

**Arranging Multiple Types of Enzymes in Defined
Space by Modular Adaptors**

NGUYEN MINH THANG

Kyoto University

Preface and Acknowledgements

The studies in this thesis have been carried out under the direction of Professor Takashi Morii at the Graduate School of Energy Science, Kyoto University, from Oct. 2014 to Dec. 2018. The studies focus on the development of orthogonal modular adaptors for locating multiple types of enzymes on a DNA nanoscaffold.

First of all, I wish to express my grateful acknowledgment to Professor Takashi Morii for his guidance, valuable suggestions, fruitful discussions and encouragement throughout the work. I am deeply grateful to Associate Professor Eiji Nakata, Dr. Masayuki Saimura, Assistant Professor Shun Nakano, Junior Associate Professor Arivazhagan Rajendran, and Dr. Ngô Anh Tiến for their kind advice and helpful discussions.

I also would like to express my appreciation to Professor Isao Saito for his helpful suggestions and encouragements.

I am greatly indebted to Professor Tsutomu Kodaki for his vectors encoding enzymes (xylose reductase, xylitol dehydrogenase, xylulose kinase) as a special gift.

I thank the past and the present members of Professor Morii's laboratory for their suggestions, cooperation, and discussions. In particular, I express my appreciation to Associate Professor Eiji Nakata, Dr. Masayuki Saimura, and Assistant Professor Shun Nakano, Junior Associate Professor Arivazhagan Rajendran, Dr. Ngô Anh Tiến and Ms Chiemi Nakata for their technical supports, kind guidance, fruitful discussions, and encouragements. I also thank Dr. Tomoki Tamura, Ms. Đinh Thị Thu Huyền, Dr. Chiara Annoni, Mr. Kenta Sasaki, Mr. Takahiro Toda, Mr. Yuki Yoshimura, Mr. Keisuke

Nishida, and Mr. Kenta Ariyama, Mr. Shunsuke Tajima, Mr. Hiroaki Konishi, Ms. Peng Lin, Mr. Zhengxiao Zhang, Mr. Musashi Shimizu, Mr Yoshitaka Toyoshima, Mr Tomohiko Wakisaki, and Ms Khongorzul Gerelbaatar for their helpful comments and discussions. I also express my great appreciation to my colleagues, past and present, in the Graduate School of Energy Science.

Heartfelt gratitude to Ms. Yasuko Morii, Ms. Yuki Tashiro, Ms. Kaori Hashimoto, Ms. Yukie Kajikawa whose kindness and generosity of spirit made my stay and study in Japan unforgettable and pleasant.

Finally, I would like to express my deep appreciation to my parents Nguyễn Công Xô and Nguyễn Thị Bình, my parents-in-law Nguyễn Đình Mạnh and Nguyễn Thị Hồng, my brother Nguyễn Hồng Ân, sister-in-law Nguyễn Thị Hải Yến, brother-in-law Nguyễn Đình Hữu, my wife Nguyễn Thị Hoa for their constant encouragements and moral support.

NGUYEN MINH THANG

Laboratory of Biofunctional Chemistry

Department of Fundamental Energy Science

Graduate School of Energy Science

Kyoto University

December 2018

Table of Contents

Chapter 1

General introduction

1.1. Introduction	2
1.2. DNA nanotechnology	5
1.3. Method for the assembly of a protein on a DNA scaffold	10
1.3.1 Methods for the noncovalent assembly of POIs	10
1.3.1.1 NI-NTA-hexahistidine interaction	10
1.3.1.2 Biotin-avidin interaction	11
1.3.1.3 Method based on antigen-antibody interaction and protein-aptamer interaction	15
1.3.1.4 DNA-binding proteins	16
1.3.2 Covalent methods for the assembly of protein on a DNA scaffold	16
1.3.2.1 Direct chemical modification of POIs	16
1.3.2.2 Method base on self-ligating protein-tag	18
1.3.2.3 Method base on HUH-tag	20
1.3.2.4 Modular adaptor	21
1.4 Aim of the research in this thesis	24
1.5 References	28

Chapter 2

Design of modular protein-tags for the orthogonal covalent bond formation at specific DNA sequences

2.1 Introduction	38
2.2 Results	41
2.2.1 Design of modular adaptors and ODNs modified with the substrates for protein-tags	41
2.2.2 Formation of covalent bonds between the modular adaptors and the substrate modified ODNs	49

2.2.3 Formation of covalent bonds between the modular adaptors and the substrate modified addresses on the DNA scaffolds	56
2.3 Discussion	61
2.4 Conclusion	66
2.5 Materials and methods	67
2.5.1 Materials	67
2.5.2 Preparation of the protein-tag derivatives	68
2.5.3 Overexpression and purification of protein-tag derivatives (AZ-SNAP, AZ-CLIP, and AZ-Halo)	69
2.5.4 Preparation of substrate modified ODNs	70
2.5.5 Preparation of ³² P-end-labeled ODNs and the analysis of covalent-linkage formation between ³² P-end-labeled ODNs and modular adaptors by denaturing gel shift assay	70
2.5.6 Preparation of the DNA origami scaffold with three cavities	71
2.5.7 Preparation of DNA origami scaffold assembled with modular adaptors	77
2.5.8 AFM imaging and statistical analysis	77
2.6 References	79

Chapter 3

Development of the sequence specific modular adaptors

3.1 Introduction	85
3.2 Results	88
3.2.1 Construction of the modular adaptor-fused enzymes and assembly of them on the DNA scaffolds	88
3.2.2 Orthogonal study of the sequence-specific modular adaptors on DNA scaffold	122
3.3 Discussion	130
3.4 Conclusion	132
3.5 Materials and methods	133
3.5.1 Materials	133
3.5.2 Preparation of sequence specific modular adaptors	134
3.5.3 Overexpression and purification of protein-tag derivatives	134
3.5.4 Overexpression and Purification of ZF-CLIP	135

3.5.5 Preparation of ODN-S and Alexa 488 modified ODNs	137
3.5.6 Preparation of ³² P-end-labeled ODNs and the Analysis of Covalent-linkage Formation between ³² P-end-labeled ODNs and Modular Adaptors by Denaturing Gel Shift Assay	139
3.5.7 Preparation of the DNA Origami Scaffold with Three Cavities	139
3.5.8 Preparation of DNA Origami Scaffold Assembled with Modular Adaptors	140
3.5.9 AFM Imaging and Statistical Analysis	140
3.5.10 Fluorescence polarization spectroscopy	141
3.6 References	143

Chapter 4

A model of three enzymes cascade reaction on DNA scaffold

4.1 Introduction	149
4.2 Results	151
4.2.1 Construction of the modular adaptor-fused enzymes and assembly of them on the DNA scaffolds	151
4.2.2 Construction of a sequential enzyme cascade and the comparison of the effect of the assistance of the DNA scaffold	157
4.3 Discussion	163
4.4 Conclusion	164
4.5 Materials and methods	164
4.5.1 Materials	164
4.5.2 Preparation of the modular adaptor fused enzymes	166
4.5.3 Overexpression and purification of protein-tag derivatives	167
4.5.4 Preparation of substrate modified ODNs	167
4.5.5 Preparation of the DNA origami scaffold with three cavities	168
4.5.6 Preparation of DNA origami scaffold assembled with modular adaptor fused enzymes	169
4.5.7 AFM imaging and statistical analysis	169
4.5.8 Xylulose kinase assay	170
4.5.9 Analysis of the enzyme cascade reaction	170
4.6 References	171

Chapter 5	
General conclusions	177
List of publications	180
List of presentations	181
List of awards	184

CHAPTER 1

General introduction

1.1 Introduction

The energy crisis is the challenge for every country in the world. Depletion of fossil fuel (oil, coal and natural gas) leading to the searching for the renewable energy such as hydroelectricity, solar energy, wind energy, wave power, geothermal energy, tidal power, and bioenergy. Among them, bioenergy is the promising source of energy because, under billions of years of evolution, living organisms are well-built for uptaking, transferring, and utilizing energy and substances to maintaining its life activities.

Proteins as enzymes play a central role in biochemical reactions which are crucial to physiological activities at cellular level, however, concentration of proteins is surprisingly low ranging from 10 to 1000 molecules per cell (10 nM - 1 μ M).¹ To gain the function, protein itself develops not only high selectivity and reactivity but also its localization within the cell in the so-called compartmentalization process in many levels such as enzymes clustering, enzymes cascading, multi-enzyme system, and organelles at nanometer-scale (Figure 1.1).²⁻⁴ Understanding of *in vivo* localization and function of enzymes would provide a basic principle for the development of enzyme technology. The enzyme cascade has been constructed on several of bio-friendly nanoscaffolds such as liposomes/polymersomes, protein scaffolds, and DNA scaffolds to study the efficiency of the cascade reaction at the close proximity (Figure 1.2).⁵⁻⁸

Thanks to the development of DNA nanotechnology (Figure 1.3), the DNA nanostructures are the ideal scaffold for immobilizing or scaffolding enzymes.⁸⁻¹⁰ DNA technology offers the controllable size and shape of the nanostructures. The number and distance of assembled enzymes also can be fully controlled. Most proteins of interest (POIs) are fragile; the very mild condition is needed for handling, especially assembling on a DNA nanoscaffold. Current methods for proteins or enzymes assembly on a DNA scaffold show many disadvantages including the loss of enzymatic activity, time and

money consuming, or difficulty in assembly of multiple enzymes in one step. The study in this thesis, I have developed a new method which is fast, efficient and orthogonal to locate multiple enzymes on the designed addresses of a DNA scaffold.

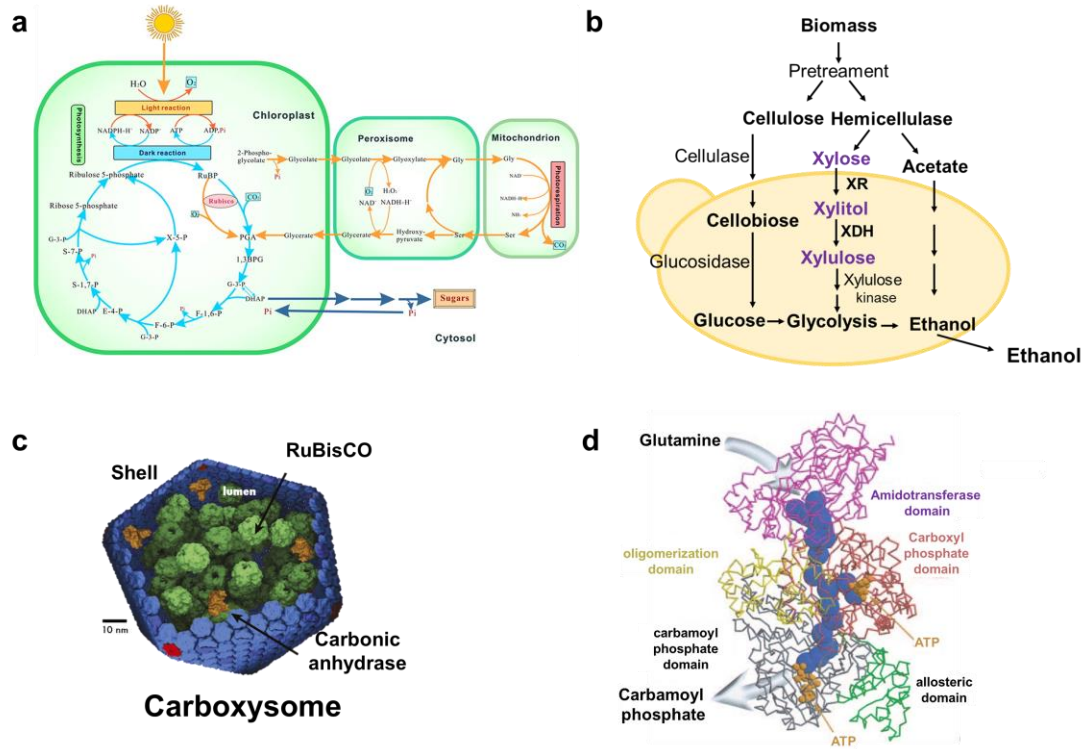


Figure 1.1 Metabolic pathway in compartmentalization. (a) Photosynthetic reactions in the chloroplasts of C3 plants.¹¹ 1,3BPG: biphosphoglycerate; G-3-P: glyceraldehyde 3-phosphate; DHAP: dihydroxyacetone phosphate; F-1,6-P: fructose 1,6-bisphosphate; F-6-P: fructose 6-phosphate; X-5-P: xylose 5-phosphate; E-4-P: erythrose 4-phosphate; S-1,7-P: sedoheptulose 1,7-bisphosphate; S-7-P: sedoheptulose 7-phosphate. (b) Schematic overview of biofuel production through co-utilization of cellobiose, xylose, and acetic acid from lignocellulosic biomass by engineered yeast.¹² XR: xylose reductase. XDH: xylitol dehydrogenase. XK: xylulose kinase. ACS: acetyl-CoA synthetase. AADH: acetylating acetaldehyde dehydrogenase. ADH: alcohol dehydrogenase. (c) The carboxysome contains two cascaded enzymes carbonic anhydrase and RuBisCO for carbon assimilation¹³ (d) Substrate channeling in carbamoyl phosphate synthase complex.¹⁴

1.2 DNA nanotechnology

From the idea of abiological utilization of the Watson-Crick base pairing of DNA sequences, the DNA nanotechnology was first developed by Nadrian Ned Seeman approximately 30 years ago.¹⁸ Many symmetric nanostructures varying from size and shape were proposed using difference DNA tiles (Figure 1.3). With the discovery of DNA origami techniques,¹⁹ the DNA scaffold has become more sophisticated and programmable. When a long single strand of DNA is hybridized with hundreds of short strands called "staple strands", the DNA origami nanostructure is formed in a well-defined manner with high folding yield. The 3D nanostructures based on DNA origami technique are also easily constructed, many methods to fold a DNA single strand are well-established (Figure 1.4, 1.5). Thanks to the flexible and dynamic properties of DNA hybridization, the DNA scaffold becomes more attractive with a moveable component which is potential for fabricating the nanodevices, nanomotors, and nanorobots (Figure 1.5).²⁰

Although DNA nanotechnology has been well developed, the DNA nanostructures are relatively inert and its beauty is only expressed when functionalizing with other materials such as nanoparticles, fluorophores, proteins, and enzymes.^{21,22} Among them, enzymes are the ideal materials to assemble on DNA scaffold because of its biological values. Study on the spatial and stoichiometric organization of cascade enzymes on DNA scaffold would shed light in the real behavior of enzyme *in vivo* conditions and provides potential application in the field of enzyme technology (Figure 1.6). Therefore, it is needed to develop a convenient method for assembling multiple enzymes on a DNA scaffold.

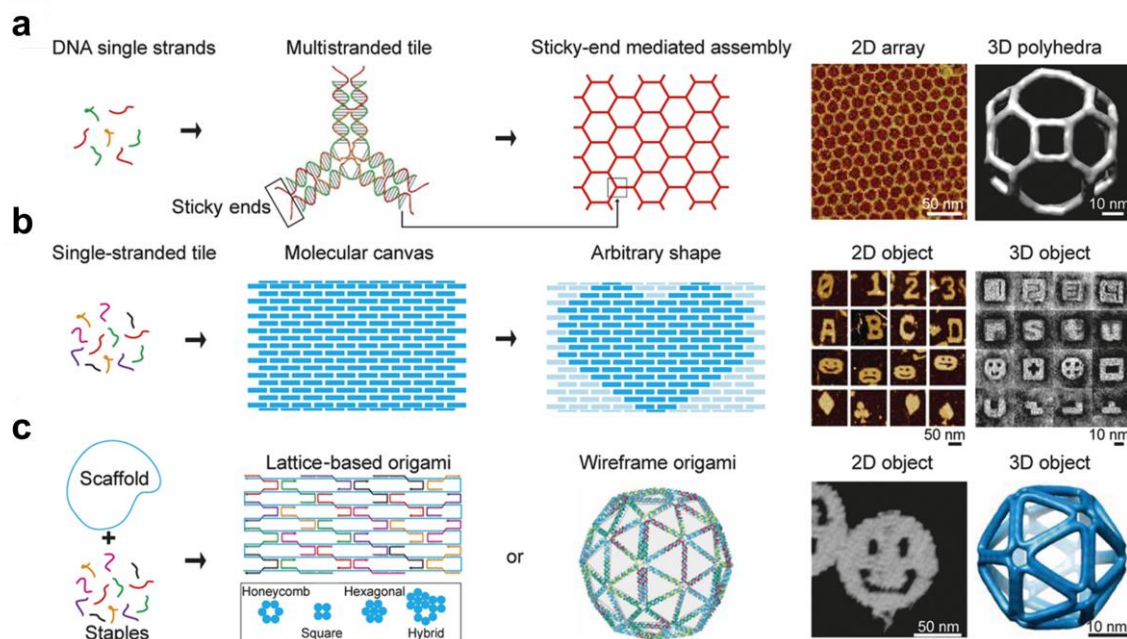


Figure 1.3. Summary of DNA tile and DNA origami.²³ (a) Multistranded DNA tile. Each tile is composed of several short DNA strands. Strands of different colors have different sequences. Sticky-end mediated cohesion between tiles leads to the assembly of larger structures such as 2D arrays or 3D polyhedra. (b) Single-stranded DNA (ssDNA) tile. A number of ssDNAs with unique sequences assemble into prescribed objects, which can serve as 2D and 3D canvases to construct arbitrary objects. (c) Scaffolded DNA origami. A long ssDNA template serves as a scaffold that is folded by a number of shorter ssDNAs (staples) into both delicate 2D and 3D objects. DNA duplexes can be organized using parallel lattices or wireframe structures.

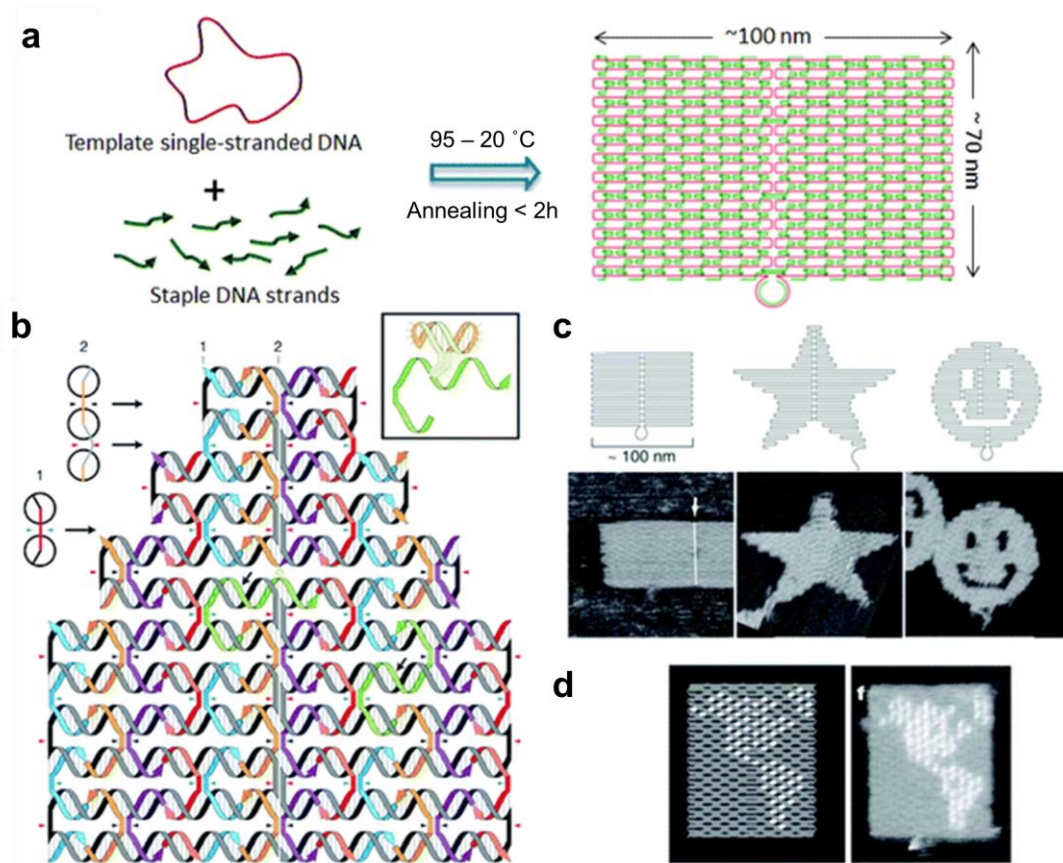


Figure 1.4 DNA origami structure.²⁴ (a) The method employed to prepare a DNA origami structure from the template single-stranded DNA and staple strands. (b) Design of a self-assembled DNA origami structure and geometry of the incorporated dsDNAs. Colored strands and a gray/black strand represent staple strands and template single-stranded DNA, respectively. Staple strands connect the adjacent duplexes with crossovers. Inset: structure of hairpin DNA for a topological marker. (c) Design and AFM images of self-assembled DNA origami structures. (d) Drawing of a hemisphere on the DNA origami with hairpin DNAs (white dots) and an AFM image of the assembled DNA origami.

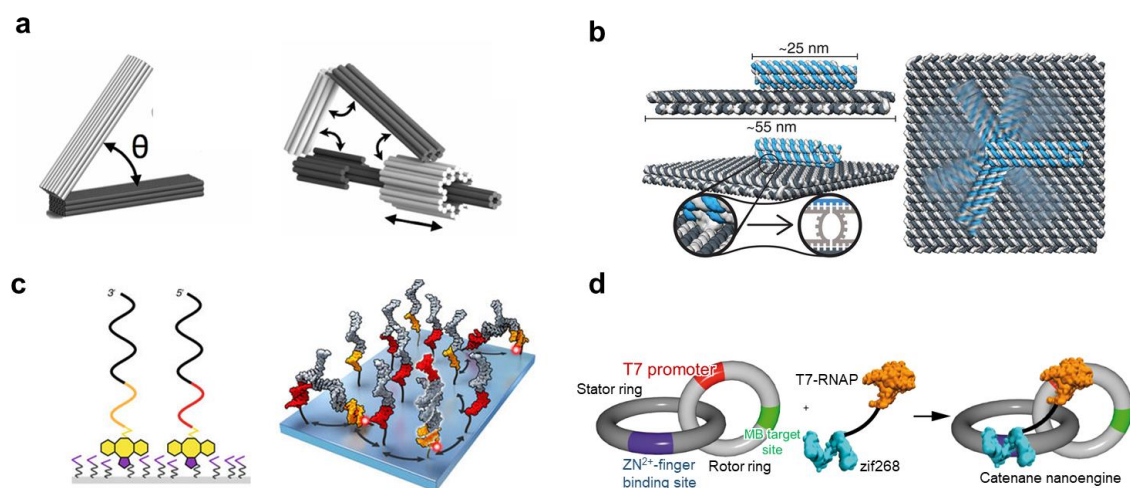


Figure 1.5 DNA nanoscaffold with moveable parts. (a) DNA origami with the stiff components and flexible joints, the motion caused by the geometry of the stiff component and the arrangement and degrees of freedom of the joints.²⁵ (b) A DNA based molecular platform with an integrated robotic arm that actuated with the external applied electrical field. Precise, computer-controlled switching of the arm between arbitrary positions on the platform can be achieved within milliseconds.²⁶ (c) The DNA walker using toehold mechanism with stepping kinetic for enhancing the velocity.²⁷ (d) A DNA catenane rotary motor. The engine consists of an engineered T7 RNA polymerase attached to a dsDNA nanoring that is catenated to a rigid rotating dsDNA wheel. The wheel motor produces long, repetitive RNA transcripts that remain attached to the engine and are used to guide its movement along predefined ssDNA tracks arranged on a DNA nanotube.²⁸

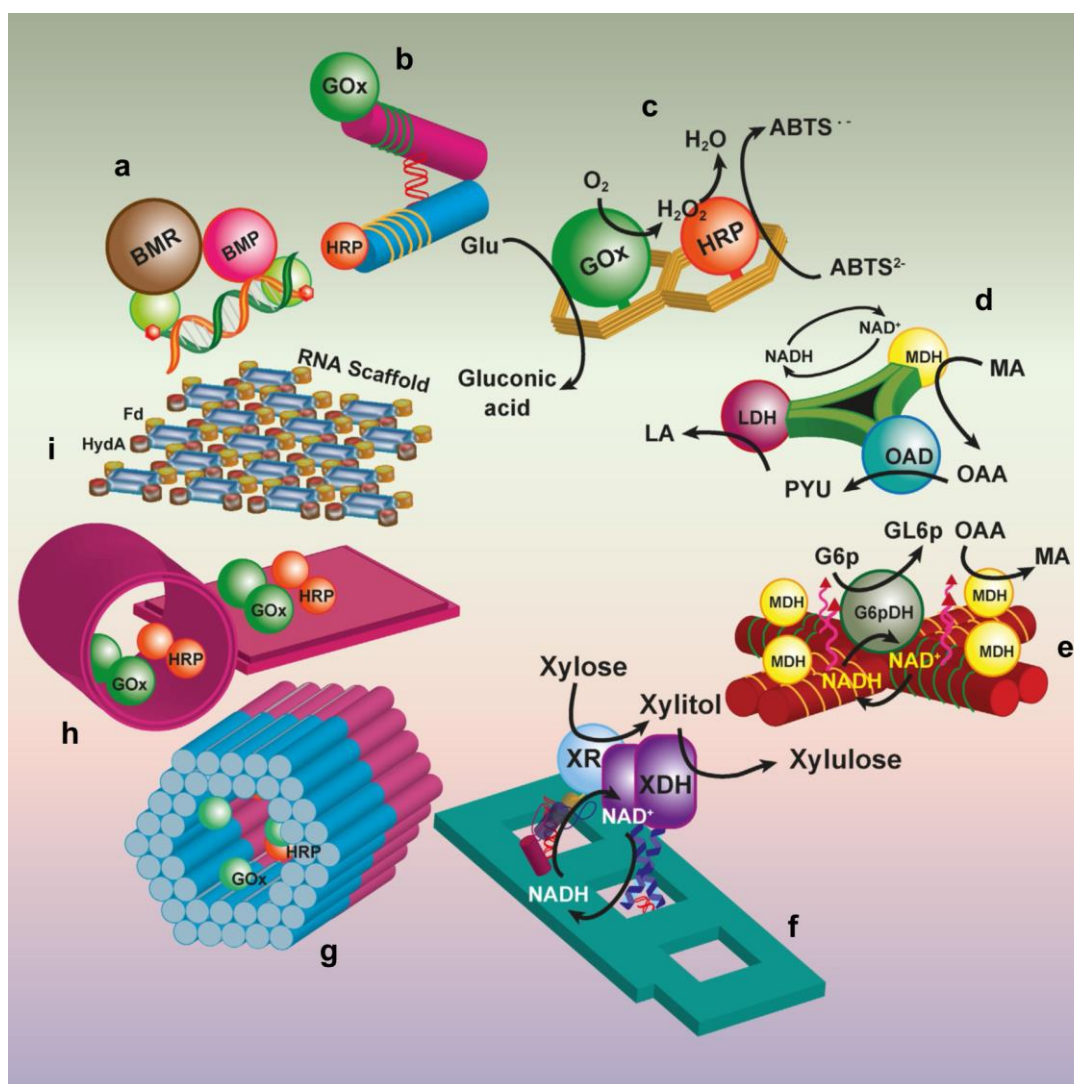


Figure 1.6 Artificial cascade enzyme reaction assembled on DNA scaffold.⁸ (a)²⁹ Assembly of two sub-domain of cytochrome P450 BM3 on a doubled strain DNA scaffold. (b,³⁰ c,³¹ e,³² f¹⁷) Two enzymes cascade reaction on a 2D scaffold. (d)³³ A three enzymes cascade reaction on a DNA scaffold. (g,³⁴ h³⁵) Two enzymes cascade reaction in a confined tubular DNA scaffold. (i)³⁶ The arrangement of two enzymes cascade on a grid-like 2D ARN scaffold. BMR: cytochrome P450 BM3 reductase domain; BMP: cytochrome P450 BM3 porphyrin domain; GOx: glucose oxidase; HRP: horse radix peroxidase; Glu: glucose; ABTS²⁺: 2,2'-azinobis-(3-ethylbenzthiazoline-6-sulfonate) dianion; MDH: malate dehydrogenase; OAD: Oxaloacetate decarboxylase; LDH: Lactate dehydrogenase; MA: malate; OAA: oxaloacetate; LA: lactate; G6pDH: glucose-6-phosphate dehydrogenase; G6p: glucose-6-phosphate; GL6p: 6-Phosphogluconolactone; XR: xylose reductase, XDH: xylitol dehydrogenase; HydA: [FeFe] hydrogenase; Fd: ferredoxin

1.3 Methods for the assembly of a protein on a DNA scaffold

The ideal method for assembly of functional enzymes on a DNA scaffold should be fast, stable, and precise in a mild condition such as at ambient temperature and natural pH. Many of methods were developed to assemble POIs on the defined addresses on a DNA scaffold.^{37–39} These methods could be classified into two strategies: the methods based on the noncovalent interaction (Ni-NTA-hexahistidine interaction, biotin-avidin interaction, antigen-antibody interaction, protein-aptamer interaction and DNA binding protein adaptor) and the methods based on the covalent linkage formation between POI and DNA scaffold (direct covalent modification of POI with staple strand, method based on self-ligating protein-tags, method based on HUH-tags, and the orthogonal modular adaptors which are presented in this thesis). The principles, advantages, disadvantages, and applications of each method described in the following paragraphs.

1.3.1 Methods for the noncovalent assembly of POIs

1.3.1.1 Ni-NTA-hexahistidine interaction

The Ni-NTA-polyhistidine has been applied in biochemical assays such as protein purification, protein immobilization, and protein labeling for many years. The interaction relies on the divalent transition metal Ni^{2+} where the polyhistidine-tagged POI could be immobilized on a surface using conjugated chelators such as nitrilotriacetic acid (NTA). The micromolar affinity interaction provides a moderately immobilization yield of the POI to a surface. Hexahistidine-tagged proteins are able to bind to multi NTA groups tethered ODN at sub-micromolar affinity. The trisNTA-ODN with three groups of NTA tethered to ODN provides the tight binding of the His₆-tag protein with the equilibrium dissociation constant (K_D) is as low as 6 nM.⁴⁰ Other divalent ions such as Co^{2+} , Cu^{2+} and Zn^{2+} could replace Ni^{2+} in the Ni-NTA complex

with the same affinity. Research also indicated that Co^{3+} serves as a mediator creating a more stable complex.⁴¹ Ni-NTA interaction method was used for assembling the EGFP on a DNA scaffold as shown in Figure 1.7.⁴²

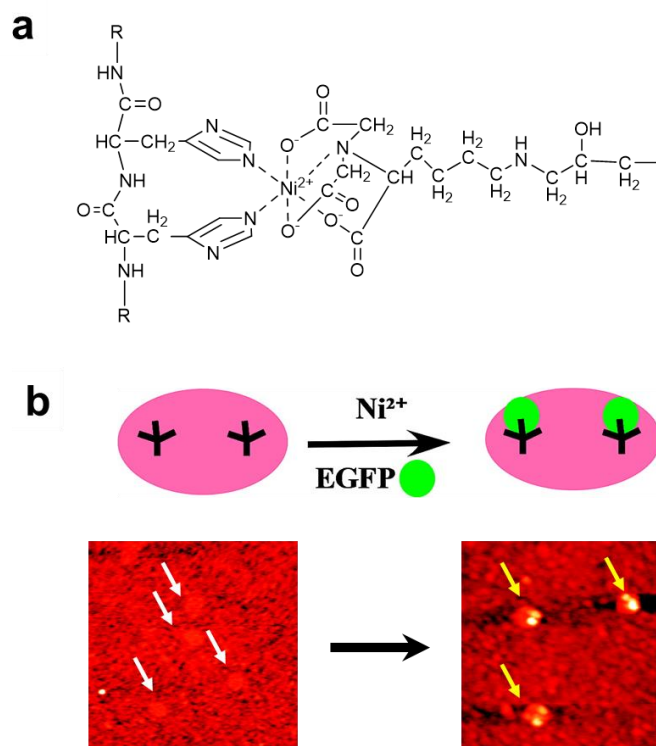


Figure 1.7 a) Mechanism of Ni-NTA interaction. b) Assembly of EGFP on the DNA scaffold⁴². White arrows and yellow arrows indicated the DNA scaffold before and after assembled with EGFP, respectively.

1.3.1.2 Biotin-avidin interaction

The biotin-avidin interaction is one of the strongest non-covalent bonds with femtomolar affinity.⁴³ Avidin is a tetrameric protein with four identical subunits allowing the binding of four biotin derivatives. Using this interaction, avidin conjugated materials were able to assemble to the defined position on a DNA scaffold containing the biotinylated DNA strands.

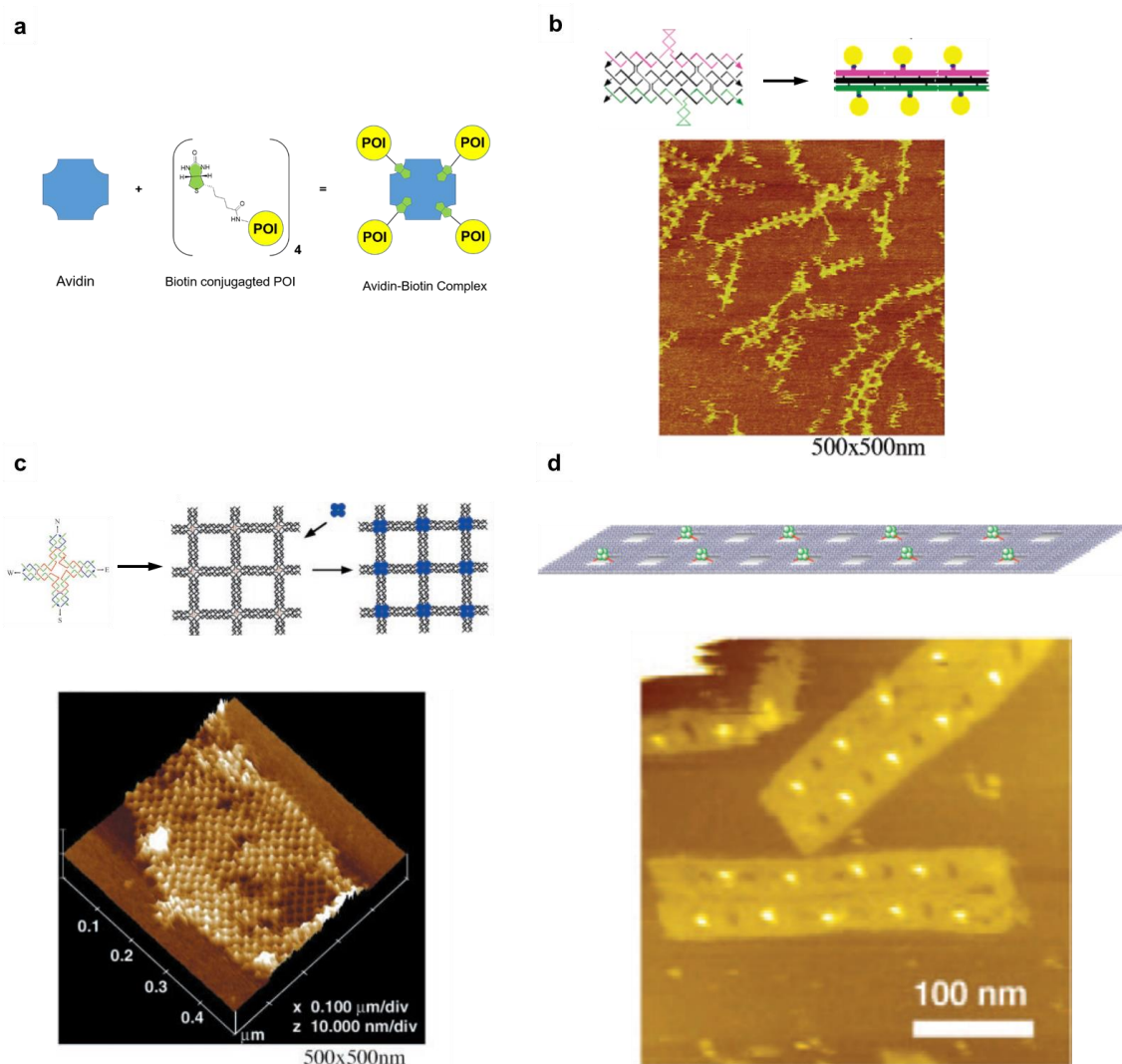


Figure 1.8 a) Principle of biotin-avidin interaction, four molecules of biotin derivatives could binding to one homotetrameric avidin molecule. Streptavidin assembled on a DNA scaffold that base on (b) the triple crossover tiles,⁴⁴ (c) four-arm junctions tiles,³⁶ and (d) the punched rectangle DNA origami.³⁸

Due to its simple but strong interaction, the biotin-avidin system is generally used for the assembly of a functional POI on a newly developed DNA nanoscaffold. Streptavidin, an avidin derivative from *Streptomyces* bacteria, was attached on a linear array of DNA structure⁴⁴ derived from the DNA triple crossover tile (Figure 1.8b) or a 2D nanogrid array acquired from the four-arm junction tile (Figure 1.8c).^{45,46} The well-defined 2D array based on DNA origami technique allows full control of the position of assembled streptavidin molecules (Figure 1.8d).^{47,48}

Functional POIs assembled on DNA scaffold with biotin-avidin system were first reported by Niemeyer group in the study the cascade reaction of NAD(P)H:FMN Oxidoreductase (NFOR) and Luciferase (Luc).²⁹ In the research, two biotinylated enzymes NFOR and LUC were assembled to two short streptavidin tethered DNA strands, respectively. The enzyme assembled strands were allowed to hybridize to one biotinylated long strand that previously immobilized to the STV-coated microplate. By this way, the two reaction center were co-located at a nanometer distance and the cascade reaction driven by two enzymes was characterized. The method is also useful for attaching enzymes on the 3D scaffold in which the enzymatic cascades reaction was enhanced not only by the close proximity and by the confined space (Figure 1.9b).³⁴

Although the biotin-avidin-based assembly of POIs is simple, it offers the highest yield among the methods based on the non-covalent interactions, up to 84% of the addresses are possibly filled with POIs. Since this approach only allows one type of POI to be assembled each time, two-step assembly is critical for the orthogonal coassembly of two POIs on the scaffold^{6,31-34} (Figure 1.9). Multi-step assembly requires much caution on loading functional proteins and enzymes. The conjugation of biotin to an enzyme is another limiting factor of the method. Although this process can be done *in vivo*, the yield of enzyme biotinylation varies among cultures and reaches only about 30%.²⁹ As a result, this method has been received much less attention than other methods for assembling multiple types POIs.

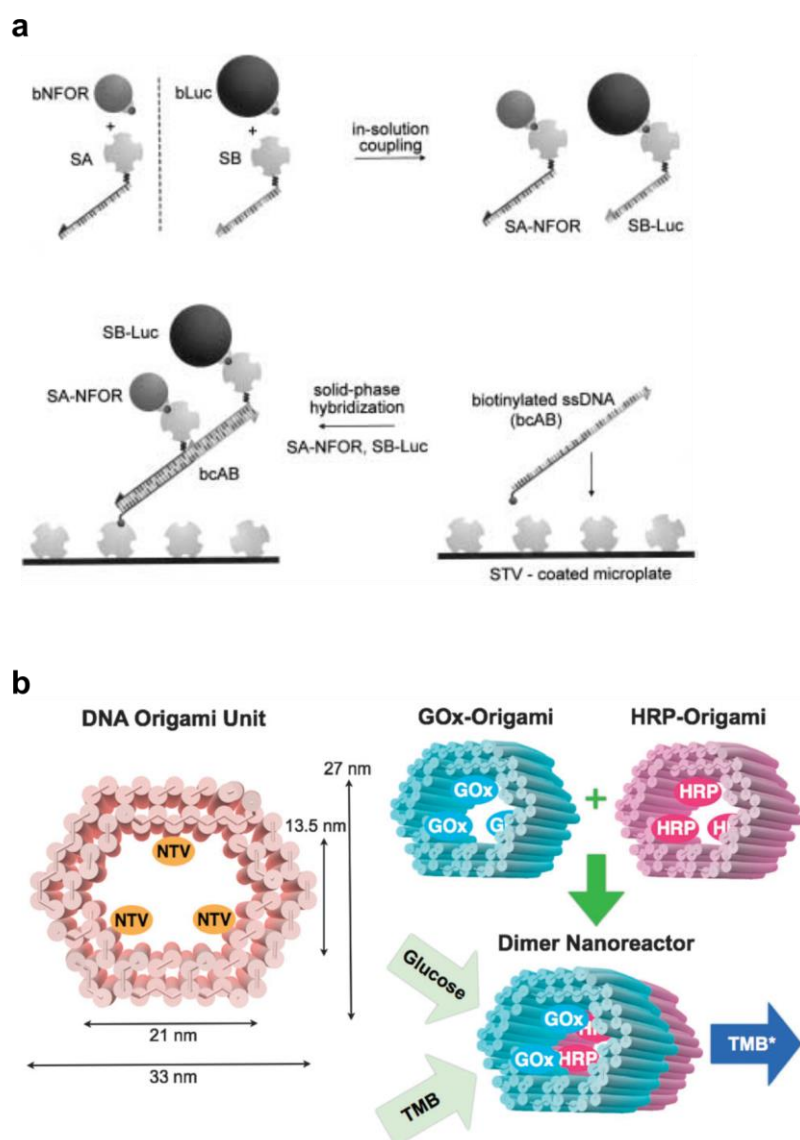


Figure 1.9 Stepwise assembly of coupled enzymes for the cascade reaction on (a) a simple DNA scaffold³¹ or (b) a tubular DNA scaffold using biotin-avidin interaction.³² a) NAD(P)H:FMN oxidoreductase (NFOR) and luciferase (Luc) were firstly assembled on two short single DNA strands by using biotin-avidin interaction. In the next step, the two short enzyme-assembled DNA strands were assembled on the long DNA strand by using the strand hybridization. b) Glucose oxidase (GOx) and horseradish peroxidase (HRP) were separately assembled on two half of the DNA origami nanoreactor by using the biotin-avidin interaction. In the next step, the two half of the DNA origami were assembled by using the strand hybridization.

1.3.1.3 Method based on antigen-antibody interaction and protein-aptamer interaction

These methods are based on the biological selectivity of the interaction of POIs with antibody or aptamer. By using the *in vivo* (for antibody) or *in vitro* selection (for aptamer), the antibody or aptamer for specific POIs could be generated, respectively. The antibody^{51,53} or aptamer^{52,54} could be used for locating POIs on DNA scaffold. These methods are resulting in orthogonal assemble of POIs in a DNA scaffold (Figure 1.10). However, these methods are time-consuming for designing and constructing the specific antibody or aptamer, not convenient for studying the cascade reaction of multiple enzymes.

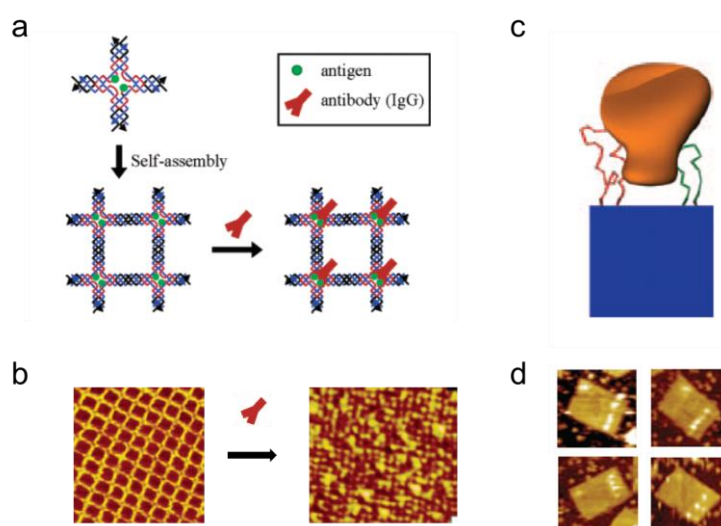


Figure 1.10 Assembly of POIs on DNA scaffold using antigen-antibody interaction (a, b) and protein-aptamer interaction (c, d). Scheme (a) and AFM image (b) of assembly antibody (IgG) on the two-dimensional DNA scaffolded antigen arrays.⁵¹ Schematic (c) showing a rigid DNA tile (blue) that can spatially separate two aptamers (red and green) at a controlled distance, with each aptamer attached to a different part of the target molecule (orange) for bivalent binding and AFM images (d) of the thrombin assembled on a DNA scaffold.⁵²

1.3.1.4 DNA binding proteins

The DNA binding proteins^{57–59} were used to locate POIs on the specific position on a DNA scaffold simply by fused the POIs to it (Figure 1.11). DNA binding proteins bind to the specific DNA sequences with approximately ten base pairs long, the protein adaptors based on the DNA binding protein allow orthogonal assembly of a number of POIs on a DNA scaffold.^{55,56,60} POIs can be located quickly on a DNA scaffold by fusing them to appropriate DNA binding proteins. The disadvantage of this technique is that the loading yield of POIs on a DNA scaffold is not quantitative. The moderate loading yield is due to the nonvalent interaction of the DNA-protein complex, even though the affinity of DNA binding protein can reach sub-nanomolar K_D .

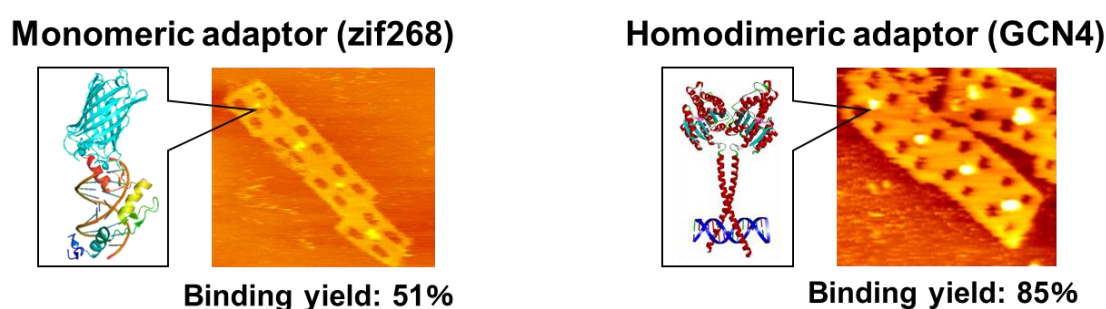


Figure 1.11 Assembly of POIs on DNA scaffold using DNA binding protein. EGFP assembled on a five-well DNA scaffold by using zinc finger protein zif268 (left).⁵⁵ Xylitol dehydrogenase (XDH) assembled on a five-well DNA scaffold by using basic leucine zipper GCN4 (G-XDH) (right).⁵⁶

1.3.2 Covalent methods for the assembly of protein on a DNA scaffold

1.3.2.1 Direct chemical modifications of POIs

Several methods covalently conjugating POIs to DNA scaffold have been developed.^{61–64} Direct tethering POIs to the DNA staples strand was the first choice because the techniques of chemical modification of protein have been well established (Figure 1.12).⁶⁵ In these techniques, several positions of POIs can be modified specifically at the terminus or the side chain of several types of amino acid residues. In

general, the modified POIs need to be purified from the un-reactant before assembled on a DNA scaffold. The post-translational protein chemical modification may significantly reduce the activity of the interested enzyme.^{32,62–64} Therefore, this method may not be suitable for the study of unstable enzymes.

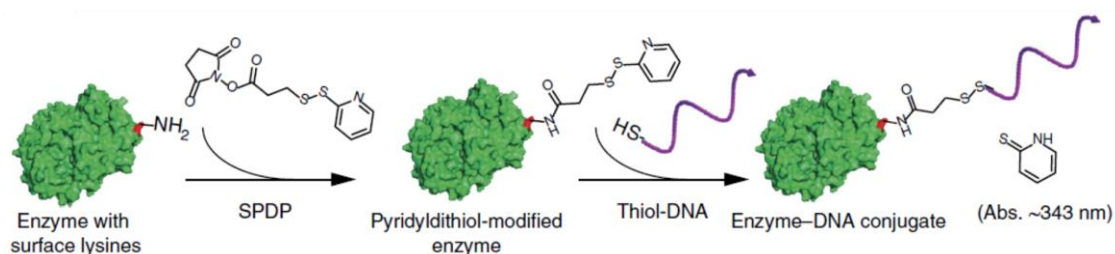


Figure 1.12. Example of chemical conjugation of oligonucleotides with enzymes. *N*-succinimidyl 3-(2-pyridyldithio) propionate (SPDP) was used for cross-linking a thiol-modified oligonucleotide to the lysine group of an enzyme, the oligonucleotide-modified enzyme needs to purify using anion-exchange FPLC before assembling on DNA scaffold.⁶¹

Gothelf and co-workers (2014)⁶⁶ showed a generic approach for the covalent attachment of single-stranded ODN to metal-coordinating proteins in a site-selective manner. The method utilizes three components: a POI to be labeled possessing a metal ion binding site, an ODN modified with a metal ligand, such as the NTA group, and its complementary ODN that carries an activated ester to react with the ϵ -amino group of lysines. To avoid the non-specific reaction between the activated ester and the lysine residues on the surface of POI, these components must be kept in diluted conditions. By tuning pH, salt concentrations and concentrations of the components, the template-directed ODN conjugated to POI was obtained in 38% isolated yield with avoiding the non-templated side reactions.

