-thiol conjugates, Bioconjug. Chem. 26 (2015) 145–152. https://doi.org/10.1021/bc5005262.

[142] N.J. Agard, J.A. Prescher, C.R. Bertozzi, A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems, J. Am. Chem. Soc. 126 (2004) 15046–15047. https://doi.org/10.1021/ja044996f.

[143] T. Wada, A. Mochizuki, S. Higashiya, H. Tsuruoka, S.I. Kawahara, M. Ishikawa, M. Sekine, Synthesis and properties of 2-azidodeoxyadenosine and its incorporation into

oligodeoxynucleotides, Tetrahedron Lett. 42 (2001) 9215–9219. https://doi.org/10.1016/S0040-4039(01)02028-7.

[144] A. Dallmann, A.H. El-Sagheer, L. Dehmel, C. Mügge, C. Griesinger, N.P. Ernsting, T. Brown, Structure and dynamics of triazole-linked DNA: Biocompatibility explained, Chem. - A Eur. J. 17 (2011) 14714–14717. https://doi.org/10.1002/chem.201102979.

[145] P. Ketterer, A.N. Ananth, D.S. Laman Trip, A. Mishra, E. Bertosin, M. Ganji, J. Van Der Torre, P. Onck, H. Dietz, C. Dekker, DNA origami scaffold for studying intrinsically disordered proteins of the nuclear pore complex, Nat. Commun. 9 (2018) 1–8. https://doi.org/10.1038/s41467-018-03313-w.

[146] N. Venkatesan, B.H. Kim, Peptide conjugates of oligonucleotides: Synthesis and applications, Chem. Rev. 106 (2006) 3712–3761. https://doi.org/10.1021/cr0502448.

[147] J. Lotze, U. Reinhardt, O. Seitz, A.G. Beck-Sickinger, Peptide-tags for site-specific protein labelling: In vitro and in vivo, Mol. Biosyst. 12 (2016) 1731–1745. https://doi.org/10.1039/c6mb00023a.

[148] D. Guan, Y. Kurra, W. Liu, Z. Chen, A click chemistry approach to site-specific immobilization of a small laccase enables efficient direct electron transfer in a biocathode, Chem. Commun. 51 (2015) 2522–2525. https://doi.org/10.1039/c4cc09179e.

[149] G. V. Los, L.P. Encell, M.G. McDougall, D.D. Hartzell, N. Karassina, C. Zimprich, M.G. Wood, R. Learish, R.F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D.H. Klaubert, R.F. Bulleit, K. V. Wood, HaloTag: A novel protein labeling technology for cell imaging and protein analysis, ACS Chem. Biol. 3 (2008) 373–382. https://doi.org/10.1021/cb800025k.

[150] C. Timm, C.M. Niemeyer, Assembly and Purification of Enzyme-Functionalized DNA Origami Structures, Angew. Chemie. 127 (2015) 6849–6854. https://doi.org/10.1002/ange.201500175.

[151] M. Breugst, H.U. Reissig, The Huisgen Reaction: Milestones of the 1,3-Dipolar Cycloaddition, Angew. Chemie - Int. Ed. 59 (2020) 12293–12307. https://doi.org/10.1002/anie.202003115.

[152] V. V. Rostovtsev, L.G. Green, V. V. Fokin, K.B. Sharpless, A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes, Angew. Chemie - Int. Ed. 41 (2002) 2596–2599. https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4.

[153] J.M. Baskin, C.R. Bertozzi, Bioorthogonal click chemistry: Covalent labeling in living systems, QSAR Comb. Sci. 26 (2007) 1211–1219. https://doi.org/10.1002/qsar.200740086.

[154] M.D. Best, Click chemistry and bioorthogonal reactions: Unprecedented selectivity in the labeling of biological molecules, Biochemistry. 48 (2009) 6571–6584. https://doi.org/10.1021/bi9007726.

[155] B. Le Droumaguet, K. Velonia, Click chemistry: A powerful tool to create polymer-based macromolecular chimeras, Macromol. Rapid Commun. 29 (2008) 1073–1089. https://doi.org/10.1002/marc.200800155.

[156] J.R. Requena, D. Groth, G. Legname, E.R. Stadtman, S.B. Prusiner, R.L. Levine, Copper-

catalyzed oxidation of the recombinant SHa(29-231) prion protein, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 7170–7175. https://doi.org/10.1073/pnas.121190898.