1.3.2.2 Method base on self-ligating protein-tag

A safe way to avoid reduction of enzyme activity is to genetically fuse the POIs to the self-ligating proteins.⁶⁷ The self-ligating protein tags allow the covalent bond formation to its specific substrate derivatives. When the substrate modified DNA strand was designed to hybridize at the specific position on DNA scaffold, the location of the assembled POIs is tunable in nanometer precision. With the available of self-ligating protein tags system (SNAP-tag,⁶⁸ and Halo-tag⁶⁹) which have high substrate selectivity towards O⁶-benzylguanine (BG) and 5-chlorohexane (CH) (Figure 1.13), respectively, orthogonal assembly of several POIs on DNA scaffold is possible in one pot reaction (Figure 1.14).

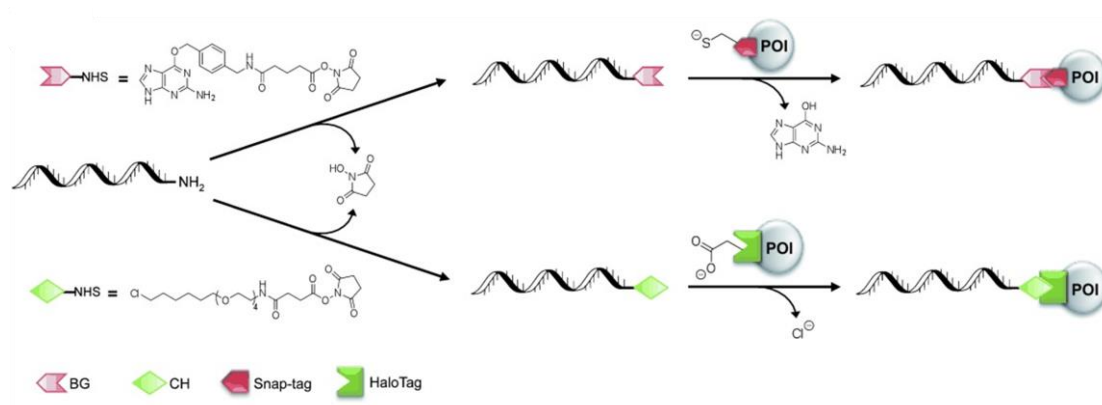


Figure 1.13 Substrates selectivity of the protein-tag system. SNAP-tag and Halo-tag selectively form the covalent bond with benzylguanine (BG), and chlorohexane (CH), respectively. The ODN modified with BG or CH could react orthogonally with POI fused SNAP-tag or POI fused Halo-tag to form a covalent bond between POI and ODN.

A recent report described this elegant method by using SNAP-tag, Halo-tag, and the avidin-biotin interaction for the assembly three POIs on a DNA scaffold.⁶⁷ Orthogonal assembly of three kinds of proteins using SNAP-tag, Halo-tag, and the avidin-biotin interaction was achieved. Although the specific reactions of SNAP-tag and Halo-tag with its respective substrates are quite fast, the performance of POIs fused

SNAP-tag and Halo-tag on the ODN conjugated substrates was poor, resulting in a very low loading yields, only about 40 – 60% (Figure 1.14). As a result, the coassembly yield of three POIs was only several percents. In addition, the assembly required an excess amount of protein-tag fused POIs (1000 times molar equivalent concentration of protein-tag fused POIs were incubation with DNA scaffold) and overnight incubation is required. The reason for this poor performance was likely that the electrostatic repulsion between the DNA scaffold and the protein-tag that prevented the protein-tag from accessing its substrates. The recent work showed an improvement of the assembly yield up to 80% by rationally designing the positive binding interface of Halo-tag.⁷⁰ Therefore, the protein tag for the efficient assembly multiple enzymes on a DNA scaffold should be very carefully selected among the variety of protein tags, such as CLIP-tag, PYP-tag, and TMP-tag.^{71,72}

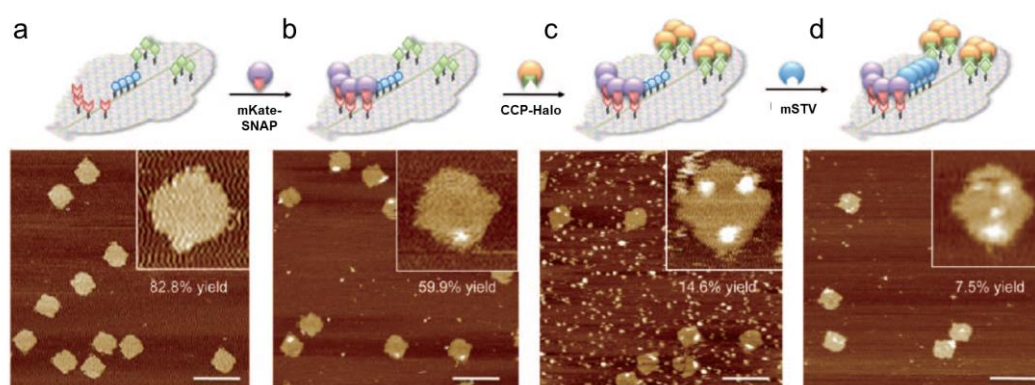


Figure 1.14 Co-assembly of three POIs on a DNA scaffold using SNAP-tag, Halo-tag, and biotin-streptavidin system.⁶⁷ a) Face-like DNA scaffold with the targets for mKate-SNAP (mouth), CCP-Halo (eyes), and mSTV (nose). b-d) Sequential assembly of mKate-SNAP (b), CCP-Halo (c), and mSVT (d) on the DNA scaffold with the overall coassembly of 7.5%. mKate: far-red fluorescent protein; CCP: cytochrome C peroxidase; mSTV: monovalent streptavidin.

1.3.2.3 Method base on HUH-tag proteins

Other approaching using the HUH-tags (or HUH endonuclease domains)^{73,74} to covalently locating POIs on a DNA scaffold (Figure 1.15). The *in vivo* endonuclease activity of HUH-tags is the nicking of the single-strand DNA at a specific sequence and following by the formation of the phosphotyrosine intermediate, specific tyrosine in the HUH-tag protein becomes covalently linked to the 5' end of the DNA strand. *In vitro*, the covalent linkage becomes permanent, HUH-tag fused POIs can be stably located on a DNA scaffold. The advantage of HUH-tags based method is that the method does not require substrate modification on the target DNA strands. With the abundance of sequence-specific HUH-tags,⁷⁵ this strategy can be used for assembling multiple enzymes on the design position of a DNA scaffold. However, the loading yield is not significantly high enough when compared with that of the other current methods.

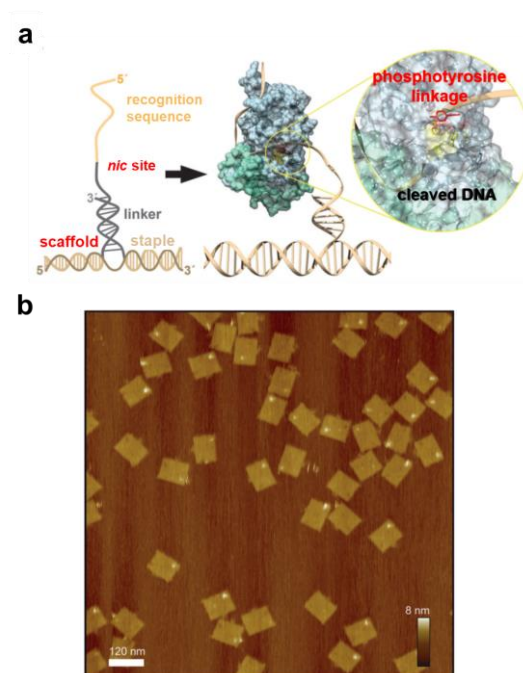


Figure 1.15. Covalent bond formation of the HUH-tag (relaxase) on DNA scaffold. a) Extended staples contain the recognition sequence of a relaxase and several base pairs linker that allows the target to extrude from the nanostructure. Binding of the relaxase to its cognate site involves bending of the ssDNA to fit the nic site into the relaxase active center. Inset: the

catalytic tyrosine carries out a transesterification reaction that links it covalently to the 5'-phosphate between the scissile nucleotides. As a result, the relaxase becomes covalently bound to the extended staple. b) AFM image of DNA nanostructures assembled with relaxases.

1.3.2.4 Modular adaptors

A modular adaptor based on the combination of a DNA binding protein and a self-ligating protein showed excellent improvement of the covalent assembly of POIs on the defined position of the DNA scaffold in which the yield of assembly reach almost unity under very mild incubation condition.^{76,77} In the design, SNAP-tag was fused to the C-terminus of the well-known zinc finger protein zif268 (ZF)⁷⁸ to form the modular adaptor ZF-SNAP that has both characteristics of the two modules to create the stable covalent bond at the designed position of the DNA scaffold (Figure 1.16). The zinc finger protein provides a fast assembly of POIs to the staple strand containing a specific sequence for ZF of the scaffold. The covalent bond is formed by the SNAP-tag modules at the targeted ODN in which the SNAP-tag substrate was tethered on (ODN-ZF-BG). In fact, the study indicates the fast and tight binding of zinc finger protein toward the sequence-specific ODN accelerated the covalent bond formation of the protein-tags with the ODN modified substrates. The assembly yield reached more than 90% after 30 min incubation) (Figure 1.17).

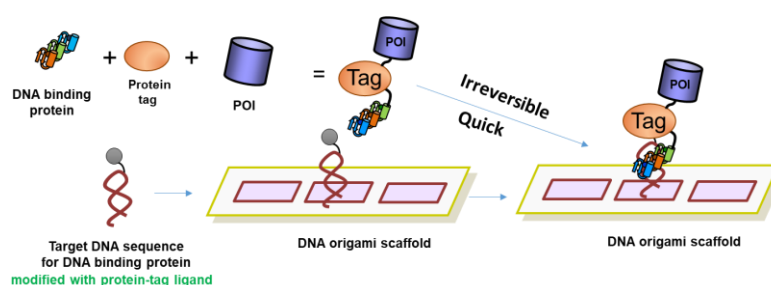


Figure 1.16 Design of modular adaptor for covalent assembly of POIs on DNA scaffold.

The modular adaptor consists of two modules, protein-tag and DNA binding protein. The

modular adaptors fused POIs can be located at the designed positions on the DNA scaffold containing the substrate modified DNA within the specific binding sequence for the DNA binding protein.

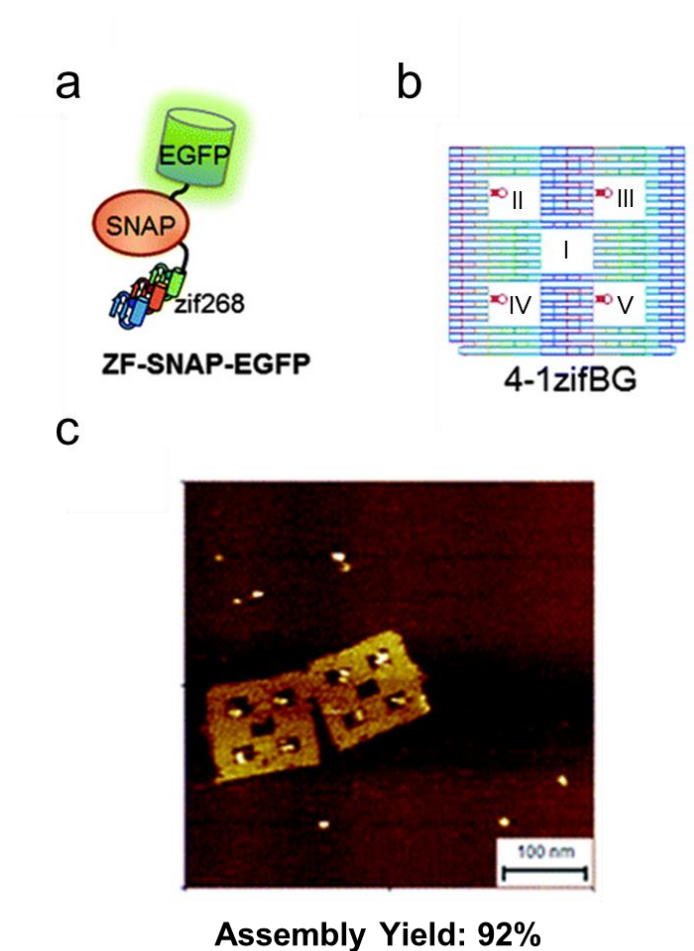


Figure 1.17. Assemble of EGPF on DNA scaffold using modular adaptor ZF-SNAP.⁷⁶ a) An illustration of EGFP fused to ZF-SNAP (ZF-SNAP-EGFP) b) The 5-well DNA scaffold containing a single binding site for ZF-SNAP-EGFP at each of the cavities II, III, IV and V. c) AFM image of ZF-SNAP-EGFP assembled on DNA scaffold.

The modular adaptor ZF-SNAP was used to load xylose reductase (XR), an enzyme from xylose metabolic pathway,⁷⁹ on a DNA scaffold.⁷⁷ The cascade reaction by ZF-SNAP fused XR (ZS-XR) and G-XDH⁵⁶ was artificially activated on a DNA scaffold and the dependency of the cascade reaction on the interenzyme distance was investigated (Figure 1.18).

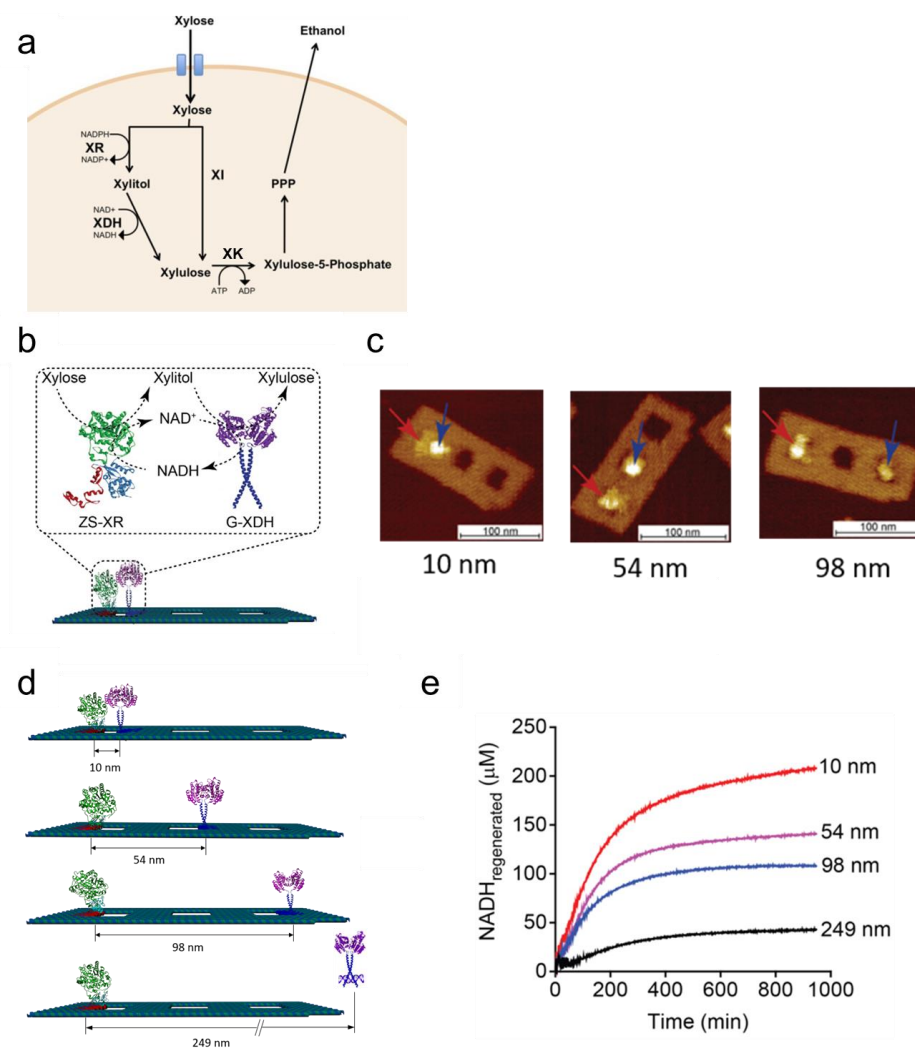


Figure 1.18 Artificial metabolic system on a DNA scaffold.¹⁷ a) Natural metabolism pathway for utilizing xylose in fungi and yeast.⁷⁹ XR: xylose reductase, XDH: xylitol dehydrogenase, XI: xylulose isomerase, XK: xylose kinase, PPP: pentose phosphate pathway b) A schematic illustration of the cascade reaction by two adaptor-fused enzymes (ZS-XR and G-XDH) coassembled on the DNA scaffold. d) Illustrations of the assembly of ZS-XR with or without G-XDH on DNA scaffold with different interenzyme distances. e) AFM images indicate the coassembly of ZS-XR and G-XDH on DNA scaffolds with different interenzyme distances. f) Time-course reaction profiles for the amount of NADH when two enzymes were coassembled with interenzyme distances of 10, 54, and 98 nm and for the free diffusion system, in which ZS-XR was located on DNA scaffold while G-XDH was free in solution with a theoretically estimated interspacing distances of 249 nm.

1.4 Aim of the research in this thesis

This thesis aims to develop a system of modular adaptors with fast loading and high orthogonality to assemble several types of enzymes on defined positions of a DNA scaffold. The modular design of adaptor possesses two characteristics, the sequence-specific of DNA binding proteins and the substrate selectivity of protein-tags showed significant improvement of the assembly yield of POIs on a DNA scaffold. Understanding the kinetic aspect of these molecular recognition and chemical reaction process is necessary for developing new modular adaptors. This study discussed the mechanism of the modular adaptors to establish a general strategy for expansion of the number of orthogonal modular adaptors system (Figure 1.19). In Chapter 2, I developed a series of modular adaptors that enabled the orthogonal assembly several POIs on DNA scaffold by investigating the combination of two DNA binding zinc finger proteins (zif268 and AZP4) and three available protein-tags (SNAP-tag, CLIP-tag, and Halo-tag) (Figure 1.20b).

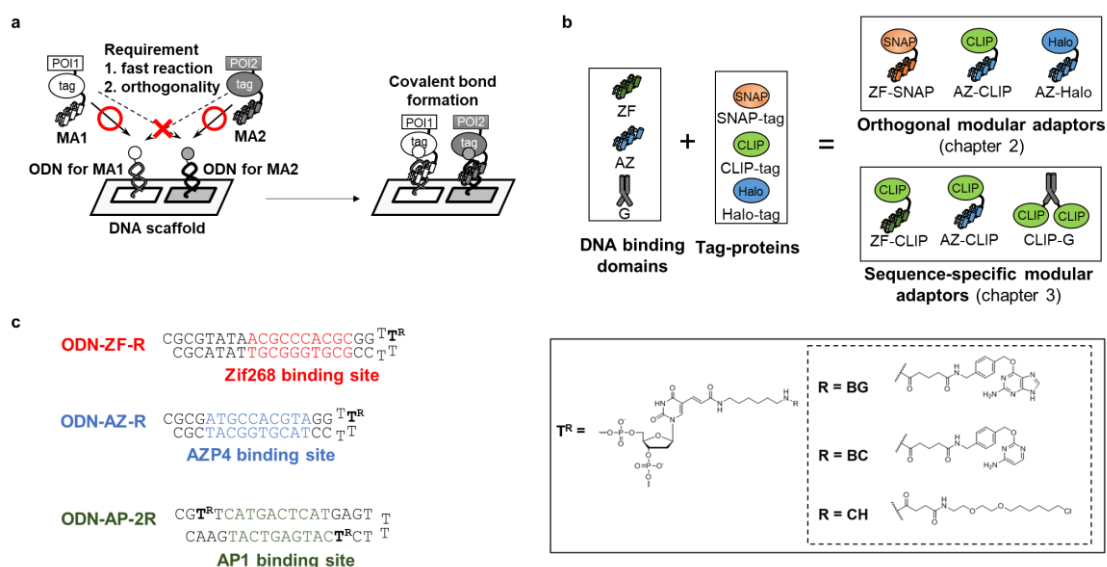


Figure 1.19 Design of the orthogonal modular adaptors. a) The modular adaptors system requires orthogonality and fast reaction at the specific positions on a DNA scaffold. b) Combination of DNA binding proteins and protein tags for the possible modular adaptors. The

1st generation: the orthogonal modular adaptors, of which the orthogonality is depended on both the sequences specificity of the DNA binding modules and substrates selectivity of protein-tags (Chapter 2). The 2nd generation: the sequence-specific modular adaptors, of which the orthogonality is depended exclusively on the sequence specificity of the DNA binding modules (Chapter 3). c) Design of substrate modified ODN containing the binding sites for zif268 (ZF), AZP4 (AZ), and GCN4 (G).

In Chapter 3, kinetic aspects of the covalent bonding formation of the modular adaptor were studied to develop a general design of the next generation of the orthogonal modular adaptors named sequence-specific modular adaptors. In this design, the orthogonality of modular adaptors is governed exclusively by the DNA binding modules, so that different DNA binding proteins could be used with only one type of protein-tag to create a series of sequence-specific modular adaptors (Figure 1.20b).

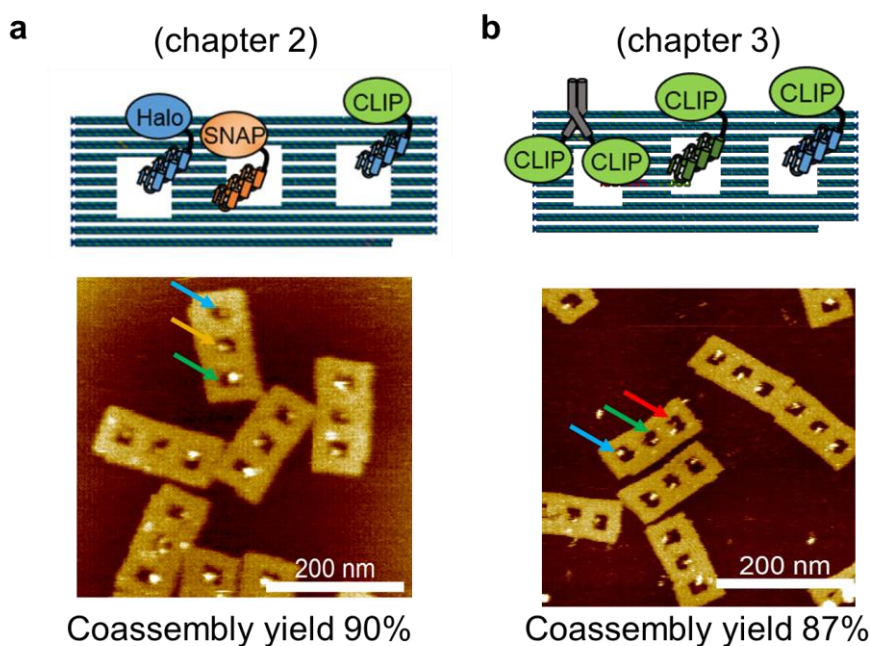


Figure 1.20. Coassembly of (a) orthogonal modular adaptors and (b) sequence-specific modular adaptors on DNA scaffold. (a) An illustration associated with AFM image of a DNA scaffold coassembled with three modular adaptors: AZ-Halo (green arrow), ZF-SNAP (orange arrow), and AZ-CLIP (blue arrow). (b) An illustration associated with AFM image of a DNA

scaffold coassembled with three modular adaptors: CLIP-G (red arrow), ZF-SNAP (green arrow), and AZ-CLIP (blue arrow).

The robustness of the orthogonal modular adaptors was next tested by using AZ-CLIP (AC) and AZ-Halo (AH) to assemble the unstable enzyme xylulose kinase (XK) on a DNA scaffold (Chapter 4). Indeed, XK was successfully constructed in the form of adaptor fused enzymes (AC-XK and AH-XK), and AC-XK and AH-XK were assembled on a DNA scaffold. I further constructed the artificial xylose metabolism pathway of three adaptor fused enzymes: ZS-XR (ZF-SNAP fused xylose reductase), G-XDH (GCN4 fused xylitol dehydrogenase), and AC-XK on a DNA scaffold. Several configurations of the assembly of three enzymes on DNA scaffold with the interenzyme distance vary from approximately 50 nm to 10 nm were constructed. The efficiency of the three enzymes cascade reaction was significantly enhanced when these three enzymes were placed at 10 nm proximity. This is the first study to investigate the cascades reaction derive from three enzymes of the natural metabolic pathway on a DNA scaffold (Chapter 4) (Figure 1.21).

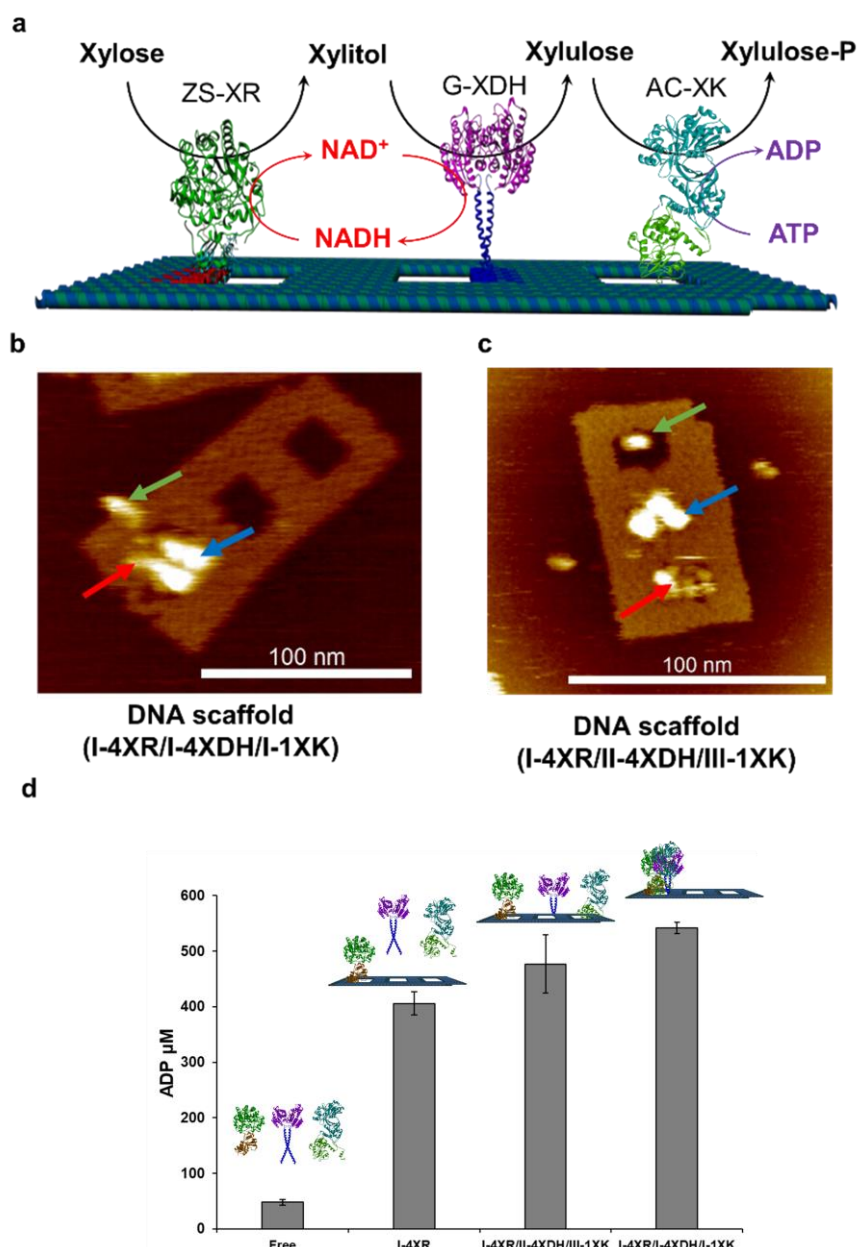


Figure 1.21. The three enzymes cascade reaction on a DNA scaffold (Chapter 4). (a)

An illustration of the cascade reaction by three enzymes ZS-XR, G-XDH and AC-XK on the DNA scaffold. (b, c) AFM images of the three enzymes bound on the DNA scaffolds (b) I-4XR/II-4XDH/III-1XK and (c) I-4XR/I-4XDH/I-1XK. The arrows in red, blue, and green indicate ZS-XR, G-XDH, and AC-XK, respectively. (d) Production of ADP from the cascade reaction all enzyme in bulk solution (Free), with the one enzyme assembled on DNA scaffold (I-4XR: ZS-XR located on the scaffold and G-XDH and AC-XK in bulk solution), and with the three enzymes assembled on DNA scaffold I-4XR/II-4XDH/III-1XK and I-4XR/I-4XDH/I-1XK.

1.5 References

- (1) Moran, U.; Phillips, R.; Milo, R. SnapShot: Key Numbers in Biology. *Cell* **2010**, *141*, 1–2.
- (2) Chen, A. H.; Silver, P. A. Designing Biological Compartmentalization. *Trends Cell Biol.* **2012**, *22*, 662–670.
- (3) Giessen, T. W.; Silver, P. A. Encapsulation as a Strategy for the Design of Biological Compartmentalization. *J. Mol. Biol.* **2016**, *428*, 916–927.
- (4) Zecchin, A.; Stapor, P. C.; Goveia, J.; Carmeliet, P. Metabolic Pathway Compartmentalization: An Underappreciated Opportunity? *Curr. Opin. Biotechnol.* **2015**, *34*, 73–81.
- (5) Itel, F.; Schattling, P. S.; Zhang, Y.; Städler, B. Enzymes as Key Features in Therapeutic Cell Mimicry. *Adv. Drug Deliv. Rev.* **2017**, *118*, 94–108.
- (6) Maity, B.; Fujita, K.; Ueno, T. Use of the Confined Spaces of Apo-Ferritin and Virus Capsids as Nanoreactors for Catalytic Reactions. *Curr. Opin. Chem. Biol.* **2015**, *25*, 88–97.
- (7) Kuchler, A.; Yoshimoto, M.; Luginbühl, S.; Mavelli, F.; Walde, P. Enzymatic Reactions in Confined Environments. *Nat. Nanotechnol.* **2016**, *11*, 409–420.
- (8) Rajendran, A.; Nakata, E.; Nakano, S.; Morii, T. Nucleic-Acid-Templated Enzyme Cascades. *ChemBioChem* **2017**, *18*, 696–716.
- (9) Linko, V.; Nummelin, S.; Aarnos, L.; Tapio, K.; Toppari, J.; Kostiainen, M. DNA-Based Enzyme Reactors and Systems. *Nanomaterials* **2016**, *6*, 139.
- (10) Kuchler, A.; Yoshimoto, M.; Luginbühl, S.; Mavelli, F.; Walde, P. Enzymatic Reactions in Confined Environments. *Nat. Nanotechnol.* **2016**, *11*, 409–420.
- (11) Luo, R.; Wei, H.; Ye, L.; Wang, K.; Chen, F.; Luo, L.; Liu, L.; Li, Y.; Crabbe, M. J. C.; Jin, L.; et al. Photosynthetic Metabolism of C3 Plants Shows Highly Cooperative Regulation under Changing Environments: A Systems Biological

- Analysis. *Proc. Natl. Acad. Sci.* **2009**, *106*, 847–852.
- (12) Wei, N.; Oh, E. J.; Million, G.; Cate, J. H. D.; Jin, Y. S. Simultaneous Utilization of Cellobiose, Xylose, and Acetic Acid from Lignocellulosic Biomass for Biofuel Production by an Engineered Yeast Platform. *ACS Synth. Biol.* **2015**, *4*, 707–713.
- (13) Bonacci, W.; Teng, P. K.; Afonso, B.; Niederholtmeyer, H.; Grob, P.; Silver, P. A.; Savage, D. F. Modularity of a Carbon-Fixing Protein Organelle. *Proc. Natl. Acad. Sci.* **2012**, *109*, 478–483.
- (14) Watford, M. Mini-Series : Modern Metabolic Concepts. *Biochem Mol. Biol. Educ.* **2003**, *31*, 289–297.
- (15) Klermund, L.; Poschenrieder, S. T.; Castiglione, K. Biocatalysis in Polymersomes: Improving Multienzyme Cascades with Incompatible Reaction Steps by Compartmentalization. *ACS Catal.* **2017**, *7*, 3900–3904.
- (16) Zhang, G.; Quin, M. B.; Schmidt-Dannert, C. Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes. *ACS Catal.* **2018**, *8*, 5611–5620.
- (17) Ngo, T. A.; Nakata, E.; Saimura, M.; Morii, T. Spatially Organized Enzymes Drive Cofactor-Coupled Cascade Reactions. *J. Am. Chem. Soc.* **2016**, *138*, 3012–3021.
- (18) Seeman, N. C. DNA in a Material World. *Nature* **2003**, *421*, 427–431.
- (19) Rothmund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.
- (20) Linko, V.; Ora, A.; Kostianen, M. A. DNA Nanostructures as Smart Drug-Delivery Vehicles and Molecular Devices. *Trends Biotechnol.* **2015**, *33*, 586–594.
- (21) Linko, V.; Dietz, H. The Enabled State of DNA Nanotechnology. *Curr. Opin. Biotechnol.* **2013**, *24*, 555–561.

- (22) Pinheiro, A. V.; Han, D.; Shih, W. M.; Yan, H. Challenges and Opportunities for Structural DNA Nanotechnology. *Nat. Nanotechnol.* **2011**, *6*, 763–772.
- (23) Wang, P.; Chatterjee, G.; Yan, H.; Labean, T. H.; Turberfield, A. J.; Castro, C. E.; Seelig, G.; Ke, Y. Practical Aspects of Structural and Dynamic DNA Nanotechnology. *MRS Bull.* **2017**, *42*, 889–896.
- (24) Endo, M.; Yang, Y.; Sugiyama, H. DNA Origami Technology for Biomaterials Applications. *Biomater. Sci.* **2013**, *1*, 347–360.
- (25) Marras, A. E.; Zhou, L.; Su, H.-J.; Castro, C. E. Programmable Motion of DNA Origami Mechanisms. *Proc. Natl. Acad. Sci.* **2015**, *112*, 713–718.
- (26) Kopperger, E.; List, J.; Madhira, S.; Rothfischer, F.; Lamb, D. C.; Simmel, F. C. A Self-Assembled Nanoscale Robotic Arm Controlled by Electric Fields. *Science.* **2018**, *359*, 296–301.
- (27) Li, J.; Johnson-Buck, A.; Yang, Y. R.; Shih, W. M.; Yan, H.; Walter, N. G. Exploring the Speed Limit of Toehold Exchange with a Cartwheeling DNA Acrobat. *Nat. Nanotechnol.* **2018**, *13*, 723–729.
- (28) Valero, J.; Pal, N.; Dhakal, S.; Walter, N. G.; Famulok, M. A Bio-Hybrid DNA Rotor–stator Nanoengine That Moves along Predefined Tracks. *Nat. Nanotechnol.* **2018**, *13*, 496–503.
- (29) Niemeyer, C. M.; Koehler, J.; Wuerdemann, C. DNA-Directed Assembly of Bienzymic Complexes from in Vivo Biotinylated NAD(P)H:FMN Oxidoreductase and Luciferase. *ChemBioChem* **2002**, *3*, 242–245.
- (30) Xin, L.; Zhou, C.; Yang, Z.; Liu, D. Regulation of an Enzyme Cascade Reaction by a DNA Machine. *Small* **2013**, *9*, 3088–3091.
- (31) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, *4*, 249–254.

- (32) Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter, N. G.; Woodbury, N. W.; Yan, H. Multi-Enzyme Complexes on DNA Scaffolds Capable of Substrate Channelling with an Artificial Swinging Arm. *Nat. Nanotechnol.* **2014**, *9*, 531–536.
- (33) Liu, M.; Fu, J.; Qi, X.; Wootten, S.; Woodbury, N. W.; Liu, Y.; Yan, H. A Three-Enzyme Pathway with an Optimised Geometric Arrangement to Facilitate Substrate Transfer. *ChemBioChem* **2016**, *17*, 1097–1101.
- (34) Linko, V.; Eerikäinen, M.; Kostianen, M. A. A Modular DNA Origami-Based Enzyme Cascade Nanoreactor. *Chem. Commun.* **2015**, *51*, 5351–5354.
- (35) Fu, Y.; Zeng, D.; Chao, J.; Jin, Y.; Zhang, Z.; Liu, H.; Li, D.; Ma, H.; Huang, Q.; Gothelf, K. V.; Fan, C. Single-Step Rapid Assembly of DNA Origami Nanostructures for Addressable Nanoscale Bioreactors. *J. Am. Chem. Soc.* **2013**, *135*, 696–702.
- (36) Delebecque, C. J.; Lindner, A. B.; Silver, P. A.; Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*. **2011**, *333*, 470–474.
- (37) Yang, Y. R.; Liu, Y.; Yan, H. DNA Nanostructures as Programmable Biomolecular Scaffolds. *Bioconjug. Chem.* **2015**, *26*, 1381–1395.
- (38) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chemie., Int. Ed.* **2009**, *48*, 6974–6998.
- (39) Niemeyer, C. M. Semisynthetic DNA-Protein Conjugates for Biosensing and Nanofabrication. *Angew. Chemie., Int. Ed.* **2010**, *49*, 1200–1216.
- (40) Goodman, R. P.; Erben, C. M.; Malo, J.; Ho, W. M.; Mckee, M. L.; Kapanidis, A. N.; Turberfield, A. J. A Facile Method for Reversibly Linking a Recombinant Protein to DNA. *ChemBioChem* **2009**, *10*, 1551–1557.
- (41) Wegner, S. V.; Spatz, J. P. Cobalt(III) as a Stable and Inert Mediator Ion between

- NTA and His6-Tagged Proteins. *Angew. Chemie - Int. Ed.* **2013**, *52*, 7593–7596.
- (42) Shen, W.; Zhong, H.; Neff, D.; Norton, M. L.; V, M. U.; Virginia, W. NTA Directed Protein Nanopatterning on DNA Origami Nanoconstructs. *J. Am. Chem. Soc.* **2009**, *131*, 6660–6661.
- (43) Wilchek, M.; Bayer, E. A. The avidin-biotin complex in bioanalytical applications. *Anal. Biochem.* **1988**, *171*, 1–32.
- (44) Li, H.; Park, S. H.; Reif, J. H.; LaBean, T. H.; Yan, H. DNA-Templated Self-Assembly of Protein and Nanoparticle Linear Arrays. *J. Am. Chem. Soc.* **2004**, *126*, 418–419.
- (45) Lund, K.; Liu, Y.; Lindsay, S.; Yan, H. Self-Assembling a Molecular Pegboard. *J. Am. Chem. Soc.* **2005**, *127*, 17606–17607.
- (46) Yan, H. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science*. **2003**, *301*, 1882–1884.
- (47) Kuzuya, A.; Kimura, M.; Numajiri, K.; Koshi, N.; Ohnishi, T.; Okada, F.; Komiyama, M. Precisely Programmed and Robust 2D Streptavidin Nanoarrays by Using Periodical Nanometer-Scale Wells Embedded in DNA Origami Assembly. *ChemBioChem* **2009**, *10*, 1811–1815.
- (48) Voigt, N. V.; Tørring, T.; Rotaru, A.; Jacobsen, M. F.; Ravnsbæk, J. B.; Subramani, R.; Mamdouh, W.; Kjems, J.; Mokhir, A.; Besenbacher, F.; Gothelf, K. V. Single-Molecule Chemical Reactions on DNA Origami. *Nat. Nanotechnol.* **2010**, *5*, 200–203.
- (49) Numajiri, K.; Kimura, M.; Kuzuya, A.; Komiyama, M. Stepwise and Reversible Nanopatterning of Proteins on a DNA Origami Scaffold. *Chem. Commun.* **2010**, *46*, 5127–5129.
- (50) Numajiri, K.; Yamazaki, T.; Kimura, M.; Kuzuya, A.; Komiyama, M. Discrete and Active Enzyme Nanoarrays on DNA Origami Scaffolds Purified by Affinity

- Tag Separation. *J. Am. Chem. Soc.* **2010**, *132*, 9937–9939.
- (51) Yu He; Ye Tian; Alexander E. Ribbe, A.; Mao, C. Antibody Nanoarrays With Pitch of ~20 nm. *J. Am. Chem. Soc.* **2006**, *128*, 1–2.
- (52) Chhabra, R.; Sharma, J.; Ke, Y.; Liu, Y.; Rinker, S.; Lindsay, S.; Yan, H. Spatially Addressable Multiprotein Nanoarrays Templated by Aptamer-Tagged DNA Nanoarchitectures. *J. Am. Chem. Soc.* **2007**, *129*, 10304–10305.
- (53) Williams, B. A. R.; Lund, K.; Liu, Y.; Yan, H.; Chaput, J. C. Self-Assembled Peptide Nanoarrays: An Approach to Studying Protein-Protein Interactions. *Angew. Chemie., Int. Ed.* **2007**, *46*, 3051–3054.
- (54) Rinker, S.; Ke, Y.; Liu, Y.; Chhabra, R.; Yan, H. Self-Assembled DNA Nanostructures for Distance-Dependent Multivalent Ligand-Protein Binding. *Nat. Nanotechnol.* **2008**, *3*, 418–422.
- (55) Nakata, E.; Liew, F. F.; Uwatoko, C.; Kiyonaka, S.; Mori, Y.; Katsuda, Y.; Endo, M.; Sugiyama, H.; Morii, T. Zinc-Finger Proteins for Site-Specific Protein Positioning on DNA-Origami Structures. *Angew. Chemie., Int. Ed.* **2012**, *51*, 2421–2424.
- (56) Ngo, T. A.; Nakata, E.; Saimura, M.; Kodaki, T.; Morii, T. A Protein Adaptor to Locate a Functional Protein Dimer on Molecular Switchboard. *Methods* **2014**, *67*, 142–150.
- (57) Pabo, C. O.; Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053–1095.
- (58) Prabakaran, P.; Siebers, J. G.; Ahmad, S.; Gromiha, M. M.; Singarayan, M. G.; Sarai, A. Classification of Protein-DNA Complexes Based on Structural Descriptors. *Structure* **2006**, *14*, 1355–1367.
- (59) Hudson, W. H.; Ortlund, E. A. The Structure, Function and Evolution of Proteins That Bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 749–760.

- (60) Kurokawa, T.; Kiyonaka, S.; Nakata, E.; Endo, M.; Koyama, S.; Mori, E.; Tran, N. H.; Dinh, H.; Suzuki, Y.; Hidaka, K.; et al. DNA Origami Scaffolds as Templates for Functional Tetrameric Kir3 K⁺ Channels. *Angew. Chemie., Int. Ed.* **2018**, *57*, 2586–2591.
- (61) Fu, J.; Yang, Y. R.; Dhakal, S.; Zhao, Z.; Liu, M.; Zhang, T.; Walter, N. G.; Yan, H. Assembly of Multienzyme Complexes on DNA Nanostructures. *Nat. Protoc.* **2016**, *11*, 2243–2273.
- (62) Rosen, C. B.; Kodal, A. L. B.; Nielsen, J. S.; Schaffert, D. H.; Scavenius, C.; Okholm, A. H.; Voigt, N. V.; Enghild, J. J.; Kjems, J.; Tørring, T.; et al. Template-Directed Covalent Conjugation of DNA to Native Antibodies, Transferrin and Other Metal-Binding Proteins. *Nat. Chem.* **2014**, *6*, 804–809.
- (63) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, *4*, 249–254.
- (64) Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **2012**, *134*, 5516–5519.
- (65) Feeney, R. E.; Yamasaki, R. B.; Geoghegan, K. F. Chemical Modification of Proteins: An Overview. *Adv. Chem.* **1982**, *198*, 3–55.
- (66) Rosen, C. B.; Kodal, A. L. B.; Nielsen, J. S.; Schaffert, D. H.; Scavenius, C.; Okholm, A. H.; Voigt, N. V.; Enghild, J. J.; Kjems, J.; Tørring, T.; Gothelf K. V. Template-Directed Covalent Conjugation of DNA to Native Antibodies, Transferrin and Other Metal-Binding Proteins. *Nat. Chem.* **2014**, *6*, 804–809.
- (67) Saccà, B.; Meyer, R.; Erkelenz, M.; Kiko, K.; Arndt, A.; Schroeder, H.; Rabe, K. S.; Niemeyer, C. M. Orthogonal Protein Decoration of DNA Origami. *Angew. Chemie., Int. Ed.* **2010**, *49*, 9378–9383.

- (68) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A General Method for the Covalent Labeling of Fusion Proteins with Small Molecules in Vivo. *Nat. Biotechnol.* **2003**, *21*, 86–89.
- (69) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; et al. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.
- (70) Koßmann, K. J.; Ziegler, C.; Angelin, A.; Meyer, R.; Skoupi, M.; Rabe, K. S.; Niemeyer, C. M. A Rationally Designed Connector for Assembly of Protein-Functionalized DNA Nanostructures. *ChemBioChem* **2016**, *17*, 1102–1106.
- (71) Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G. Peptide-Tags for Site-Specific Protein Labelling: In Vitro and in Vivo. *Mol. Biosyst.* **2016**, *12*, 1731–1745.
- (72) Li, C.; Tebo, A. G.; Gautier, A. Fluorogenic Labeling Strategies for Biological Imaging. *Int. J. Mol. Sci.* **2017**, *18*, 1473.
- (73) Sagredo, S.; Pirzer, T.; AghebatRafat, A.; Goetzfried, M. A.; Moncalian, G.; Simmel, F. C.; De La Cruz, F. Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases. *Angew. Chemie., Int. Ed.* **2016**, *55*, 4348–4352.
- (74) Lovendahl, K. N.; Hayward, A. N.; Gordon, W. R. Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. *J. Am. Chem. Soc.* **2017**, *139*, 7030–7035.
- (75) Chandler, M.; De La Cruz, F.; Dyda, F.; Hickman, A. B.; Moncalian, G.; Ton-Hoang, B. Breaking and Joining Single-Stranded DNA: The HUH Endonuclease Superfamily. *Nat. Rev. Microbiol.* **2013**, *11*, 525–538.
- (76) Nakata, E.; Dinh, H.; Ngo, T. A.; Saimura, M.; Morii, T. A Modular Zinc Finger

- Adaptor Accelerates the Covalent Linkage of Proteins at Specific Locations on DNA Nanoscaffolds. *Chem. Commun.* **2015**, 51, 1016–1019.
- (77) Ngo, T. A.; Nakata, E.; Saimura, M.; Morii, T. Spatially Organized Enzymes Drive Cofactor-Coupled Cascade Reactions. *J. Am. Chem. Soc.* **2016**, 138, 3012–3021.
- (78) Pavletich, N. P.; Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science*. **1991**, 252, 809–817.
- (79) Kricka, W.; Fitzpatrick, J.; Bond, U. Metabolic Engineering of Yeasts by Heterologous Enzyme Production for Degradation of Cellulose and Hemicellulose from Biomass: A Perspective. *Front. Microbiol.* **2014**, 5, 174.

CHAPTER 2

**Design of modular protein-tags for the
orthogonal covalent bond formation at specific
DNA sequences**

2.1 Introduction

DNA nanostructures¹⁻⁴ provide an ideal scaffold to assemble the proteins of interest (POI), such as enzymes participating in sequential enzymatic reactions, in a spatially well-defined manner.⁵⁻⁹ Construction of a given sequential enzymatic reaction *in vitro* accelerates our understanding of the spatial factor and the chemistry behind the highly efficient cascade reaction and provides a principle for further application of current enzyme technology.⁵⁻⁹ In many cases, POIs were located on the DNA scaffold through the hybridization of oligodeoxyribonucleotides (ODNs) that were chemically-modified on POIs.¹⁰⁻¹¹ This approach could locate multiple POIs on the DNA scaffold, but the activity of POIs tends to be reduced upon the chemical modification with ODN.¹¹⁻¹⁵ Other methods to stably locate a POI on the DNA scaffold, such as the one relies on the avidin-biotin interaction, may not be suitable for assembling more than two kinds of POIs on the same scaffold.¹⁰⁻¹¹ Sequence-specific DNA binding proteins, the zinc finger protein¹⁶ and the basic-leucine zipper protein¹⁷, have utilized by the Morii laboratory as the protein-based DNA binding adaptors to locate a POI at a target position on the DNA scaffold. Upon simple construction of adaptor-fused POIs by genetic modification, the sequence-selective DNA binding of these adaptors enables orthogonal arrangements of POIs at the specific DNA sequences on the DNA scaffold with fast binding kinetics.¹⁶ A drawback of this adaptor system could be its reversible nature, which causes incomplete loading of POIs to the defined DNA addresses on the DNA scaffold. To overcome this issue, a chemoselective crosslinking domain was conjugated to the adaptor to afford a modular adaptor, which would convert the noncovalent adaptor-DNA complex into a covalently linked complex. A modular adaptor (ZF-SNAP)¹⁸ consisting of the zinc finger protein (zif268)¹⁹ and a self-ligating protein-tag (SNAP-tag)²⁰ expedites a covalent linkage formation between

the SNAP-tag domain and its substrate modified at a target DNA sequence on the DNA scaffold to near saturation. A monomeric enzyme xylose reductase (XR) fused to ZF-SNAP (ZS-XR) was specifically located on the DNA scaffold in fast reaction kinetics and in a high yield with fully retaining the enzymatic activity.²¹

In order to assemble several POIs on the DNA scaffold at their specific locations in functional forms, modular adaptors with orthogonality and fast reaction kinetics under a mild condition are required, especially for the thermally unstable proteins. The self-ligating protein tags²² modified with a basic peptide²³ may not be suitable for this purpose even with its improved reaction efficiency. Combination of DNA-binding proteins with distinct sequence-specificity and the same self-ligating protein tag would ideally provide a series of modular adaptors that orthogonally forms covalent linkages at different DNA sequences. However, ZF-SNAP forms a covalent linkage not only with the substrate in its specific DNA sequence but also with that in the nonspecific DNA sequence.¹⁸ This result is reminiscent of the DNA alkylation by small molecules, in which the balance of reactivity and affinity of ligands are important to induce the alkylation reaction at specific DNA sequences.²⁴ In this work, I asked a question of whether a combination of protein-tags and DNA-binding proteins could provide modular adaptors that orthogonally form covalent linkages at specific locations with defined substrates (Figure 2.1). Among the combination of two DNA-binding proteins and three protein-tags, a set of three modular adaptors was successfully chosen to orthogonally react at their respective target sites with over 90% coassembly yield.

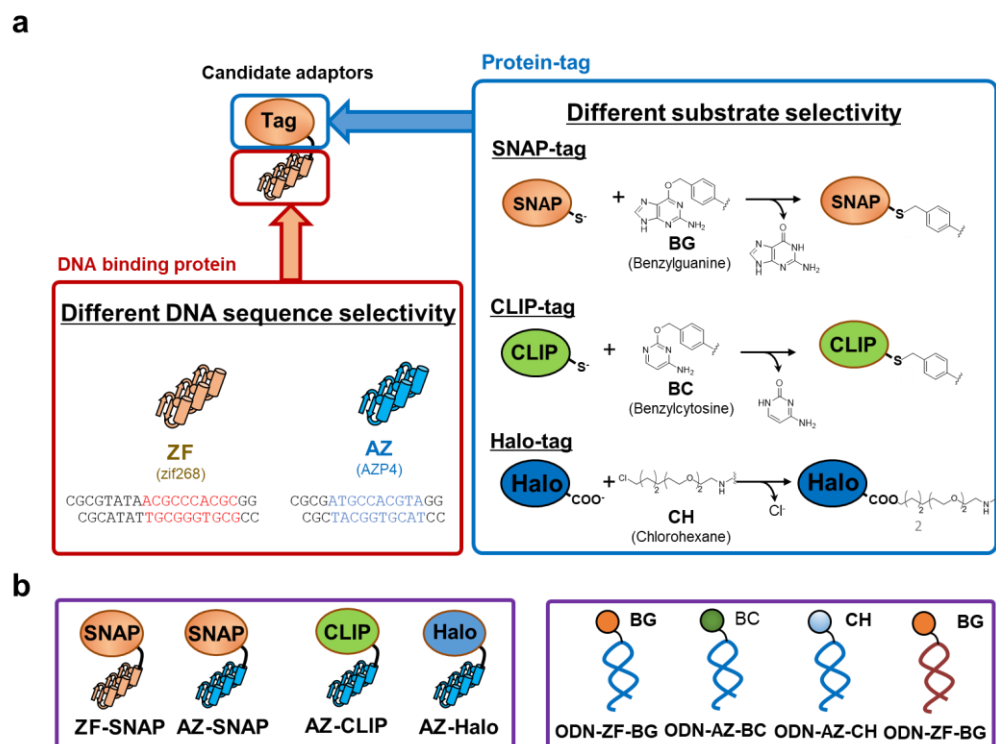


Figure 2.1 Candidates of the orthogonal modular adaptors. (a) Protein-tag modules (SNAP-tag, CLIP-tag, and Halo-tag) and DNA binding modules (ZF and AZ). (b) Candidates of the modular adaptors used in this study: ZF-SNAP, AZ-SNAP, AZ-CLIP, and AZ-Halo with the matched substrate modified ODNs (ODN-ZF-BG, ODN-AZ-BG, ODN-AZ-BC, and ODN-AZ-CH), respectively.

2.2 Results

2.2.1 Design of modular adaptors and ODNs modified with the substrates for protein-tags

A series of modular adaptors consisting of a DNA binding domain and a crosslink-forming protein-tag (Figure 2.2a) was systematically constructed to evaluate the orthogonal covalent bond formation at specific positions of DNA addresses on the DNA scaffold (Figure 2.2b). Zinc finger proteins zif268 and AZP4,²⁸ a derivative of zif268 with unique sequence selectivity, were utilized as the sequence-specific DNA binding domain (Figure 2.2c). Three protein-tags, SNAP-tag²⁰, CLIP-tag²⁹ and Halo-tag³⁰ were adapted as the crosslink-forming domain. SNAP-tag and CLIP-tag are reported to react specifically with O⁶-benzylguanine (BG) derivatives and O²-benzylcytosine (BC) derivatives, respectively, but they retain a certain degree of cross reactivities.²⁸ Halo-tag reacts specifically with 5-chlorohexane (CH) derivatives with much less cross reactivity to BG or BC. By fusing SNAP-tag, CLIP-tag and Halo-tag, respectively, to the C-terminal of AZP4 through a GGSGGS linker, a series of modular adaptors, AZ-SNAP, AZ-CLIP, and AZ-Halo, were constructed (Figures 2.3, 2.4, Table 2.1). Crosslinking reactions of these three modular adaptors were compared with that of ZF-SNAP.

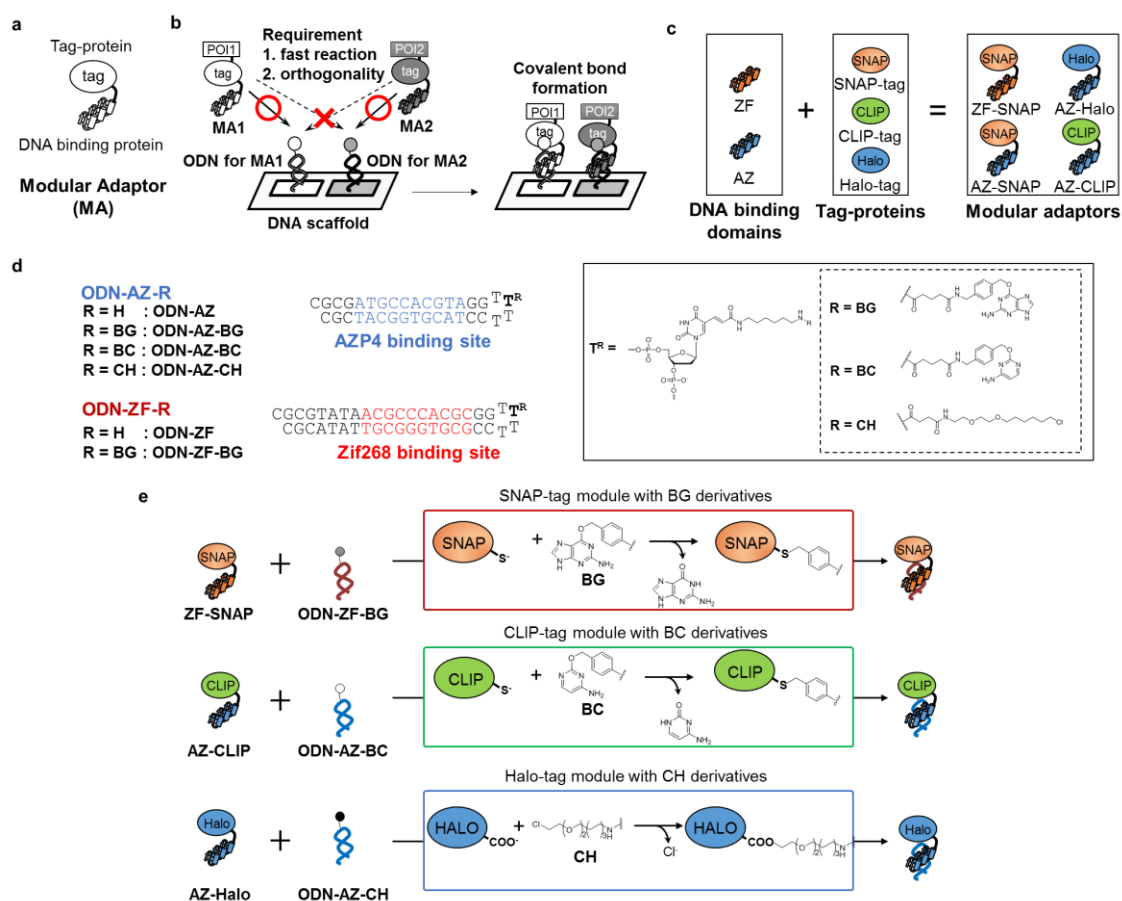


Figure 2.2. (a) An illustration of modular adaptor. (b) A scheme representing fast and orthogonal loading of POIs fused modular adaptors at the defined DNA addresses on the DNA scaffold. (c) Combination of the DNA binding domains and tag-proteins for constructing a series of modular adaptors used in this study. (d) Nucleotide sequences of the ODNs for each of modular adaptors with the substrate for protein-tags and the chemical structures of the substrates for protein-tags modified T, denoted as “T^R”. (e) Reaction schemes representing the crosslinking reactions between the modular adaptors and the substrates incorporated in the ODN.

a

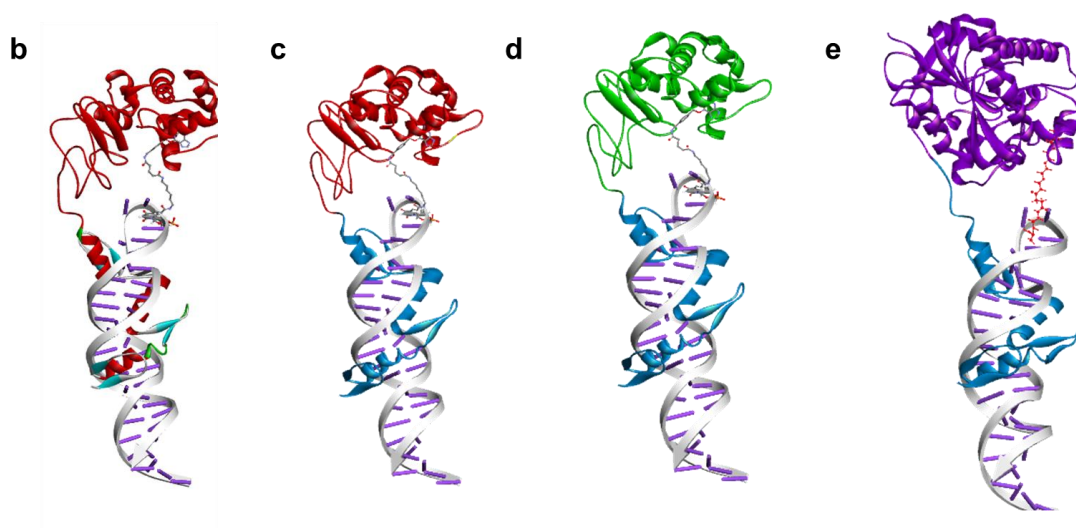
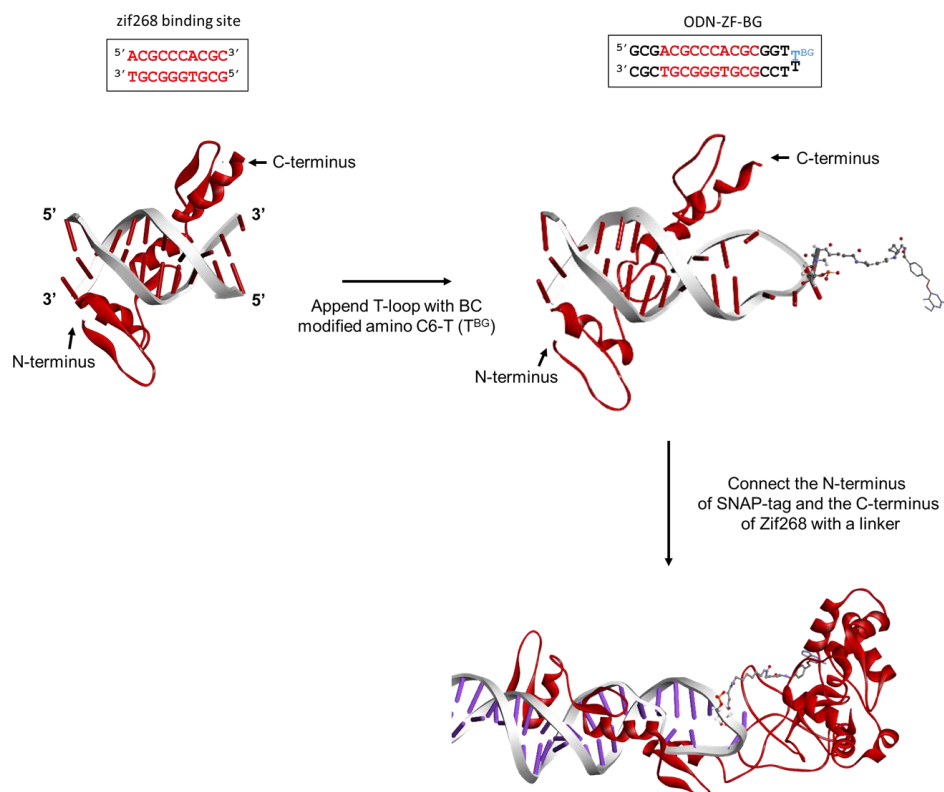


Figure 2.3. a) A scheme illustrates the design of ZF-SNAP and ODN-ZF-BG. (b-d) Models for the complex of (b) ZF-SNAP with ODN-ZF-BG, (c) AZ-SNAP with ODN-AZ-BG, (d) AZ-CLIP with ODN-AZ-BC, (e) AZ-Halo with ODN-AZ-CH, respectively, based on the crystal structures of the complex between zif268 and ODN (PDB ID : 1ZAA) and the complex of SNAP-tag with BG (PDB ID:3KZY), and Halo-tag with CH (PDB ID : 1CQW). The models were constructed by using Discovery Studio (version 3.1, Accelrys Inc.).

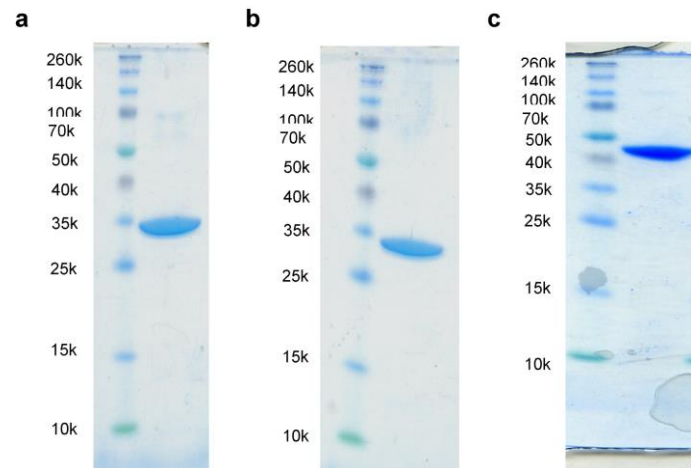


Figure 2.4. SDS-PAGE analyses of purified modular adaptors. (a) AZ-SNAP, (b) AZ-CLIP, (c) AZ-Halo. Amino acid sequences and molecular weights of these proteins were shown in Table 2.1.

Table 2.1 Amino acid sequences and molecular weights of modular adaptors.

Modular adaptors	Amino acid sequences	Mw (Da)
AZ-SNAP	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPF QCRICMRNFSRSDSLTRHIRTHTGEKPFACDICGRKFAE SDNRKTHTKIHTGEKEFGGSGGSMDKDCEMKRTTLDSP LKLELSGCEQGLHEIIIFLGKGTSAADAVEVPAPAAVLGG PEPLMQATAWLNAYFHQPEAIEEFVVPALHHPVFQQESF TRQVLWKLLKVVKFGEVISYSHLAALAGNPAATAAVKTA LSGNPVPILIPCHRUVVQGDLDVGGYEGGLAVKEWLLAHE GHRLGKPGLGGSGGSHHHHHH	32,181
AZ-CLIP	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPF QCRICMRNFSRSDSLTRHIRTHTGEKPFACDICGRKFAE SDNRKTHTKIHTGEKEFGGSGGSMDKDCEMKRTTLDSP LKLELSGCEQGLHRIIFLGKGTSAADAVEVPAPAAVLGG PEPLIQATAWLNAYFHQPEAIEEFVVPALHHPVFQQESF TRQVLWKLLKVVKFGEVISESHLAALVGNPAATAAVNTA LDGNPVPILIPCHRUVVQGDSDVGPYLGGLAVKEWLLAHE GHRLGKPGLGGSGGSHHHHHH	32,196
AZ-Halo	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPF QCRICMRNFSRSDSLTRHIRTHTGEKPFACDICGRKFAE SDNRKTHTKIHTGEKEFGGSGGSMAEIGTGFPDPHYVE VLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHV APTHRCIAPDLIGMGKSDKPDLYFFDDHVRFMDFIEA LGLLEEVVLVIHDWGSALGFHWAKRNPVERVKGI AFMEFIR PIPTWDEWPEFARETFFQAFRTTDVGRKLIIDQNVFIEGT LPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPI AGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPA EAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARW LSTLEISGEPTTEDLYFQSDNAIAHHHHHHH	47,768

Among the modular adaptors, ZF-SNAP and AZ-SNAP shared the same protein-tag (SNAP-tag) but differed in the DNA binding domain (zif268 or AZP4). Comparison of them asks whether the orthogonality is dictated by the specificity of DNA binding domain or not. AZ-SNAP, AZ-CLIP, and AZ-Halo possessed the same DNA binding domain (AZP4) but fused to different protein-tags (SNAP-tag, CLIP-tag or Halo-tag). Reactions of these adaptors clarify whether the substrate selectivity of protein-tag could control the orthogonality of the crosslinking reaction or not.

An ODN containing the AZP4 binding sequence was designed to form a loop with four T nucleotides, in which one of the T nucleotides was displaced by amino-C6-T (ODN-AZ) (Figure 2.2d). The amino-C6-T was modified with each of the substrates for the protein-tags (Figure 2.2d: ODN-AZ-BG, ODN-AZ-BC, and ODN-AZ-CH). ODN containing the zif268 binding sequence modified with BG (ODN-ZF-BG) was also prepared (Figure 2.2d).¹⁸ The binding ability of the modular adaptors AZ-SNAP, AZ-CLIP and AZ-Halo to the AZP4 binding sequence (ODN-AZ) or to the zif268 binding sequence (ODN-ZF), both lacking the substrate modification, was studied by titrations of gel mobility shift (Figure 2.5, 2.6 and Table 2.2). The equilibrium dissociation constants (K_D) for the matched complexes, AZ-SNAP, AZ-CLIP and AZ-Halo for ODN-AZ, were 61 ± 30 nM, 65 ± 2 nM, and 118 ± 6 nM, respectively (Figure 2.5, Table 2.2). For the unmatched complexes, AZ-SNAP, AZ-CLIP, and AZ-Halo did not afford measurable dissociation constants for ODN-ZF ($K_D > 1000$ nM, Figure 2.6, Table 2.2). These results confirmed the expected sequence-specific DNA binding of these modular adaptors.

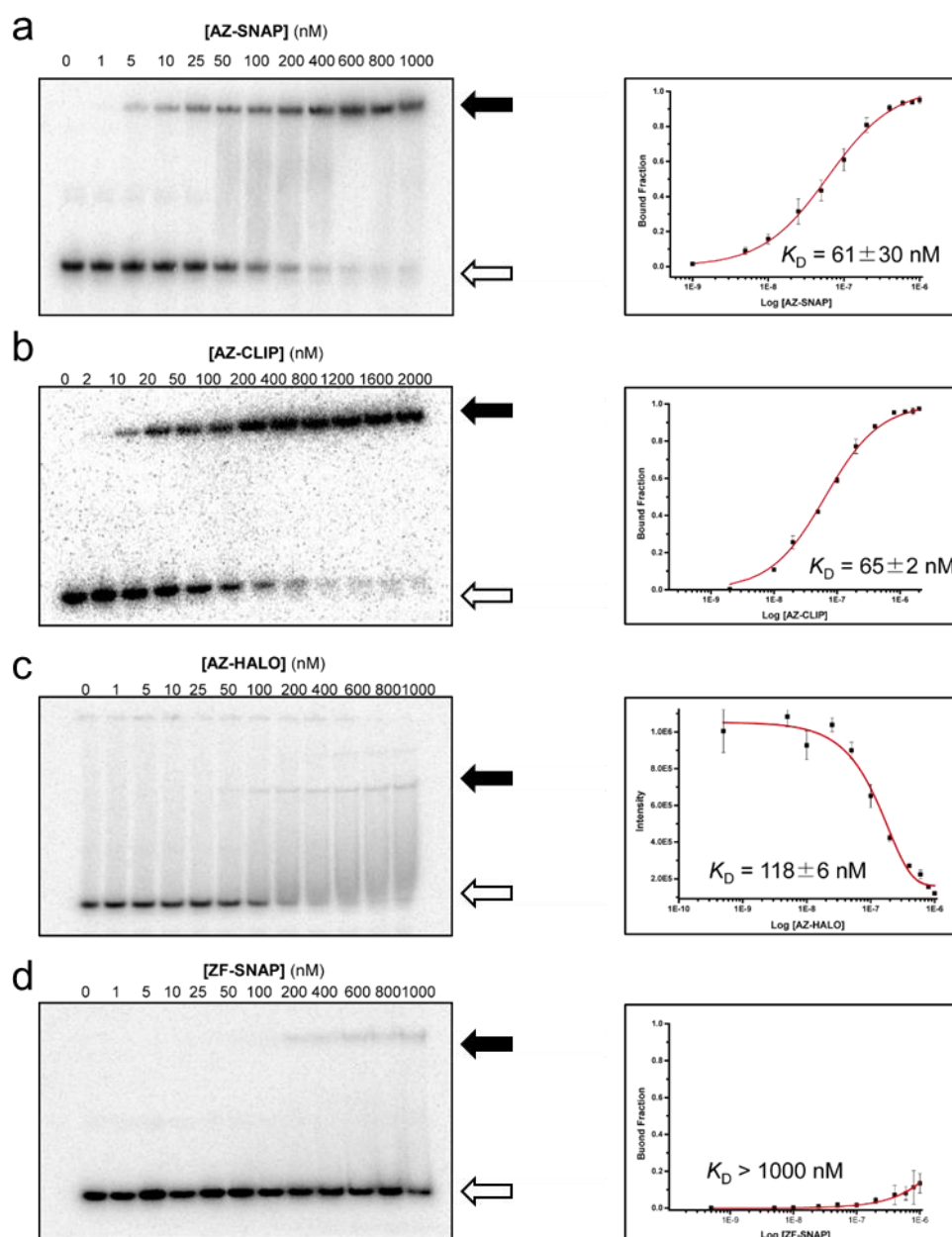


Figure 2.5. Autoradiograms show the electrophoretic mobility shift titration of (a) AZ-SNAP, (b) AZ-CLIP, (c) AZ-Halo, and (d) ZF-SNAP to 0.05 nM 5'-³²P-labeled ODN-AZ in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrows and filled arrows denote free ODN-AZ and modular adaptor bound ODN-AZ, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 2.2

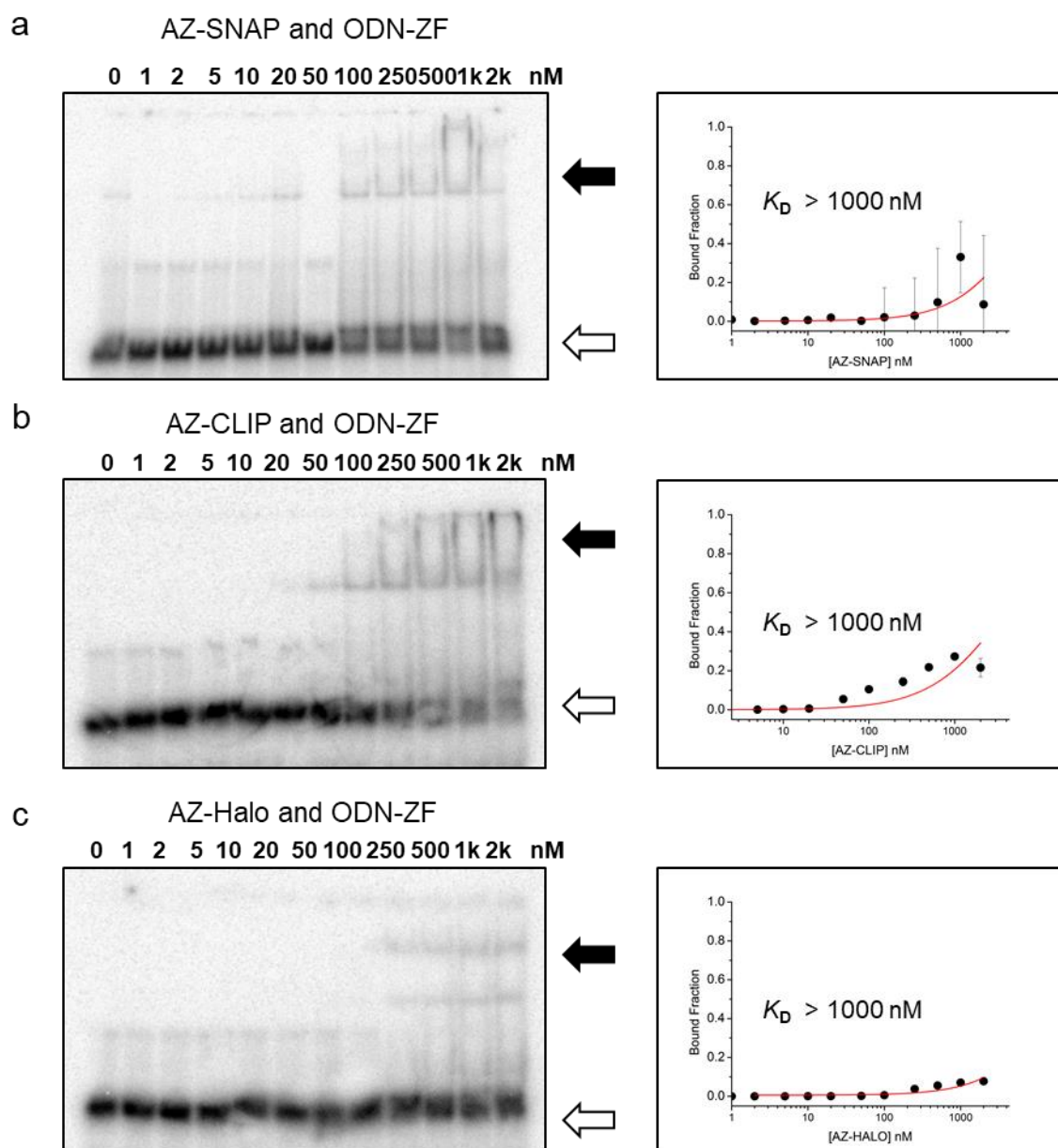


Figure 2.6. Autoradiograms show the electrophoretic mobility shift titration of (a) AZ-SNAP, (b) AZ-CLIP, (c) AZ-Halo, and to 0.05 nM 5'-³²P-labeled ODN-ZF in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrow and filled arrow denote free ODN-ZF and modular adaptor bound ODN-ZF, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 2.2.

Table 2.2. Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs.

Modular Adaptors	K_D (nM)	
	ODN-AZ	ODN-ZF
ZF-SNAP	> 1000	63 ± 18*
AZ-SNAP	61 ± 30	> 1000
AZ-CLIP	65 ± 2	> 1000
AZ-Halo	118 ± 6	> 1000

*the value is taken from ref. 18

2.2.2 Formation of covalent bonds between the modular adaptors and the substrate modified ODNs

As reported previously, the slow kinetics of the covalent bond formation of SNAP-tag by itself²² to its substrate on ODN was dramatically improved by using ZF-SNAP due to the fast kinetics of zif268 binding to its target DNA sequence and successive increase in the effective molarity of substrate.¹⁸ Formation of a covalent linkage between each of the modular adaptors and the substrate modified ODNs was monitored by denaturing polyacrylamide gel electrophoreses (PAGE) as shown in Figure 7. For all the reactions of matched pairs, ZF-SNAP with ODN-ZF-BG, AZ-SNAP with ODN-AZ-BG, AZ-CLIP with ODN-AZ-BC, and AZ-Halo with ODN-AZ-CH, the crosslinking yields exceeded 90% (Figure 2.7b). The second-order rate constants of these reactions were determined to be $10^5 - 10^6 \text{ M}^{-1}\text{s}^{-1}$ (Figure 2.8-2.11 and Table 2.3). These values were similar to each other within the four modular adaptors despite the fact that the parent protein-tags showed several orders of difference in the rate constants ($10^3 - 10^6 \text{ M}^{-1}\text{s}^{-1}$).^{20,29,30} The reactions of unmatched pairs in the same conditions were evaluated next (Figure 2.7b). Among them, the cross-linking yields of ZF-SNAP to ODN-AZ-BG and AZ-SNAP to ODN-ZF-BG also reached over

90%. In addition, measurable cross-linking reactions were observed for AZ-SNAP to ODN-AZ-BC (15%) and AZ-CLIP to ODN-AZ-BG (14%). No detectable reaction was observed for the unmatched pairs of AZ-Halo.

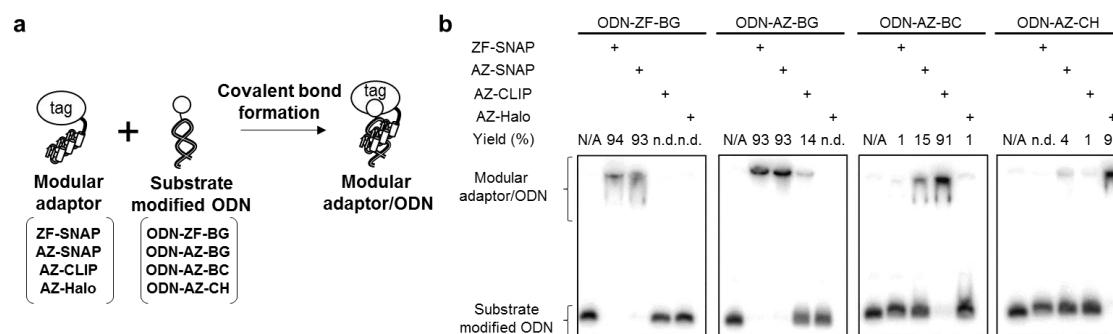


Figure 2.7. Analyses of crosslinking reactions between the modular adaptors and the substrate modified ODNs. (a) A scheme illustrates the crosslinking reactions of the modular adaptors to the substrate modified ODNs. (b) Denaturing PAGE analyses of the crosslinking reaction by modular adaptors (ZF-SNAP, AZ-SNAP, AZ-CLIP, and AZ-Halo) and each of the substrates modified ODNs, (ODN-ZF-BG, ODN-AZ-BG, ODN-AZ-BC, and ODN-AZ-CH), respectively. A substrate modified ODN was 5'-³²P-end-labeled (0.5 nM) and incubated with a modular adaptor (100 nM) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, and 100 nM calf thymus DNA. Reactions were carried out for 30 min at ambient temperature. N/A : not applicable, n.d. : not detectable.

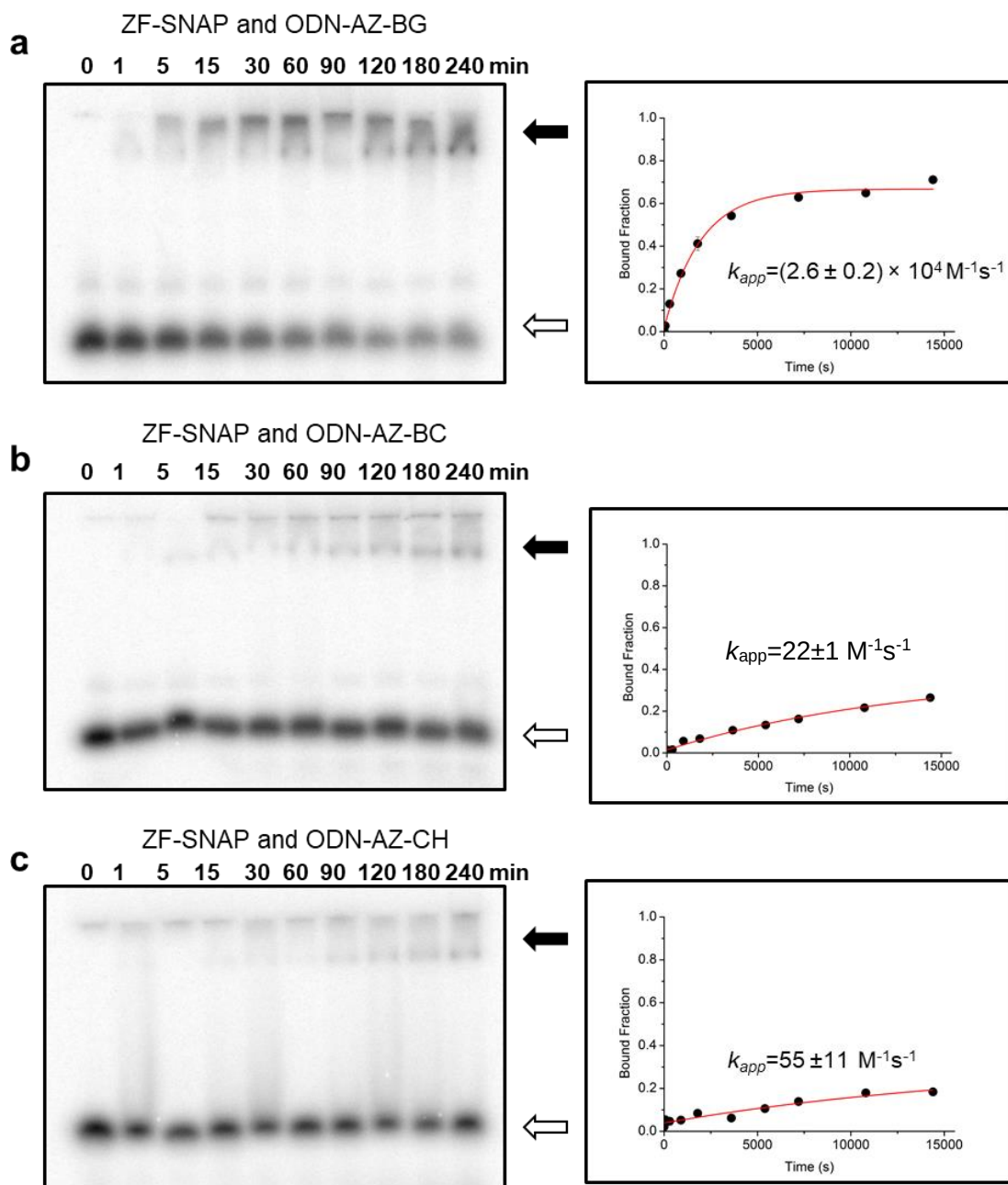


Figure 2.8. Autoradiograms shows denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-AZ-BG, 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC, and 0.5 nM 5'-³²P-end-labeled ODN-AZ-CH with 10 nM, 1000 nM, and 1000 nM ZF-SNAP, respectively (left). Open arrows and filled arrows denote unmodified ODN and ZF-SNAP bound ODN, respectively. Time-course plots for the crosslinking reactions of ODNs and ZF-SNAP to obtain the rate constants (k_{app}) (right). The estimated rate constants are listed in Table 2.3.

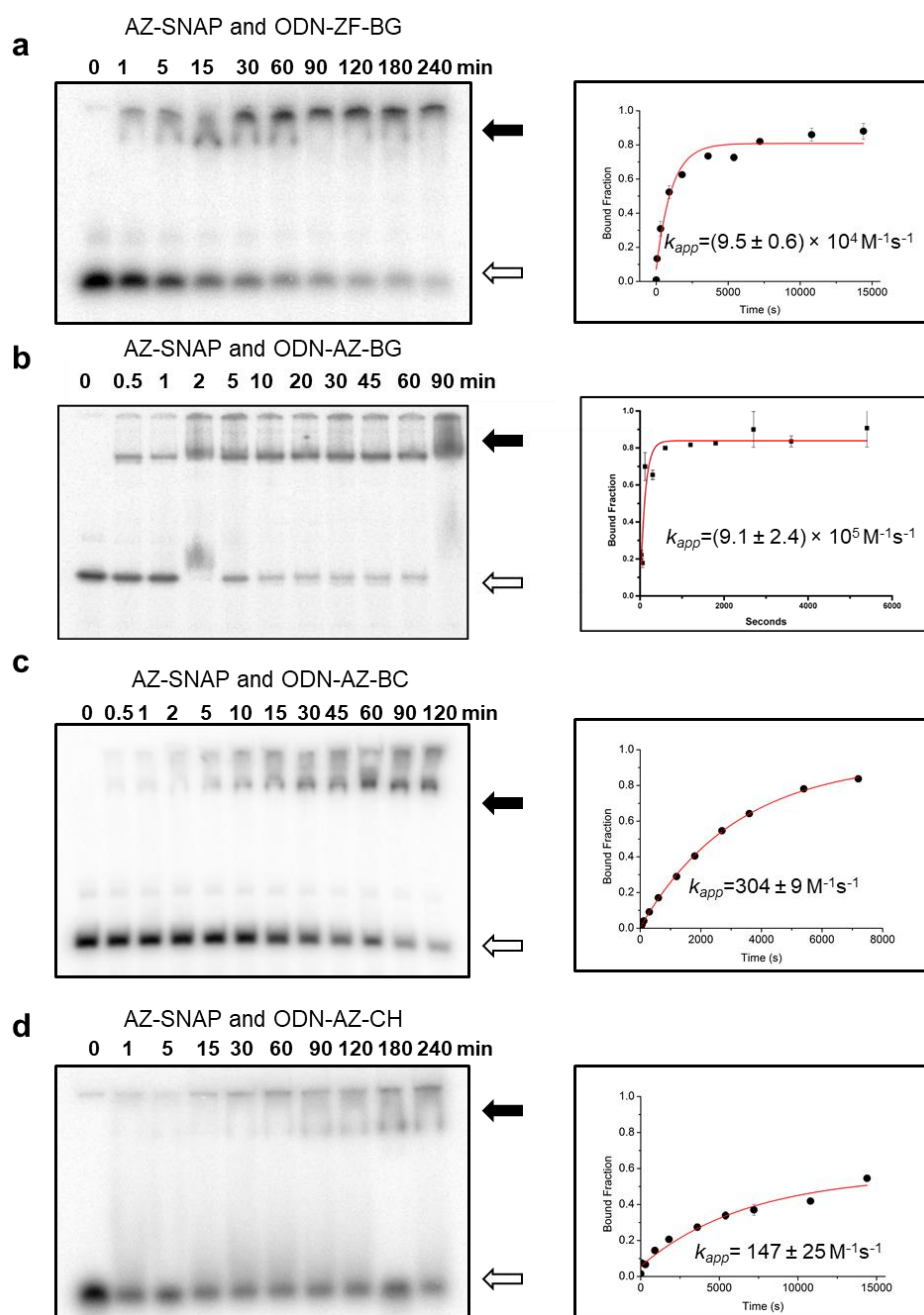


Figure 2.9. Autoradiograms show denaturing gel electrophoretic analysis of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BG, 0.5 nM 5'-³²P-end-labeled ODN-AZ-BG, ODN-AZ-BC, and 0.5 nM 5'-³²P-end-labeled ODN-AZ-CH with 10 nM, 10 nM, 1000 nM, and 1000 nM AZ-SNAP, respectively (left). Open arrow and filled arrow denote unmodified ODN and AZ-SNAP bound ODN, respectively. Time-course plots for the crosslinking reactions of ODNs and AZ-SNAP to obtain the rate constants (k_{app}) (right). The estimated rate constants are listed in Table 2.3.

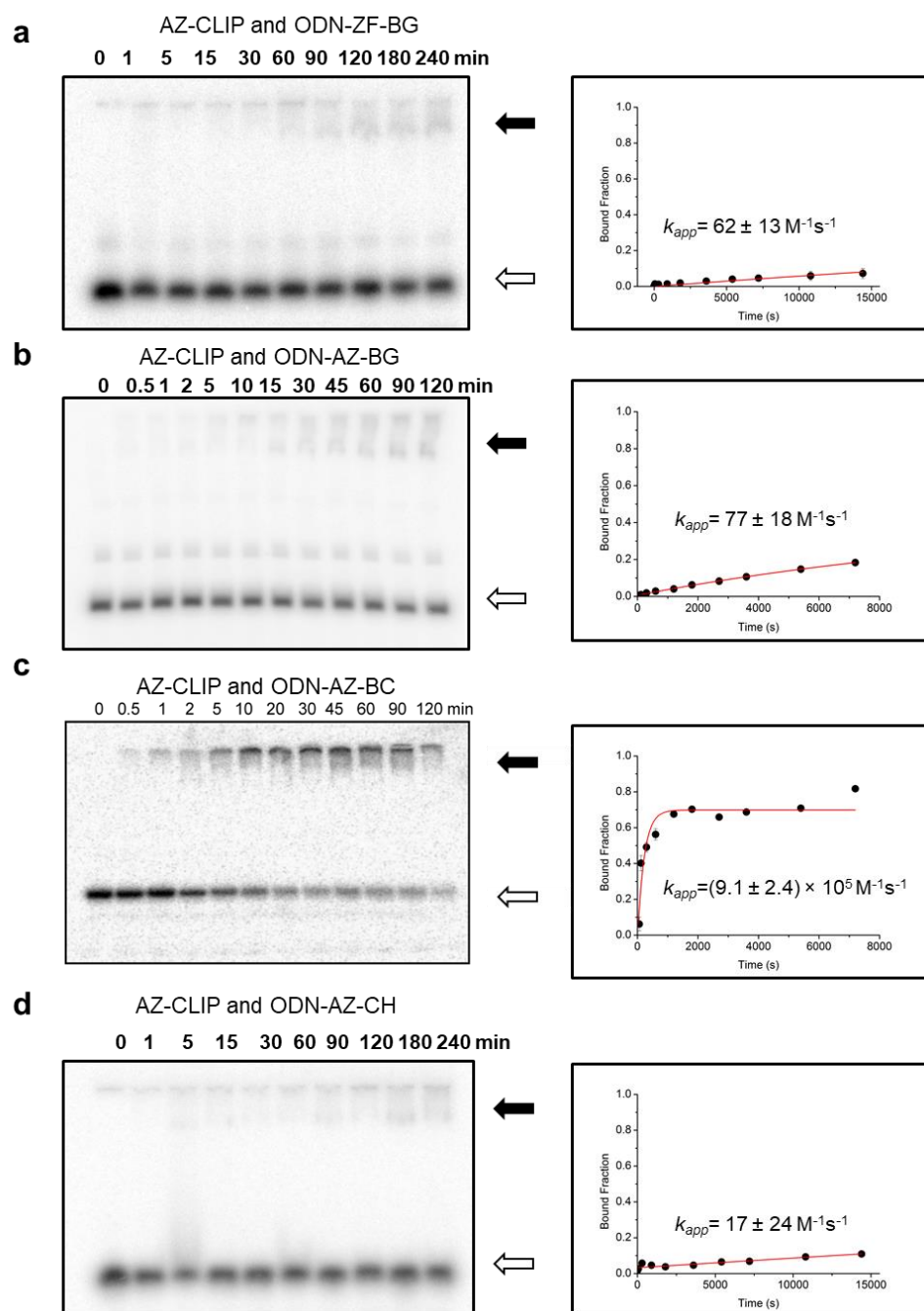


Figure 2.10. Autoradiograms show denaturing gel electrophoretic analysis of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BG, 0.5 nM 5'-³²P-end-labeled ODN-AZ-BG, ODN-AZ-BC, and 0.5 nM 5'-³²P-end-labeled ODN-AZ-CH with 1000 nM, 1000 nM, 10 nM, and 1000 nM AZ-CLIP, respectively (left). Open arrow and filled arrow denote unmodified ODN and AZ-CLIP bound ODN, respectively. Time-course plots for the crosslinking reactions of ODNs and AZ-CLIP to obtain the rate constants (k_{app}) (right). The estimated rate constants are listed in Table 2.3.

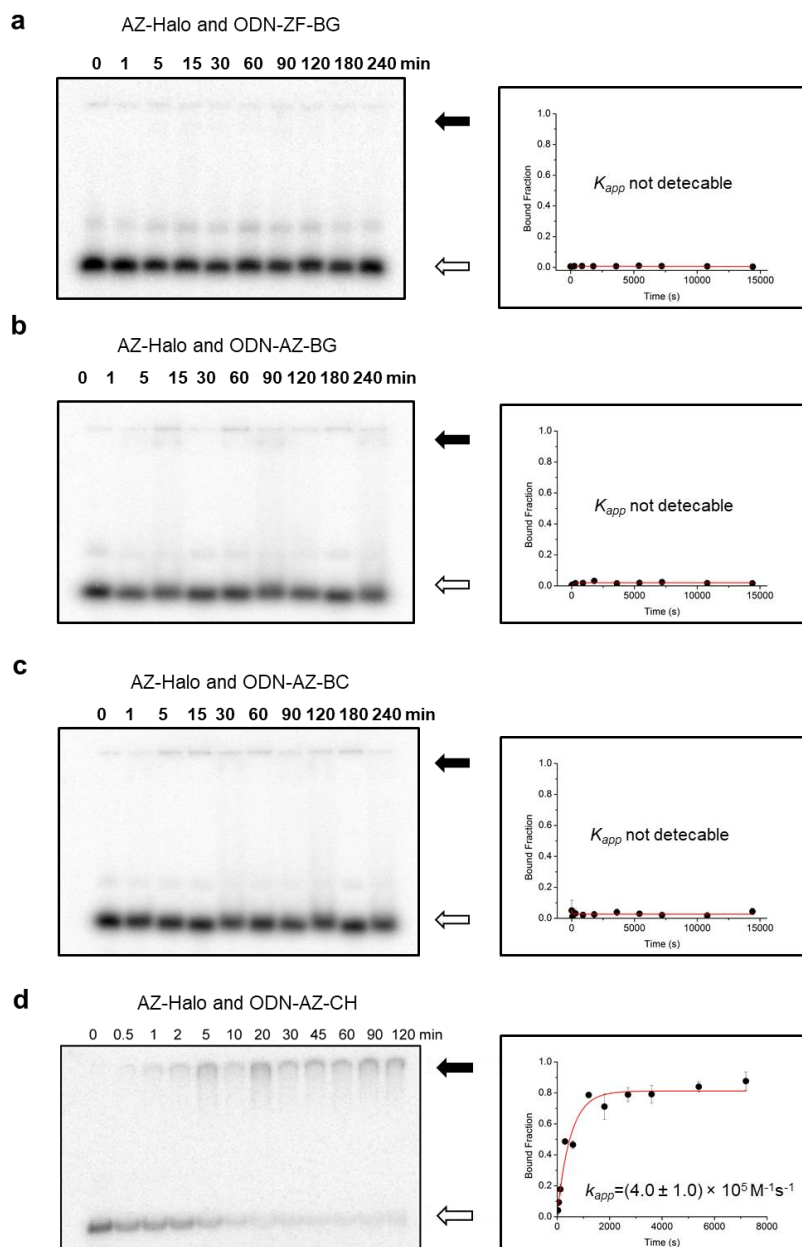


Figure 2.11. Autoradiograms show denaturing gel electrophoretic analysis of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BG, 0.5 nM 5'-³²P-end-labeled ODN-AZ-BG, ODN-AZ-BC, and 0.5 nM 5'-³²P-end-labeled ODN-AZ-CH with 1000 nM, 1000 nM, 1000 nM, and 10 nM AZ-Halo, respectively (left). Open arrows and filled arrows denote unmodified ODN and AZ-Halo bound ODN, respectively. Time-course plots for the crosslinking reactions of ODNs and AZ-Halo to obtain the rate constants (k_{app}) (right). The estimated rate constants are listed in Table 2.3.

Table 2.3. Kinetic parameters for the cross-linking reaction between 5'-³²P-end-labeled ODN derivatives and modular adaptors

Rate constant k_{app} ($M^{-1}s^{-1}$)		Modular adaptors			
		ZF-SNAP	AZ-SNAP	AZ-CLIP	AZ-Halo
ODN derivatives	ODN-ZF-BG	$(3.8 \pm 1.2) \times 10^5$ *	$(9.5 \pm 0.6) \times 10^4$	62 ± 13	n.d.
	ODN-AZ-BG	$(2.6 \pm 0.2) \times 10^4$	$(9.1 \pm 2.4) \times 10^5$	77 ± 18	n.d.
	ODN-AZ-BC	22 ± 1	304 ± 9	$(5.0 \pm 0.7) \times 10^5$	n.d.
	ODN-AZ-CH	55 ± 11	147 ± 25	17 ± 24	$(4.0 \pm 1.0) \times 10^5$
n.d. : not detectable					

The modular adaptors possessing the same protein-tag with different DNA binding domains, such as ZF-SNAP and AZ-SNAP, are not suitable for a pair of orthogonal adaptors. When modular adaptors shared the same DNA binding domain but differed in the protein tag, combination of AZ-SNAP and AZ-Halo or AZ-CLIP and AZ-Halo reduced the crosslinking reaction for unmatched pairs due to the lower reactivity of SNAP-tag and CLIP-tag to the Halo-tag substrate (CH) or *vice versa*. On the other hand, crosslinking reactions for the unmatched pairs of both AZ-SNAP to ODN-AZ-BC and AZ-CLIP to ODN-AZ-BG were observed due to the inherent crosslinking ability of SNAP-tag for BC and CLIP-tag for BG.²⁹ Therefore, a combination of ZF-SNAP with ODN-AZ-BC and AZ-CLIP with ODN-ZF-BG was preferred to reduce the unmatched crosslinking reaction. As shown in Figure 2.7b, no crosslinking reaction was observed for the unmatched pairs of these combinations. Taken all together, complete orthogonality for the crosslinking reaction of three modular adaptors, ZF-SNAP, AZ-CLIP, and AZ-Halo, to each matched ODN was established as shown in Figure 2.12.