[157] C.J. Burrows, J.G. Muller, Oxidative nucleobase modifications leading to strand scission, Chem. Rev. 98 (1998) 1109–1151. https://doi.org/10.1021/cr960421s.

[158] M. Pitié, C.J. Burrows, B. Meunier, Mechanisms of DNA cleavage by copper complexes of 3-Clip-Phen and of its conjugate with a distamycin analogue, Nucleic Acids Res. 28 (2000) 4856–4864. https://doi.org/10.1093/nar/28.24.4856.

[159] R.B. Turner, W.S. Lindsay, V. Boekelheide, Heats of hydrogenation, Tetrahedron. 27 (1971) 3341–3344. https://doi.org/10.1016/s0040-4020(01)97746-4.

[160] T. Sano, C.L. Smith, C.R. Cantor, Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates, Science (80-.). 258 (1992) 120–122. https://doi.org/10.1126/science.1439758.

[161] F. Diezmann, O. Seitz, DNA-guided display of proteins and protein ligands for the interrogation of biology, Chem. Soc. Rev. 40 (2011) 5789–5801. https://doi.org/10.1039/c1cs15054e.

[162] B.P. Duckworth, Y. Chen, J.W. Wollack, Y. Sham, J.D. Mueller, T.A. Taton, M.D. Distefano, A universal method for the preparation of covalent protein-DNA conjugates for use in creating protein nanostructures, Angew. Chemie - Int. Ed. 46 (2007) 8819–8822. https://doi.org/10.1002/anie.200701942.

[163] H. Yan, S.H. Park, G. Finkelstein, J.H. Reif, T.H. LaBean, DNA-templated self-assembly of protein arrays and highly conductive nanowires, Science (80-.). 301 (2003) 1882–1884. https://doi.org/10.1126/science.1089389.

[164] S. Rinker, Y. Ke, Y. Liu, R. Chhabra, H. Yan, Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding, Nat. Nanotechnol. 3 (2008) 418–422. https://doi.org/10.1038/nnano.2008.164.

[165] E. Wallin, G. Von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, Protein Sci. 7 (1998) 1029–1038. https://doi.org/10.1002/pro.5560070420.

[166] J. Drews, Drug discovery: A historical perspective, Science (80-.). 287 (2000) 1960–1964. https://doi.org/10.1126/science.287.5460.1960.

[167] R. van Grondelle, J.P. Dekker, T. Gillbro, V. Sundstrom, Energy transfer and trapping in photosynthesis, BBA - Bioenerg. 1187 (1994) 1–65. https://doi.org/10.1016/0005-2728(94)90166-X.

[168] D.J. Mothersole, D.A. Farmer, A. Hitchcock, C.N. Hunter, Photosynthetic apparatus in purple bacteria, 2018. https://doi.org/10.1201/9781351242899-6.

[169] V.I. Novoderezhkin, D. Rutkauskas, R. Van Grondelle, Dynamics of the emission spectrum of a single LH2 complex: Interplay of slow and fast nuclear motions, Biophys. J. 90 (2006) 2890–2902. https://doi.org/10.1529/biophysj.105.072652.

[170] R.H. Goldsmith, W.E. Moerner, Watching conformational- and photodynamics of single fluorescent proteins in solution, Nat. Chem. 2 (2010) 179–186.

https://doi.org/10.1038/nchem.545.

[171]   J.D. Pédelacq, S. Cabantous, T. Tran, T.C. Terwilliger, G.S. Waldo, Engineering and
         characterization of a superfolder green fluorescent protein, Nat. Biotechnol. 24 (2006) 79–88.
         https://doi.org/10.1038/nbt1172.

[172]   A. Spriestersbach, J. Kubicek, F. Schäfer, H. Block, B. Maertens, Purification of His-Tagged
         Proteins, Methods Enzymol. 559 (2015) 1–15. https://doi.org/10.1016/bs.mie.2014.11.003.

[173]   S.H. HUTNER, Anaerobic and aerobic growth of purple bacteria (Athiorhodaceae) in
         chemically defined media., J. Gen. Microbiol. 4 (1950) 286–293.
         https://doi.org/10.1099/00221287-4-3-286.

[174]   K. Duquesne, J.N. Sturgis, Membrane Protein Solubilization BT - Recombinant Protein
         Production in Yeast, Recomb. Protein Prod. Yeast. 601 (2009) 205–217.
         https://doi.org/10.1007/978-1-60761-344-2.