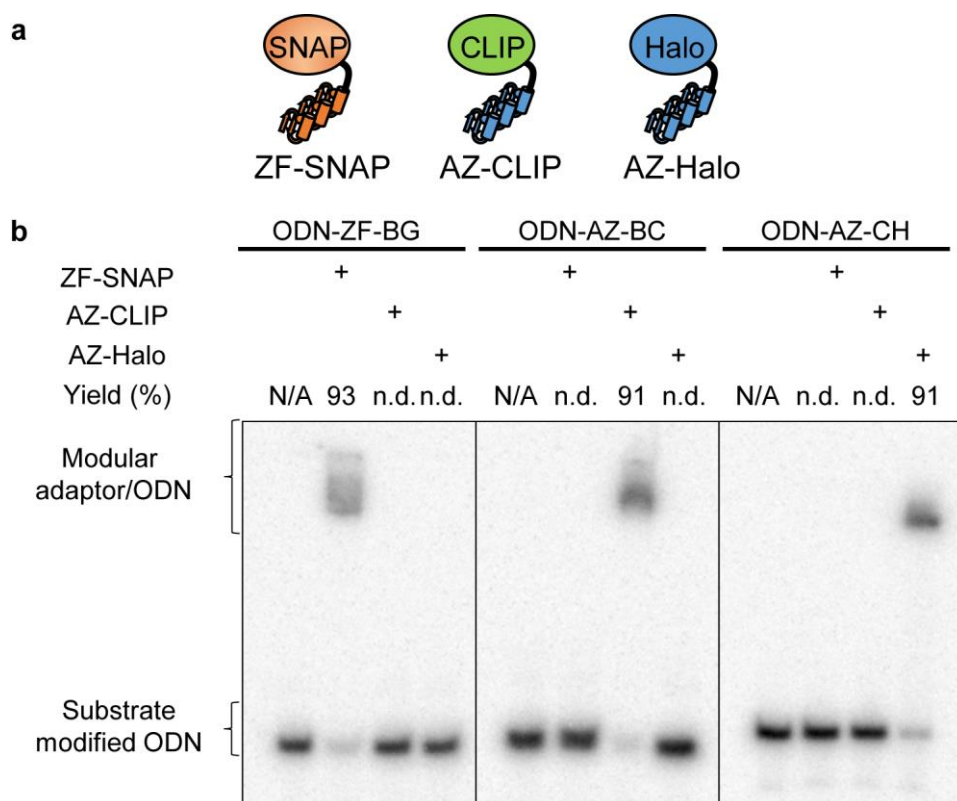


Figure 2.12. (a) Combination of modular adaptors to validate their orthogonal reactions to target sites. (b) Denaturing PAGE analyses of the crosslinking reaction by modular adaptors (ZF-SNAP, AZ-CLIP, and AZ-Halo) and the substrate modified ODN (ODN-ZF-BG, ODN-AZ-BC, and ODN-AZ-CH), respectively. Each 5'-³²P-end-labeled ODN (0.5nM) (ODN-ZF-BG, ODN-AZ-BC or ODN-AZ-CH) was incubated with a modular adaptor (100 nM : ZF-SNAP, AZ-CLIP or AZ-Halo) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature. N/A : not applicable, n.d: not detectable.

2.2.3 Formation of covalent bonds between the modular adaptors and the substrate modified addresses on the DNA scaffolds

Orthogonal crosslinking reactions of three modular adaptors (ZF-SNAP, AZ-Halo, and AZ-CLIP) were analyzed at the designed positions of DNA scaffold. A rectangular DNA scaffold (I-1AH/II-1ZS/III-1AC) containing three cavities (Figure

2.13, Table 2.4, 2.6) was prepared as described previously.²¹ Briefly, 10nM of the M13mp18 single-stranded DNA M13mp18 was incubation with 50 nM of the staple strands (Table 2.4 and 2.6) for 30 min and stored at 4°C in a thermal cycler.

Table 2.4. Nucleotide sequences for the staple strands including zif268 and AZP4 binding sites with substrates modified T (T^{BG} , T^{CH} , T^{BC} for T modified with BG, BC, and CH, respectively). The zif268 and AZP4 binding sites on the staple strands were colored in red and blue, respectively.

Oligo name	Sequence (from 5' to 3')
8E-AZ-CH	CTACTAA CTTATGCCACGTAGCGTT ^{CH} TTCGCTACGTGGCATAAG TGACCATTAGATACAACGAGTAGA
11D-ZF-BG	AACAGGTC CTTACGCCACGCGGCG TT ^{BG} TT CGCGCGTGGCGTAAG GAACCAGACCGGAAGATTGAGC
24D-AZ-BC	GGACAGAT CTTATGCCACGTAGCGTT ^{BC} TTCGCTACGTGGCATAAG AAATTGTGTGCGAAATCTGTATCAT

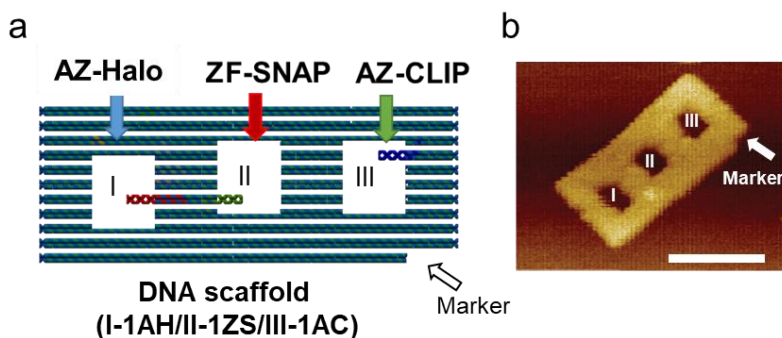


Figure 2.13. An illustration shows the shape and addresses of the DNA origami scaffold used in this study. (a) DNA scaffold (I-1AH/II-1ZS/III-1AC) containing the single binding site for AZ-Halo, ZF-SNAP, and AZ-CLIP at the cavity I, I and III, respectively. (b) AFM image of the DNA scaffold (I-1AH/II-1ZS/III-1AC). The scale bars represent 100 nm.

Upon reaction of the DNA scaffold (5 nM) and each of the modular adaptors (100 nM) for 30 min, the yields of crosslinking reaction at matched and unmatched sites were estimated by AFM images (Figures 2.14b-d, Table 2.5). The yields of covalent bond formation by each of the modular adaptors at the matched positions were over 93% with the highest yield of 96% for AZ-Halo (Figure 2.14b). Little or no crosslinking reaction was observed for the unmatched sites on DNA scaffold. These results confirmed nearly perfect orthogonal assembly of these three modular adaptors.

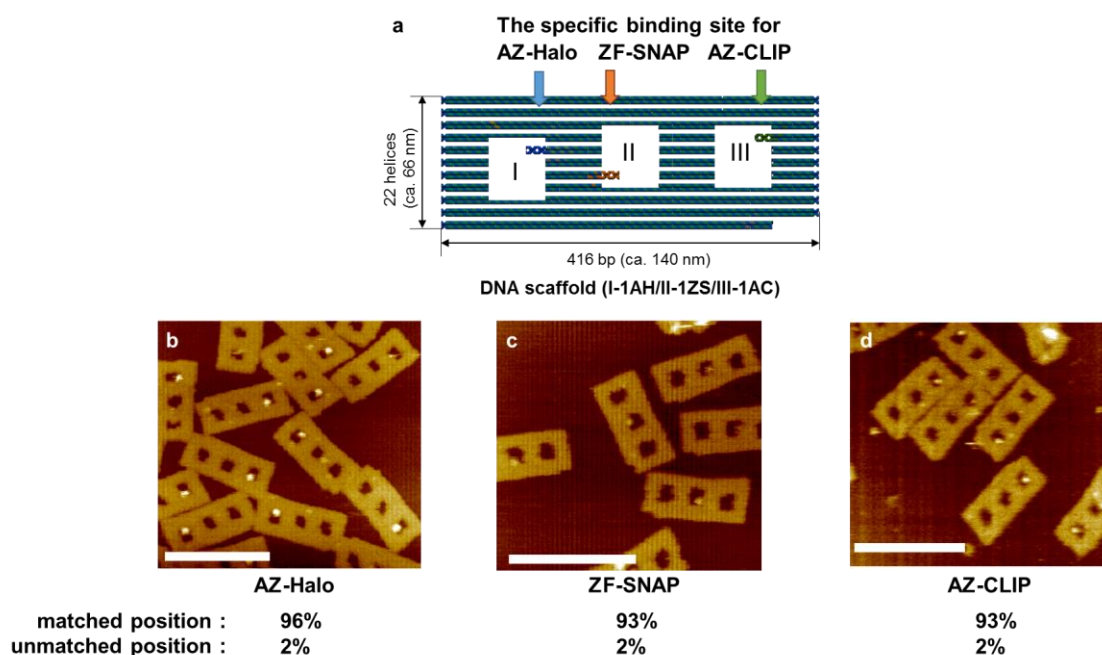


Figure 2.14. (a) An illustration of DNA origami scaffold (I-1AH/II-1ZS/III-1AC), in which each cavity contained a single address for one of the modular adaptors. The stem loops in blue, orange and green denote the binding site for AZ-Halo (AH) in cavity I, ZF-SNAP (ZS) in cavity II, and AZ-CLIP (AC) in cavity III, respectively. (b-d) AFM images of the DNA scaffold reacted by the modular adaptor (b) AZ-Halo, (c) ZF-SNAP, or (d) AZ-CLIP at the predesigned specific position. [I-1AH/II-1ZS/III-1AC] = 5 nM and [AZ-Halo], [ZF-SNAP] or [AZ-CLIP] = 100 nM. The reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20 for 30 min at ambient temperature. Yields for the matched and the unmatched crosslinking reactions were estimated by

counting the number of cavities occupied by the modular adaptors (Table 2.5). The scale bars represent 200 nm.

Sequential assembly of these modular adaptors to the DNA scaffold was next tested (Figure 2.15). DNA scaffold (5 nM) was first incubated with AZ-Halo (100 nM), then with ZF-SNAP (100 nM), and finally with AZ-CLIP (100 nM). After each step of the reaction, an aliquot of reaction mixture was taken to purify by a size-exclusion chromatography for AFM analysis. Statistical analyses of AFM images showed that the loading yield of the modular adaptor to its designed position at each step was over 93%. The overall co-assembly yield of three modular adaptors after three steps of reactions was 90% (Figure 2.15), which represents the highest co-assembly yield to date for loading three different proteins to specific sites through covalent linkages on the DNA scaffold.²²

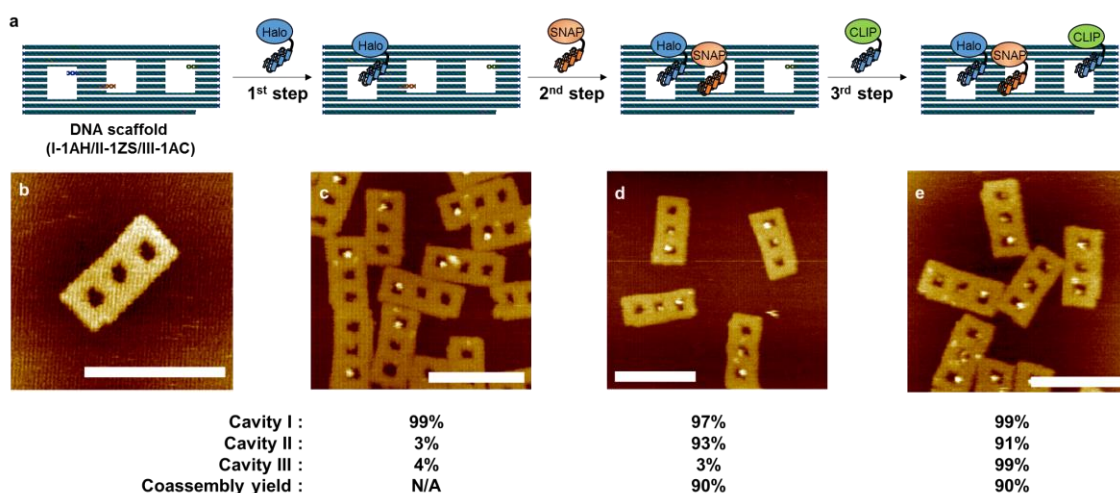


Figure 2.15. (a) An illustration of the sequential assembly of modular adaptors on the DNA scaffold (I-1AH/II-1ZS/III-1AC) and (b-e) AFM images of the DNA scaffold reacted by the modular adaptors. The DNA scaffold (5 nM) was incubated (c) with AZ-Halo (100 nM) for 10 min at ambient temperature (1st step), and then (d) with the ZF-SNAP for additional 10 min

(2nd step), and finally (e) with AZ-CLIP for further 10 min (3rd step). The reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20. At each step, an aliquot of the reaction mixture was purified by size-exclusion chromatography and then analyzed by AFM. Yields at each cavity and co-assembly yields after the 2nd step and the 3rd step were estimated by counting the number of cavities occupied by the modular adaptors (Table 2.5). The scale bars represent 200 nm.

One pot co-assembly reaction of the modular adaptors for their target positions on DNA scaffold was further assessed (Figure 2.16). The DNA scaffold (5 nM) was incubated with three modular adaptors (100 nM of each) at ambient temperature or on ice. The co-assembly yields of the modular adaptors were estimated from the AFM images. The yield of co-assembly reaction reached 93% in 5 min at ambient temperature or 87% in 20 min on ice, respectively.

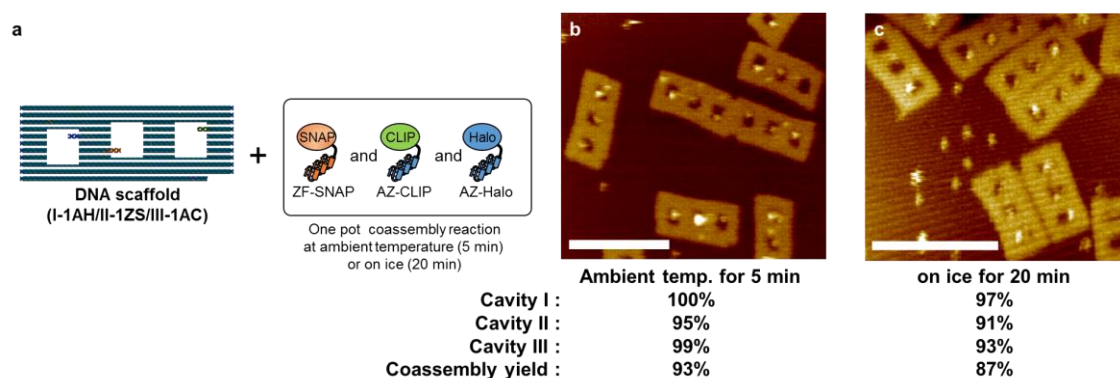


Figure 2.16 (a) One pot co-assembly reaction of three adaptors on the DNA scaffold (I-1AH/II-1ZS/III-1AC) under different incubation times and temperatures. (b-c) AFM images of the DNA scaffold reacted by the modular adaptors. A reaction mixture of 5 nM DNA scaffold and 100 nM each of adaptors (AZ-Halo, ZF-SNAP and AZ-CLIP) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and

0.02% Tween20 was incubated for (b) 5 min at ambient temperature or (c) 20 min on ice. Yields were estimated by counting the number of cavities occupied by the modular adaptors (Table 2.5). The scale bars represent 200 nm.

Table 2.5. Total numbers and yields of the DNA scaffolds (I-1AH/II-1ZS/III-1AC) assembled with the modular adaptors analyzed by AFM.

Modular adaptor derivatives	Number of well-formed DNA scaffold	Numbers and yields of the modified cavities			Co-assembly yield	AFM image	Temp.
		Cavity I	Cavity II	Cavity III			
AZ-Halo	151	145 [96%]	1 [1%]	4 [3%]	n.a.	Figure 2.14b	25°C
ZF-SNAP	179	1 [1%]	166 [93%]	5 [3%]	n.a.	Figure 2.14c	25°C
AZ-CLIP	147	1 [96%]	5 [3%]	137 [93%]	n.a.	Figure 2.14d	25°C
AZ-Halo	157	155 [99%]	5 [3%]	6 [4%]	n.a.	Figure 2.15 1 st step	25°C
AZ-Halo and ZF-SNAP	129	125 [97%]	120 [93%]	4 [3%]	118 [91%]	Figure 2.15 2 nd step	25°C
AZ-Halo, ZF-SNAP and AZ-CLIP	101	100 [99%]	92 [91%]	100 [99%]	92 [91%]	Figure 2.15 3 rd step	25°C
AZ-Halo, ZF-SNAP and AZ-CLIP	153	153 [100%]	146 [95%]	151 [98%]	142 [93%]	Figure 2.16b	25°C
AZ-Halo, ZF-SNAP and AZ-CLIP	153	149 [97%]	139 [91%]	143 [93%]	133 [87%]	Figure 2.16c	4°C

2.3 Discussion

DNA nanotechnology enables the rapid production of fully addressable nanostructures that are suitable for the scaffold of site-directed assembly of functional materials in the nanometer precision. Proteins are a particularly interesting class of molecules to assemble in such high precision because such protein assemblies act as in vitro models for the cellular protein assemblies. Proteins, in contrast to the stable small

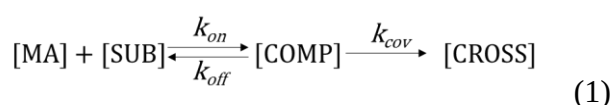
molecules and nanoparticles, usually require much more caution not to suppress or abolish their activity during the assembly formation on DNA scaffold. In addition, stable assembly of proteins on the DNA scaffold through covalent linkages promotes its reliable and reproducible functional analyses. Previous methods to locate POIs on the DNA scaffold through covalent linkages required a large excess amount of modified proteins to the target on the DNA scaffold with long incubation times,²² which is not preferable to retain the enzyme activity. When POIs chemically modified with ODN were used to locate on the DNA scaffold through hybridization, an annealing process, for example, an incubation from 37 to 4 °C for 1-2 h, was required to maximize the assembling yield.¹¹ Such a process often reduces the activity of thermally unstable enzymes. In addition, multiple redundant binding sites for locating POIs were introduced in many cases on the DNA scaffold to enhance the apparent loading yield.^{12-15,22,23,32} Such a scaffold design causes difficulty in controlling the accurate stoichiometry of located proteins.

Unlike these methods, site-specific crosslinking reactions by the modular adaptors are efficient enough to saturate a unique binding site on a DNA scaffold. The one-pot coassembly yield of three different modular adaptors reached almost 90% at ambient temperature within 5 min or on ice for 20 min. The apparent rate constants for the crosslinking reaction of three different modular adaptors described here are at the same order of magnitude. Such characteristics of multiple crosslinking reactions are very useful to simultaneously assemble various POIs on the DNA scaffold by a one-pot co-assembly reaction. The short reaction time for assembly formation is suitable to load the thermally unstable enzyme on the DNA scaffold with maintaining the activity. These newly designed orthogonal modular adaptors are suitable for stably assembling various kinds of enzymes in their active forms on the DNA scaffold. In fact, the enzyme XK,

which is reported to be thermally unstable,²⁷ was successfully loaded on the DNA scaffold to drive the enzyme cascade reaction (Chapter 4). It is also worth mentioning that the high loading yield achieved by the modular adaptor fused enzyme to a single binding site on the DNA scaffold allows us to control the distinct stoichiometry of enzymes on the DNA scaffold.

The set of three orthogonal modular adaptors described here now enables successful location of three POIs at their respective positions on the DNA scaffold. To assemble more than three kinds of POIs, further expansion of the members of orthogonal modular adaptors, which may consist of different types of DNA binding domains³³ and the chemoselective cross-linking domains³³⁻³⁴ are required.

The kinetics and selectivity studies described in this study revealed important characteristics for the crosslinking reaction of modular adaptors that are useful for the design of the orthogonal modular adaptors. To verify the kinetic aspect of crosslinking reaction between the modular adaptor (MA) and the substrate modified ODN (SUB), the reaction was considered as represented by scheme (1), where a reversible complex of DNA binding domain of modular adaptor and ODN was formed as a transient complex (COMP):³⁶



The first step of scheme (1) is the formation of reversible complex, COMP, by MA and SUB. The second step is an intramolecular crosslinking reaction within COMP between MA and SUB to form a crosslinking product CROSS. The apparent rate constant of the crosslinking reaction (k_{app}) ($M^{-1}s^{-1}$) is represented as an equation (2):

$$k_{app} = \frac{k_{cov} * k_{on}}{k_{off} + k_{cov}} \quad (2)$$

where k_{on} ($M^{-1}s^{-1}$) and k_{off} (s^{-1}) are the association and dissociation rate constants of the DNA binding domain with SUB, and $k_{cov} \times M$ (s^{-1}) is the rate constant of covalent bond formation in close proximity, which is defined by the product of the rate constant of intermolecular covalent formation ($M^{-1}s^{-1}$) and the effective concentration (M).²⁴ When the rate constant of covalent bond formation is much larger than the dissociation rate constant ($k_{cov} \times M \gg k_{off}$), the equation (2) leads to an equation (3):

$$k_{app} \approx k_{on} \quad (3)$$

On the other hand, when the rate constant of covalent bond formation is much smaller than the dissociation rate constant ($k_{cov} \times M \ll k_{off}$), the equation (2) leads to an equation (4):

$$k_{app} \approx k_{cov} \times M * k_{on} / k_{off} = k_{cov} \times M / K_D \quad (4)$$

where K_D (M) is the equilibrium dissociation constant. I first consider the reaction of matched pairs of the modular adaptor and the substrate modified ODN. The association rate constants of DNA binding domains including the zinc finger motif to DNA are reported to be quite large (10^6 to $10^8 M^{-1}s^{-1}$).⁴⁴ From the equilibrium dissociation constant obtained for the matched complex of the zinc finger motif (~ 60 nM) as shown in Table 2.2 and the association rate constant of 10^6 to $10^8 M^{-1}s^{-1}$, the dissociation rate

constant of COMP is expected to be 6 to 0.06 s^{-1} . The rate constant of covalent bond formation for the protein-tag and its substrate ranged from 10^3 to $10^6 \text{ M}^{-1}\text{s}^{-1}$ (e.g. SNAP-tag and BG derivative: $\sim 10^4 \text{ M}^{-1}\text{s}^{-1}$),^{20,29,30} which is much larger than the dissociation rate constant of COMP (0.06 s^{-1}) by assuming that the effective concentration of the reactant within COMP is over 1 mM as previously reported for such complexes.³⁷⁻³⁹ Therefore, the apparent rate constant for the crosslinking reaction (k_{app}) corresponds to the association rate constant for the complex formation between the DNA binding domain and ODN as in the equation (3). This explains the results that k_{app} obtained for the crosslinking reaction of matched complexes are around $10^5 - 10^6 \text{ M}^{-1}\text{s}^{-1}$ (Table 2.3). We next consider the reaction of unmatched pairs of the modular adaptor and the substrate modified ODN. The sequence selectivity of the zinc finger motif is governed by the dissociation rate constants, not by the association rate constants for the DNA complex formation.⁴⁰⁻⁴⁴ Because the equilibrium dissociation constant of unmatched complex is larger than 10^{-6} M (Table 2.2) and the association rate constant of unmatched complex is similar to that of matched complex, the association rate constant and dissociation rate constant of unmatched complex are expected to be $10^6 \text{ M}^{-1}\text{s}^{-1}$ and 1 s^{-1} , respectively. Even though the dissociation rate constant for the unmatched complex (1 s^{-1}) is larger than that for the matched complex (0.06 s^{-1}), the condition of $k_{\text{cov}} \times M \gg k_{\text{off}}$ again stands for the reaction of the protein-tag and its substrate in the un-matched complex. These considerations explain the results that ZF-SNAP and AZ-SNAP formed covalent linkages at the BG modified sites both in the matched and the unmatched sequences. The modular adaptors possessing the same protein-tag with different DNA binding domains, such as ZF-SNAP and AZ-SNAP, are not suitable for a pair of orthogonal adaptors because these modular adaptors possess characteristics leading the equation (3). Due to the inherent side reactions of SNAP-tag

for BC and CLIP-tag for BG, minor crosslinking reactions (less than 15% yield) for the unmatched pairs of both AZ-SNAP to ODN-AZ-BC and AZ-CLIP to ODN-AZ-BG were observed (Figure 2.7b). In the reaction of AZ-SNAP to ODN-AZ-BC, AZ-SNAP specifically binds ODN-AZ-BC with fast associate rate constant but the crosslinking kinetic constant for SNAP-tag with BC was reported to be $26 \text{ M}^{-1}\text{s}^{-1}$.³⁷⁻³⁹ For this unmatched complex, the condition for the rate constant of covalent bond formation and the dissociation rate constant reaction observed in the case of matched complex such as ZF-SNAP and ODN-ZF-BG ($k_{\text{cov}} \times M \gg k_{\text{off}}$) does not stand. Assuming that the effective concentration of the reactant within COMP is over 1 mM as previously reported for such complexes,³⁷⁻³⁹ $k_{\text{cov}} \times M$ for AZ-SNAP and ODN-AZ-BC is estimated to be $2.6 \times 10^{-2} \text{ s}^{-1}$, which is in the same order to the dissociation constant of COMP (0.06 s^{-1}). The kinetic aspect of this side reaction cannot be simplified as equation (3) or equation (4). Thus, reducing the rate constant of covalent bond formation by the protein-tag to follow the conditions of equation (4) would provide an effective strategy to prevent the crosslinking reaction at the unmatched DNA sequences by the modular adaptors bearing the same chemoselective crosslinking domains. Base on the equation (4), the higher the equilibrium dissociation constant (K_D) for the unmatched complex of zinc finger motif and ODN is, the lesser the undesired crosslinking reactions of unmatched complexes would proceed. A wide variation of sequence-selective DNA binding proteins has been well characterized to date.³³ These DNA-binding proteins reveal a wide range of affinity and sequence selectivity. The chemoselective crosslinking domains³⁴⁻³⁵ could be tuned to show the different rate constant of covalent bond formation. Further investigation on the combination of DNA binding domain and chemoselective crosslinking domain will provide us various sets of orthogonal modular adaptors, in which the same chemoselective crosslinking domain is shared by different DNA binding domains, to

complete the crosslinking reactions at different DNA sequences.

2.4 Conclusion

I have systematically investigated a series of modular adaptors that differed in the sequence-specific DNA binding domain and the self-ligating protein tag to achieve not only the fast kinetics and the high loading yield for the crosslinking reaction but also the high orthogonality to the individual target addresses. This study suggested an important guideline for the design of modular protein-tags for the orthogonal formation of covalent bonds at specific DNA sequences. Further application and functional evaluation of the combination of these modular adaptors are currently underway.

2.5 Materials and methods

2.5.1 Materials

The single-stranded M13mp18, restriction enzymes (NdeI and HindIII), pSNAP-tag (T7)-2 Vector (N9181S), BG-GLA-NHS (S9151S), BC-GLA-NHS (S9237S) were purchased from New England Biolabs. pCLIPf(T7)-2 Vector was generously donated from New England Biolabs. HaloTag 7 vector (pFN18K : G2681) and HaloTag Succinimidyl Ester (O2) Ligand (P1691) were purchased from Promega. Purified oligonucleotides as the staple strands for DNA origami, oligonucleotide primers for gene construction and all other oligonucleotides were obtained from Sigma-Aldrich (St. Louis, MO), Thermo Fisher Scientific Inc. (Waltham, MA, USA) or Gene Design Inc. (Osaka, Japan). *Escherichia coli* BL21 (DE3) competent cells were purchased from Invitrogen (Carlsbad, CA). Mini Elute Gel Extraction Kit was prepared from QIAGEN (Tokyo, Japan). HiTrap SP XL cation exchange column (5 ml), HisTrap HP column (5 ml), and Sephacryl S-400 were from GE Healthcare Japan Inc. (Tokyo,

Japan). PrimeSTAR HS DNA polymerase, T4 DNA ligase, and E. Coli DH5 α competent cells were obtained from TaKaRa Bio Inc. (Shiga, Japan). Ultrafree-MC-DV was obtained from Merck Millipore (Darmstadt, Germany). Cosmosil 5C18-MS II column (4.6 i.d. $\times$ 150 mm) and HPLC-grade acetonitrile were purchased from Nacalai Tesque (Kyoto, Japan). Gel electrophoresis grade acrylamide, bis(acrylamide), phenol, and all other chemicals and reagents were purchased from Wako Chemicals (Tokyo, Japan) or Nacalai Tesque (Kyoto, Japan). Phosphate buffer (PB) was prepared as 20 mM Na₂HPO₄ and 20 mM NaH₂PO₄. DNA origami buffer (pH 8.0) contained 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂.

2.5.2 Preparation of the protein-tag derivatives (AZ-SNAP, AZ-CLIP, and AZ-Halo)

Overexpression and purification of the ZF-SNAP were described in the previous report.¹⁸ As a typical example, preparation of AZ-SNAP is shown below. All of the vectors encoding protein-tag derivatives (pET-30a-AZ-CLIP, pET-30a-AZ-Halo,) were prepared and the protein-tag derivatives (AZ-CLIP, AZ-Halo) were overexpressed and purified in the same manner.

Construction of Vectors for AZ-SNAP (pET-30a-AZ-SNAP). SNAP-tag in pSNAP-tag (T7)-2 vector was amplified by PCR using the following primer pairs.

Forward primer (F_EcoRI_SNAP):

TAATAAGAATTCGGCGGCTCCGGCGGCTCCGACAAAGATTGCGAAA

Reverse primer (R_SNAP_HindIII):

TTATTAAAGCTTTTAATGATGGTGATGATGATGGTGATGATGGTGGGT
ACCATTAACCTCGAGCCCGGGG

The PCR products were run on a 1% agarose gel (TAE) and were purified by

Mini Elute Gel Extraction Kit. The PCR products and pET-30a-cys-AZP46 were digested with EcoRI and HindIII and were purified in the same manner. These products were incubated with T4-DNA-ligase. The mixture was transformed into *E. coli* DH5 α competent cells for amplification. The vector encoding AZ-SNAP (termed as pET-30a-AZ-SNAP) was purified, sequence checked, then transformed into *E. coli* BL21(DE3) competent cells.

2.5.3 Overexpression and purification of protein-tag derivatives (AZ-SNAP, AZ-CLIP, and AZ-Halo)

The transformed cells were grown at 37 °C until OD₆₀₀ reached 0.5, and protein expression was induced with 1 mM IPTG for 24 h at 25 °C. The soluble fraction of the cell lysate containing recombinant protein was loaded to HisTrap HP column equilibrated with a 50 mM phosphate buffer (pH 8.0) containing 200 mM NaCl and 1 mM DTT, and eluted by imidazole gradient. The main fractions containing the target protein were collected, loaded to HiTrap SP HP column equilibrated with a 50 mM phosphate buffer (pH 7.0) containing 1 mM DTT and eluted by NaCl gradient. The purified protein was dialyzed by using a 50 mM phosphate buffer (pH 8.0) containing 1 mM DTT, 50 μ M ZnCl₂, and 50% glycerol and stocked at -20°C. The purity of the target protein was checked by SDS-PAGE. The major band in SDS-PAGE corresponded to the calculated molecular weight with estimated purity of over 95% (Figure 2.4). The modular adaptors were characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, SA matrix). AZ-SNAP: m/z calcd 32181, found 32183; AZ-CLIP: m/z calcd 32196, found 32249; AZ-Halo: m/z calcd 47768, found 47736. Amino acid sequences and calculated molecular weights of recombinant proteins used in this study are shown in Table 2.1.

2.5.4 Preparation of substrate modified ODNs

A coupling reaction between amino modified ODNs (100 μ M) and succinimidyl derivative of protein-tag substrates (1 mM) was carried out in a 100 mM phosphate buffer (pH 8.0) for 8 hours at ambient temperature. The substrate modified ODNs were purified by reversed-phase HPLC on a Cosmosil 5C18-MS II column (4.6 x 150 mm, elution with 100 mM triethylammonium acetate buffer, pH 7.0, linear gradient over 30 min from 2.5 to 30% acetonitrile at flow rate of 1.0 mL.min⁻¹), and characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, HPA matrix). ODN-AZ-BG: m/z calcd 11236, found 11237; ODN-AZ-BC: m/z calcd 11195, found 11193; ODN-AZ-CH: m/z calcd 11175, found 11172; ODN-8e-AZ-CH: m/z calcd 21084, found 21085; ODN-24D-AZ-BC: m/z calcd 21471, found 21471.

2.5.5 Preparation of ³²P-end-labeled ODNs and the analysis of covalent-linkage formation between ³²P-end-labeled ODNs and modular adaptors by denaturing gel shift assay

The ODNs (Figure 2.2d) were 5'-³²P-end-labeled as previously described.¹⁸ In a typical experiment for kinetic analysis, ³²P-ODN-AZ-CH (less than 0.1 nM) was incubated with AZ-Halo (10 nM) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 0.02% Tween20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature and aliquots taken at defined reaction times were quenched by addition of formamide. The aliquots were analyzed by 8 M Urea PAGE and the intensities of the bands on the gel were analyzed by using Storm 860 Molecular imager (Amersham). The kinetics data were fitted to a reaction model assuming the first-order kinetics, and then the second-order rate constants were determined. For checking the selectivity, each 5'-³²P-end-labeled ODNs

(ODN-ZF-BG, ODN-AZ-BG, ODN-AZ-BC or ODN-AZ-CH) was incubated with a modular adaptor (100 nM) (ZF-SNAP, AZ-SNAP, AZ-CLIP, AZ-Halo) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 0.02% Tween 20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature.

2.5.6 Preparation of the DNA origami scaffold with three cavities

A solution (50 μ L) containing M13mp18 single-stranded DNA (New England Biolabs, 10 nM), staple DNA strands (5 equiv, 50 nM, the nucleotide sequences of all the staple DNA strands were shown in our previous report¹¹ and Table S4) in a buffer (40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, pH 8.0) was heated at 95 °C for 1 min, annealed at 53 °C for 30 min, and kept at 4 °C by using a thermal cycler. The samples were purified by size-exclusion chromatography (400 μ L volume of Sephacryl S-400, GE Healthcare) equilibrated with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂ in Ultrafree-MC-DV (Millipore).

Table 2.6. Unmodified staple strand for assembly of DNA scaffold

Name	Address	Sequence
3W-type2-001	1a	GCGGGAGGTTTTGAAGCGAGGCGTTTTAGCGACATGTTCA
3W-type2-002	1bc	GCTAATGCATTCTGTCCAGACGACTCTTCTGACCTAAATTT TTTTCAA
3W-type2-003	1d	ATATATTTAGATGATGAAACAAACTATTCATTTCAATTACCT TTTACA
3W-type2-004	1ef	TCGGGAGAACGTCAGATGAATATAGAAGGGTTAGAACCTA ATTGTTTG
3W-type2-005	1g	GATTATACAAAGAAACCACCAGAATTTGAGTAACATTATCA TTTAGAA
3W-type2-006	1hi	GTATTAGAGAGCCGTCAATAGATAGTTGGCAAATCAACAGT ATCAAAC
3W-type2-007	1j	CCTCAATCAAAGGGACATTCTGGCAGATTCAC
3W-type2-008	1k	CAGTCACAAACGGTACGCCAGAATGAGGCCGATTAAAGG G

3W-type2-009	2ab	GTAAAGTAAGAACGCGCCTGTTTCGGTATTCTAAGAACGCCTTAAA
3W-type2-010	2c	AAACAAAAAAGAACGCGAGAAAATAATGGTTTGAAATAGACAAAAG
3W-type2-011	2de	CAGGTTTAAACAATAACGGATTCTCGCGCAGAGGCGAATATCAAGA
3W-type2-012	2f	AATTATCACATCAATATAATCCTGCCATATCAAAATTATTAGATTTT
3W-type2-013	2gh	ATAGATTACTTTACAAACAATTCTTATTAATTTTAAAAGGGAACGG
3W-type2-014	2i	GATAGAACCTTGCTGAACCTCAAATTGAAAGGAATTGAGACAACTA
3W-type2-015	2jk	CTAAACAGCCTGAGAAGTGTTTTGATTATTTACATTGGCCAACAGA
3W-type2-016	3a	TCAAGATTAGTTGCTATTTTGCACTATAGAAGGCTTATCATCAACAA
3W-type2-017	3b	TAGATAAGAGAATATAAAGTACCCCGACCGT
3W-type2-018	3c	GTGATAAATCCAATCGCAAGACATTAATTAC
3W-type2-019	3d	ATTTAACATACCAAGTTACAAAAGCCTGATT
3W-type2-020	3e	GCTTTGAAAATAAAGAAATTGCGTTGCACGT
3W-type2-021	3f	AAAACAGACAGATGATGGCAATTTTCATATTC
3W-type2-022	3g	CTGATTATTCCTTTGCCCCAACGGACAACCTC
3W-type2-023	3h	GTATTAAAATCTTTAGGAGCACTGAAGGTTA
3W-type2-024	3i	TCTAAAATAATCTAAAGCATCACCTTCTGA
3W-type2-025	3j	CCTGAAAGCAATCGTCTGAAATGTATAATCA
3W-type2-026	3k	GTGAGGCCCTCGTTAGAAATCAGAGCGGGAG
3W-type2-027	4a	AAATCAGACCAGCTACAATTTTATCCTGAATC
3W-type2-028	4b	TAAATAAGTCGAGCCAGTAATAAGTCCTGAACAAGAAAAATAGCAAGC
3W-type2-029	4c	ATTTCAATTGAATTACTAAATGCTGATGCAAATAAGGCGT
3W-type2-030	4j	TTGACGCTCGTAAGAATACGTGGCCCAGCAGCAAATGAAA
3W-type2-031	4k	CGTGCTTTACCGAGTAAAAGAGTCCCTACATT
3W-type2-032	5a	TTACCAACCATTACCGCGCCCAATAATATCC
3W-type2-033	5b	CATCCTAATAGGCAGAGGCATTTAATAAACA
3W-type2-034	5c	CCGGAATCTTATATAACTATATGCTTTTTTTAA
3W-type2-035	5j	GCTGAGAGACAGACAATATTTTAAACGCTCATGGAAATA
3W-type2-036	5k	TGTCCATCACGCAAAGGTTGCTTTGACGAGCACGTATAA
3W-type2-037	6a	GTAGGAATGCTAACGAGCGTCTTTCCAGAGC
3W-type2-038	6b	ATGTAATTTTACGAGCATGTAGTTTTCATC
3W-type2-039	6c	TGGAAACAGTACATAAATCAATCCGGCTTAGGTTGGGATAATTACTAGAAAAACGCCAAC
3W-type2-040	6j	AGCTGCATTGCCGAAACCAGGCACGGCGGA
3W-type2-041	6k	CGCCGCTACAGGGCGCGTACTATTTAACCGT
3W-type2-042	7a	CTAATTTGAAGCAAGCCGTTTTTAAAACCAAT
3W-type2-043	7b	CAATAATCATCGCCATATTTAACAAGCCTGTT
3W-type2-044	7c	TAGTATCAGACTACCTTTTTTAACCATATGTGAGTGAATAA
3W-type2-045	7j	TGAGGCGGTCAGTATTTAATGCGCGAACTGACAATATTA

3W-type2-046	7k	CCGCCAGCTACTTCTTTGATTAGTGCTTAATG
3W-type2-047	8a	CGAGAACCCAGTTACAAAATAAACAGCCAT
3W-type2-048	8b	ATTGAGAGGCTGTCTTTCCTTACGCACTCAT
3W-type2-049	8c	TCTGAGATATGCGTTATACAAATAGGGCTTA
3W-type2-050	8d	TTTAGTTCCTTGCTTCTGTAAAAATCATAGG
3W-type2-051	8e	CTACTAATGACCATTAGATACAACGAGTAGA
3W-type2-052	8f	AATCATATAGTAGTAGCATTAAAGCATCAATT
3W-type2-053	8g	TAATACTCAGGCAAGGCAAAGACATCCAATA
3W-type2-054	8h	ACGCAAGTTTGCGGGAGAAGCCATGACCCTG
3W-type2-055	8i	ACAGAGGGATAAAAAATTTTTAGTTTATTTC
3W-type2-056	8j	TCCAGAATAGCCCTAAAACATCGAAGATAAA
3W-type2-057	8k	CCGCCGCAATAACATCACTTGCCTGGTAATA
3W-type2-058	9a	ATTATTTATATTAAACCAAGTACTCATTCCA
3W-type2-059	9b	AGAACGGGGCCAACGCTCAACAGTTCTTACC
3W-type2-060	9c	AGTATAAATAGTGAATTTATCAATCGTCGCT
3W-type2-061	9d	ATTAATTATTCCCAATTCTGCGATTTGCGAA
3W-type2-062	9e	ATGGTCAACGAGCTGAAAAGGTGTTTGCGGA
3W-type2-063	9f	TGGCTTAGTGATAAGAGGTCATTATTAGCAA
3W-type2-064	9g	AATTAAGCTGTACCAAAAACATTGAAGCCCG
3W-type2-065	9h	AAAGACTTCAAAAAGATTAAGAGAACCCTCA
3W-type2-066	9i	TATATTTTAACGAACCACCAGCAGCCATTAA
3W-type2-067	9j	AAATACCGAAACTATCGGCCTTGCTGAGTAG
3W-type2-068	9k	AAGAACTCTCACGCTGCGCGTAACCACCACAC
3W-type2-069	10ab	TTGAGTAAAAGGTGAATTATCACCCGAAATTATTCATTAT CCCAATC
3W-type2-070	10c	TTAGAATCCGCTGAGAAGAGTCAACAGTGCCCGTATAAAC TCAGTGCC
3W-type2-071	10de	TTGGGGCGTAACCTGTTTAGCTATTCCATATAACAGTTGAA TTTTCCC
3W-type2-072	10f	CTCAGAGCCTTTAATTGCTCCTTTAGCTTAATTGCTGAATAT TTTCAT
3W-type2-073	10gh	GATTGCATCAAATATCGCGTTTTAATAAAGCTAAATCGGTA ATAAAGC
3W-type2-074	10i	CAGTATCGTCTGCCAGTTTGAGGGAAATGCAATGCCTGAG GCAAAGCG
3W-type2-075	10jk	GTGTAGCGGTTTCACCAGTGAGACGCCAGGGTGGTTTTTC TGACGACGA
3W-type2-076	11a	CAAATAAGAAACGATTTTTTGTGTTGAAGGTAAATATTGAGT CACCGA
3W-type2-077	11b	CTTGAGCCTAAGTTTTTAACGGGGAGTTAATG
3W-type2-078	11c	CCCCCTGCTAGCTTAGATTAAGACTTGAAAA
3W-type2-079	11d	CATAGCGAGTCTGGAAGTTTCATGCAACTAA
3W-type2-080	11e	AGTACGGTACATGTTTTAAATATATAATGCT
3W-type2-081	11f	GTAGCTCAAGGATTAGAGAGTACCAAACCTCC
3W-type2-082	11g	AACAGGTCTGAACCAGACCGGAAGATTTCGAGC
3W-type2-083	11h	TTCAAAGCCTATTATAGTCAGAATAATGTGT

3W-type2-084	11i	AGGTAAAGCATCGTAACCGTGCAGCCTCAGG
3W-type2-085	11j	AAGATCGCTTTGCGTATTGGGCGGGCAACAG
3W-type2-086	11k	CTGATTGCGCGGGCGCTAGGGCGCTGGCAA
3W-type2-087	12a	ATTAGAGAACCGATTGAGGGAGGAACGTCAAAAATGAA
3W-type2-088	12b	CTATTTTCGGAACCTAAGGAGTGTACTGGTAAATTTGGGA
3W-type2-089	12i	GATGGGCGATTCAAAAGGGTGAGCAGGTCTTTACCCTGA
3W-type2-090	12j	AGAGGCGGACTCCAGCCAGCTTTTTGGTGTA
3W-type2-091	12k	CGAAAGGACCTTCACCGCCTGGCGCGCGGGG
3W-type2-092	13a	AATAGCAGCCTTTACACAAAAGGGCGACATTCCCAGCAAA
3W-type2-093	13b	ATCACCAGTGGCTTTTGATGATACTTATTCTGAAACATGA
3W-type2-094	13i	CGAGAATGACCATAAATCAAAAATAAAGGCCGGAGACAGT AATGGGAT
3W-type2-095	13j	AGGTCACGCCGGCACCGCTTCTGGTAATGAATCGGCCAAC CCTGAGAG
3W-type2-096	13k	AGTTGCAGTGGCGAGAAAGGAAGGGAAGAAAG
3W-type2-097	14a	TTACCATTTACCAGCGCCAAAGAGAGAGAATAACATAAA
3W-type2-098	14b	AAGTATTAAGAGGCTGAGTAAGCGTCATACATAGCACCA
3W-type2-099	14i	TTGACCGTCAAATCACCATCAATTCAGAAAA
3W-type2-100	14j	CAATTCCAGTTGGGTAACGCCAGATCAGCTC
3W-type2-101	14k	GGCGAACGCAAGCGGTCCACGCTGTCGTGCC
3W-type2-102	15a	AACAGGGAAGCGCATGAAAATTCATATGGTTAGCAAGGC
3W-type2-103	15b	CGGAAACCTGAATTTACCGTTCCAGACTCCTCAAGAGA
3W-type2-104	15i	AATGCTTTAAACAGTATGATATTCAACCGTTTCTCCGTGGG AACAAAAAGCGCC
3W-type2-105	15j	ATTCGCCACCAGTCGGGAAACCTGGTTTGCC
3W-type2-106	15k	CCAGCAGGTTTAGAGCTTGACGGGGAAAGCC
3W-type2-107	16a	TGAAACCATTTGTCACAATCAATATAGACGGG
3W-type2-108	16b	GATTAGCGATGGAAAGCGCAGTCTGTCACCAA
3W-type2-109	16c	GGTTTAGTATAGGTGTATCACCGTAGGATTAG
3W-type2-110	16d	CTCAGAACACCGCCACCCTCAGAAACTCAGGA
3W-type2-111	16e	TACAAACTCGTAACACTGAGTTTCCCGCCACC
3W-type2-112	16f	CAGCCCTCACAACGCCTGTAGCATGTCACCAG
3W-type2-113	16g	AACAGTTTATGGGATTTTGCTAAATCCACAGA
3W-type2-114	16h	TAAATTAACATTGAATCCCCCTCACAACCTTC
3W-type2-115	16i	GCGCAACTTAACAACCCGTCGGATCTAGCTGA
3W-type2-116	16jk	GCCCCGACGAAAATCCTGTTTGAGCTCACTGCCCCGCTTT TTCAGGCT
3W-type2-117	17a	AGAATTAACCTGAACACCCTGAACCACGGAA
3W-type2-118	17b	TAAGTTTATTTCGATAGCAGCACCGCCTCATT
3W-type2-119	17c	AAAGCCAGAGGGTTTTGCTCAGTAATAAGTA
3W-type2-120	17d	TAGCCCGGAGATCGTCACCCTCAGAAGGCCG
3W-type2-121	17e	CTTTTGCGGCGCCACCCTCAGAGCATAGGAA
3W-type2-122	17f	CCCATGTACCGAGGTGAATTTCTTTTTATCA
3W-type2-123	17g	GCTTGCTTTATAGTTAGCGTAACGTAAATGA
3W-type2-124	17h	ATTTTCTGTCCAGCGGAGTGAGAATAATCGTC

3W-type2-125	17i	ATAAATATTTGCCGGAGAGGGTAGATTAAAT
3W-type2-126	17j	GTGAGCGAGGTTGGGAAGGGCGATACATTAA
3W-type2-127	17k	TTGCGTTGCTGGTGGTTCCGAAATCACTAAATCGGAACCCT AAAGGGA
3W-type2-128	18a	AAAGACACAAAGTCAGAGGGTAATTGAGCGCT
3W-type2-129	18b	AAATAAATTAATCAGTAGCGACAGAGAAACGC
3W-type2-130	18c	GGGTTGATCCAGGCGGATAAGTGCTCACAAAC
3W-type2-131	18d	GGGAGTTACAGCGAAAGACAGCATCGTCGAGA
3W-type2-132	18e	CAAGCCCACACCACCCTCATTTTCGGCTTGCA
3W-type2-133	18f	TGTATCGGAAACAGCTTGATACCGAGGGATAG
3W-type2-134	18g	GACGTTAGATCTAAAGTTTTGTGCGCCTTTAAT
3W-type2-135	18h	TACTGCGGAGAAAGGAACAATACTTTCCA
3W-type2-136	18i	TCATCAACCTATTTTTGAGAGATCGCGTCCAA
3W-type2-137	18j	GCTAACTCCGGTGCGGGCCTCTTCGCCAGCTT
3W-type2-138	18k	CGTAAAGCGGC AAAATCCCTTATATGAGTGA
3W-type2-139	19a	AATATCAGGTGGCAACATATAAAAATCAAGT
3W-type2-140	19b	TTGCCTTTGATTGGCCTTGATATTTGAGGAC
3W-type2-141	19c	TAAAGACTACGGCTACAGAGGCTCGGAACGA
3W-type2-142	19d	GGGTAGCAATATTCCGTGCTGACCCACGCA
3W-type2-143	19e	TAACCGATTGACAACAACCATCGATAGTTGC
3W-type2-144	19f	GCCGACAAAGGCTCCAAAAGGAGAAAATCTC
3W-type2-145	19g	CAAAAAAAAAATTTTTTCACGTTGAGGAATTG
3W-type2-146	19h	CGAATAATTGTTTAGACTGGATATACAAAGG
3W-type2-147	19i	CTATCAGGTCTGGCCTTCCTGTAGCTATTAC
3W-type2-148	19j	GCCAGCTGAGCCTGGGGTGCCTAAAATCAAA
3W-type2-149	19k	AGAATAGCTCAAGTTTTTTGGGGTCGAGGTGC
3W-type2-150	20a	ACTGTAGAAATACATACATAAAGAGAGATAA
3W-type2-151	20b	TTTTTCATGAGGAAGTTTGAGGCAGGTCAGACAGCGTCAG
3W-type2-152	20i	AATTCGCGTCATTGCCTGAGAGTGGGGGTAATAGTAAAA
3W-type2-153	20j	AAGTGTAAGCGAAAGGGGGATGTTCAAAAAT
3W-type2-154	20k	CACCCAAACCGAGATAGGGTTGAGAAGCATA
3W-type2-155	21a	CCCACAAGAATTGAGTTAAGCCCTGTTAGCA
3W-type2-156	21b	AACGTAGACGCGTTTTTCATCGGCACAGCATTGACAGGAGG TTCCATTA
3W-type2-157	21i	TTTTGCAAAAGAAGTTTTGCCAGACTGGAGCAAACAAGA GACCAATAGGAACGCCAGCTGCAAG
3W-type2-158	21j	GCGATTAACACAACATACGAGCCGGTGTTGTT
3W-type2-159	21k	CCAGTTTGCGCATGGCCCACTACGTGAACCAT
3W-type2-160	22a	CATAGCCCCCTTATTACGCAGTAAATAATAAGAGCAAGAAA CAATGA
3W-type2-161	22b	AACGGGTAAAATACGTAATGCCACCACCAGAGCCGCCGCT TTTCGGT
3W-type2-162	22i	ATTTTTTAAATCGATGAACGGTACGAGAGGC
3W-type2-163	22j	TGTAGCAACATTGCAACAGGAAAGAATGGCTATTAGTCTA ACACCGCCTGCAACAGTGCCAC
3W-type2-164	22k	CTATCAGGGAACAAGAGTCCACTTCCGCTCA

3W-type2-165	23a	AATAGCAATGGCATGATTAAGACTCCTTATTA
3W-type2-166	23b	GCGTTTGCCCGCCACCAGAACCACTACGAAGGCACCAAC C
3W-type2-167	23i	TAAAAACCAAAATAGATCGTAAAACTAGCATTAAATTTTT GTAAAGGTTTTCC
3W-type2-168	23j	CAGTCACGTGTGTGAAATTGTAAATTAAAGA
3W-type2-169	23k	ACGTGGACAACGTCAAAGGGCGAAAAACCGT
3W-type2-170	24a	AAAAGAACTAGCTATCTTACCGAAGCCCTTTT
3W-type2-171	24b	CCTCAGAGCATCTTTTCATAATCAGAATACCC
3W-type2-172	24c	CGCCTGATTAAAACGAAAGAGGCAGCCACCAC
3W-type2-173	24d	GGACAGATAAATTGTGTGCGAAATCTGTATCAT
3W-type2-174	24e	TGCTCATTGAACGGTGTACAGACCTTGAAAGA
3W-type2-175	24f	TGCGATTTCAAGTGAATAAGGCTTGAACAAAGC
3W-type2-176	24g	ATCAGTTGTAAGAACTGGCTCATTTTACCTTA
3W-type2-177	24h	GACGACGAAGATTTAGGAATACCAAAAGATTC
3W-type2-178	24i	AATTCGCAGTCAATCATATGTACCGTTTACCA
3W-type2-179	24j	CTGTTTCCACGTTGTAAAACGACGTTTGTAA
3W-type2-180	25a	TAAGAAAAACGCAATAATAACGAAATCACC
3W-type2-181	25b	GGAACCAGACCGCCACCCTCAGAAAAGAATA
3W-type2-182	25c	CACTAAAAAGTACAACGGAGATTCGCGACCT
3W-type2-183	25d	GCTCCATGCCGAACCTGACCAACTAGGCGCAT
3W-type2-184	25e	AGGCTGGCTACCCAAATCAACGTCCCTGACG
3W-type2-185	25f	AGAAACACTTTAATCATTGTGAAATACCAGT
3W-type2-186	25g	CAGGACGTCATTATTACAGGTAGCATTCAAC
3W-type2-187	25h	TAATGCAGACTATCATAACCCTCCCGGTTGA
3W-type2-188	25i	TAATCAGATGTAAACGTTAATATGCCAGTGC
3W-type2-189	25j	CAAGCTTGCGAATTCGTAATCATGGTCATAG
3W-type2-190	26a	ACCGAGGAGTAAGCAGATAGCCG
3W-type2-191	26b	CCCTCAGAAGCCACCACCGGAACAGAAGGAA
3W-type2-192	26c	CGAAACAACACTCATCTTTGACCAGCCGCCA
3W-type2-193	26d	TAAGGGAATTACTTAGCCGGAACACCAAGCG
3W-type2-194	26e	ATATTCATTGACCTTCATCAAGAGTCAATCA
3W-type2-195	26f	ATTTCAACCAGAACGAGTAGTAAAGAACCGG
3W-type2-196	26g	CGGAACAATGGGAAGAAAAATCTATGGTTTA
3W-type2-197	26h	AGAGCAACATACATAACGCCAAACGAACTAA
3W-type2-198	26i	ATTTAAATAAAGCCCCAAAAACAGCATAGTA
3W-type2-199	26j	CCCCGGGTACCGAGCTCATGCCTGCAGGTCGAAGCAAAT

2.5.7 Preparation of DNA origami scaffold assembled with modular adaptors

A DNA origami scaffold (I-1AH/II-1ZS/III-1AC) was incubated with ZF-SNAP, AZ-CLIP, and/or AZ-Halo under the conditions shown in the caption of figures or tables. In the typical condition, the crosslinking reaction of each of modular adaptor was carried out for 30 minutes at ambient temperature with 5 nM DNA scaffold and 100 nM modular adaptor (ZF-SNAP, AZ-CLIP or AZ-Halo) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, and 0.02% Tween20. The mixture was purified by size-exclusion chromatography (400 μ L volume of Sephacryl S-400 in Ultrafree-MC-DV) equilibrated with the buffer used for the reaction to remove the unbound modular adaptors. The fractions containing DNA origami scaffold were utilized for AFM analyses. The preparation of DNA origami scaffolds assembled with modular adaptor-fused enzymes was performed in the same manner.

2.5.8 AFM imaging and statistical analysis

The sample was deposited on freshly cleaved mica surface (1.5 mm ϕ) and adsorbed for 5 min at ambient temperature, then washed three times with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂. The sample was scanned in tapping mode using a fast-scanning AFM system (Nano Live Vision, RIBM Co. Ltd., Tsukuba, Japan) with a silicon nitride cantilever (Olympus BL-AC10DS-A2). At least three independent preparations of each sample were analyzed by AFM and several images were acquired from different regions of the mica surface. The total number of DNA scaffolds corresponds to the number of expected rectangular shapes possessing three cavities observed by AFM. The specific and nonspecific binding of modular adaptors were counted for only the modular adaptor

bound to the perfectly folded DNA scaffold. The yield of DNA-scaffold-assembled modular adaptor was calculated as reported previously.

The yield of modular adaptor - DNA origami complex ($P_{matched\ posi}$) was calculated as the percentage of attached structures, *i.e.*, the number of modified DNA origami bearing the proteins at the expected position divided by the total number of cavities containing specific binding sites for modular adaptor of the well-formed DNA origami structures. In this case, the total number of the well-formed DNA origami structures (N_{total}) is equal to the total number of cavities containing the ZF-SNAP, AZ-CLIP or AZ-Halo binding sites:

$$P = (N_{matched\ posi} / N_{total}) \times 100$$

The yield of modular adaptor found in nonspecific positions ($P_{unmatched}$) was also calculated as the percentage of attached structures, *i.e.* the number of cavities on which modular adaptor bound nonspecifically ($N_{unmatched\ posi}$) divided by the total number of cavities without the modular adaptor binding sites in the well-formed DNA origami structures:

$$P_{unmatched} = (N_{unmatched} / N_{total}) \times 100$$

The co-assembly yield of two or three modular adaptors found in matched positions ($P_{coassembly}$) was also calculated as the percentage of attached structures, *i.e.* the number of cavities on which two or three modular adaptors bound at the matched position ($N_{coassembly}$) divided by the total number of cavities without the modular adaptor binding sites in the well-formed DNA origami structures:

$$P_{coassembly} = (N_{coassembly} / N_{total}) \times 100$$

2.6 References

- (1) Seeman N. C. DNA in a material world. *Nature* **2003**, 421, 427-431.
- (2) Pinheiro A. V.; Han D.; Shih W. M.; Yan H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **2011**, 6, 763-772,
- (3) Chandrasekaran A. R.; Anderson N.; Kizer M., Halvorsen K., Wang X., Beyond the fold: Emerging biological applications of DNA origami. *ChemBioChem.* **2016**, 17, 1081-1089,
- (4) Rothemund P. W. K., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, 440, 297-302.
- (5) Kazenwadel F.; Franzreb M.; Rapp B. E.; Synthetic enzyme supercomplexes: Co-immobilization of enzyme cascades. *Anal. Methods.* **2015**, 7, 4030-4037.
- (6) Linko V.; Nummelin S.; Aarnos L.; Tapio K.; Toppari J. J.; Kostianen M. A. DNA-based enzyme reactors and systems. *Nanomaterials* **2016**, 6, 139.
- (7) Wheeldon I.; Minter S. D.; Banta S.; Barton S. C.; Atanassov P.; Sigman M.; Substrate channeling as an approach to cascade reactions. *Nat. Chem.* **2016**, 8, 299-309.
- (8) Küchler A.; Yoshimoto M.; Luginbühl S.; Mavelli F.; Walde P. Enzymatic reactions in confined environments. *Nat. Nanotechnol.* **2016**, 11, 409-420.
- (9) Rajendran A.; Nakata E.; Nakano S.; Morii T. Nucleic-acid-templated enzyme cascades. *ChemBioChem* **2017**, 18, 696-716.
- (10) Niemeyer C. M. Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angew. Chem. Int., Ed.* **2010**, 49, 1200-1216.
- (11) Yang Y. R.; Liu Y.; Yan H. DNA nanostructures as programmable biomolecular scaffolds. *Bioconjugate Chem.* **2015**, 26, 1381-1395.
- (12) Wilner O. I.; Weizmann Y.; Gill R.; Lioubashevski O.; Freeman R.; Willner I.

- Enzyme cascades activated on topologically programmed DNA scaffolds. *Nat. Nanotechnol.* **2009**, *4*, 249-254.
- (13) Fu J.; Liu M.; Liu Y.; Woodbury N. W.; Yan H. Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *J. Am. Chem. Soc.* **2012**, *134*, 5516-5519.
- (14) Fu J.; Yang Y. R.; Johnson-Buck A.; Liu M.; Liu Y.; Walter N. G.; Woodbury N. W.; Yan H.; Multi-enzyme complexes on DNA scaffolds capable of substrate channeling with an artificial swinging arm. *Nat. Nanotechnol.* **2014**, *9*, 531-536.
- (15) Rosen C. B.; Kodal A. L. B.; Nielsen J. S.; Schaffert D. H.; Scavenius C.; Okholm A. H.; Voigt N. V.; Enghild J. J.; Kjems J.; Tørring T.; Gothelf K. V.; Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. *Nat. Chem.* **2014**, *6*, 804-809.
- (16) Nakata E.; Liew F. F.; Uwatoko C.; Kiyonaka S.; Mori Y.; Katsuda Y.; Endo M.; Sugiyama H.; Morii T. Zinc-finger proteins for site-specific protein positioning on DNA-origami structures. *Angew. Chem., Int. Ed.* **2012**, *51*, 2421-2424.
- (17) Ngo T. A.; Nakata E.; Saimura M.; Kodaki T.; Morii T. A protein adaptor to locate a functional protein dimer on molecular switchboard, *Methods* **2014**, *67*, 142-150.
- (18) Nakata E.; Dinh H.; Ngo T. A.; Saimura M.; Morii T. A modular zinc finger adaptor accelerates the covalent linkage of proteins at specific locations on DNA nanoscaffolds, *Chem. Commun.* **2015**, *51*, 1016-1019.
- (19) Pavletich N. P.; Pabo C. O. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å, *Science* **1991**, *252*, 809-817.
- (20) Keppler A.; Gendreizig S.; Gronemeyer T.; Pick H.; Vogel H.; Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*, *Nat. Biotechnol.* **2003**, *21*, 86-89.

- (21) Ngo T. A.; Nakata E.; Saimura M.; Morii T. Spatially organized enzymes drive cofactor-coupled cascade reactions. *J. Am. Chem. Soc.* **2016**, *138*, 3012-3021,
- (22) Saccà B.; Meyer R.; Erkelenz M.; Kiko K.; Arndt A.; Schroeder H.; Rabe K. S.; Niemeyer C. M. Orthogonal protein decoration of DNA origami. *Angew. Chem., Int. Ed.* **2010**, *49*, 9378-9383.
- (23) Koßmann K. J.; Ziegler C.; Angelin A.; Meyer R.; Skoupi M.; Rabe K. S.; Niemeyer C. M. A rationally designed connector for assembly of protein-functionalized DNA nanostructures. *ChemBioChem* **2016**, *17*, 1102-1106,
- (24) Bando T.; Sugiyama H.; Synthesis and biological properties of sequence-specific DNA-alkylating pyrrole-imidazole polyamides. *Acc. Chem. Res.* **2006**, *39*, 935-944.
- (25) Jeffries T. W. Engineering yeasts for xylose metabolism. *Curr. Opin. Biotechnol.* **2006**, *17*, 320-325.
- (26) Matsushika A.; Inoue H.; Kodaki T.; Sawayama S. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **2009**, *84*, 37-53.
- (27) Pival S. L.; Birner-Gruenberger R.; Krump C.; Nidetzky B. D-xylulose kinase from *Saccharomyces cerevisiae*, Isolation and characterization of the highly unstable enzyme, recombinantly produced in *Escherichia coli*. *Protein. Expr. Purif.* **2011**, *79*, 223-230.
- (28) Sera T.; Uranga C. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry* **2002**, *41*, 7074-7081.
- (29) Gautier A.; Juillerat A.; Heinis C.; Corrêa Jr, I. R. Kindermann M., Beaufils F., Johnsson K., An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **2008**, *15*, 128-136.

- (30) Los G. V.; Encell L. P.; McDougall M. G.; Hartzell D. D.; Karassina N.; Zimprich C.; Wood M. G.; Learish R.; Ohana R. F.; Urh M.; Simpson D.; Mendez J.; Zimmerman K.; Otto P.; Vidugiris G.; Zhu J.; Darzins A.; Klaubert D. H.; Bulleit R. F.; Wood K. V. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **2008**, 3, 373-382.
- (31) England C. G.; Luo H.; Cai W. HaloTag technology: A versatile platform for biomedical applications. *Bioconjugate Chem.* **2015**, 26, 975-986.
- (32) Sagredo S.; Pirzer T.; Aghebat Rafat A.; Goetzfried M. A.; Moncalian G.; Simmel F. C.; De La Cruz F. Orthogonal protein assembly on DNA nanostructures using relaxases. *Angew. Chem. Int., Ed.* **2016**, 55, 4348-4352.
- (33) Gamsjaeger R.; Liew C. K.; Loughlin F. E.; Crossley M.; Mackay J. P. Sticky fingers: Zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* **2007**, 32, 63-70.
- (34) Hinner M. J.; Johnsson K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **2010**, 21, 766-776.
- (35) Lang K.; Chin J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **2014**, 9, 16-20.
- (36) Schreiber G.; Haran G.; Zhou H. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* **2009**, 109, 839-860.
- (37) Gargano J. M.; Ngo T.; Kim J. Y.; Acheson D. W. K.; Lees W. J. Multivalent inhibition of AB5 toxins. *J. Am. Chem. Soc.* **2001**, 123, 12909-12910.
- (38) Krishnamurthy V. M.; Semetey V.; Bracher P. J.; Shen N.; Whitesides G. M. Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J. Am. Chem. Soc.* **2007**, 129, 1312-1320.
- (39) Hobert E. M.; Doerner A. E.; Walker A. S.; Schepartz A. Effective molarity redux,

Proximity as a guiding force in chemistry and biology. *Isr. J. Chem.* **2013**, 53, 567-576.

- (40) Berg O. G.; Winter R. B.; von Hippel P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **1981**, 20, 6929-6948.
- (41) Kim J.; Pabo C. O. Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 2812-2817.
- (42) Halford S. E.; Marko J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **2004**, 32, 3040-3052.
- (43) Halford S. E. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* **2009**, 37, 343-348.
- (44) Geertz M.; Shore D.; Maerkl S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci. USA.* **2012**, 109, 16540-16545.

CHAPTER 3

Development of the sequence-specific modular adaptors

3.1 Introduction

In the previous chapter, I have demonstrated the advantages of the orthogonal modular adaptors for assembling multiple enzymes on a DNA scaffold. Further combination of DNA binding proteins and protein-tags would provide the increasing the number of the orthogonal modular adaptors. In this chapter, I developed a general strategy for the simple, yet effective method to expand the number of orthogonal modular adaptors. The aimed strategy is to design the modular adaptor in the way that the same protein-tag is combined with different DNA binding proteins and that the orthogonality is governed exclusively by the DNA binding proteins. These next generation of modular adaptors are termed as the sequence-specific modular adaptors.

Site-specific chemical reactions to modify nucleic acids by ligands rely on the specific molecular recognition of target sites and selective reactions.¹⁻⁵ Sequence-selective chemical modification of DNA by DNA binding molecules is one of the long-standing challenges in the field of chemistry.^{4,5} Various sequence-specific DNA binders, ranging from small molecules^{6,7} to short peptides,⁸ proteins,⁹ and synthetic oligonucleotides,¹⁰ have been coupled with a reactive group to specifically modify the nucleic acid bases, the sugar moiety or the phosphodiester linkage. However, coupling a reactive group to a sequence-specific DNA binding domain often accompanied with off-target reactions. A basic principle to design DNA modifiers exclusively governed by the DNA sequence recognition remains to be established.

The fully addressable nano-architectures of various shapes and geometries provide ideal scaffolds to locate a molecule of interest with a nanometre spatial precision.¹¹ DNA scaffolds prepared by DNA origami¹² are readily incorporated additional DNA sequences on the surface by simple extension of the staple strand. These extended DNA strands are further chemically modified to serve as binding sites for various molecules.¹³⁻¹⁵ Proteins are a particularly interesting class of molecules to

assemble on the DNA scaffold because of their huge functional variability. To assemble on the DNA scaffold through the hybridization, a protein of interest (POI) is chemically modified with a short DNA strand that is complementary to the attached sequence on the DNA scaffold. Though this method is convenient, it is suffered from drawbacks: the method often limits the modification conditions and causes the loss of activity.¹⁶⁻¹⁹ It would be ideal to locate the POI to a specific position on DNA scaffold without the need of post-chemical modification of POI to prepare the precisely controlled assembly of POI that is suitable for further investigation.²⁰⁻²⁶ To realize the efficient and quick location of POI at specific DNA sites, we have developed modular adaptors (Figure 3.2a). The modular adaptor consists of a DNA binding protein (zif268) as DNA sequence-selective recognition module and a self-ligating protein tag (SNAP-tag) that forms a chemoselective covalent linkage between the protein tag and its substrate (BG) modified at the given DNA sequence (Figure 3.2a).²⁷ A modular adaptor fused enzyme was specifically located at the target position of DNA scaffold in fast reaction kinetics and in high yield with fully retaining the original enzymatic activity.^{27,28} To assemble three types of proteins at unique positions on DNA scaffold by taking the advantage of chemoselective characteristics for three types of self-ligating protein tag (SNAP-tag, CLIP-tag, and Halo-tag), a set of modular adaptors with orthogonality and retaining fast reaction kinetics under mild conditions has been selected from the systematic combination of DNA binding domains (zif268 or AZP4) and protein tags.²⁸ Three types of orthogonal modular adaptors simultaneously locate three types of proteins at their respective positions on the DNA scaffold.

Because the reaction of modular adaptor to its substrate locating in the specific DNA sequence proceeds through the specific recognition of DNA sequence and the alkylation reaction between the protein-tag and its substrate, the factors governing specific reaction of modular adaptors are well correlated to the basic principle for

designing selective DNA modifiers driven by the sequence-specific DNA recognition. In this study, I constructed a sequence-specific modular adaptors by tuning the kinetics of the reaction between the self-ligating protein tag and its substrate. Based on the results of proof-of-principle experiments, I propose a general principle on the kinetic parameters for the DNA binding and alkylation reaction to design a site-specific modifier driven by the target recognition (Figure 3.1).

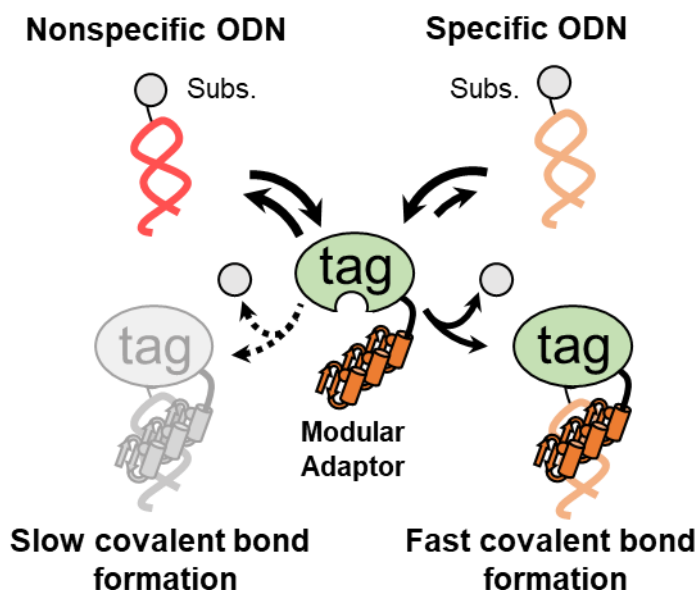


Figure 3.1. Model of sequence-specific modular adaptor. Covalent bond formation only occurs at the sequence-specific ODN of the DNA binding module.

In Chapter 3, with an aim to increase the number of the modular adaptors, I investigated the three sequence-specific modular adaptors, ZF-CLIP, AZ-CLIP, and CLIP-G, containing CLIP-tag and three different DNA binding proteins zif268, AZP4, and GCN4,²⁹⁻³¹ respectively.

3.2 Results

3.2.1 Design of sequence-specific modular adaptor and the substrate modified ODNs

Previous study on the reaction of modular adaptors described in Chapter 2 suggested important factors affecting the kinetics and the selectivity of cross-linking reaction. The covalent bond formation between the modular adaptor and its substrate-modified ODN is described by the following two steps (Figure 3.2b): the first step forms a reversible complex between the DNA binding domain and substrate-modified ODN, and the second step is an intramolecular crosslinking reaction between the self-ligating protein tag domain and its substrate on substrate-modified ODN.²⁸ The apparent rate constant of cross-linking reaction (k_{app}) ($M^{-1}s^{-1}$) between the modular adaptor and the substrate-modified DNA is represented by an equation shown in Figure 3.2b. From the equation of k_{app} , a modular adaptor with a highly reactive protein tag to its substrate falls into case 1 ($k_{cov} \gg k_{off}$). Because the association rates (k_{on}) of DNA binding proteins to its specific and non-specific site are almost identical,³²⁻³⁴ the modular adaptor is expected to react substrate-modified ODN regardless the sequence of DNA (the case 1 in Figure 3.2b). This is actually the case for the reaction of SNAP-tag and BG (Table 3.1, $\sim 10^4 M^{-1} s^{-1}$).²⁸ The modular adaptors ZF-SNAP and AZ-SNAP consisting of the same SNAP-tag and different DNA binding proteins zif268 and AZP4, respectively, showed the crosslinking reaction at the expected sequence, but the reactions with the substrate located at the unmatched sequence proceeded to the same extents (Figure 3.3 and Table 3.1).

Therefore, reducing the reactivity of the protein tag to the substrate to fit in the condition of case 2 (Figure 3.2b) would be an effective strategy to prevent the cross-linking reaction at the unmatched DNA sequences by the modular adaptors. In case 2, the higher K_D for the unmatched complex of zinc finger motif and ODN is, the

lesser cross-linking reactions of the unmatched complexes proceed. The crosslinking reactions for the matched and unmatched pairs are described as in Figure 3.2c. By following the conditions of case 2, the sequence-selective crosslinking reaction will take place when $k_{app} \gg k'_{app}$.

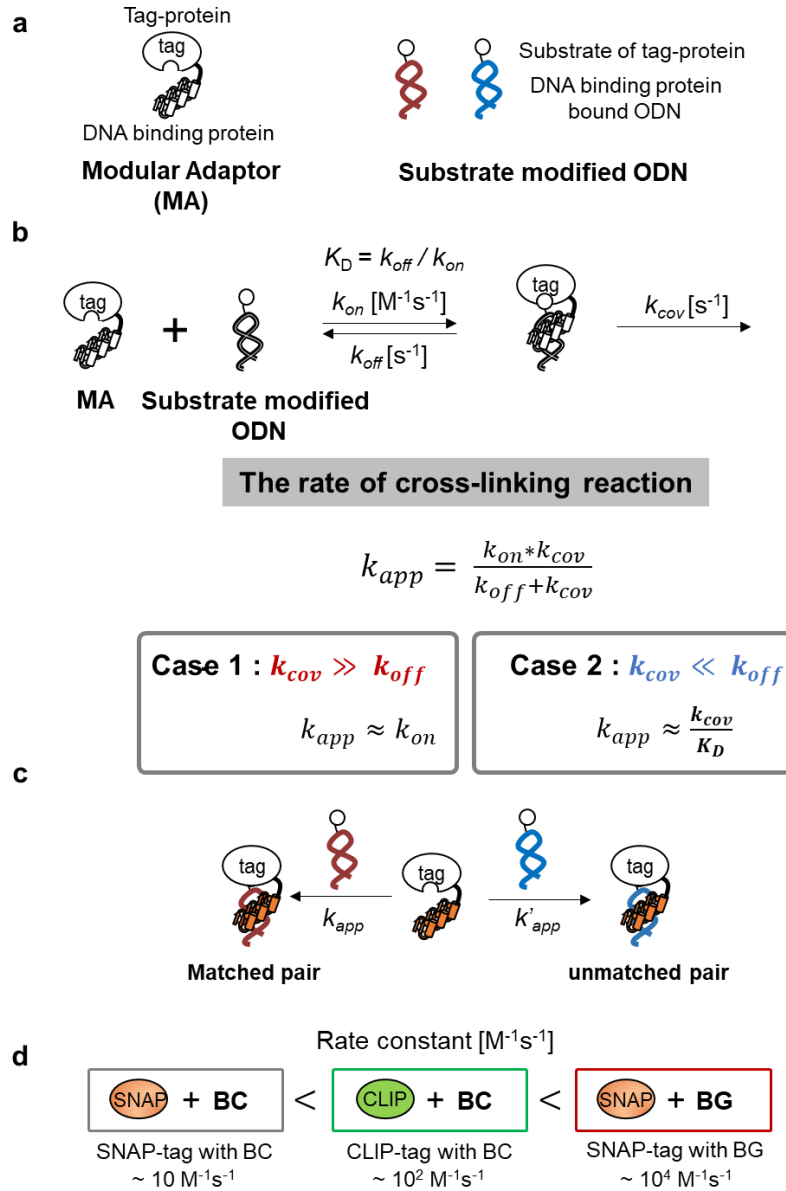


Figure 3.2. Design of sequence-specific modular adaptors. (a) Illustrations for the structures and elements of modular adaptor and substrate modified ODN. (b) A scheme illustrates the cross-linking reaction between modular adaptor and its substrate-modified ODN. The complex formation between modular adaptor and its substrate-modified ODN is governed by

k_{on} and k_{off} . The equilibrium dissociation constant (K_D) for the complex of MA and substrate-modified ODN is $K_D = k_{off} / k_{on}$. The apparent rate constant (k_{app}) is described by K_D and the rate constant k_{cov} [s^{-1}] for the proximity-driven covalent bond formation reaction between MA and substrate-modified ODN. (c) The sequence-selective and chemo-selective cross-linking reaction between a matched pair is governed by k_{app} , and non-sequence-specific but chemoselective reaction for an unmatched pair is by k'_{app} . Sequence-selective crosslinking reaction proceeds when $k_{app} \gg k'_{app}$. (d) Comparison of the rate constant of protein-tag with their substrates. Chemical structures of the substrates are shown in Figure 3.33a

Table 3.1 Kinetic parameters for the cross-linking reaction between 5'-³²P-end-labeled ODN derivatives with BG and modular adaptors containing SNAP-tag. (bold: Rate constant for the matched pair (k_{app}), plane: Rate constant for the unmatched pair (k'_{app})).

Rate constant ($M^{-1} s^{-1}$)	ZF-SNAP	AZ-SNAP
ODN-ZF-BG	$(3.8 \pm 1.2) \times 10^5$	$(9.5 \pm 0.6) \times 10^4$
ODN-AZ-BG	$(2.6 \pm 0.2) \times 10^4$	$(9.1 \pm 2.4) \times 10^5$

The values were taken from Chapter 2.

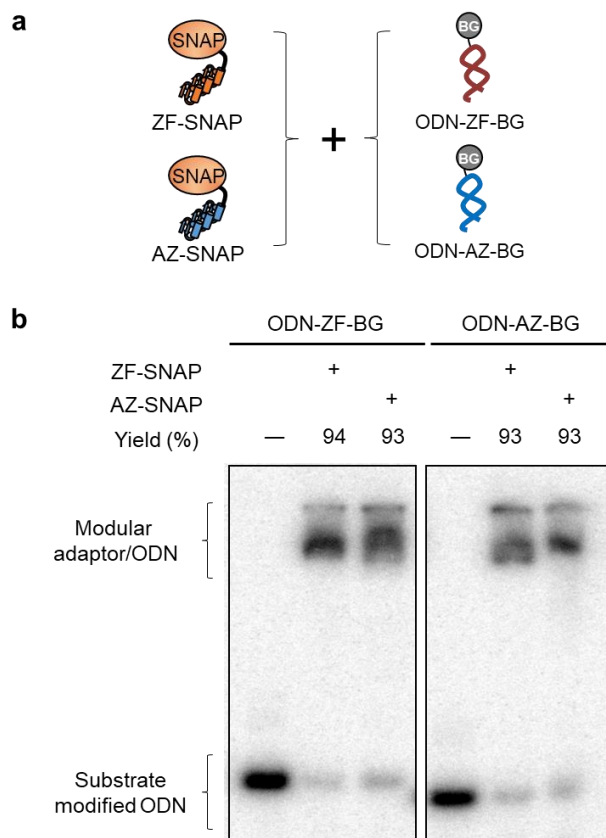


Figure 3.3 (a) Analyses of crosslinking reactions between the modular adaptors (ZF-SNAP and AZ-SNAP) and the substrate modified ODNs (ODN-ZF-BG and ODN-AZ-BG). (b) Denaturing PAGE analyses of the crosslinking reaction by modular adaptors (ZF-SNAP and AZ-SNAP) and the substrate (BG) modified ODN (ODN-ZF-BG and ODN-AZ-BG), respectively. Each 5'-³²P-end-labeled ODN (0.5 nM) (ODN-ZF-BG or ODN-AZ-BG) was incubated with a modular adaptor (100 nM : ZF-SNAP or AZ-SNAP) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature.

Based on the simulation of reactions by varying these rate constants (Figure 3.4 and Figure 3.5), the orthogonal crosslinking reaction will take place when k_{app} is two orders higher than k'_{app} . The apparent reaction rates of matched pair (k_{app}) and unmatched pair (k'_{app}) is very changed depending on k_{cov} . In the case of Line 1 ($k_{cov} = 1 \text{ s}^{-1}$) in Figure 3.5, k_{app} is almost $1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and k'_{app} is near $1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, thus the cross-linking reaction of matched pair and unmatched pair is progressed in the same time scale. In the case of Line 3 ($k_{cov} = 1 \times 10^{-5} \text{ s}^{-1}$), k_{app} is almost $1 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ and k'_{app} is near $10 \text{ M}^{-1}\text{s}^{-1}$, which is the maximum difference in this situation, but the long time incubation is required to complete the reaction (see also Figure 3.4). In the case of Line 2 ($k_{cov} = 4 \times 10^{-3} \text{ s}^{-1}$), k_{app} is almost $3 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$ and k'_{app} is near $4 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$, which showed almost two order difference with keeping reactivity of the matched pair. These simulation results strongly indicated that the tuning of the rate constant of covalent bond formation is very important to appear the sequence selectivity between matched pair and unmatched pair.

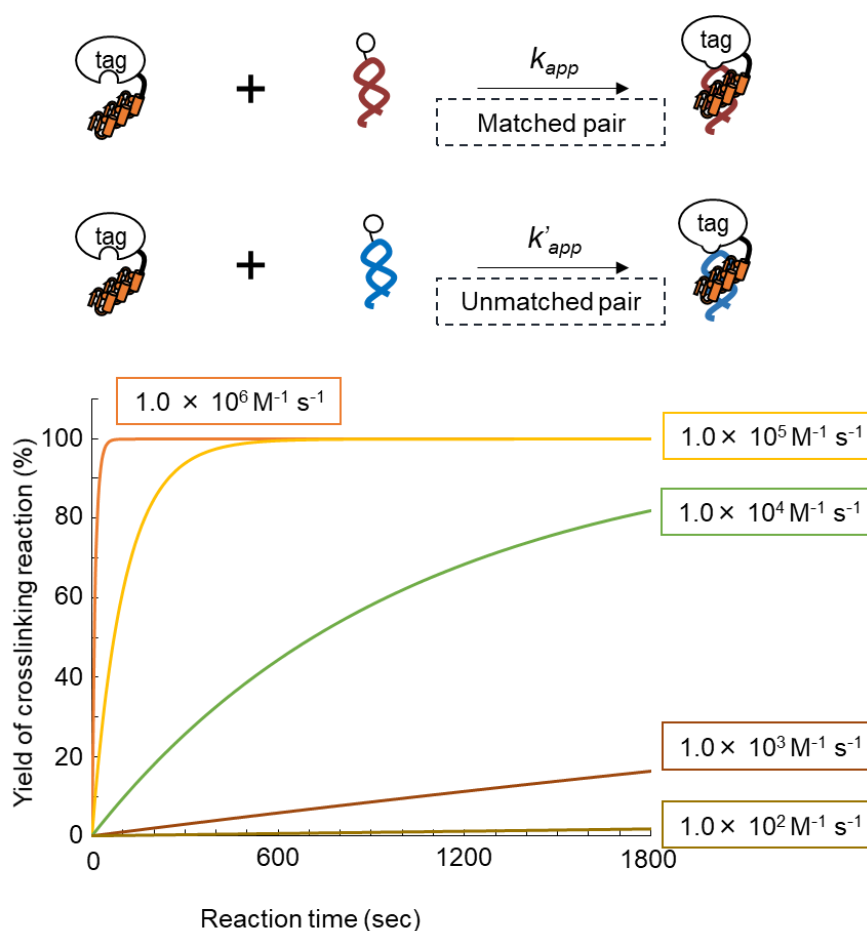


Figure 3.4 Rate simulation. (a) Schematic illustration of cross-linking reaction of matched pair and unmatched pair. The apparent reaction rate of the matched pair and the unmatched pair are written as k_{app} [$M^{-1} s^{-1}$] and k'_{app} [$M^{-1} s^{-1}$], respectively. (b) Simulation data for the time-course yields of cross-linking reaction at different time point. The yields of cross-linking reaction under the apparent reaction rates (10^2 - $10^6 M^{-1} s^{-1}$) was simulated by using Tenua³⁵⁻³⁷ as the kinetic simulator. As the simulation condition, the concentrations of substrate tethered ODN and modular adaptor were set as 10 nM and 100 nM, respectively.

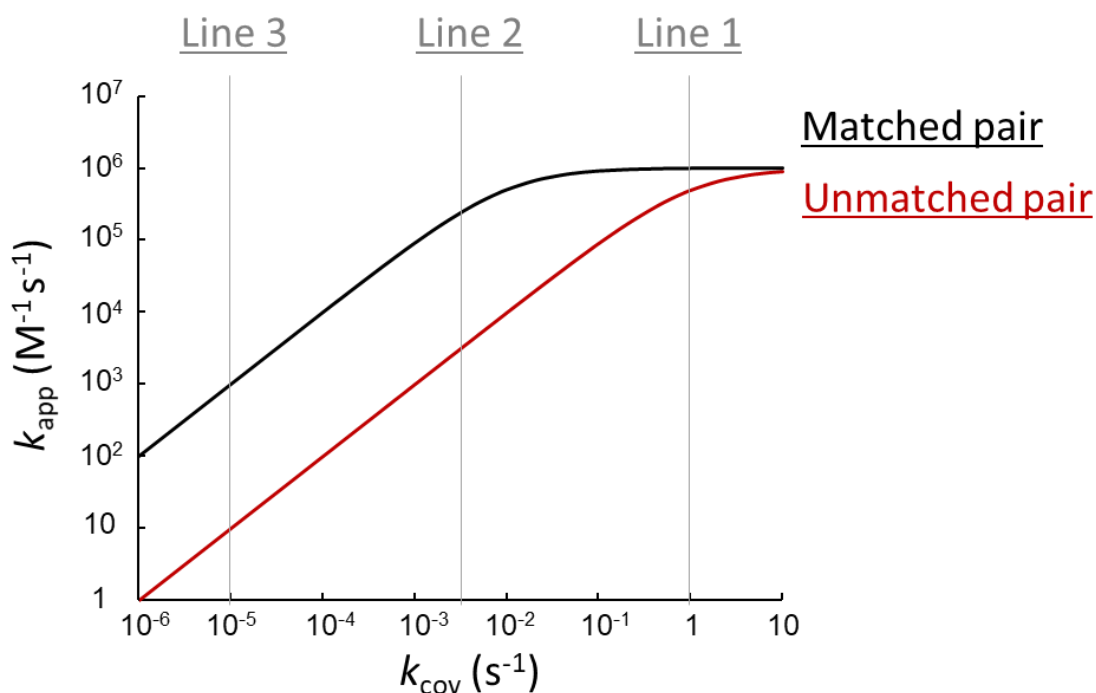


Figure 3.5 The apparent reaction rates (k_{app}) ($M^{-1} s^{-1}$) with various rate constants of covalent bond formation (k_{cov}) (10^{-6} - $10 s^{-1}$) were simulated. The association rate constant was set as $1 \times 10^6 M^{-1} s^{-1}$ for both matched and unmatched pair and the dissociation rate constant was set as $0.01 s^{-1}$ for matched pair and $1 s^{-1}$ for unmatched pair, respectively. It should be noted that these parameters are consistent with the actually determined values as shown in Figure 3.8, 3.9 and Table 3.1-3.4.

To test the hypothesis, two combinations of protein tag and substrate, CLIP-tag with *O*²-benzylcytosine (BC) ($\sim 10^2 M^{-1} s^{-1}$) and SNAP-tag with BC ($< 10 M^{-1} s^{-1}$) (Figure 3.6, Table 3.2) were applied as the cross-link-forming domain of modular adaptor. Both expected to show lower reactivity than the combination of SNAP-tag and BG ($\sim 10^4 M^{-1} s^{-1}$) and k_{cov} values that are smaller than k_{off} . Two DNA binding proteins, zif268 and AZP4, were utilized for the DNA-binding domain of modular adaptor. ZF-CLIP was constructed by fusing CLIP-tag to the C-terminal of zif268 through a GGSGGS linker. ZF-SNAP, AZ-SNAP, and AZ-CLIP were prepared as described previously.²⁸ ODNs containing the zif268 or AZP4 binding sequence were designed to

form a loop with four T nucleotides, in which one of the T nucleotides was replaced by amino-C6-T (ODN-ZF or ODN-AZ). The amino-C6-T was modified with each of the substrates BG or BC.

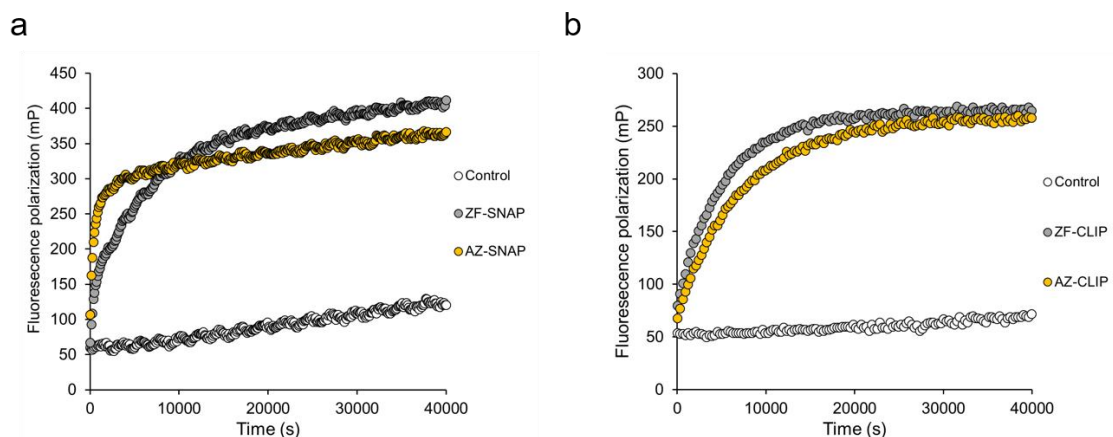


Figure 3.6. Time course formation of covalent bond of (a) SNAP-tag derivatives and (b) CLIP derivatives to fluorophore modified BC and BG, respectively. The reaction was conducted in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl_2 , 1 mM DTT, 1 μM ZnCl_2 , and 0.02% Tween20 at ambient temperature. (a) A 250 nM SNAP-tag derivative was added to the reaction buffer containing 1 nM fluorophore modified BG to start the reaction. (b) A 2 μM CLIP-tag derivative was added to a reaction buffer containing 10 nM fluorophore modified BC to start the reaction.

Table 3.2 Kinetic parameters for the cross-linking reactions between an adaptor containing SNAP-tag or CLIP-tag with ATTO488-BC or ATTO488-BG.

k_{cov} ($\text{M}^{-1}\text{s}^{-1}$)	ATTO488-BG	ATTO488-BC
ZF-SNAP	$(6.0 \pm 0.1) \times 10^3$	< 10
AZ-SNAP	$(2.4 \pm 0.1) \times 10^4$	< 10
ZF-CLIP	< 10	87 ± 7.0
AZ-CLIP	< 10	70 ± 1.7

Cross-linking reactions of these modular adaptors to substrate-modified ODN containing the respective matched or unmatched sequence were analyzed by means of denaturing polyacrylamide gel electrophoresis (PAGE) as shown in Figures 3.8 and 3.9. In the combination of CLIP-tag and BC, the apparent second-order rate constants (k_{app}) for the reactions of matched pair ODN-ZF-BC with ZF-CLIP and ODN-AZ-BC with AZ-CLIP were 4.0×10^6 and $5.0 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, respectively (Figure 3.9 and Table 3.4). In the combination of SNAP-tag with BC, k_{app} for the reactions of matched pair ODN-ZF-BC with ZF-SNAP and ODN-AZ-BC with AZ-SNAP were 1.7×10^3 and $3.0 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$, respectively (Figure 3.8 and Table 3.3). The cross-linking reactions between ZF-CLIP with ODN-ZF-BC and AZ-CLIP with ODN-AZ-BC proceeded in 92 and 90% yield, respectively (Figure 3.9). The reactions of ZF-SNAP with ODN-ZF-BC and AZ-SNAP with ODN-AZ-BC were very slow with the yield of 40 and 46%, respectively, after 30 min (Figure 3.8).

Table 3.3 Kinetic parameters for the cross-linking reactions between 5'-³²P-end-labeled ODN modified with BC and modular adaptors containing SNAP-tag. (bold: Rate constant for the matched pair (k_{app}), plane: Rate constant for the unmatched pair (k'_{app})).

Rate constant ($M^{-1} s^{-1}$)	ZF-SNAP	AZ-SNAP
ODN-ZF-BC	$(1.7 \pm 0.2) \times 10^3$	11 ± 3
ODN-AZ-BC	$22 \pm 1^*$	$(3.0 \pm 0.1) \times 10^2$

*The value was taken from Chapter 2.

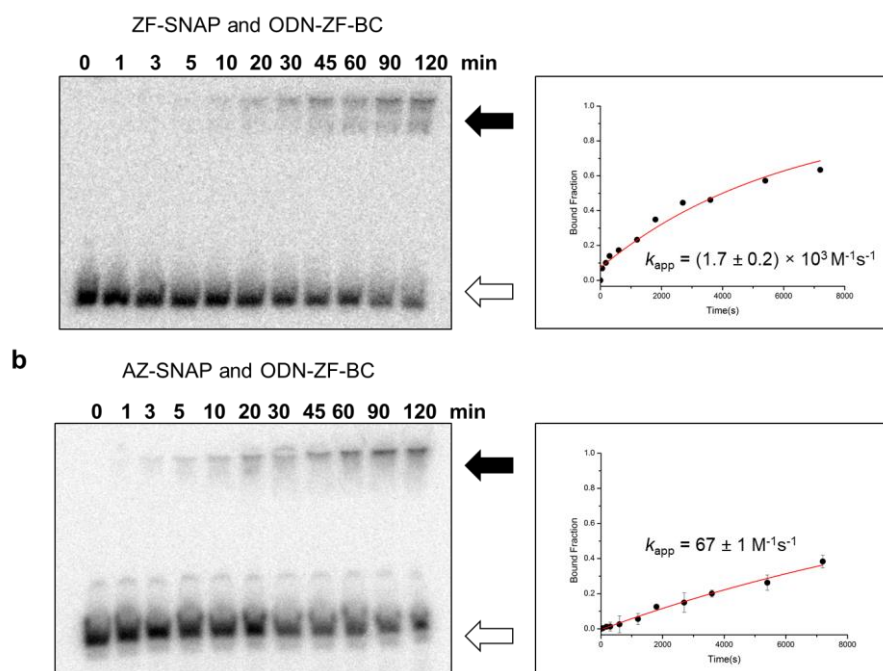


Figure 3.7 Autoradiograms show denaturing gel electrophoretic analyse of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC with 100 nM ZF-SNAP (a) and 100 nM AZ-SNAP (b). Open arrows and filled arrows denote the free ODNs and the covalent complex with modular adaptors, respectively (left). Time-course plot for the crosslinking reaction to estimate the rate constants (k) (right). The estimated rate constants are listed in Table 3.2. Reaction conditions: 40 mM Tris-HCl (pH8.0), 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$, 0.02% Tween20, 200 nM BSA, and 100 nM calf thymus DNA at ambient temperature.

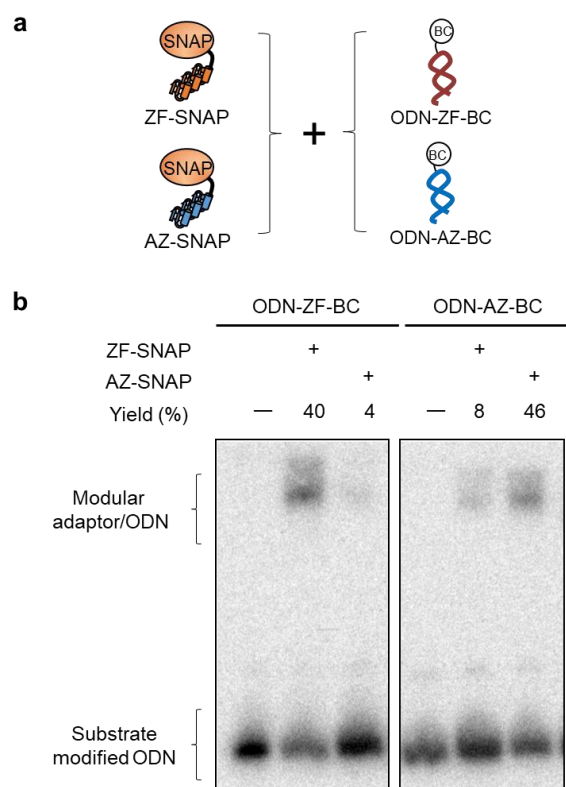


Figure 3.8 (a) Analyses of crosslinking reactions between the modular adaptors (ZF-SNAP and AZ-SNAP) and the substrate (BC) modified ODNs (ODN-ZF-BC and ODN-AZ-BC). (b) Denaturing PAGE analyses of the crosslinking reactions by modular adaptors (ZF-SNAP and AZ-SNAP) and the substrate modified ODNs, ODN-ZF-BC and ODN-AZ-BC, respectively. Each 5'-³²P-end-labeled ODN (0.5 nM) (ODN-ZF-BC or ODN-AZ-BC) was incubated with a modular adaptor (100 nM : ZF-SNAP or AZ-SNAP) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, and 100 nM calf thymus DNA at ambient temperature.