[175]   A. Kotecha, T. Georgiou, M.Z. Papiz, Evolution of low-light adapted peripheral light-harvesting
         complexes in strains of Rhodopseudomonas palustris, Photosynth. Res. 114 (2013) 155–164.
         https://doi.org/10.1007/s11120-012-9791-0.

[176]   M. Watanabe, K. Kawahara, K. Sasaki, N. Noparatnaraporn, Biosorption of cadmium ions
         using a photosynthetic bacterium, Rhodobacter sphaeroides S and a marine photosynthetic
         bacterium, Rhodovulum sp. and Their biosorption kinetics, J. Biosci. Bioeng. 95 (2003) 374–
         378. https://doi.org/10.1263/jbb.95.374.

[177]   R.J. Cogdell, N.W. Isaacs, T.D. Howard, K. McLuskey, N.J. Fraser, S.M. Prince, How
         photosynthetic bacteria harvest solar energy, J. Bacteriol. 181 (1999) 3869–3879.
         https://doi.org/10.1128/jb.181.13.3869-3879.1999.

[178]   V. Yurkov, J.T. Beatty, Isolation of aerobic anoxygenic photosynthetic bacteria from black
         smoker plume waters of the Juan de Fuca Ridge in the Pacific Ocean, Appl. Environ.
         Microbiol. 64 (1998) 337–341. https://doi.org/10.1128/aem.64.1.337-341.1998.

[179]   D.G. Rhodes, T.M. Laue, Chapter 38 Determination of Protein Purity, Methods Enzymol. 463
         (2009) 677–689. https://doi.org/10.1016/S0076-6879(09)63038-X.

[180]   X. Shu, N.C. Shaner, C.A. Yarbrough, R.Y. Tsien, S.J. Remington, Novel chromophores and
         buried charges control color in mFruits, Biochemistry. 45 (2006) 9639–9647.
         https://doi.org/10.1021/bi060773l.

[181]   A.J. Lam, F. St-Pierre, Y. Gong, J.D. Marshall, P.J. Cranfill, M.A. Baird, M.R. McKeown, J.
         Wiedenmann, M.W. Davidson, M.J. Schnitzer, R.Y. Tsien, M.Z. Lin, Improving FRET dynamic
         range with bright green and red fluorescent proteins, Nat. Methods. 9 (2012) 1005–1012.
         https://doi.org/10.1038/nmeth.2171.

[182]   S. Kesgin-Schaefer, J. Heidemann, A. Puchert, K. Koelbel, B.A. Yorke, N. Huse, A.R.
         Pearson, C. Uetrecht, H. Tidow, Crystal structure of a domain-swapped photoactivatable
         sfGFP variant provides evidence for GFP folding pathway, FEBS J. 286 (2019) 2329–2340.
         https://doi.org/10.1111/febs.14797.

[183]   O. V. Stepanenko, O. V. Stepanenko, I.M. Kuznetsova, V. V. Verkhusha, K.K. Turoverov,

Sensitivity of superfolder GFP to ionic agents, PLoS One. 9 (2014).
https://doi.org/10.1371/journal.pone.0110750.

[184] R.J. Cogdell, A. Gall, J. Köhler, The architecture and function of the light-harvesting apparatus
of purple bacteria: From single molecules to in vivo membranes, Q. Rev. Biophys. 39 (2006)
227–324. https://doi.org/10.1017/S0033583506004434.

[185] P. Qian, C.A. Siebert, P. Wang, D.P. Canniffe, C.N. Hunter, Cryo-EM structure of the
Blastochloris viridis LH1-RC complex at 2.9 Å, Nature. 556 (2018) 203–208.
https://doi.org/10.1038/s41586-018-0014-5.

[186] H. Gadgil, L.A. Jurado, H.W. Jarrett, DNA affinity chromatography of transcription factors,
Anal. Biochem. 290 (2001) 147–178. https://doi.org/10.1006/abio.2000.4912.

[187] N. Kotagiri, Z. Li, X. Xu, S. Mondal, A. Nehorai, S. Achilefu, Antibody quantum dot conjugates
developed via copper-free click chemistry for rapid analysis of biological samples using a
microfluidic microsphere array system, Bioconjug. Chem. 25 (2014) 1272–1281.
https://doi.org/10.1021/bc500139u.

[188] N.J. Russell, J.L. Harwood, Changes in the acyl lipid composition of photosynthetic bacteria
grown under photosynthetic and non-photosynthetic conditions., Biochem. J. 181 (1979) 339–
345. https://doi.org/10.1042/bj1810339.