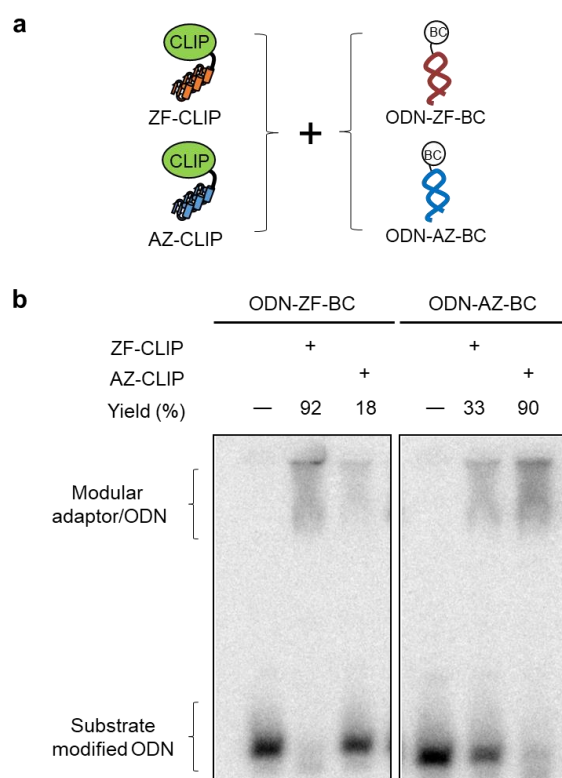


Figure 3.9 (a) Analyses of crosslinking reactions between the modular adaptors (ZF-CLIP and AZ-CLIP) and the substrate (BC) modified ODNs (ODN-ZF-BC and ODN-AZ-BC). (b) Denaturing PAGE analyses of the crosslinking reaction by modular adaptors, ZF-SNAP and AZ-SNAP, and the substrate (BC) modified ODNs, ODN-ZF-BC and ODN-AZ-BC, respectively. Each 5'-³²P-end-labeled ODN (0.5 nM) (ODN-ZF-BC or ODN-AZ-BC) was incubated with a modular adaptor (100 nM : ZF-SNAP or AZ-SNAP) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature.

The reactions of unmatched pairs for each combination of the protein tag and the substrate were evaluated next (Figures 3.8 and 3.9). The apparent second-order rate constants for the reaction of unmatched pairs of ZF-CLIP and AZ-CLIP were two orders lower than those of the matched pairs (Table 3.4). The observed differences in the apparent second-order rate constants were parallel to the difference in the dissociation

constants for the matched and unmatched complexes of ZF-CLIP and AZ-CLIP as expected from the equation for case 2 (Figure 3.2b). These results indicate that the crosslinking reaction of the modular adaptors containing the CLIP-tag to the substrate BC on ODN proceeded in case 2 of Figure 3.2b. Because the larger k_{app} for the matched pair was preferred as long as it falls in case 2, the pair of CLIP-tag and BC was selected as the self-ligating protein tag domain for modular adaptor that would form a covalent linkage at specific sequence driven by the DNA recognition domain.

Three types of modular adaptors bearing the same self-ligating protein tag (CLIP-tag), but differed in the DNA binding domain (ZF-CLIP, AZ-CLIP, and CLIP-G) were tested for the sequence-selective crosslinking reaction with ODN modified with BC. CLIP-G consisted of a basic-leucine zipper (bZIP) class of protein GCN4 as a homodimeric DNA binding module and the CLIP-tag fused to the N-terminus of GCN4 through a GGSGGS linker. An ODN containing AP1 sequence as the GCN4 binding site was designed for the matched target for CLIP-G, in which two T nucleotides at either end of AP1 sequence were displaced by amino-C6-T (ODN-AP), ODN-AP was further modified with BC (ODN-AP-2BC) as the substrate of CLIP-G (Figure 3.10).

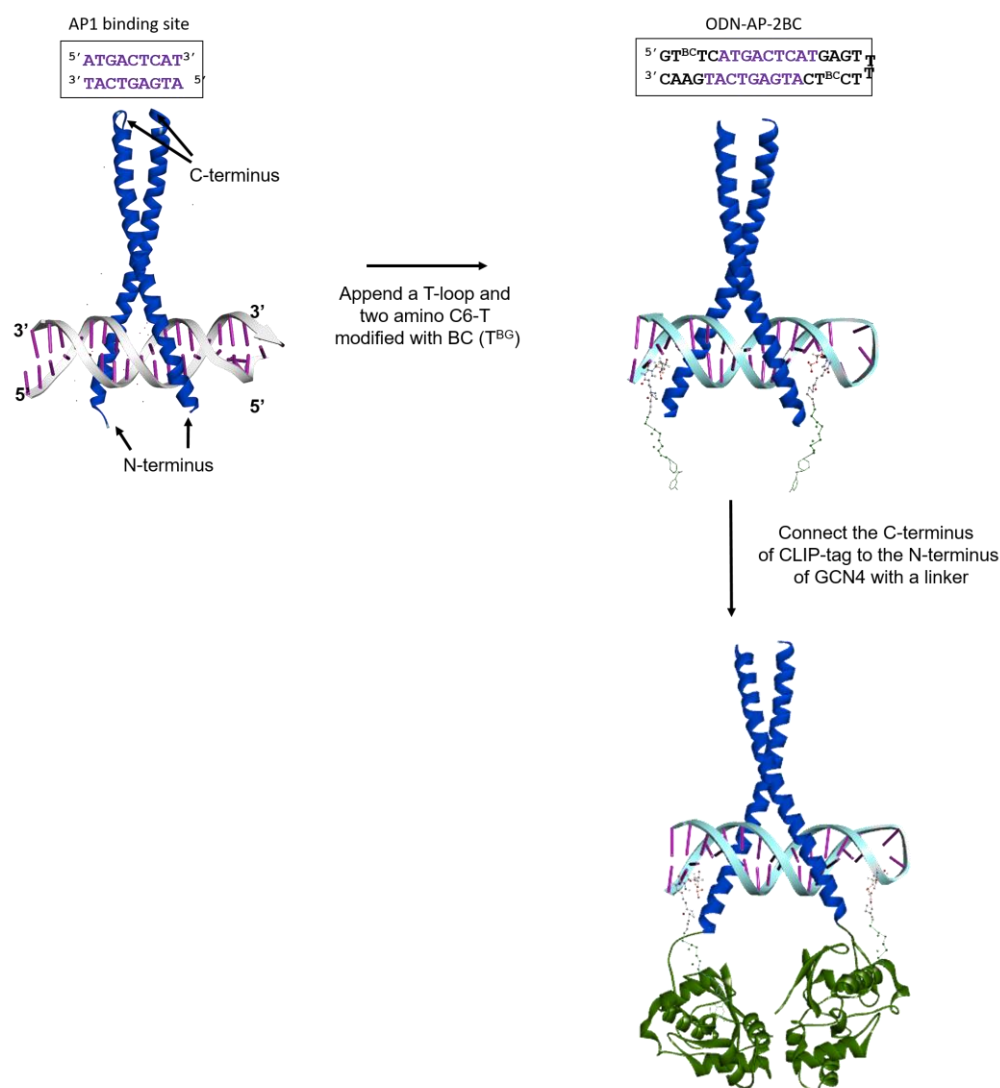


Figure 3.10 A scheme illustrates the design of CLIP-G and ODN-AP-2BC based on the crystal structures of the complex between GCN4 and DNA containing AP1 sequence (PDB ID : 1DGC) and SNAP-tag (PDB ID : 3KZY). Two BCs were appended on ODN-AP with 10 bases inter-distance, which allow access of the introduced BCs to the CLIP-tag of modular adaptor CLIP-G. The molecular model was constructed by using Discovery Studio (version 3.1, Accelrys Inc.).

Table 3.4. Kinetic parameters for the cross-linking reactions between 5'-³²P-end-labeled ODN derivatives and modular adaptors. (bold: Rate constants for the matched pair (k_{app}), plane: Rate constants for the unmatched pair (k'_{app}))

Rate constant k_{app} ($M^{-1}s^{-1}$)	ZF-CLIP	AZ-CLIP	CLIP-G
ODN-ZF-BC	$(4.0 \pm 0.5) \times 10^6$	$(8.1 \pm 2.7) \times 10^2$	$(1.7 \pm 0.4) \times 10^3$
ODN-AZ-BC	$(5.3 \pm 0.7) \times 10^3$	$(5.0 \pm 0.7) \times 10^5$ *	$(5.61 \pm 0.4) \times 10^3$
ODN-AP-2BC	$(7.0 \pm 0.9) \times 10^2$	$(5.8 \pm 1.5) \times 10^2$	$(2.1 \pm 0.1) \times 10^6$

*The value was taken from Chapter 2.

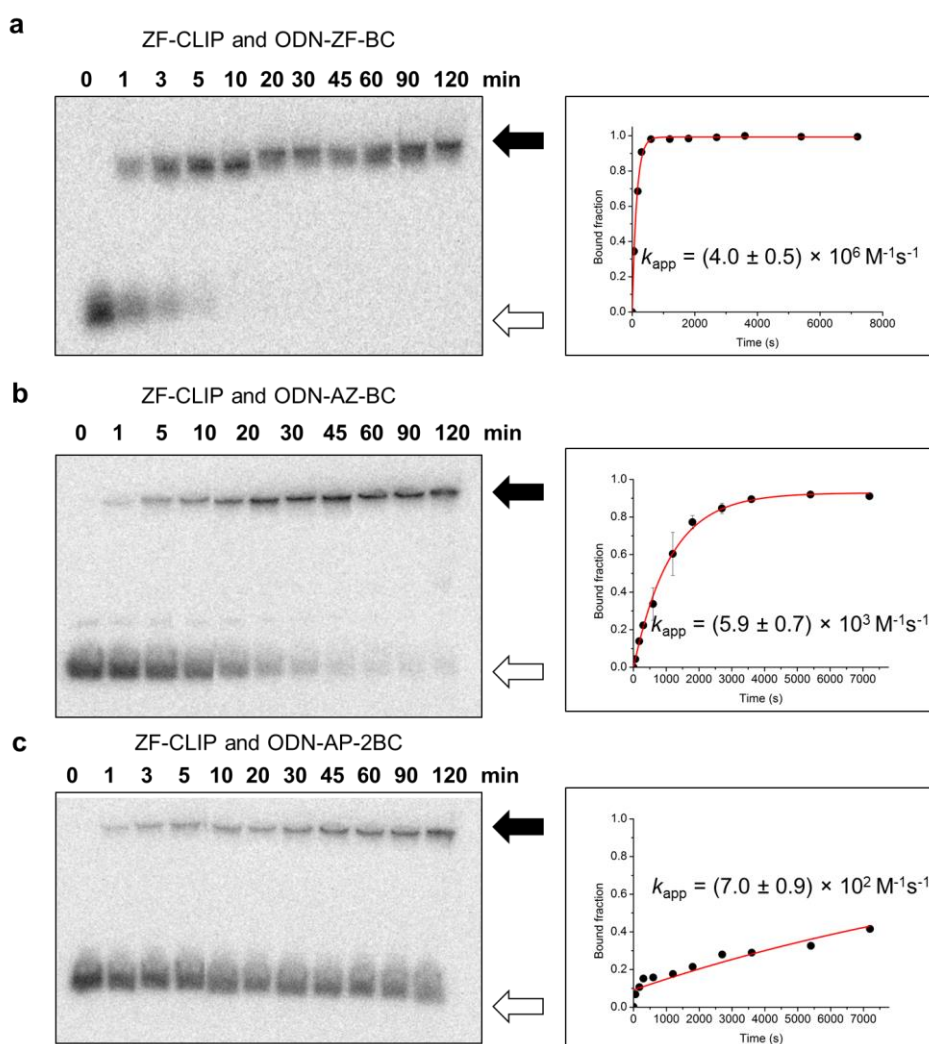


Figure 3.11 Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC (a), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC (b), and 0.5 nM 5'-³²P-end-labeled ODN-AP-2BC (c) with 10 nM ZF-CLIP (a), 100 nM ZF-CLIP (b, c) respectively. Open arrows and filled arrows denote the free ODNs and

covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (k) (right). The estimated rate constants are listed in Table 3.4. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl_2 , 1 mM DTT, 1 μM ZnCl_2 , 0.02% Tween 20, 200 nM BSA, and 100 nM calf thymus DNA at ambient temperature.

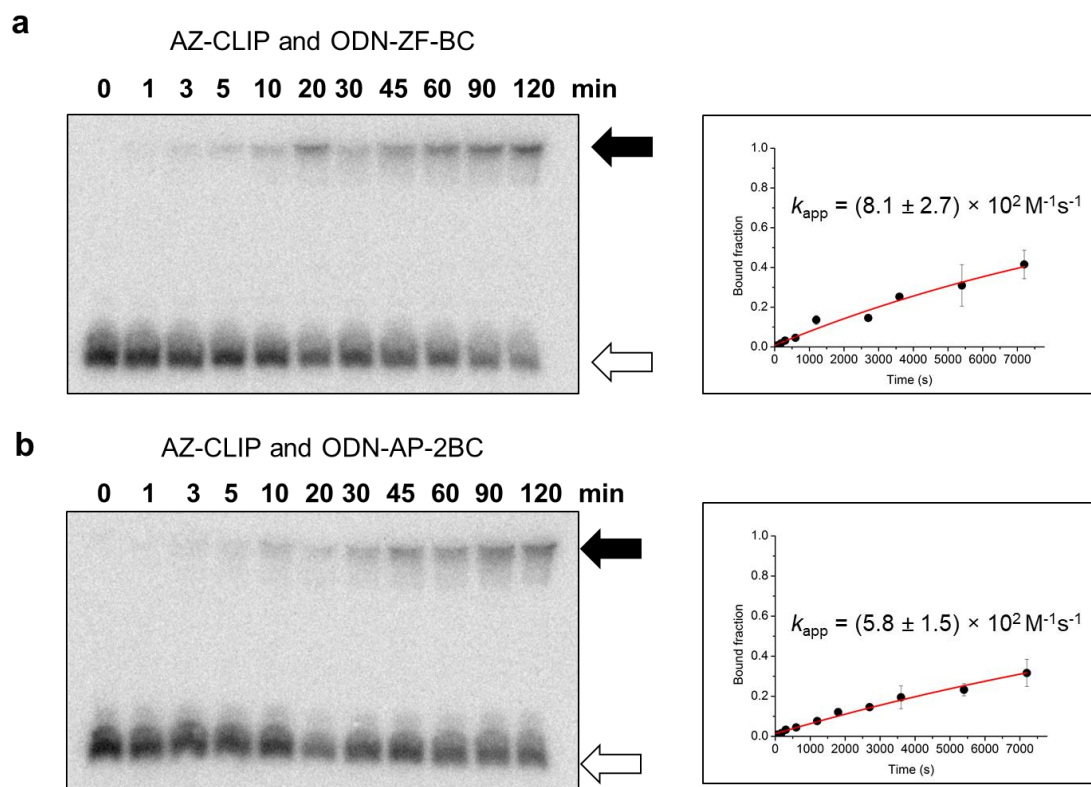


Figure 3.12 Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'- ^{32}P -end-labeled ODN-ZF-BC (a), and 0.5 nM 5'- ^{32}P -end-labeled ODN-AP-2BC (b) with 100 nM AZ-CLIP, respectively. Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plot for the crosslinking reactions to estimate the rate constants (k) (right). The estimated rate constants are listed in Table 3.4. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl_2 , 1 mM DTT, 1 μM ZnCl_2 , 0.02% Tween 20, 200 nM BSA, and 100 nM calf thymus DNA at ambient temperature.

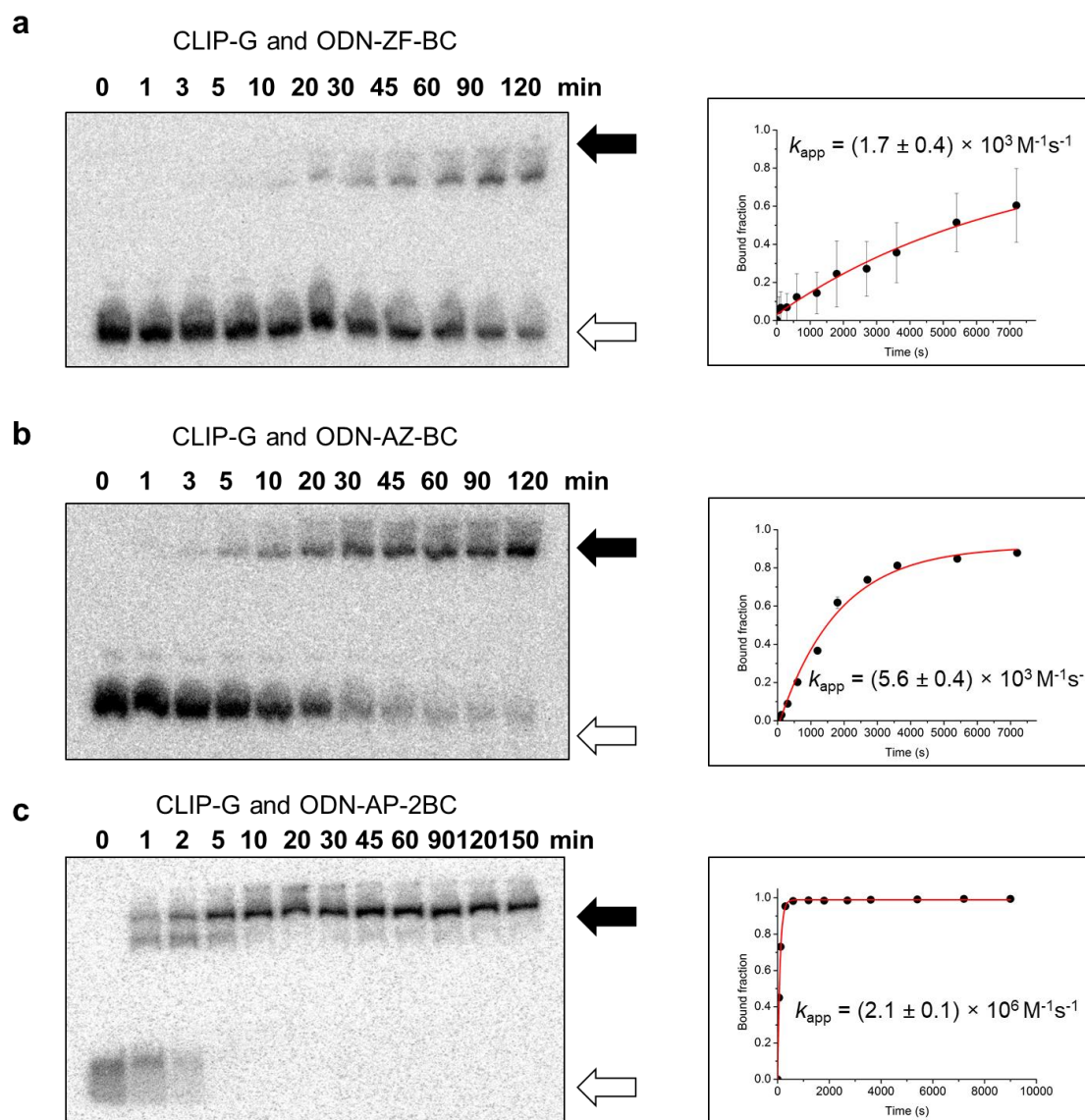


Figure 3.13 Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC (a), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC (b), and 0.5 nM 5'-³²P-end-labeled ODN-AP-2BC (c) with 100 nM CLIP-G, respectively. Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (*k*) (right). The estimated rate constants are listed in Table 3.4. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, and 100 nM calf thymus DNA at ambient temperature.

The cross-linking reactions of ZF-CLIP, AZ-CLIP, and CLIP-G to each respective matched ODN were nearly orthogonal as shown in Figure 3.14b with more than 90% yield for the matched pairs and less than 12% covalent bond formation for the unmatched pairs. The observed sequence-specific cross-linking reactions by ZF-CLIP, AZ-CLIP, and CLIP-G were driven by the specific DNA binding of each modular adaptor (Table 3.4). The second-order rate constant (k_{app}) for the reaction of CLIP-G and its matched sequence ODN-AP-2BC was determined to be $2.1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and those of the unmatched pairs (k'_{app}) were ranged from 10^3 to $10^4 \text{ M}^{-1} \text{ s}^{-1}$. For the reactions of unmatched pairs, ZF-CLIP or AZ-CLIP with ODN-AP-2BC, k'_{app} were ranged from 10^2 to $10^3 \text{ M}^{-1} \text{ s}^{-1}$.

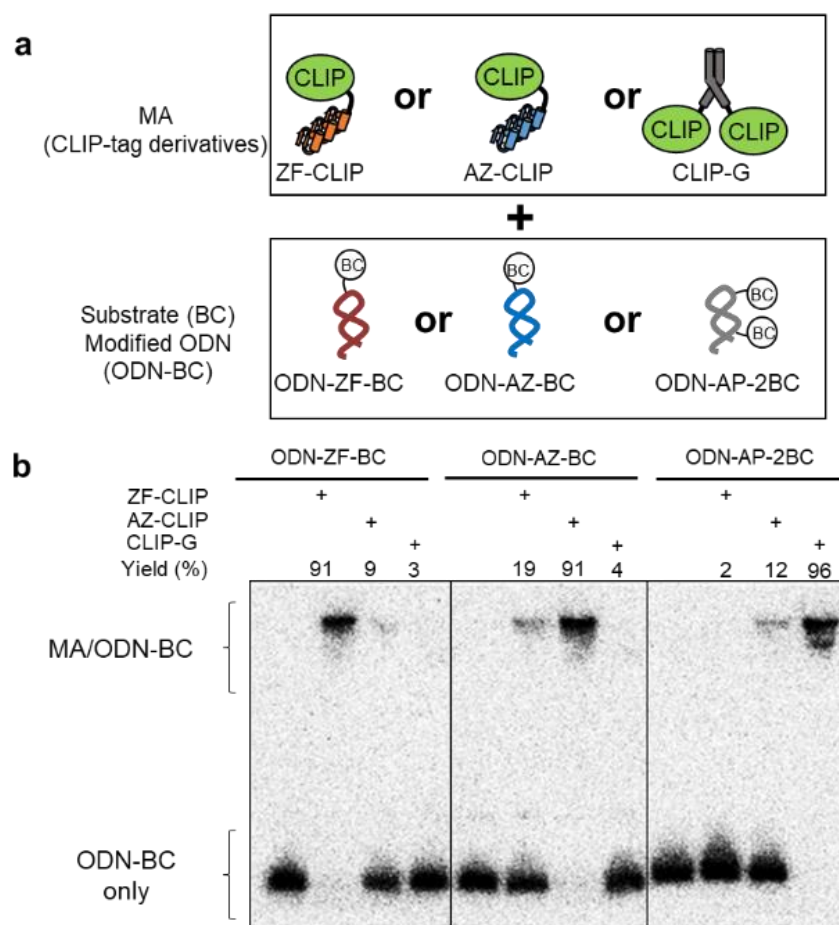


Figure 3.14 (a) Illustrations of the optimized sequence-specific modular adaptors (MAs) and the substrate modified ODNs. (b) Autoradiogram show denaturing PAGE analyses of the

crosslinking reactions by modular adaptor (ZF-CLIP, AZ-CLIP and CLIP-G) and substrate-modified ODN (ODN-ZF-BC, ODN-AZ-BC, and ODN-AP-2BC). Each 5'-³²P-end-labeled substrate-modified ODN (0.5 nM) was incubated with 50 nM each of modular adaptors. These reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, and 0.02% Tween20 at ambient temperature.

In the presence of 200 mM NaCl, the non-specific DNA binding was reduced¹⁵ (Figure 3.15, Table 3.6, 3.7) and the covalent bond formations for the unmatched pairs were further reduced to less than 5% (Figure 3.15), while the yields for the matched pairs were reduced for AZ-CLIP (64%) and ZF-CLIP (88%) although the higher concentrations of AZ-CLIP and ZF-CLIP were used in this study (Figure 3.15). The kinetic constants of the matched and unmatched pairs in the presence of 200 mM NaCl were shown in Table 3.5 for modular adaptors containing SNAP-tag and Table 3.6 for modular adaptors containing CLIP-tag. The k_{app} of the reactions in the presence of 200 mM NaCl also indicated the dramatically reducing of the activity of AZ-CLIP (Table 3.6). Table 3.6 showed the difference of the k_{app} and k'_{app} was almost three order of magnitudes, therefore, is it possible to increase the concentration of modular adaptors or the incubation time to increase the yield at the matched pairs with low reactivity at the unmatched pairs.

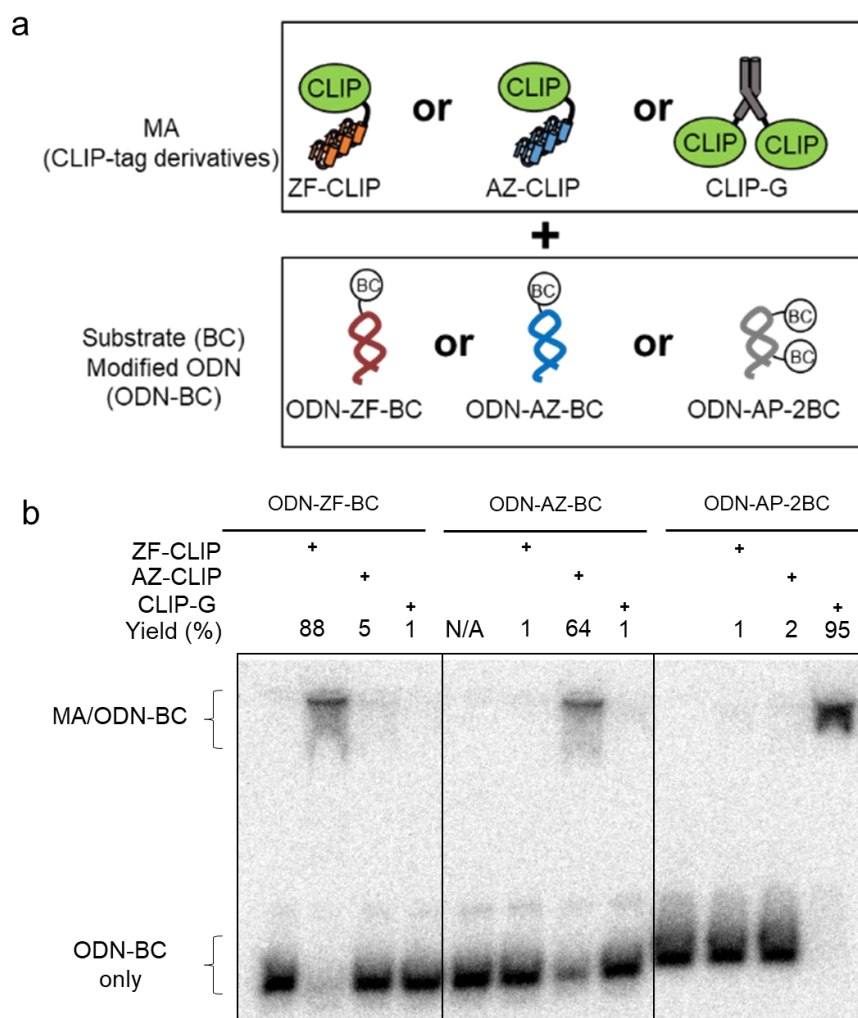


Figure 3.15 (a) Illustrations of the optimized sequence-specific modular adaptors (MAs) and the substrate modified ODNs. (b) Autoradiogram show denaturing PAGE analyses of the crosslinking reactions by modular adaptor (ZF-CLIP, AZ-CLIP, and CLIP-G) and substrate-modified ODN (ODN-ZF-BC, ODN-AZ-BC, and ODN-AP-2BC). Each 5'-³²P-end-labeled substrate-modified ODN (0.5 nM) was incubated in 60 min with 200 nM ZF-CLIP, 500 nM AZ-CLIP, and 100 nM CLIP-G. These reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween20, and 200 mM NaCl at ambient temperature.

Table 3.5 Kinetic parameters for the cross-linking reactions between 5'-³²P-end-labeled ODN derivatives modified with BG and modular adaptors containing SNAP-tag in the presence of 200 mM NaCl. (bold: Rate constants for the matched pair (k_{app}), plane: Rate constants for the unmatched pair (k'_{app})).

Rate constant ($M^{-1} s^{-1}$)	ZF-SNAP	AZ-SNAP
ODN-ZF-BG	$(1.3 \pm 0.2) \times 10^4$	$(8.5 \pm 0.3) \times 10^2$
ODN-AZ-BG	$(6.1 \pm 0.3) \times 10^2$	$(2.2 \pm 0.2) \times 10^4$

Table 3.6 Kinetic parameters for the cross-linking reaction between 5'-³²P-end-labeled ODN derivatives modified with BC and modular adaptors containing CLIP-tag in the presence of 200 mM NaCl. (bold: Rate constant for the matched pair (k_{app}), plane: Rate constant for the unmatched pair (k'_{app})).

Rate constant k_{app} ($M^{-1} s^{-1}$)	ZF-CLIP	AZ-CLIP	CLIP-G
ODN-ZF-BC	$(3.1 \pm 0.2) \times 10^3$	< 10	< 10
ODN-AZ-BC	< 10	$(3.1 \pm 0.2) \times 10^2$	< 10
ODN-AP-2BC	< 10	< 10	$(1.8 \pm 0.3) \times 10^4$

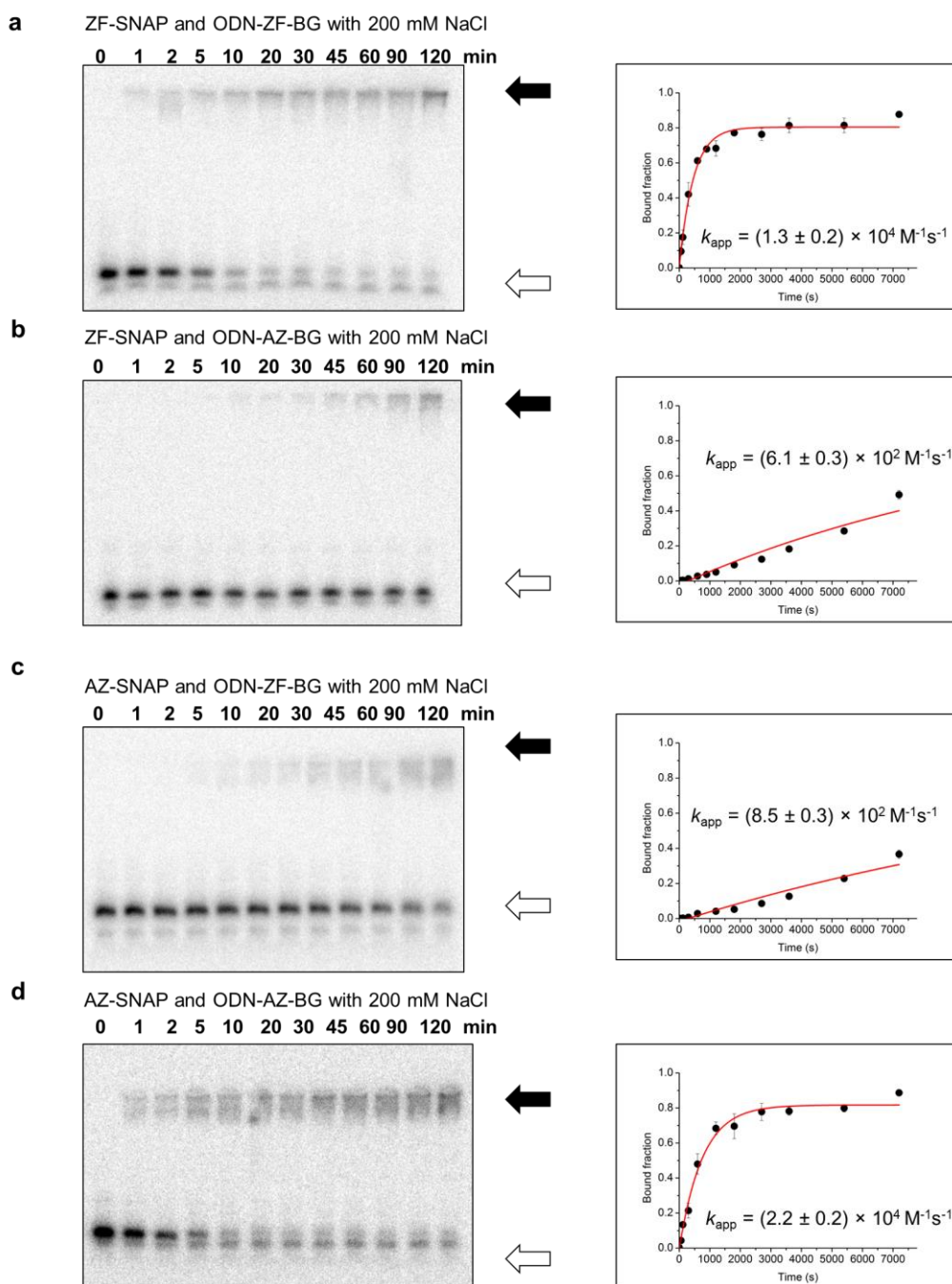


Figure 3.16 Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BG (a, c), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BG (b, d) with ZF-SNAP (a, b), and AZ-SNAP (c, d). Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (**k**) (right). The estimated rate constants are listed in Table 3.5. Reaction condition: buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, 100 nM calf thymus DNA, and 200 mM NaCl at ambient temperature.

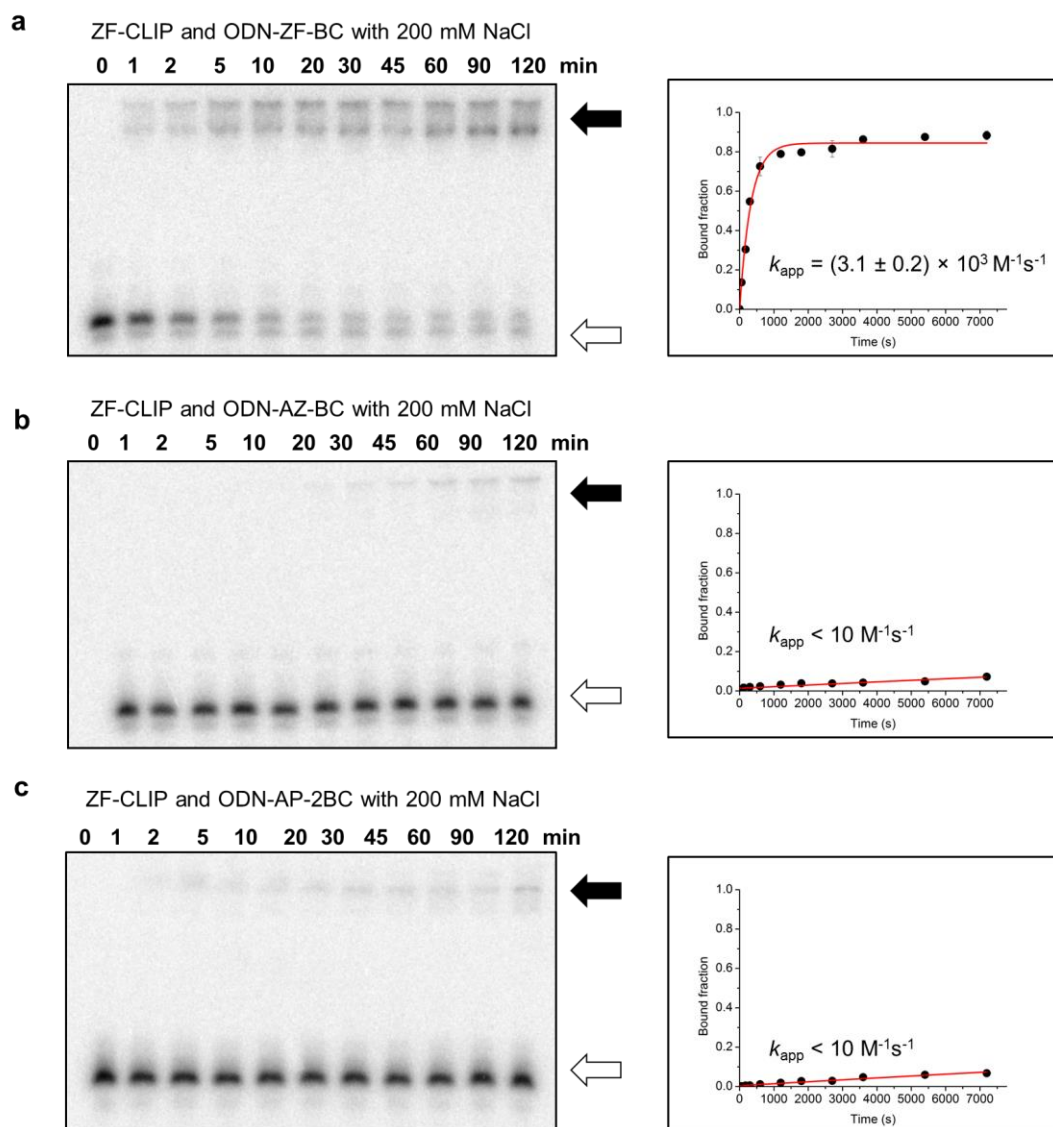


Figure 3.17. Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC (a), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC (b), and 0.5 nM 5'-³²P-end-labeled ODN-AP-2BC (c) with 100 nM ZF-CLIP, respectively. Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (*k*) (right). The estimated rate constants are listed in Table 3.6. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, 100 nM calf thymus DNA, and 200 mM NaCl at ambient temperature.

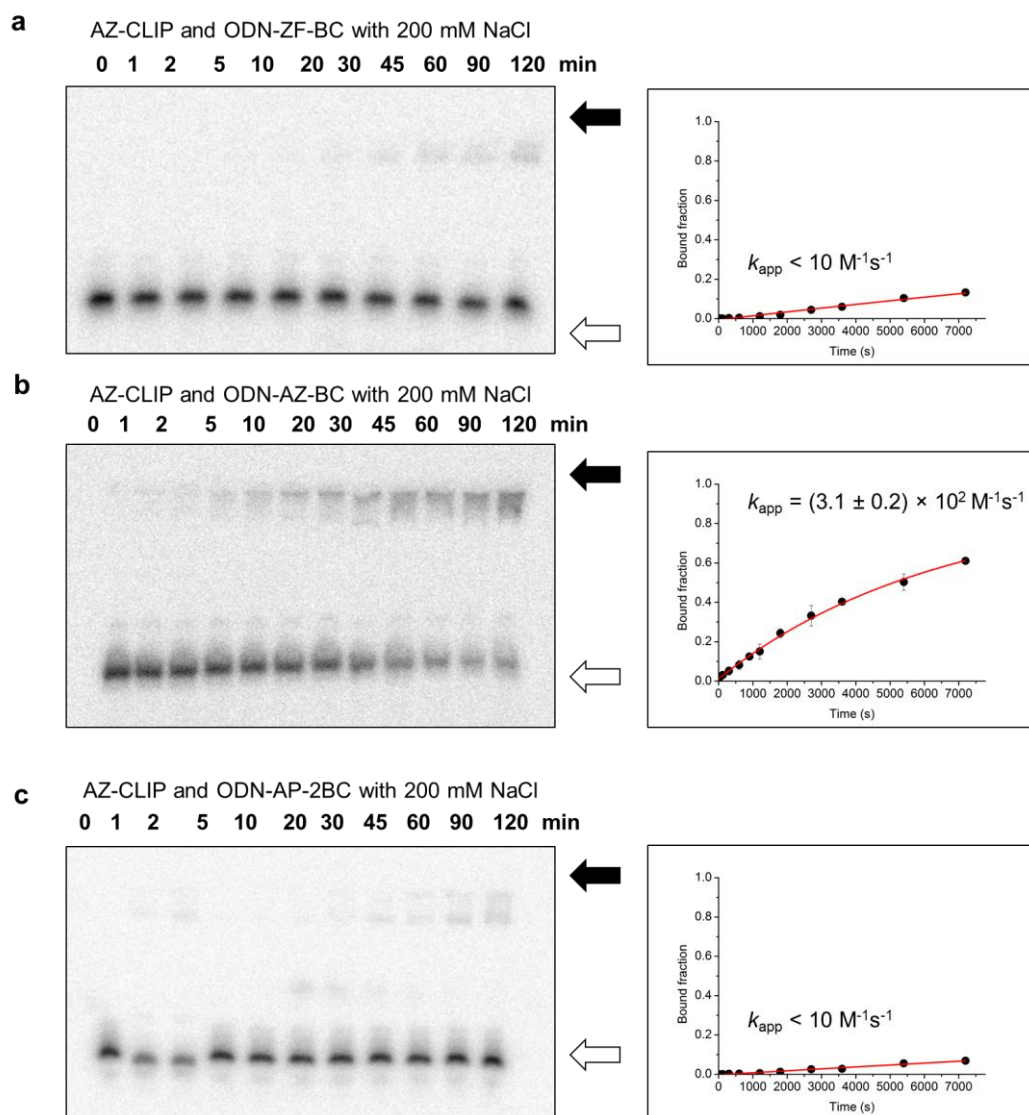


Figure 3.18. Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC (a), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC (b), and 0.5 nM 5'-³²P-end-labeled ODN-AP-2BC (c) with 100 nM AZ-CLIP, respectively. Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (*k*) (right). The estimated rate constants are listed in Table 3.6. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, 200 nM BSA, 100 nM calf thymus DNA, and 200 mM NaCl at ambient temperature.

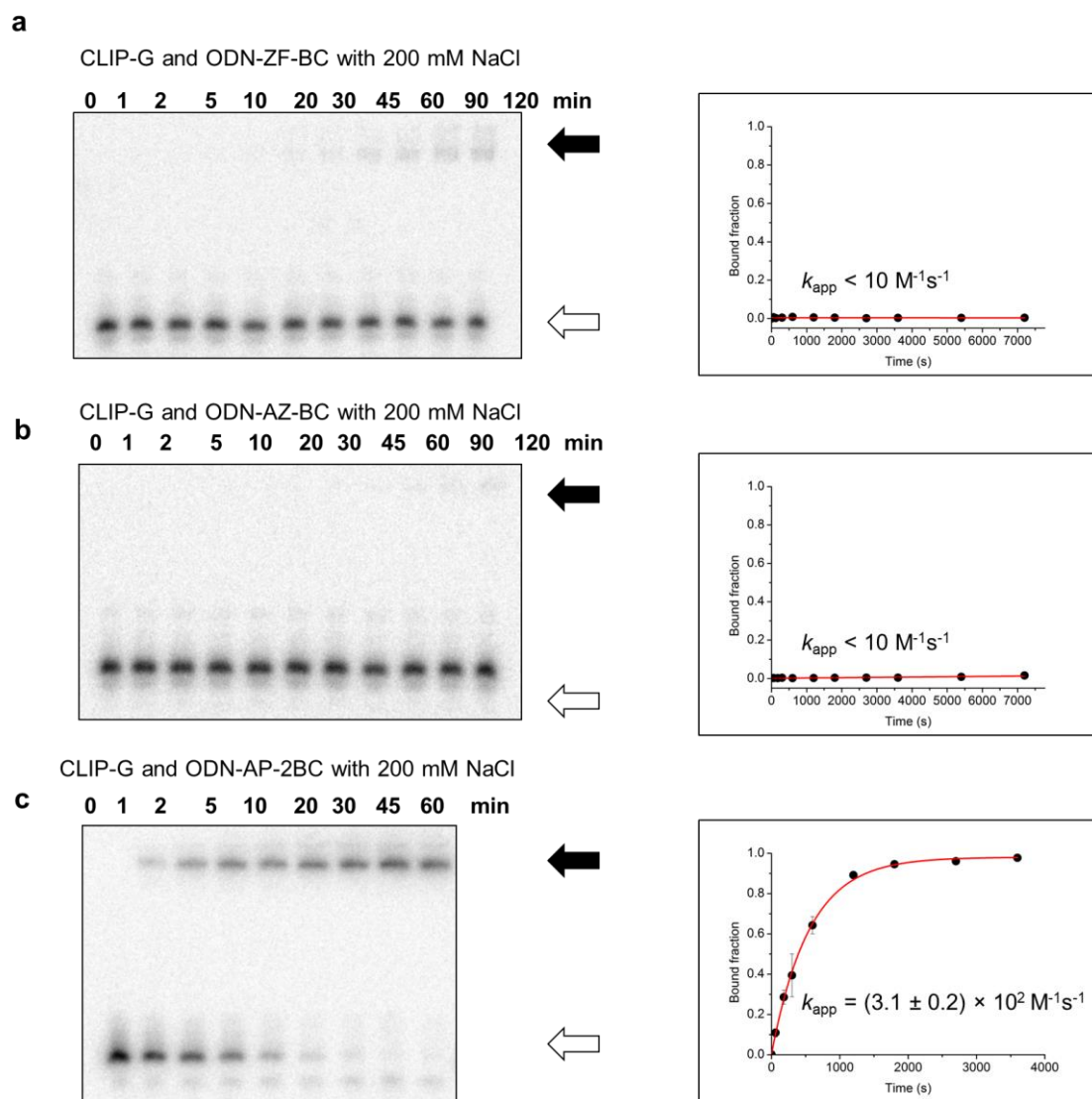


Figure 3.19. Autoradiograms show denaturing gel electrophoretic analyses of the cross-linking reactions of 0.5 nM 5'-³²P-end-labeled ODN-ZF-BC (a), 0.5 nM 5'-³²P-end-labeled ODN-AZ-BC (b), and 0.5 nM 5'-³²P-end-labeled ODN-AP-2BC (c) with 100 nM CLIP-G, respectively. Open arrows and filled arrows denote the free ODNs and covalent complexes with modular adaptors, respectively (left). Time-course plots for the crosslinking reactions to estimate the rate constants (k) (right). The estimated rate constants are listed in Table 3.6. Reaction conditions: 40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 0.02% Tween 20, 200 nM BSA, 100 nM calf thymus DNA, and 200 mM NaCl at ambient temperature.

The dissociation constants of the modular adaptors and three ^{32}P -labelled-ODNs (ODN-ZF, ODN-AZ, and ODN-AP) were studied by using gel mobility shift assay (Figure 3.20-3.25) and listed in Table 3.7 (in the absence of 200 mM NaCl), and Table 3.8 (in the presence of 200 mM NaCl). The results indicated the increasing of K_D when 200 mM NaCl was used, however, the increasing of the obtained K_D was not significant when compared with the decreasing of k_{app} . (Table 3.1 and Table 3.7, Table 3.4 and Table 3.8). The reason of the irrelevant of the obtained K_D and k_{app} may due to the inaccurate estimation of K_D by using the gel mobility shift assay, the concentration of NaCl was diluted during loading and running the PAGE analyses. Therefore, fluorescence polarization was used to study K_D of the modular adaptors and the fluorophore modified ODNs. Fluorescence polarization also indicated enhanced specificity for the matched and matched pairs in the presence of 200 mM NaCl (Figure 3.26, Table 3.9, 3.10). Although the titration of concentration-dependent of binding the modular adaptors to the fluorophore modified ODNs did not reach the saturation point indicating the complete formation of the modular adaptor/ODN complex (the highest possible concentration of the modular adaptor in the reaction was only 1024 nM) (Figure 3.26), the assay showed the dissociation constants of the modular adaptors and matched and unmatched ODNs were increased orders of magnitudes in the presence of 200mM NaCl. These results indicated that the rates of covalent bond formation (k_{app} and k'_{app}) depended on K_D of the complexes between modular adaptors and ODNs as in case 2 of Figure 3.2b.

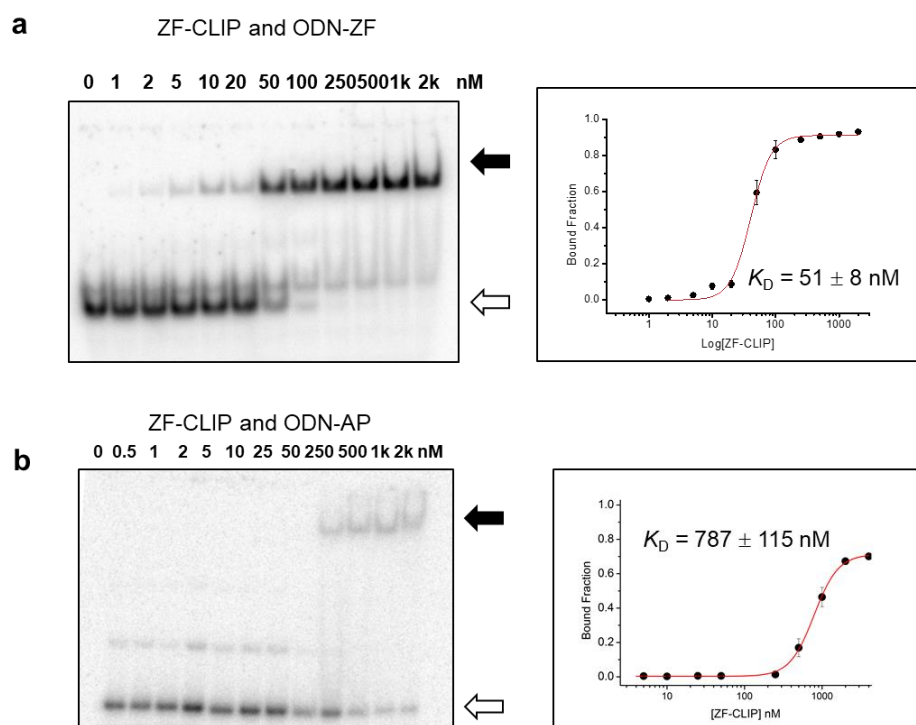


Figure 3.20 Autoradiograms show the electrophoretic mobility shift titration of ZF-CLIP to (a) 0.05nM 5'-³²P-labeled ODN-AZ and (b) 0.05nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrows and filled arrows denote free ODN and modular adaptor bound ODN, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.7

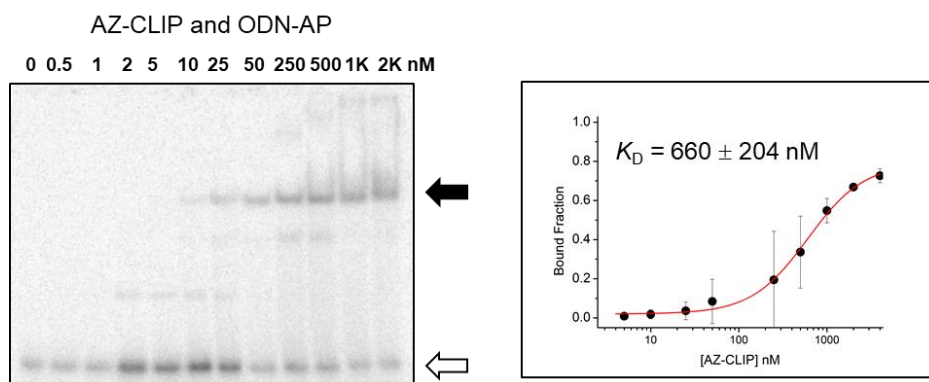


Figure 3.21 An autoradiogram shows the electrophoretic mobility shift titration of AZ-CLIP to 0.05 nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrows and filled arrows denote free ODN-AP and modular adaptor bound ODN-AZ, respectively. A semilogarithmic plot shows the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.7

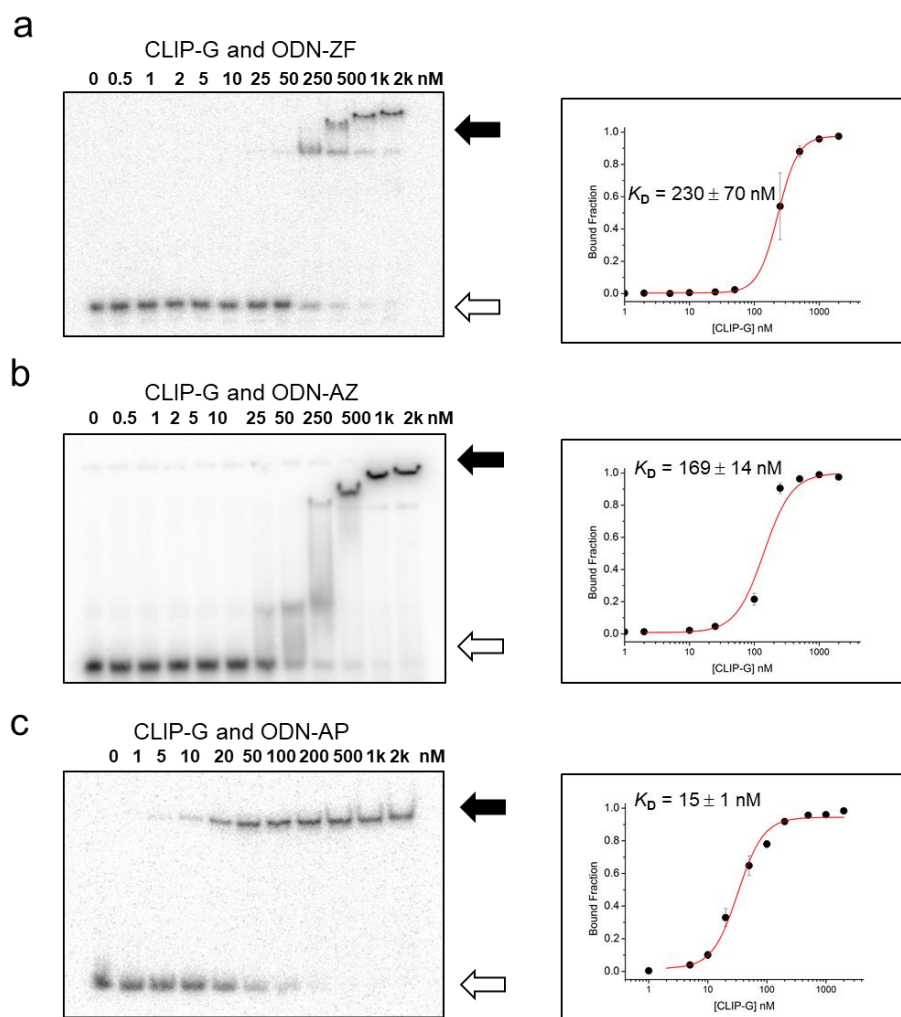


Figure 3.22 Autoradiograms show the electrophoretic mobility shift titration of CLIP-G to (a) 0.05 nM 5'-³²P-labeled ODN-ZF, (b) 0.05 nM 5'-³²P-labeled ODN-AZ, and (c) 0.05nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrows and filled arrows denote free ODN and modular adaptor bound ODN, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.7

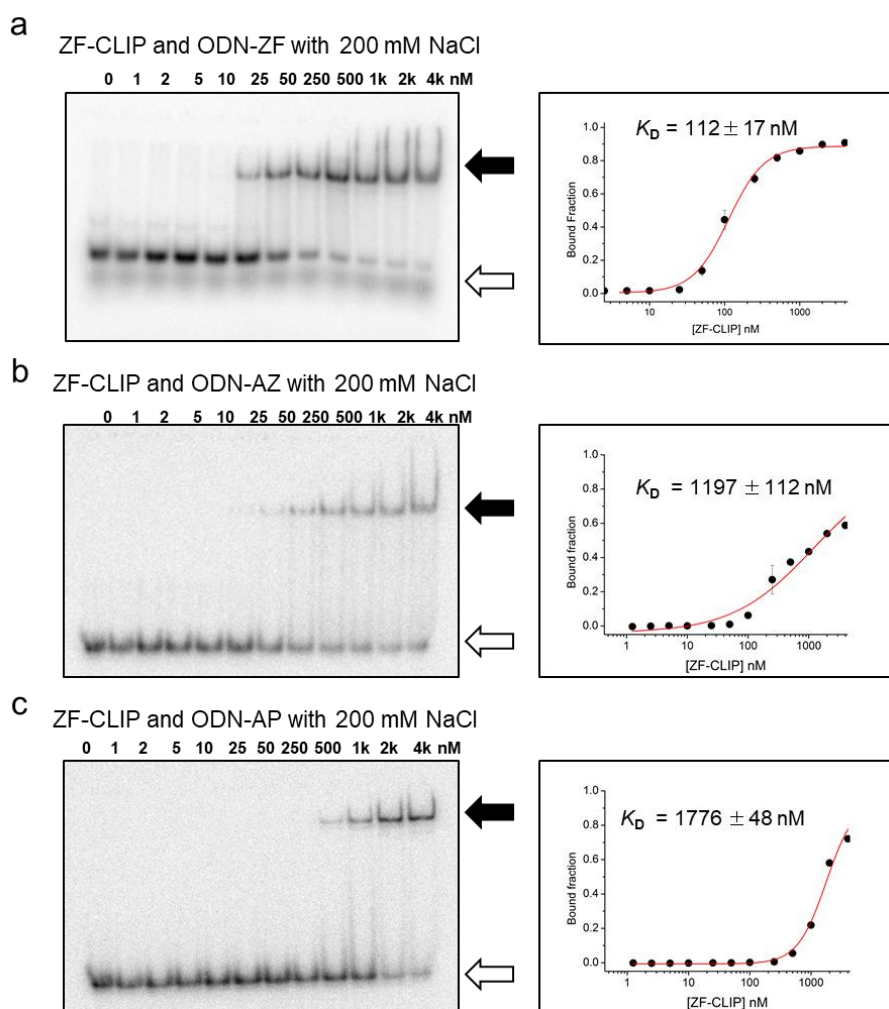


Figure 3.23 Autoradiograms show the electrophoretic mobility shift titration of ZF-CLIP to (a) 0.05nM 5'-³²P-labeled ODN-ZF, (b) 0.05 nM 5'-³²P-labeled ODN-AZ, and (c) 0.05nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, 200 nM BSA, and 200 mM NaCl at ambient temperature (left). Open arrows and filled arrows denote free ODN and modular adaptor bound ODN, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.8

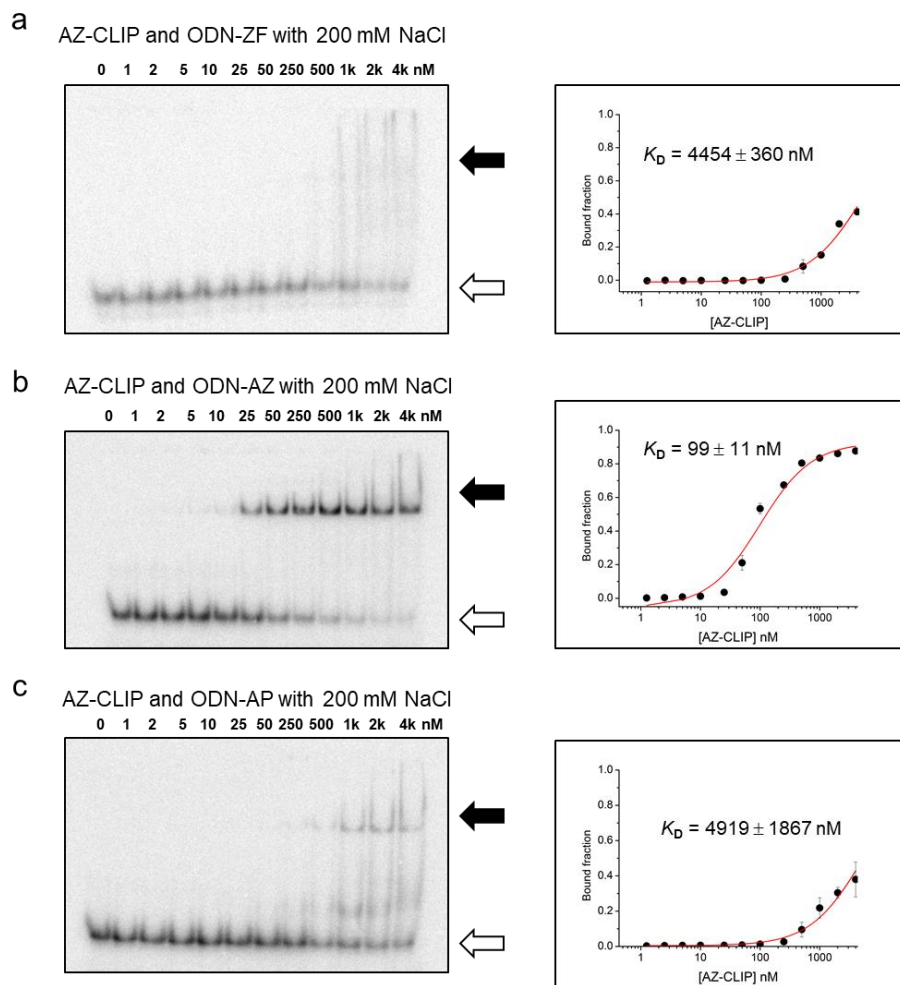


Figure 3.24 Autoradiograms show the electrophoretic mobility shift titration of AZ-CLIP to (a) 0.05 nM 5'-³²P-labeled ODN-ZF, (b) 0.05 nM 5'-³²P-labeled ODN-AZ, and (c) 0.05 nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, 200 nM BSA, and 200 mM NaCl at ambient temperature (left). Open arrows and filled arrows denote free ODN and modular adaptor bound ODN, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.8

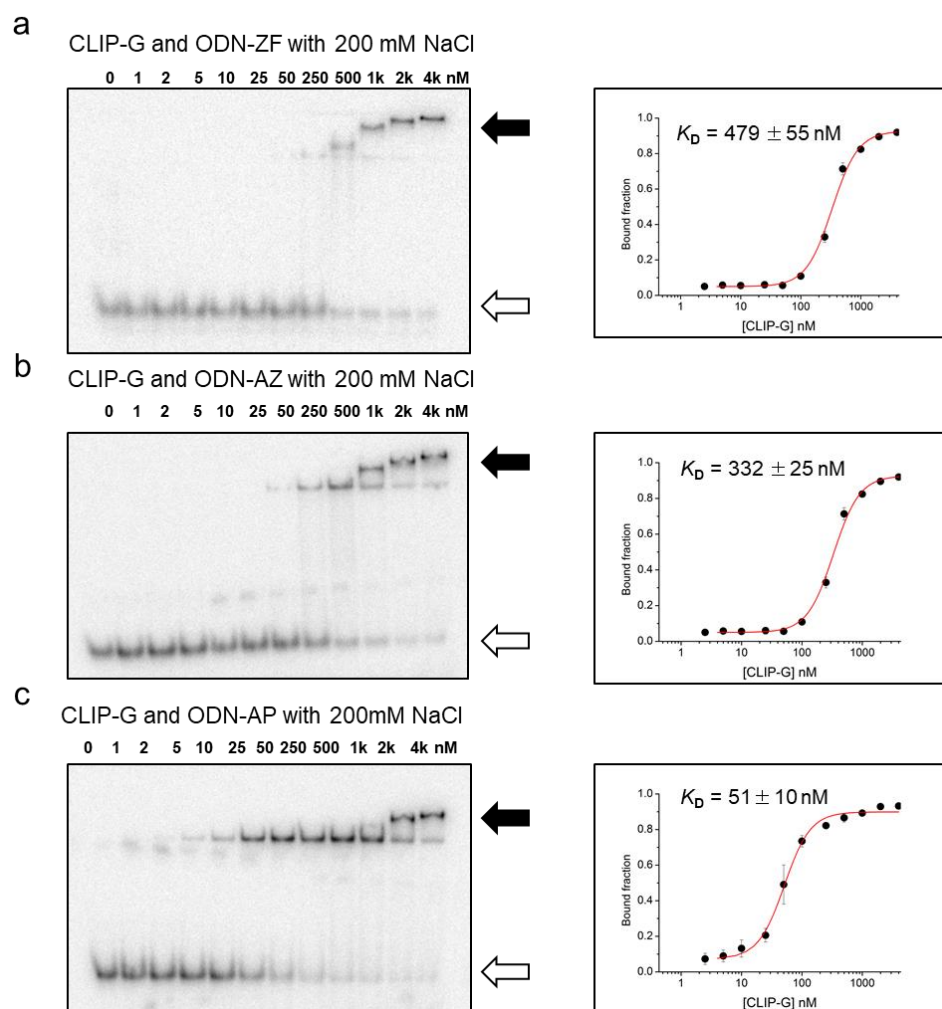


Figure 3.25 Autoradiograms show the electrophoretic mobility shift titration of CLIP-G to (a) 0.05 nM 5'-³²P-labeled ODN-ZF, (b) 0.05 nM 5'-³²P-labeled ODN-AZ, and (c) 0.05 nM 5'-³²P-labeled ODN-AP in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1mM DTT, 1μM ZnCl₂ 0.02 % Tween 20, 200 nM BSA, and 200 mM NaCl at ambient temperature (left). Open arrows and filled arrows denote free ODN and modular adaptor bound ODN, respectively. Semilogarithmic plots show the fractions of 5'-³²P-labeled ODN bound to modular adaptors (right). The estimated equilibrium constants are listed in Table 3.8

Table 3.7 Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs determined by gel-shift assay. (bold: K_D for the matched pair, plane: K_D for the unmatched pair).

Recognition module (without NaCl)	K_D (nM)		
	ODN-ZF	ODN-AZ	ODN-AP1
Zif268 (ZF)	$63 \pm 18^*$	$> 1000^*$	787 ± 115
AZP4 (AZ)	$> 1000^*$	$61 \pm 30^*$	660 ± 240
GCN4 (G)	231 ± 70	169 ± 14	15 ± 1

* The values were taken from Chapter 2

Table 3.8 Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs determined by titration of gel-shift assay. (bold: K_D for the matched pair, plane: K_D for the unmatched pair).

Recognition module (200 mM NaCl)	K_D (nM)		
	ODN-ZF	ODN-AZ	ODN-AP1
Zif268 (ZF)	112 ± 17	4454 ± 360	1776 ± 48
AZP4 (AZ)	1197 ± 112	99 ± 11	4919 ± 1867
GCN4 (G)	479 ± 55	332 ± 25	51 ± 10

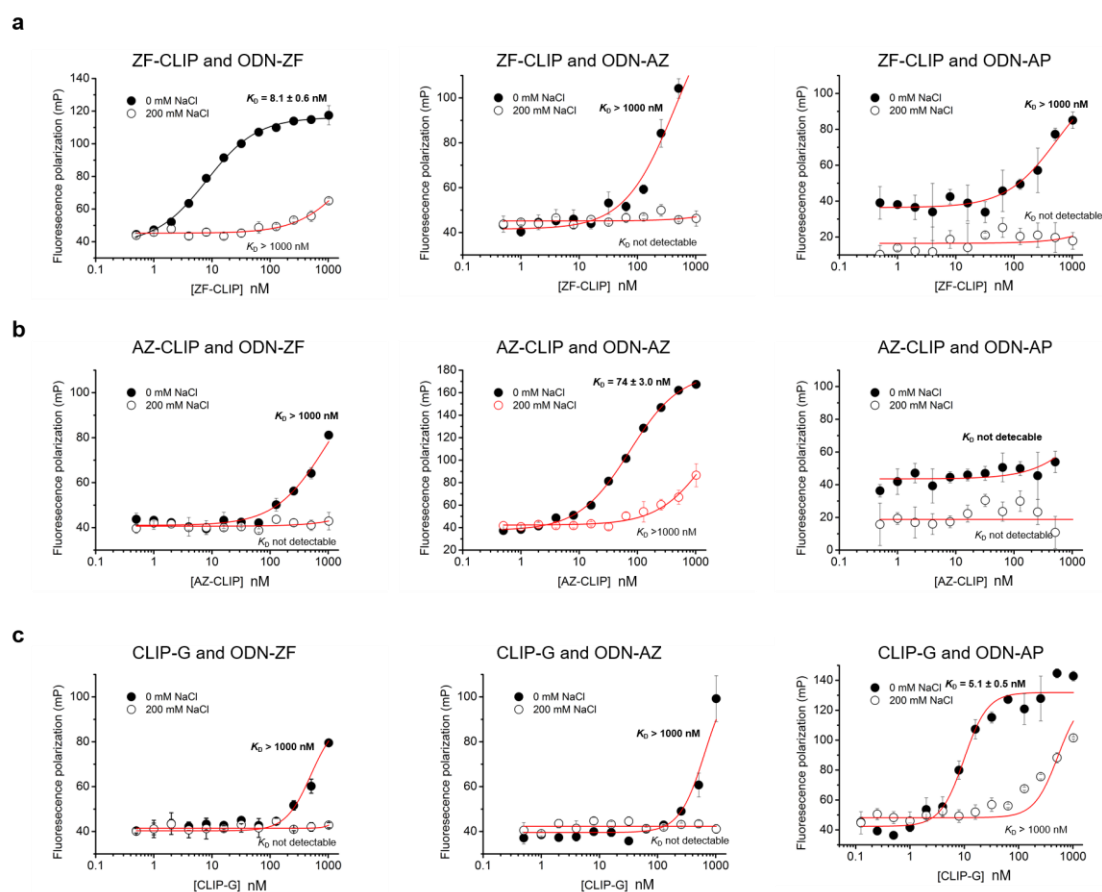


Figure 3.26 Fluorescence polarization analyses for study of the binding of modular adaptors (a) ZF-CLIP, (b) AZ-CLIP and (c) CLIP-G to ODN-ZF, ODN-AZ, and ODN-AP, without 200 mM NaCl (filled circle) and with 200 mM NaCl (open circle). Reaction was conducted in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μ M $ZnCl_2$, 100 nM calf thymus DNA and 0.02% Tween20, and 0.1 nM fluorophore modified ODN with various concentrations of modular adaptor (0.25 nM ~ 1024 nM).

Table 3.9 Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs determined by titration of fluorescent polarization. (bold: K_D for the matched pair, plane: K_D for the unmatched pair).

Recognition module (without NaCl)	K_D (nM)		
	ODN-ZF	ODN-AZ	ODN-AP
Zif268 (ZF)	8.1 ± 0.6	> 1000	> 1000
AZP4 (AZ)	> 1000	74 ± 3.0	n.d.
GCN4 (G)	> 1000	> 1000	5.1 ± 0.5
n.d. not detectable			

Table 3.10 Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs determined by titration of fluorescent polarization in the presence of 200 mM NaCl. (bold: K_D for the matched pair, plane: K_D for the unmatched pair).

Recognition module (200 mM NaCl)	K_D (nM)		
	ODN-ZF	ODN-AZ	ODN-AP
Zif268 (ZF)	> 1000	n.d.	n.d.
AZP4 (AZ)	n.d.	> 1000	n.d.
GCN4 (G)	n.d.	n.d.	> 1000
n.d. not detectable			

3.2.2 Orthogonality of the sequence-specific modular adaptors on DNA scaffold

Orthogonality was studied on a 3-well DNA scaffold (I-1CG/II-1ZC/III-1AC) containing a single binding site for CLIP-G, ZF-CLIP and AZ-CLIP at cavity I, II and III, respectively (Figure 3.27a). In this study, 50 nM each of the modular adaptors was incubation 5 min, 10 min and 15 min with a DNA scaffold (I-1CG/II-1ZC/III-1AC) containing three single binding sites for CLIP-G, ZF-CLIP, and AZ-CLIP at cavity I, II and III, respectively. At the matched addresses, all the three modular adaptors could be loaded efficiently in 5 min incubation with over 93% yield. At the unmatched addresses, CLIP-G show orthogonality with thin 15 min incubation with less than 6% yield,

AZ-CLIP show less than 8% yield at the unmatched addresses after 10min incubation. ZF-CLIP showed poor orthogonality toward the address designed for AZ-CLIP with the yield from 21% (5 min incubation) to 47% (15 min incubation) (Figure 3.27). The orthogonal study indicated the sequential assembly of these three modular adaptors (firstly assemble CLIP-G, follows by the assembly of AZ-CLIP, and ZF-CLIP at the last step), could realize the good orthogonality.

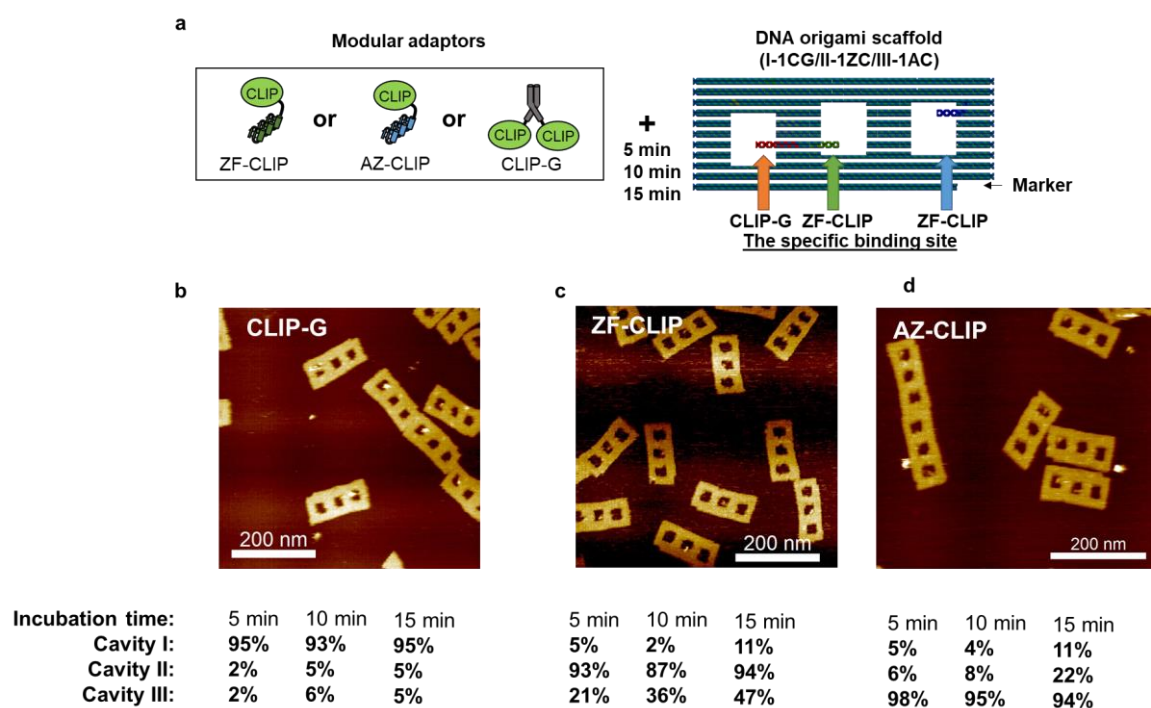


Figure 3.27. Orthogonal study of modular adaptors on DNA scaffold. 50 nM of eachof modular adaptors was incubation with the DNA scaffold (I-1CG/II-1ZC/III-1AC) contain three single binding sites for CLIP-G, ZF-CLIP and AZ-CLIP at the cavity I, II and III respectively at ambient temperature. After 5 min, 10 min and 15 min of incubation, an aliquot of the reaction mixture was purified by size-exclusion chromatography and then analyzed by AFM. Reaction was conducted in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20 at ambient temperature. Yields for the matched and the unmatched cross-linking reactions were estimated by counting the number of cavities occupied by the modular adaptors (Table 2.11).

We performed the sequential assembly of three modular adaptors on the same DNA scaffold (I-1CG/II-1ZC/III-1AC). Basically, the sequential assembly consisted of three 5 min incubation steps assembly of three modular adaptors on DNA scaffold. At the first step, the DNA scaffold (I-1CG/II-1ZC/III-1AC) was incubated with CLIP-G, after 5 min the reaction continues with AZ-CLIP, and finally, ZF-CLIP were added. The result is showed in Figure 3.28, the sequential assembly method allows me to maintain the high assembly yield at matched address and reduced the yield at the unmatched addressed for ZF-CLIP to below 11 %. The co-assembly at matched address is almost 90%. The modular adaptors system is fast enough to finish the sequential assembly of three modular adaptors with thin 15 min reaction, which is difficult to archive using other methods.

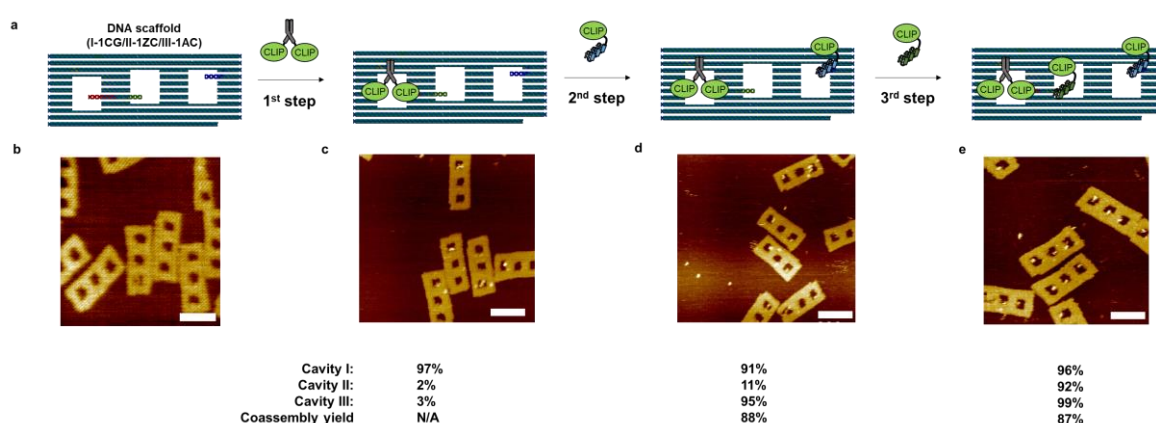


Figure 3.28. Sequential assembling of three modular adaptors on a DNA scaffold. (a) An illustration of the sequential assembly of modular adaptors on the DNA scaffold (I-1AH/II-1ZS/III-1AC) and (b–e) AFM images of the DNA scaffold reacted with the modular adaptors. The DNA scaffold (5 nM) was incubated (c) with CLIP-G (50 nM) for 5 min at ambient temperature (1st step), then (d) with the AZ-CLIP for additional 5 min (2nd step), and finally (e) with ZF-CLIP for further 5 min (3rd step). The reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20. At each step, an aliquot of the reaction mixture was purified by

size-exclusion chromatography and then analyzed by AFM. Yields at each cavity and coassembly yields after the 2nd step and 3rd step were estimated by counting the number of cavities occupied by the modular adaptors. The scale bar represents 100 nm.

Addition of 200 mM NaCl to the assembly solution show the significant decreasing of the assembly yield at the unmatched position (less than 7%), while the three modular adaptors were loaded to the matched positions over 90% yield in 30 min, (Figure 3.29). Interestingly, the assembly of AZ-CLIP to the design position of the DNA scaffold (Figure 3.30) was significantly faster than the covalent bond formation of AZ-CLIP to the BC modified ODN (Figure 3.15). Figure 3.30 showed the assembly of AZ-CLIP (250 nM) on DNA scaffold (5 nM) took only 20 min to reach to more than 90%. The larger DNA scaffold may allow the faster binding and searching for the specific sequence of the DNA binding protein module of the modular scaffold.³⁸⁻⁴¹

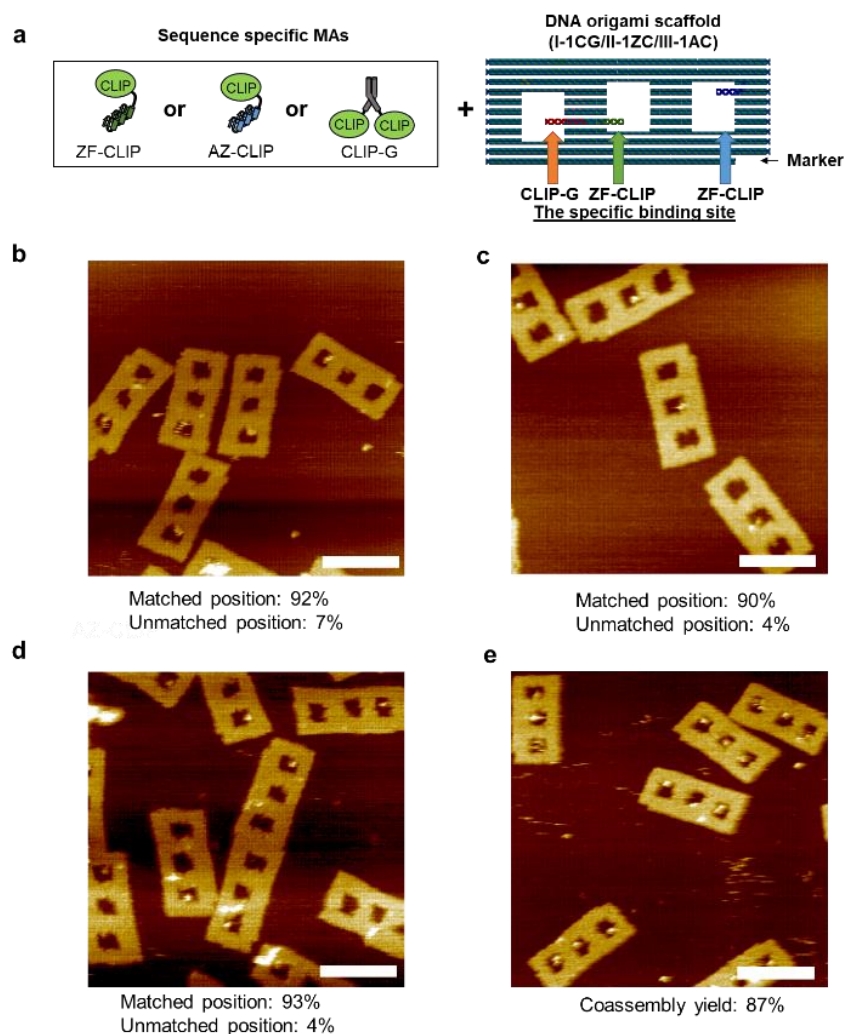


Figure 3.29. Three types of the sequence-specific modular adaptors orthogonally react at the predesigned positions on DNA scaffold in the presence of 200 mM NaCl. (a) An illustration shows three types of modular adaptors and DNA origami scaffold (I-1CG/II-1ZC/III-1AC), in which each cavity contained a single binding site for one of the modular adaptors. The stem loops in red, green, and blue denote the binding site for CLIP-G (CG) in cavity I, ZF-CLIP (ZC) in cavity II, and AZ-CLIP (AC) in cavity III, respectively. (b–e) AFM images of the DNA scaffold reacted with modular adaptors. (b–d) The DNA scaffold (5 nM) was incubated with MA (250 nM) (b) CLIP-G, (c) ZF-CLIP or (d) AZ-CLIP for 30 min at ambient temperature. (e) One-pot co-assembly reaction of three MAs on DNA scaffold (I-1CG/II-1ZC/III-1AC). The DNA scaffold (5 nM) was incubated with CLIP-G, ZF-CLIP, and AZ-CLIP (each of MA: 250 nM) for 30 min at ambient temperature. These reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM

MgCl₂, 200 mM NaCl, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween 20, and 200 mM NaCl. The reaction mixture was purified by size-exclusion chromatography and then analyzed by AFM. Yields at each cavity and co-assembly yields were estimated by counting the number of cavities occupied by the modular adaptors (Table 3.11). The scale bars represent 100 nm.

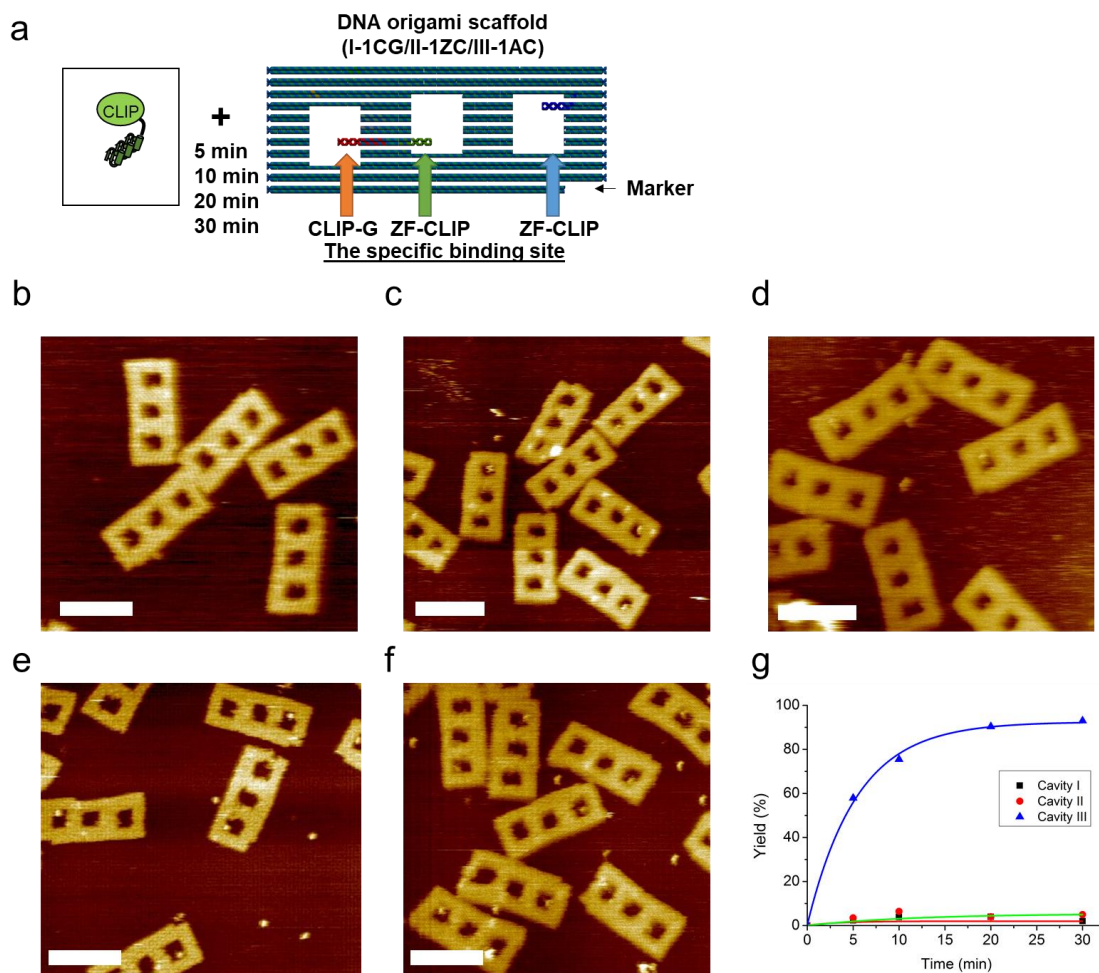


Figure 3.30 (a) illustration of the time dependence assembly yield of AZ-CLIP on the matched cavity (Cavity III) and unmatched cavities (Cavity I and Cavity II) on 3-well DNA scaffold (I-1AH/II-1ZS/III-1AC). (b) AFM image of DNA scaffold I-1AH/II-1ZS/III-1AC. (c-f) 5 nM DNA scaffold was incubated with 250 nM AZ-CLIP, after 5 min, 10 min, 20 min, and 30 min incubation, an aliquot of the reaction mixture was purified by size-exclusion chromatography and then analyzed by AFM, respectively. The reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween20, and 200 mM NaCl. (g) Yields were estimated by counting the number of

cavities occupied by the AZ-CLIP after 5 min, 10 min, 20 min, and 30 min incubation. The scale bars represent 100 nm.

Sequential reactions in the presence of 200 mM NaCl in the order of CLIP-G, AZ-CLIP, and ZF-CLIP reduced the modification at unmatched positions to less than 9% as showed in Figure 3.31 with the overall co-assembly yield reached 85% after 60 min sequential assembly. One pot co-assembly reaction of the modular adaptors for their respective target positions on DNA scaffold (Figure 3.29e) resulted in 87% co-assembly yield after 30 min incubation.

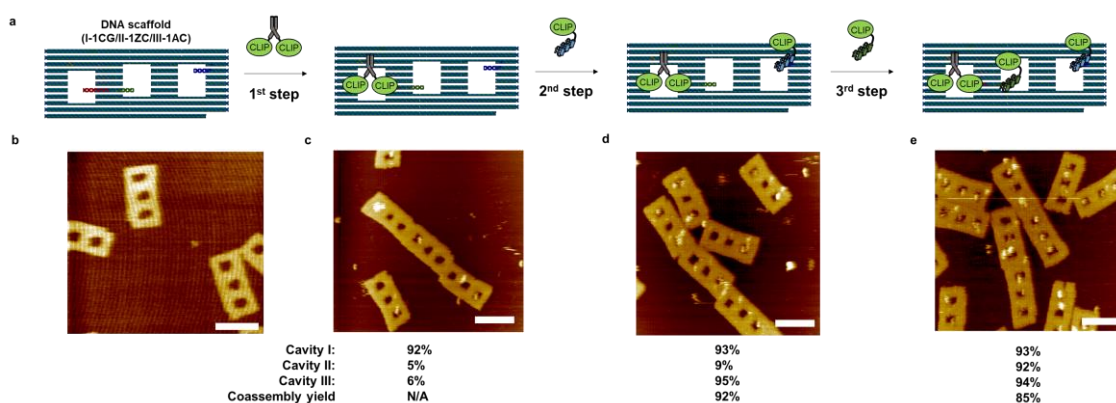


Figure 3.31. (a) An illustration of the sequential assembly of modular adaptors on the DNA scaffold (I-1AH/II-1ZS/III-1AC) in the presence of 200 mM NaCl. (b–e) AFM images of the DNA scaffold reacted with the modular adaptors. The DNA scaffold (5 nM) was incubated (c) with CLIP-G (250 nM) for 5 min at ambient temperature (1st step), then (d) with the AZ-CLIP (250 nM) for additional 5 min (2nd step), and finally (e) with ZF-CLIP (250 nM) for further 5 min (3rd step). The reactions were carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 200 mM NaCl, 1 mM DTT, 1 μ M ZnCl₂, and 0.02% Tween 20. At each step, an aliquot of the reaction mixture was purified by size-exclusion chromatography and then analysed by AFM. Yields at each cavity and co-assembly yields after the 2nd step and the 3rd step were estimated by counting the number of cavities occupied by the modular adaptors. The scale bars represent 100 nm.

Table 3.11. Total numbers and yields of the DNA scaffolds (I-1CG/II-1ZC/III-1AC) assembled with the modular adaptors deduced from the AFM images.

Incubation condition	Modular adaptor derivatives	Number of well-formed DNA scaffold	Numbers and yields of the modified cavities			Co-assembly yield	AFM image
			Cavity I	Cavity II	Cavity III		
5 min incubation	CLIP-G	166	158 [95%]	4 [2%]	4 [2%]	n.a.	Figure 3.27b
	ZF-CLIP	152	7 [5%]	142 [93%]	32 [21%]	n.a.	Figure 3.27b
	AZ-CLIP	130	7 [5%]	8 [6%]	127 [98%]	n.a.	Figure 3.27b
10 min incubation	CLIP-G	121	112 [93%]	6 [5%]	7 [6%]	n.a.	Figure 3.27c
	ZF-CLIP	105	2 [2%]	91 [87%]	38 [36%]	n.a.	Figure 3.27c
	AZ-CLIP	152	6 [4%]	13 [8%]	145 [95%]	n.a.	Figure 3.27c
15 min incubation	CLIP-G	131	125 [95%]	7 [5%]	7 [5%]	n.a.	Figure 3.27d
	ZF-CLIP	141	16 [11%]	132 [94%]	66 [47%]	n.a.	Figure 3.27d
	AZ-CLIP	170	18 [11%]	37 [22%]	160 [94%]	n.a.	Figure 3.27d
5 min sequential assembly	CLIP-G	172	166 [96%]	3 [2%]	5 [3%]	n.a.	Figure 3.28c
	CLIP-G, AZ-CLIP	168	152 [96%]	19 [11%]	159 [95%]	148 [88%]	Figure 3.28d
	CLIP-G, AZ-CLIP, ZF-CLIP	167	161 [96%]	153 [92%]	165 [99%]	146 [87%]	Figure 3.28e

n.a. : not applicable

Table 3.11. (Continued)

Incubation condition	Modular adaptor derivatives	Number of well-formed DNA scaffold	Numbers and yields of the modified cavities			Co-assembly yield	AFM image
			Cavity I	Cavity II	Cavity III		
30 min incubation with 200 mM NaCl	CLIP-G	165	151 [92%]	10 [6%]	14 [8%]	n.a.	Figure 3.29b
	ZF-CLIP	197	5 [3%]	177 [90%]	9 [5%]	n.a.	Figure 3.29c
	AZ-CLIP	153	3 [2%]	7 [5%]	143 [93%]	n.a.	Figure 3.29d Figure 3.30f
5 min incubation with 200 mM NaCl	AZ-CLIP	174	4 [2%]	6 [4%]	99 [58%]	n.a.	Figure 3.30c
10 min incubation with 200 mM NaCl	AZ-CLIP	155	6 [4%]	10 [6%]	117 [75%]	n.a.	Figure 3.30d
20 min incubation with 200 mM NaCl	AZ-CLIP	176	7 [4%]	7 [4%]	159 [90%]	n.a.	Figure 3.30e
20 min sequential assembly with 200mM NaCl	CLIP-G	142	131 [92%]	7 [5%]	8 [6%]	n.a.	Figure 3.31c
	CLIP-G, AZ-CLIP	135	126 [93%]	12 [9%]	128 [95%]	124 [92%]	Figure 3.31d
	CLIP-G, AZ-CLIP, ZF-CLIP	144	134 [93%]	132 [92%]	135 [94%]	123 [85%]	Figure 3.31e
30 min one-pot reaction with 200 mM NaCl	CLIP-G, AZ-CLIP, ZF-CLIP	144	137 [95%]	131 [91%]	139 [97%]	125 [87%]	Figure 3.29e

n.a. : not applicable

3.3 Discussion

The aim of the sequence-specific modular adaptors is to demonstrate a general strategy to develop a simple and convenient method for locating a number of POIs or more specify functional enzymes on the DNA nanoscaffold. In this strategy, the selectivity of the modular adaptors on targeted sites of a DNA scaffold should only rely on the sequence selectivity of DNA binding module, so that any sequence-specific DNA binding protein could be used to create the orthogonal modular adaptor with the same covalent bond formation module. This simple design could significantly expand the

number of the orthogonal modular adaptor because one protein-tag could be used to combine with a number of DNA binding proteins.

The first attempt to develop the sequence-specific modular adaptors used the two types of DNA binding protein zif268 and AZP4 with the same protein-tag (SNAP-tag). However, ZF-SNAP and AZ-SNAP did not show orthogonality which could be explained by the fact that the reactivity of SNAP-tag to BG modified ODN is too fast ($10^4 \text{ M}^{-1}\text{s}^{-1}$) to maintain the sequence selectivity. The modular adaptor containing SNAP-tag reacts with BG modified ODN according to case 1 in Figure 3.2b. Therefore the activity of covalent bond formation was reduced to obtain the sequence-specific modular adaptors. With changing protein-tag substrates, the reactivity could be tuned for several orders of magnitudes. The second order rate constants of SNAP-tag and CLIP-tag to fluorescence modified BC is approximately $1 \text{ M}^{-1}\text{s}^{-1}$ and $10^2 \text{ M}^{-1}\text{s}^{-1}$, respectively, indicated that the modular adaptors based on SNAP-tag or CLIP-tag to BC modified ONDs is a good candidate. In this chapter, the combinations of SNAP-tag or CLIP-tag with three sequences selective DNA binding proteins (ZF, AZ, and GCN4) have been constructed: SNAP-tag derivatives (ZF-SNAP, AZ-SNAP) and CLIP-tag derivatives (ZF-CLIP, AZ-CLIP, and CLIP-G). By reducing the reactivity of SNAP-tag from $10^4 \text{ M}^{-1}\text{s}^{-1}$ (with BG as the substrate) to approximately $1 \text{ M}^{-1}\text{s}^{-1}$ (with BC as the substrate), the reasonable orthogonality was achieved. However, the covalent bond formation yield at the matched pair for ZF-SNAP and AZ-SNAP were only 68% and 64%, respectively. To increase the performance of modular adaptors, the higher reactivity for the protein tag module was required. Indeed when the CLIP-tag derivatives were used, the reaction yields reached to more than 90% after only 10 min incubation. The addition of 200 mM NaCl in the reaction mixture further reduced the covalent bond formation yield of the modular adaptor at the unmatched sequence. As the result, the modular adaptor containing CLIP-tag derivatives in combination with the

BC modified ODN were successfully loaded at the designed positions on a DNA scaffold with almost 90% co-assembly yield and less than 10 % yield at the unmatched positions.

Modular structures of proteins are commonly found in nature to exert high performance with the lowest trade-off.^{42–45} Nowadays, the modular structure becomes important in the development of the modern drug.^{1,2,46} The modular adaptor design described here consists of two functional modules: the sequence recognition module (DNA binding module) and the covalent bond formation module (the protein-tag module). The design of modular molecule is aimed for the high performance that always depends on the balance of affinity and activity as in case 2 of Figure 3.2b. The rational design of the sequence-specific modular adaptors and the boundary conditions for the kinetics of covalent bond formation and complex formation described in this chapter serve as the general principle for the design of molecular recognition-driven modifiers.

3.4 Conclusion

In summary, I propose here a general principle to design sequence-specific DNA modifiers driven by the specific DNA recognition. The kinetic requirements for the complex formation step of DNA and the DNA binder and the successive intramolecular chemical reaction were applied to design modular adaptors by appropriate combinations of the self-ligating protein tag and the sequence-specific DNA-binding protein. The modular adaptors with the same self-ligating protein tag bearing different types of sequence-specific DNA binding proteins indeed reacted preferentially with the substrate located in the matched DNA sequence with little reactivity to the substrate in unmatched sequence. Our design principle will expand the variation of sequence-specific modular adaptors by simply conjugating various sequence-specific DNA binding proteins^{47–49} and a protein-tag with the appropriate

selections of its substrate. This orthogonal set of modular adaptors will locate a variety of enzymes or receptors on a DNA scaffold to study the chemistry of spatially arranged biomolecules. More importantly, our design principle on the kinetic characteristics would be applicable for the design of almost any kind of the recognition-driven modifiers.

3.5 Materials and methods

3.5.1 Materials

The single-stranded M13mp18, restriction enzymes (*Nde*I and *Hind*III), pCLIP-tag (T7)-2 Vector (N9181S), BC-GLA-NHS (S9237S) were purchased from New England Biolabs. pCLIP_f(T7)-2 Vector was generously donated from New England Biolabs. Purified oligonucleotides as the staple strands for DNA origami, oligonucleotide primers for gene construction and all other oligonucleotides were obtained from Sigma-Aldrich (St. Louis, MO), Thermo Fisher Scientific Inc. (Waltham, MA, USA) or Gene Design Inc. (Osaka, Japan). *Escherichia coli* BL21 (DE3) competent cells were purchased from Invitrogen (Carlsbad, CA). Mini Elute Gel Extraction Kit was prepared from QIAGEN (Tokyo, Japan). HiTrap SP XL cation exchange column (5 ml), HisTrap HP column (5 ml), and Sephacryl S-400 were from GE Healthcare Japan Inc. (Tokyo, Japan). PrimeSTAR HS DNA polymerase, T4 DNA ligase, and *E. Coli* DH5 α competent cells were obtained from TaKaRa Bio Inc. (Shiga, Japan). Ultrafree-MC-DV was obtained from Merck Millipore (Darmstadt, Germany). Cosmosil 5C18-MS II column (4.6 i.d. $\times$ 150 mm) and HPLC-grade acetonitrile were purchased from Nacalai Tesque (Kyoto, Japan). Gel electrophoresis grade acrylamide, bis(acrylamide), phenol, and all other chemicals and reagents were purchased from Wako Chemicals (Tokyo, Japan) or Nacalai Tesque (Kyoto, Japan). Phosphate buffer

(PB) was prepared as 20 mM Na₂HPO₄ and 20 mM NaH₂PO₄. DNA origami buffer (pH 8.0) contained 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂.

3.5.2 Preparation of sequence-specific modular adaptors (ZF-SNAP, AZ-SNAP, ZF-CLIP, AZ-CLIP, and CLIP-G)

Overexpression and purification of the ZF-SNAP, AZ-SNAP, and AZ-CLIP were described in the previous chapter. As a typical example, the preparation of ZF-CLIP is shown below. All of the vectors encoding protein-tag derivatives (pET-30a-ZF-CLIP, pET-30a-CLIP-G) were prepared and the protein-tag derivatives (ZF-CLIP and CLIP-G) were overexpressed and purified in the same manner.

3.5.3 Construction of vectors for ZF-CLIP (pET-30a-ZF-CLIP)

CLIP-tag in *pCLIP-tag* (T7)-2 vector was amplified by PCR using the following primer pairs.

Forward primer (F_EcoRI_CLIP):

TAATAAGAATTCGGCGGCTCCGGCGGCTCCGACAAAGATTGCGAAA

Reverse primer (R_CLIP_HindIII):

TTATTAAAGCTTTTAATGATGGTGATGATGATGGTGATGATGGTGGG
TACCATTAACCTCGAGCCCGGGG.

The PCR products were run on a 1% agarose gel (TAE) and were purified by Mini Elute Gel Extraction Kit. The PCR products and pET-30a-cys-zif268⁶ were digested with EcoRI and HindIII and were purified in the same manner. These products were incubated with T4-DNA-ligase. The mixture was transformed into *E. coli* DH5 α competent cells for amplification. The vector encoding ZF-CLIP (termed as pET-30a-ZF-CLIP) was purified, sequence checked, then transformed into *E. coli* BL21(DE3) competent cells.

3.5.4 Overexpression and purification of ZF-CLIP

The transformed cells were grown at 37 °C until OD₆₀₀ reached 0.5, and protein expression was induced with 1 mM IPTG for 24 h at 25 °C. The soluble fraction of the cell lysate containing recombinant protein was loaded to HisTrap HP column equilibrated with a 50 mM phosphate buffer (pH 8.0) containing 200 mM NaCl and 1 mM DTT, and eluted by imidazole gradient. The main fractions containing the target protein were collected, loaded to HiTrap SP HP column equilibrated with a 50 mM phosphate buffer (pH 7.0) containing 1 mM DTT and eluted by NaCl gradient. The purified protein was dialyzed by using a 50 mM phosphate buffer (pH 8.0) containing 1 mM DTT, 50 µM ZnCl₂, and 50% glycerol and stocked at -20°C. The purity of the target protein was checked by SDS-PAGE. The major band in SDS-PAGE corresponded to the calculated molecular weight with estimated purity of over 95% (Figure 3.32). The modular adaptors were characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, SA matrix). ZF-CLIP: *m/z* calcd 32462, found 32440; CLIP-G: *m/z* calcd 27108, found 27083; Amino acid sequences and calculated molecular weights of recombinant proteins used in this study are shown in Table 3.12.

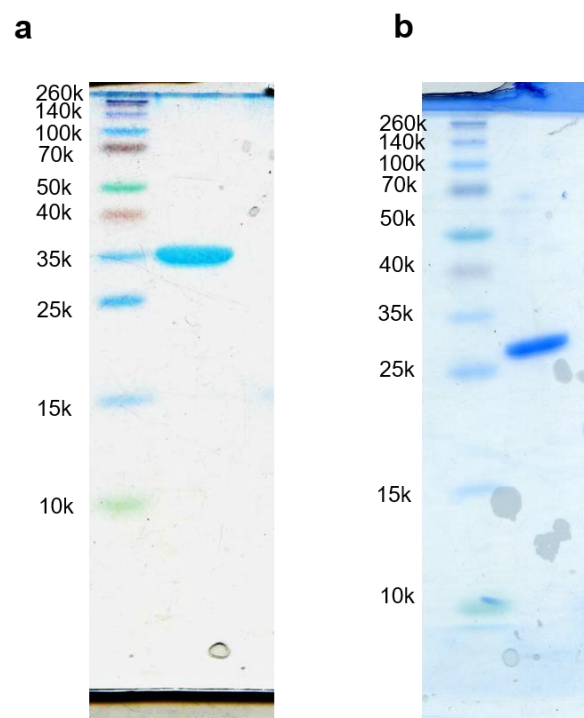


Figure 3.32. SDS-PAGE analyses of purified modular adaptors. (a) ZF-CLIP, (b) CLIP-G. Amino acid sequences and molecular weights of these proteins are shown in Table 3.12.

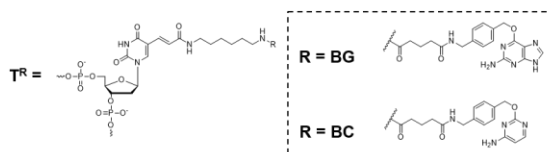
Table 3.12. Amino acid sequences and molecular weights of modular adaptor derivatives.

Modular adaptors	Amino acid sequences	Mw (Da)
ZF-CLIP	MMKTGEKRPYACPVESCDRRFRSDELTRHIRIHTGQKPFQCRICMRNFS RSDHLTTHIRTHTGEKPFACDICGRKFARSDERKRHTKIHTGEKEFGGSG GSMDKDCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKGTSAADAVEVPA PAAVLGGPEPLIQATAWLNAYFHQPEAIEEFVVPALHHPVFQQESFTRQV LWKLLKVVKFGEVISESHLAALVGNPAATAAVNTALDGNPVPILIPCHRV VQGDSDVGPYLGGLAVKEWLLAHEGHR LGKPGLGGSGGSHHHHHH	32462
CLIP-G	MDKDCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKGTSAADAVEVPAPA AVLGGPEPLIQATAWLNAYFHQPEAIEEFVVPALHHPVFQQESFTRQVLW KLLKVVKFGEVISESHLAALVGNPAATAAVNTALDGNPVPILIPCHRVVQ GDSDVGPYLGGLAVKEWLLAHEGHR LGKPGLGGSGDPAALKRARNTAA RRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGERHHHHHH	27108

3.5.5 Preparation of substrate modified ODNs and Alexa 488 modified ODNs

A coupling reaction between amino modified ODNs (100 μ M) and succinimidyl derivatives of alexa488 (1 mM) or succinimidyl derivatives of protein tag substrates was carried out in a 100 mM phosphate buffer (pH 8.0) for 8 hours at ambient temperature. The substrate modified ODNs were purified by reversed-phase HPLC on a Cosmosil 5C18-MS II column (4.6 x 150 mm, elution with 100 mM triethylammonium acetate buffer, pH 7.0, linear gradient over 30 min from 2.5 to 30% acetonitrile at flow rate of 1.0 mL.min⁻¹), and characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, HPA matrix). ODN-ZF-Alexa488: *m/z* calcd 11397, found 11395; ODN-AZ-Alexa488: *m/z* calcd 11400, found 11399; ODN-AP-Alexa488: *m/z* calcd 12011, found 12014. ODN-ZF-BC *m/z* calcd 11198, found 11200, ODN-AP-2BC: *m/z* calcd 12290, found 12290, ODN-8G-AP-2BC: *m/z* calcd 21577, found 21578, ODN-11G-ZF-BC: *m/z* calcd 21160, found 21167. Preparation of ODN-AZ-BC, ODN-AZ-BG, ODN-ZF-BC, and ODN-24D-AZ-BC were described in Chapter 2

a



b

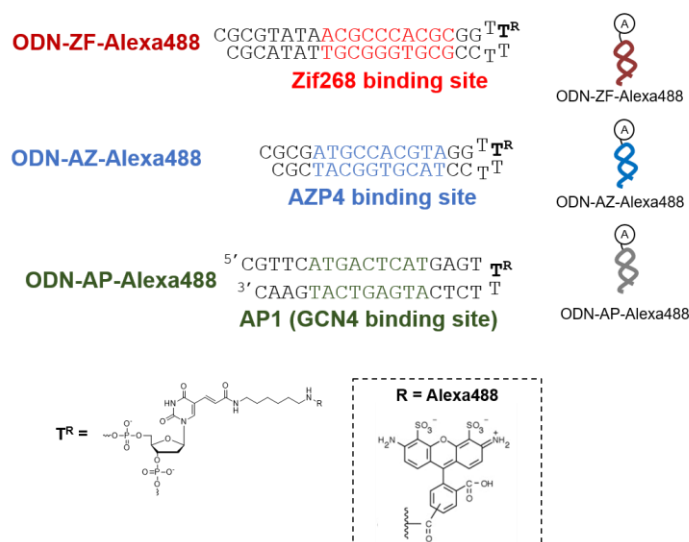


Figure 3.33 a) Nucleotide sequences of the ODNs for each of the modular adaptors with the substrate for protein tags and the chemical structures of the substrates for protein tag modified T, denoted as “T^R” or “T^{BC}”. b) Alexa488 modified ODNs for each of the modular adaptors for fluorescence polarization assay to obtain the equilibrium dissociation constant (K_D).

3.5.6 Preparation of ^{32}P -end-labeled ODNs and the analysis of covalent-linkage formation between ^{32}P -end-labeled ODNs and modular adaptors by denaturing gel shift assay

The ODNs (Figure 3.33a) were 5'- ^{32}P -end-labeled as previously described (Chapter 2). In a typical experiment for kinetic analysis, ^{32}P -ODN-ZF-BC (about 0.5 nM) was incubated with ZF-CLIP (10 nM) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl_2 , 1 mM DTT, 1 μM ZnCl_2 , 0.02% Tween20, 200 nM BSA and 100 nM calf thymus DNA and 200 mM NaCl (when present) at ambient temperature and aliquots taken at defined reaction times were quenched by addition of formamide. The aliquots were analyzed by 8 M Urea PAGE and the intensities of the bands on the gel were analyzed by using Storm 860 Molecular imager (Amersham). The kinetics data were fitted to a reaction model assuming the first-order kinetics, and then the second-order rate constants were determined.

3.5.7 Preparation of the DNA origami scaffold with three cavities

A solution (50 μL) containing M13mp18 single-stranded DNA (New England Biolabs, 10 nM), staple DNA strands (5 equiv, 50 nM, the nucleotide sequences of all the staple DNA strands were shown in Chapter 2 and Table 3.13) in a buffer (40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl_2 , pH 8.0) was heated at 95 $^{\circ}\text{C}$ for 1 min, annealed at 53 $^{\circ}\text{C}$ for 30 min, and kept at 4 $^{\circ}\text{C}$ by using a thermal cycler. The samples were purified by size-exclusion chromatography (400 μL volume of Sephacryl S-400, GE Healthcare) equilibrated with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl_2 in Ultrafree-MC-DV (Millipore).

Table 3.13. The staple strands including zif268, AZP4 and GCN4 binding sites with BC modified T^{BC}. The zif268 and AZP4 binding sites on the staple strands were colored in red and blue, respectively.

Oligo name	Sequence (from 5' to 3')
8g-AP-2BC	TAATACTGT ^{BC} TCATGAGTCATGAGTTTCT ^{BC} CATGACTCATGAACCAG GCAAGGCAAAGACATCCAATA
11g-ZF-BC	AACAGGTCCTTACGCCCACGCGCGTT ^{BC} TTTCGCGCGTGGGCGTAAGG AACCAGACCGGAAGATTCGAGC
24D-AZ-BC	GGACAGATCTTATGCCACGTAGCGTT ^{BC} TTTCGCTACGTGGCATAAGAA ATTGTGTCGAAATCTGTATCAT

3.5.8 Preparation of DNA origami scaffold assembled with modular adaptors

A DNA origami scaffold was incubated with ZF-CLIP, AZ-CLIP, and/or CLIP-G under the conditions shown in the caption of figures or tables. In a typical experiment, the crosslinking reaction of each of modular adaptor was carried out for 30 minutes at ambient temperature with 5 nM DNA scaffold and 100 nM modular adaptor (ZF-CLIP, AZ-CLIP or CLIP-G) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20. The mixture was purified by size-exclusion chromatography (400μL volume of Sephacryl S-400 in Ultrafree-MC-DV) equilibrated with the buffer used for the reaction to remove the unbound modular adaptors. The fractions containing DNA origami scaffold were utilized for AFM analyses. The preparation of DNA origami scaffolds assembled with modular adaptor-fused enzymes was performed in the same manner.

3.5.9 AFM imaging and statistical analysis

The sample was deposited on freshly cleaved mica surface (1.5 mmφ) and adsorbed for 5 min at ambient temperature, then washed three times with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂. The sample

was scanned in tapping mode using a fast-scanning AFM system (Nano Live Vision, RIBM Co. Ltd., Tsukuba, Japan) with a silicon nitride cantilever (Olympus BL-AC10DS-A2). At least three independent preparations of each sample were analyzed by AFM and several images were acquired from different regions of the mica surface. The total number of DNA scaffolds corresponds to the number of expected rectangular shapes possessing three cavities observed by AFM. The specific and nonspecific binding of modular adaptors were counted for only the modular adaptor bound to the perfectly folded DNA scaffold. The yield of DNA-scaffold-assembled modular adaptor was calculated as in the previous chapter.

3.5.10 Fluorescence polarization spectroscopy

Fluorescence polarization assay was conducted to study the binding of modular adaptors to its matched and unmatched Atto488 modified substrates using Tecan INFINITE F PLEX spectroscopy (Figure 3.34). Reaction was carried out in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 100 nM calf thymus DNA and 0.02% Tween20 at 25°C. The kinetic parameter k_{obs} (s⁻¹) were fit to the to a reaction model assuming the first-order kinetics, and then the second-order rate constants were determined.

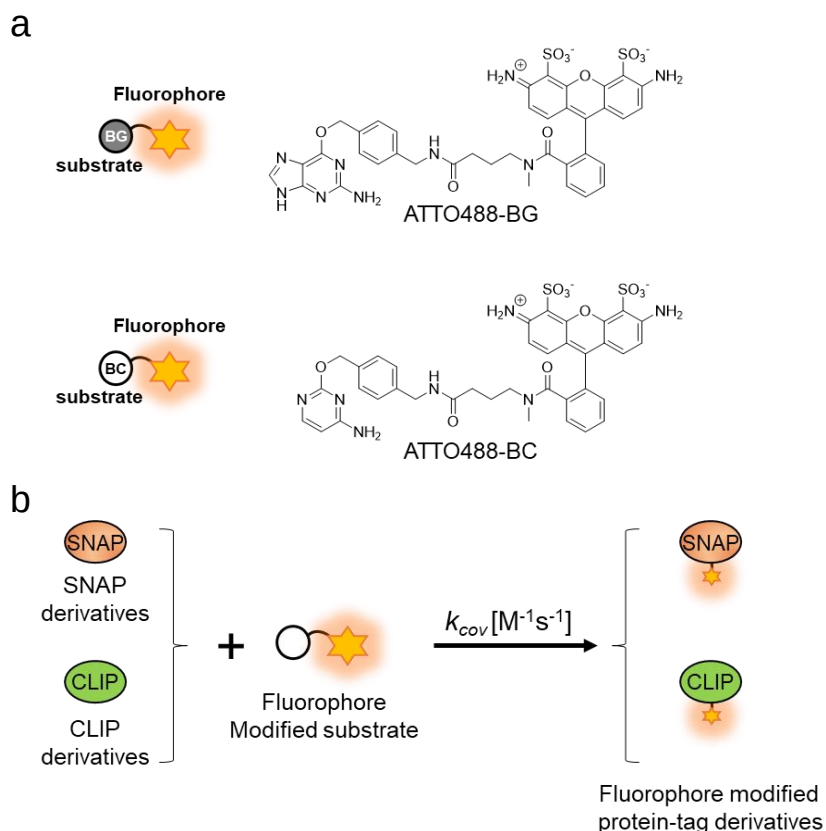


Figure 3.34 a) Chemical structures of ATTO488-BG and ATTO488-BC. b) A scheme shows crosslinking reactions between the modular adaptors (SNAP derivatives and CLIP derivatives) and ATTO488 modified substrates (ATTO488-BG and ATTO488-BC).

Fluorescence polarization assay was conducted to study the dissociation constants of modular adaptors to its matched and unmatched Alexa488 modified ODNs (Figure 3.33b) using Tecan INFINITE F PLEX spectroscopy. fluorophore modified ODNs (0.1 nM) was reacted with various concentration of modular adaptors (0.25 nM ~ 1024 nM) in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 100 nM calf thymus DNA, 0.02% Tween20, and 200 mM NaCl (when present) at 25°C.

3.5 References

- (1) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie., Int. Ed.* **2016**, *55*, 13408–13421.
- (2) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.
- (3) Foley, T. L.; Burkart, M. D. Site-Specific Protein Modification: Advances and Applications. *Curr. Opin. Chem. Biol.* **2007**, *11*, 12–19.
- (4) Warpehoski, M. A.; Hurley, L. H. Sequence Selectivity of DNA Covalent Modification. *Chem. Res. Toxicol.* **1988**, *1*, 315–333.
- (5) Tse, W. C.; Boger, D. L. Sequence-Selective DNA Recognition: Natural Products and Nature's Lessons. *Chem. Biol.* **2004**, *11*, 1607–1617.
- (6) Ali, A.; Bhattacharya, S. DNA Binders in Clinical Trials and Chemotherapy. *Bioorganic Med. Chem.* **2014**, *22*, 4506–4521.
- (7) Neidle, S. DNA Minor-Groove Recognition by Small Molecules. *Nat. Prod. Rep.* **2001**, *18*, 291–309.
- (8) Bando, T.; Sugiyama, H. Synthesis and Biological Properties of Sequence-Specific DNA-Alkylating Pyrrole-Imidazole Polyamides. *Acc. Chem. Res.* **2006**, *39*, 935–944.
- (9) Kovacic, R. T.; Welch, J. T.; Franklin, S. J. Sequence-Selective DNA Cleavage by a Chimeric Metallopeptide. *J. Am. Chem. Soc.* **2003**, *125*, 6656–6662.
- (10) Ma, D. D. F.; Rede, T.; Naqvi, N. A.; Cook, P. D. Synthetic Oligonucleotides as Therapeutics: The Coming of Age. *Biotechnol. Annu. Rev.* **2000**, *5*, 155–196.
- (11) Seeman, N. C.; Sleiman, H. F. DNA Nanotechnology. *Nat. Rev. Mater.* **2017**, *3*, 17068.
- (12) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.

- (13) Chandrasekaran, A. R.; Anderson, N.; Kizer, M.; Halvorsen, K.; Wang, X. Beyond the Fold: Emerging Biological Applications of DNA Origami. *ChemBioChem* **2016**, 1081–1089.
- (14) Linko, V.; Nummelin, S.; Aarnos, L.; Tapio, K.; Toppari, J.; Kostianen, M. DNA-Based Enzyme Reactors and Systems. *Nanomaterials* **2016**, 6, 139.
- (15) Rajendran, A.; Nakata, E.; Nakano, S.; Morii, T. Nucleic-Acid-Templated Enzyme Cascades. *ChemBioChem* **2017**, 18, 696–716.
- (16) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, 4, 249–254.
- (17) Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **2012**, 134, 5516–5519.
- (18) Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter, N. G.; Woodbury, N. W.; Yan, H. Multi-Enzyme Complexes on DNA Scaffolds Capable of Substrate Channelling with an Artificial Swinging Arm. *Nat. Nanotechnol.* **2014**, 9, 531–536.
- (19) Rosen, C. B.; Kodal, A. L. B.; Nielsen, J. S.; Schaffert, D. H.; Scavenius, C.; Okholm, A. H.; Voigt, N. V.; Enghild, J. J.; Kjems, J.; Tørring, T.; et al. Template-Directed Covalent Conjugation of DNA to Native Antibodies, Transferrin and Other Metal-Binding Proteins. *Nat. Chem.* **2014**, 6, 804–809.
- (20) Saccà, B.; Meyer, R.; Erkelenz, M.; Kiko, K.; Arndt, A.; Schroeder, H.; Rabe, K. S.; Niemeyer, C. M. Orthogonal Protein Decoration of DNA Origami. *Angew. Chemie., Int. Ed.* **2010**, 49, 9378–9383.
- (21) Nakata, E.; Liew, F. F.; Uwatoko, C.; Kiyonaka, S.; Mori, Y.; Katsuda, Y.; Endo, M.; Sugiyama, H.; Morii, T. Zinc-Finger Proteins for Site-Specific Protein

- Positioning on DNA-Origami Structures. *Angew. Chemie., Int. Ed.* **2012**, *51*, 2421–2424.
- (22) Ngo, T. A.; Nakata, E.; Saimura, M.; Kodaki, T.; Morii, T. A Protein Adaptor to Locate a Functional Protein Dimer on Molecular Switchboard. *Methods* **2014**, *67*, 142–150.
- (23) Koßmann, K. J.; Ziegler, C.; Angelin, A.; Meyer, R.; Skoupi, M.; Rabe, K. S.; Niemeyer, C. M. A Rationally Designed Connector for Assembly of Protein-Functionalized DNA Nanostructures. *ChemBioChem* **2016**, *17*, 1102–1106.
- (24) Sagredo, S.; Pirzer, T.; AghebatRafat, A.; Goetzfried, M. A.; Moncalian, G.; Simmel, F. C.; De La Cruz, F. Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases. *Angew. Chemie., Int. Ed.* **2016**, *55*, 4348–4352.
- (25) Lovendahl, K. N.; Hayward, A. N.; Gordon, W. R. Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. *J. Am. Chem. Soc.* **2017**, *139*, 7030–7035.
- (26) Kurokawa, T.; Kiyonaka, S.; Nakata, E.; Endo, M.; Koyama, S.; Mori, E.; Tran, N. H.; Dinh, H.; Suzuki, Y.; Hidaka, K.; et al. DNA Origami Scaffolds as Templates for Functional Tetrameric Kir3 K⁺ Channels. *Angew. Chemie., Int. Ed.* **2018**, *57*, 2586–2591.
- (27) Nakata, E.; Dinh, H.; Ngo, T. A.; Saimura, M.; Morii, T. A Modular Zinc Finger Adaptor Accelerates the Covalent Linkage of Proteins at Specific Locations on DNA Nanoscaffolds. *Chem. Commun.* **2015**, *51*, 1016–1019.
- (28) Nguyen, T. M.; Nakata, E.; Saimura, M.; Dinh, H.; Morii, T. Design of Modular Protein Tags for Orthogonal Covalent Bond Formation at Specific DNA Sequences. *J. Am. Chem. Soc.* **2017**, *139*, 8487–8496.
- (29) Hope, I. A.; Mahadevan, S.; Struhl, K. Structural and Functional Characterization

- of the Short Acidic Transcriptional Activation Region of Yeast GCN4 Protein. *Nature*. **1988**, 333, 635–640.
- (30) König, P.; Richmond, T. J. The X-Ray Structure of the GCN4-BZIP Bound to ATF/CREB Site DNA Shows the Complex Depends on DNA Flexibility. *J. Mol. Bio.* **1993**, 233, 139–154.
 - (31) Morii, T.; Saimei, Y.; Okagami, M.; Makino, K.; Sugiura, Y. Factors Governing the Sequence-Selective DNA Binding of Geometrically Constrained Peptide Dimers. *J. Am. Chem. Soc.* **1997**, 119, 3649–3655.
 - (32) Von Hippel, P. H.; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, 264, 675–678.
 - (33) Kim, J.-S.; Pabo, C. O. Getting a Handhold on DNA: Design of Poly-Zinc Finger Proteins with Femtomolar Dissociation Constants. *Proc. Natl. Acad. Sci.* **1998**, 95, 2812–2817.
 - (34) Geertz, M.; Shore, D.; Maerkl, S. J. Massively Parallel Measurements of Molecular Interaction Kinetics on a Microfluidic Platform. *Proc. Natl. Acad. Sci.* **2012**, 109, 16540–16545.
 - (35) Barshop, B. A.; Wrenn, R. F.; Frieden, C. Analysis of Numerical Methods for Computer Simulation of Kinetic Processes: Development of KINSIM-A Flexible, Portable System. *Anal. Biochem.* **1983**, 130, 134–145.
 - (36) Zimmerle, C. T.; Frieden, C. Analysis of Progress Curves by Simulations Generated by Numerical Integration. *Biochem. J.* **1989**, 258, 381–387.
 - (37) Tsai, Y. H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W. Selective, Rapid and Optically Switchable Regulation of Protein Function in Live Mammalian Cells. *Nat. Chem.* **2015**, 7, 554–561.
 - (38) Sheinman, M.; Bénichou, O.; Kafri, Y.; Voituriez, R. Classes of Fast and Specific Search Mechanisms for Proteins on DNA. *Reports Prog. Phys.* **2012**, 75, 026601.

- (39) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. How a Protein Searches for Its Site on DNA: The Mechanism of Facilitated Diffusion. *J. Phys. A Math. Theor.* **2009**, *42*, 434013.
- (40) Kolomeisky, A. B. Physics of Protein-DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088–2095.
- (41) Esadze, A.; Iwahara, J. Stopped-Flow Fluorescence Kinetic Study of Protein Sliding and Intersegment Transfer in the Target DNA Search Process. *J. Mol. Biol.* **2014**, *426*, 230–244.
- (42) Patthy, L. Modular Exchange Principles in Proteins. *Curr. Opin. Struct. Biol.* **1991**, *1*, 351–361.
- (43) Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W. From Molecular to Modular Cell Biology. *Nature* **1999**, *402*, C47–C52.
- (44) Clune, J.; Mouret, J. B.; Lipson, H. The Evolutionary Origins of Modularity. *Proc. R. Soc. B Biol. Sci.* **2013**, *280*, 20122863.
- (45) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chemie., Int. Ed.* **2013**, *52*, 5700–5725.
- (46) Gillis, E. P.; Burke, M. D. A Simple and Modular Strategy for Small Molecule Synthesis: Iterative Suzuki-Miyaura Coupling of B-Protected Haloboronic Acid Building Blocks. *J. Am. Chem. Soc.* **2007**, *129*, 6716–6717.
- (47) Pabo, C. O.; Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053–1095.
- (48) Prabakaran, P.; Siebers, J. G.; Ahmad, S.; Gromiha, M. M.; Singarayan, M. G.; Sarai, A. Classification of Protein-DNA Complexes Based on Structural Descriptors. *Structure* **2006**, *14*, 1355–1367.
- (49) Hudson, W. H.; Ortlund, E. A. The Structure, Function and Evolution of Proteins That Bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 749–760.

CHAPTER 4

A model of three enzymes cascade reaction on DNA scaffold

4.1 Introduction

In Chapter 2 and Chapter 3, several modular adaptors showed orthogonal sequence specific DNA modification and high efficiency for assembling proteins fused to them at the design positions on a DNA scaffold. In Chapter 2, a set of three orthogonal modular adaptors (ZF-SNAP, AZ-CLIP, and AZ-Halo) has been constructed by using two DNA binding proteins (zif268¹ and AZP4²) and three types of protein-tags (SNAP-tag³, CLIP-tag⁴ and Halo-tag⁵). Three types of modular adaptors were loaded orthogonally to defined DNA sequences with the co-assembly yield over 90%. In Chapter 3, three sequence-specific modular adaptors (ZF-CLIP, AZ-CLIP, and CLIP-G) constructed from three different DNA binding proteins (zif268, AZP4, and GCN4) and one type of protein-tag (CLIP-tag) were co-assembled on the DNA scaffold with the overall yield of 87%. In this chapter, I applied the orthogonal set of modular adaptors to construct the cascade of three enzymes derived from a cellular metabolic pathway on a DNA scaffold.

DNA nanostructures⁶⁻¹⁰ provide an ideal scaffold to assemble enzymes participating in sequential enzymatic reactions, in a spatially well-defined manner.¹¹⁻¹⁵ The cascade reaction of two model enzymes on various scaffolds was demonstrated,^{12,16-19} a model of three enzymes cascade reaction was also reported (Figure 4.1).¹³ In the previous study, the cascade reaction of two enzymes xylose reductase (XR) and xylitol dehydrogenase (XDH) from the xylose metabolism pathway^{20,21} were activated on DNA scaffold.¹⁴ The study described the sharp distance dependency for the efficiency of the cascade reaction with bimolecular transfer on a DNA scaffold. The modular adaptor system described in the previous chapters gives a good chance to expand the cascade reaction to the third enzyme of the pathway, the xylulose kinase (XK). This chapter discusses the details on construction and investigation of the modular adaptor based cascaded reaction of the three enzymes XR,

XDH and XK on a DNA scaffold. Three enzymes ZS-XR (ZF-SNAP adaptor fused XR), G-XDH (GCN4 adaptor fused XDH) and AC-XK (AZ-CLIP adaptor fused XK), that catalyze the cascade reaction of the xylose pathway from xylose to the final product xylulose 5-phosphate, were co-assembled on a DNA scaffold. The efficiency of the co-assembled enzymatic cascade reaction was significantly enhanced when compared with that of the three enzymes in bulk solution. This is the first study on the real cascades reaction of the three enzymes derived from the metabolic pathway on a DNA scaffold.

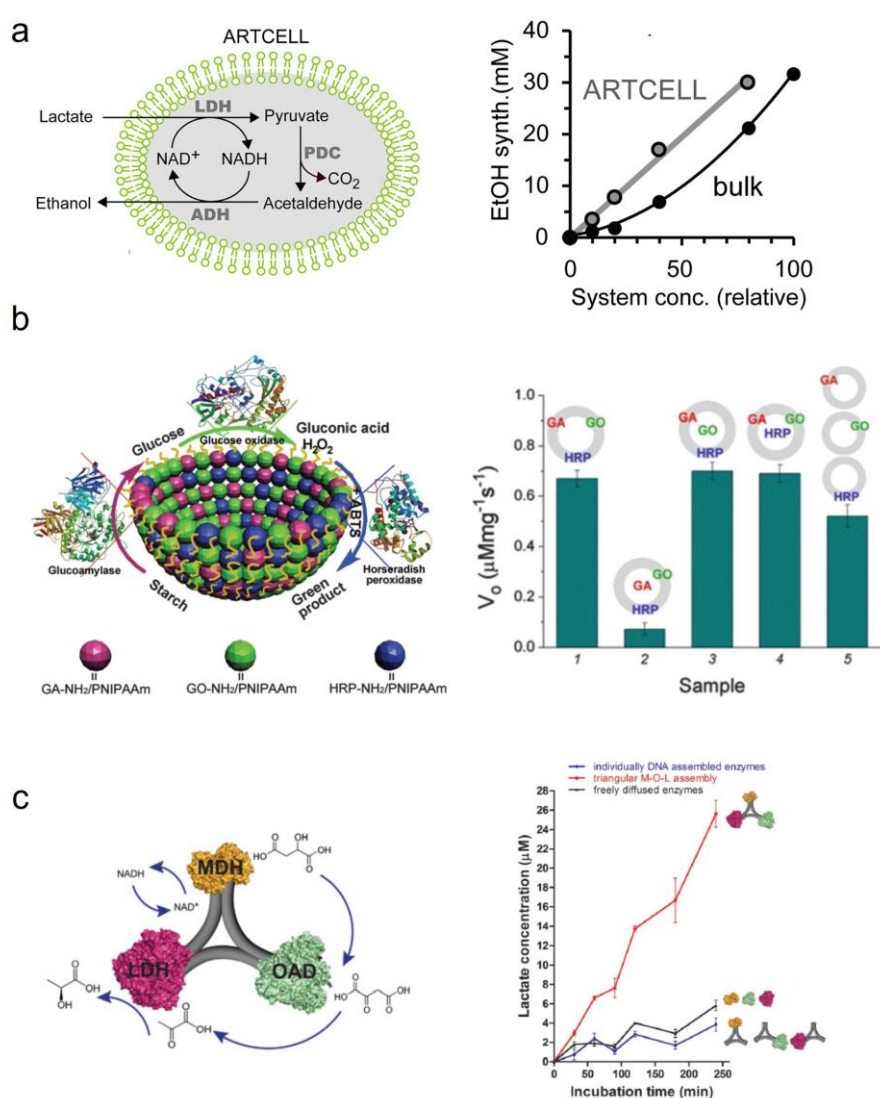


Figure 4.1 Cascade reactions of three enzyme system. a) Illustration of the three enzymes

encapsulated in a liposome (artcell) (left) and comparison of the efficiency of the cascade reaction in artcell with that in bulk solution.²² LDH: lactate dehydrogenase, PDC: pyruvate decarboxylase, ADH: alcohol dehydrogenase. b) Illustration of the assembly in a protein-polymer membrane of three enzyme glucose amylase (red sphere, GA-NH₂/PNIPAAm), Glucose oxidase (green sphere, GO-NH₂/PNIIPAAm), and horseradish peroxidase (Blue sphere, HPR-NH₂/PNIPAAm) (right) and the efficiency of cascade reaction with different arrangements of the three enzymes on a membrane.²³ c) Illustration of the assembly of malate dehydrogenase (MDH), oxaloacetate decarboxylase (OAD), and lactate dehydrogenase (LDH) on a DNA scaffold (right) and the efficiency of the cascade reaction on DNA scaffold (left).¹³

4.2 Results

4.2.1 Construction of the modular adaptor-fused enzymes and assembly of them on the DNA scaffolds

To verify the applicability of modular adaptors to assemble POIs on the DNA scaffold, xylulose kinase (XK) from *Saccharomyces cerevisiae* was chosen as a model POI. XK was reported to be a thermally unstable enzyme.²⁴ In this study, XK was fused to the C-terminus of AZ-Halo or AZ-CLIP with a linker to afford AH-XK or AC-XK (Figure 4.2, 4.3 and Table 4.1).

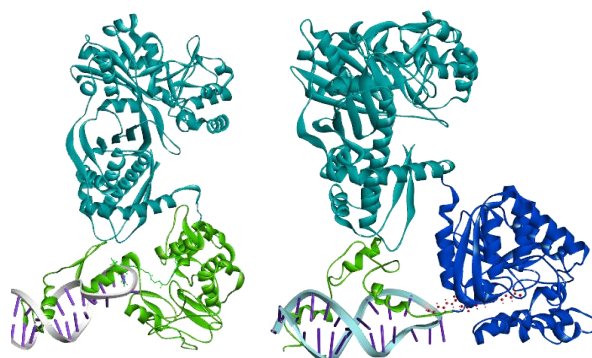


Figure 4.2. Possible models for the complex of (left) AC-XK with ODN-AZ-BC, (right) AH-XK with ODN-AZ-CH, respectively, based on the crystal structure of the complex between

zif268 and ODN (PDB ID : 1ZAA) and the complex of SNAP-tag with BG (PDB ID:3KZY), and Halo-tag with CH (PDB ID : 1CQW); the model of XK was adapted from crystal of XK from *E. coli* (PDB ID:2ITM). The models were constructed by using Discovery Studio (version 3.1, Accelrys Inc.).

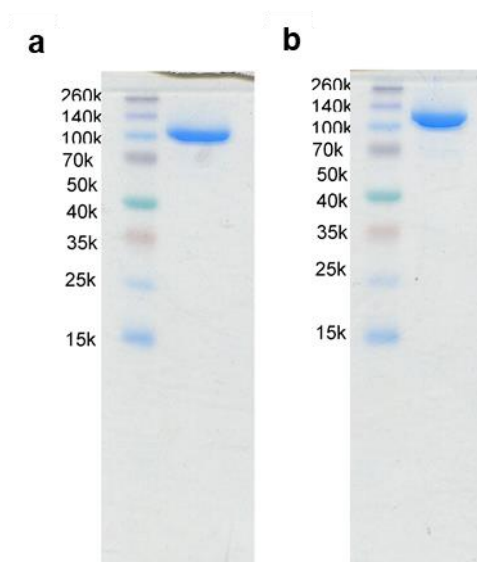


Figure 4.3. SDS-PAGE analyses of purified modular adaptor-fused XK. (a) AC-XK, (b) AH-XK. Amino acid sequences and molecular weights of these proteins are shown in Table 4.1.

Table 4.1. Amino acid sequences and molecular weights of modular adaptor derivatives.

Derivatives	Amino acid sequence	Mw (Da)
AC-XK	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRI CMRNFSRSDSLTRHIRTHTGEKPFACDICGRKFAESDNRKTH TKIHTGEKEFGGSGGSMDKDCMKRTTLDSPLGKLELSGCE QGLHRIIFLGKGTSAADAVEVPAPAAVLGGPEPLIQATAWLNA YFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLKVVKFGE VIESHLAALVGNPAATAAVNTALDGNPVPILIPCHRVVQGDS DVGPYLGGLAVKEWLLAHEGHRLGKPGLGGSGLSMLCSV IQRQTREVSNTMSLDSYYLGFDLSTQQLKCLAINQDLKIVHS ETVEFEKDLPHYHTKKGVYIHGDTIECPVAMWLEALDLVLS KYREAKFPLNKVMAVSGSCQQHGSVYWSSQAESLLEQLNK KPEKDLLHYVSSVAFARQTAPNWQDHSTAKQCQEFEECIGG PEKMAQLTGSRAHFRFTGPQILKIAQLEPEAYEKTCTISLVSN FLTSLVGHLEEEADACGMNLYDIRERKFSDELLHLIDSSS KDKTIRQKLMRAPMKNLIAGTICKYFIEKYGFNTNCKVSPM TGDNLATICSLPLRKNDVLVSLGTSTTVLLVTDKYHPSPNYH LFIHPTLPNHYMGMICYCNGSLARERIRDELNKERENNYEKT NDWTLFNQAVLDDSESENELGVYFPLGEIVPSVKAINKRIVF NPKTGMIEREVAKFKDKRHDANKNIVESQALSCRVRISPLSD SNASSQQLNEDTIVKFDYDESPLRDYLNKRPERTFFVGGAS KNDIAIVKKFAQVIGATKGNFRLETPNSCALGGCYKAMWSLL YDSNKIAVPFDKFLNDNFPWHVMESISDVDNENWDYNSKI VPLSELEKTLIGSGGSHHHHHH	100,901
AH-XK	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRI CMRNFSRSDSLTRHIRTHTGEKPFACDICGRKFAESDNRKTH TKIHTGEKEFGGSGGSMAEIGTGFPFDPHYVEVLGERMHYV DVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAPDLIG MGKSDKPDLYFFDDHVRFMDAFIEALGLEEVVLVIHDWGS ALGFHWAKRNPVERVKGIAFMEFIRPIPTWDEWPEFARETFA FRRTDVGRKLIIQNVFIEGTLPMGVVRPLTEVEMDHYREF LNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVP KLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQED NPDLIGSEIARWLSTLEISGEPTTEDLYFQSDNAIAMLCSVIQR QTREVSNTMSLDSYYLGFDLSTQQLKCLAINQDLKIVHSETV EFEKDLPHYHTKKGVYIHGDTIECPVAMWLEALDLVLSKYR EAKFPLNKVMAVSGSCQQHGSVYWSSQAESLLEQLNKKPE KDLLHYVSSVAFARQTAPNWQDHSTAKQCQEFEECIGGPEK MAQLTGSRAHFRFTGPQILKIAQLEPEAYEKTCTISLVSNFLT SLVGHLEEEADACGMNLYDIRERKFSDELLHLIDSSSKDKT IRQKLMRAPMKNLIAGTICKYFIEKYGFNTNCKVSPMTGDN LATICSLPLRKNDVLVSLGTSTTVLLVTDKYHPSPNYHHLFIHP TLPNHYMGMICYCNGSLARERIRDELNKERENNYEKTNDW TLFNQAVLDDSESENELGVYFPLGEIVPSVKAINKRIVFNPK TGMIEREVAKFKDKRHDANKNIVESQALSCRVRISPLSDSNA SSQQLNEDTIVKFDYDESPLRDYLNKRPERTFFVGGASKND AIVKKFAQVIGATKGNFRLETPNSCALGGCYKAMWSLLYDS NKIAVPFDKFLNDNFPWHVMESISDVDNENWDYNSKIVPL SELEKTLIGSGGSHHHHHH	116,473

The Michaelis constant (K_m) and the turnover number (k_{cat}) for the reaction of AH-XK and AC-XK were comparable to those of native XK (Figure 4.4 and Table 4.2). Both AC-XK and AH-XK showed similar enzymatic activity to that in the bulk solution when assembled on the DNA scaffold (Figure 4.5). The assembly yields of AH-XK, AC-XK, and ZS-XR, were 93%, 90%, and 93%, respectively, when these modular adaptor-conjugated enzymes (25 nM) were incubated for 30 min with 5 nM of the DNA scaffold (I-1AH/II-1ZS/III-1AC) on ice (Figure 4.5). Because AC-XK marked slightly higher activity than AH-XK, AC-XK was utilized for the three enzymes cascade reaction on a DNA scaffold.

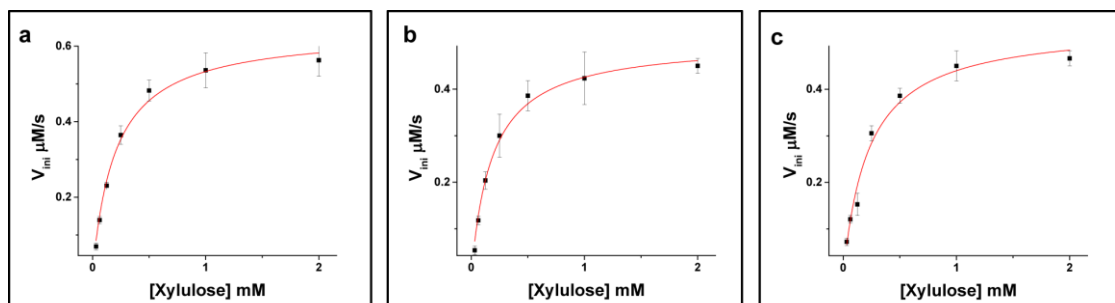


Figure 4.4. Michaelis-Menten plots for the phosphorylation of xylulose by (a) XK, (b) AC-XK, or (c) AH-XK. Enzymatic reactions with the same concentration of enzyme (2 nM) were performed at 25 °C in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween20, 100 mM NaCl, 1.1 mM ATP, 0.2 mM NADH, 2.3 mM phosphoenolpyruvate, 4.8 U/ml PK and 4.5 U/ml LDH, and the indicated concentration of xylulose (0.03 to 2mM). The reaction was started by an addition of xylulose. Kinetic parameters were listed on Table 4.2.

Table 4.2. Kinetic parameters for XK, AC-XK, and AH-XK for the phosphorylation of xylulose

	XK	AC-XK	AH-XK
K_m for xylulose (μM)	203 ± 20	180 ± 21	223 ± 25
k_{cat} (s^{-1})	257 ± 23	210 ± 15	240 ± 21
k_{cat}/K_m ($\text{mM}^{-1}.\text{s}^{-1}$)	1271 ± 55	1146 ± 194	1030 ± 84

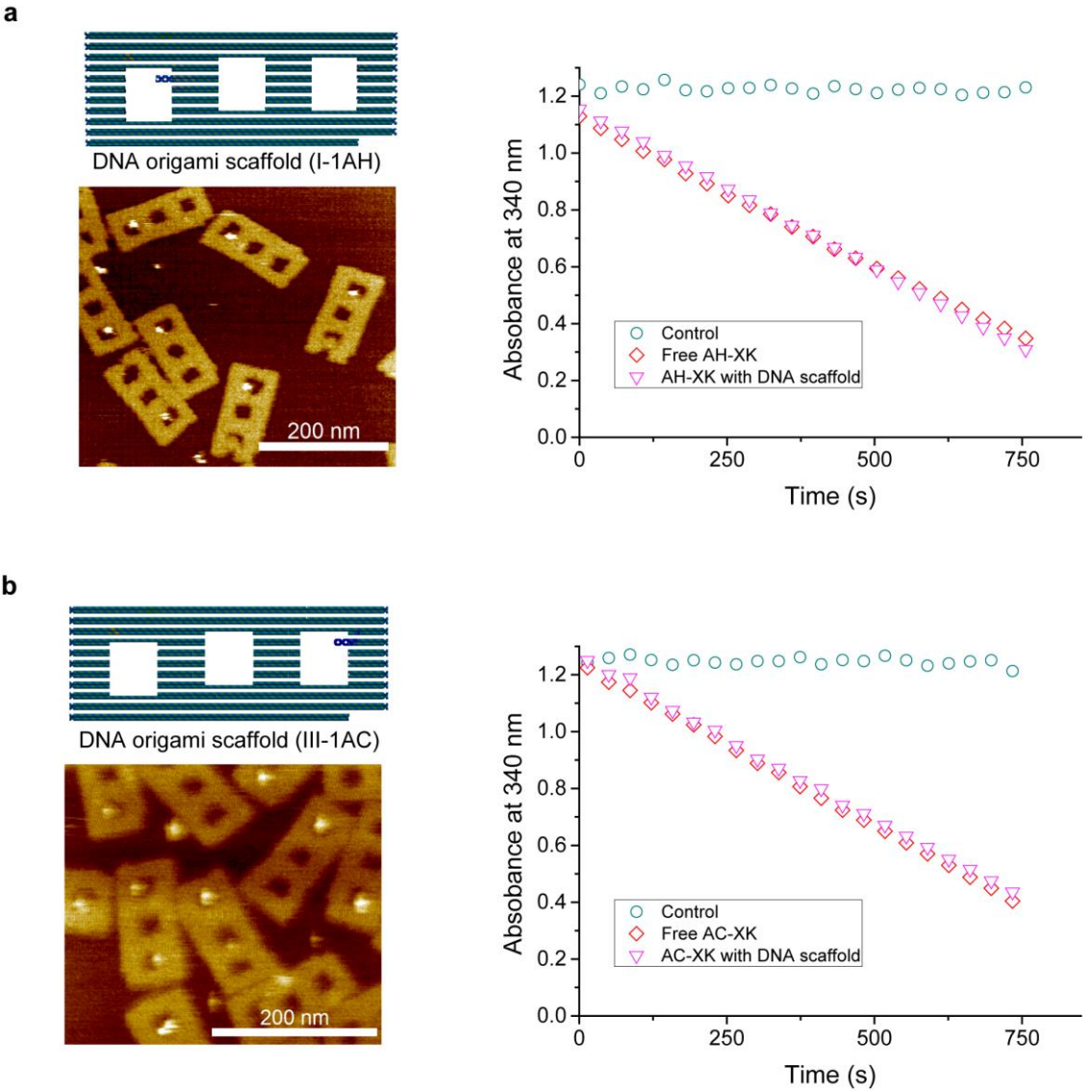


Figure 4.5. Effect of DNA scaffold on the catalytic activity of XK (a) An illustration of the DNA scaffold holding a binding site modified by CH (for AH-XK) and an AFM image of

AH-XK bound on the DNA scaffold (left). Time-course profiles of the oxidation of NADH monitored by absorbance at 340 nm (right). (b) An illustration of the DNA scaffold holding a binding site modified by BC (for AC-XK) and an AFM image of AC-XK bound on the DNA scaffold (left). Time-course profiles of the oxidation of NADH monitored by absorbance at 340 nm (right). Enzyme assay was carried out in a solution containing AH-XK or AC-XK in the absence (Free AH-XK or Free AC-XK, red diamonds) or presence of the DNA scaffold (AH-XK with the DNA scaffold or AC-XK with the DNA scaffold, pink triangles). Prior to the assay for XK activity, 2 nM AH-XK (or AC-XK) was incubated with or without 10 nM DNA scaffold for 30 minutes on ice in a buffer (pH 7.6) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, and 0.02% Tween20. Enzyme reactions were carried out with the same enzyme concentration (1 nM) at 25 °C in a buffer (pH 7.6) containing 100 mM NaCl, 1.1 mM ATP, 0.2 mM NADH, 2.3 mM PEP, 4.8 U/ml PK and 4.5 U/ml LDH. The reactions were started by the addition of xylulose (2 mM).

Orthogonal assembling reactions of three enzymes AH-XK, ZS-XR, and AC-XK were studied on DNA scaffold (I-1AH/II-1ZS/III-1AC) (Figure 4.6a). The modular adaptor fused enzymes were efficiently loaded on DNA scaffold with the yield more than 90% at the matched addressed and less than 2% assembly yield at the unmatched addressed. Assembly yields were estimated by counting the number of cavities occupied by the modular adaptor fused enzymes (Table 4.3)

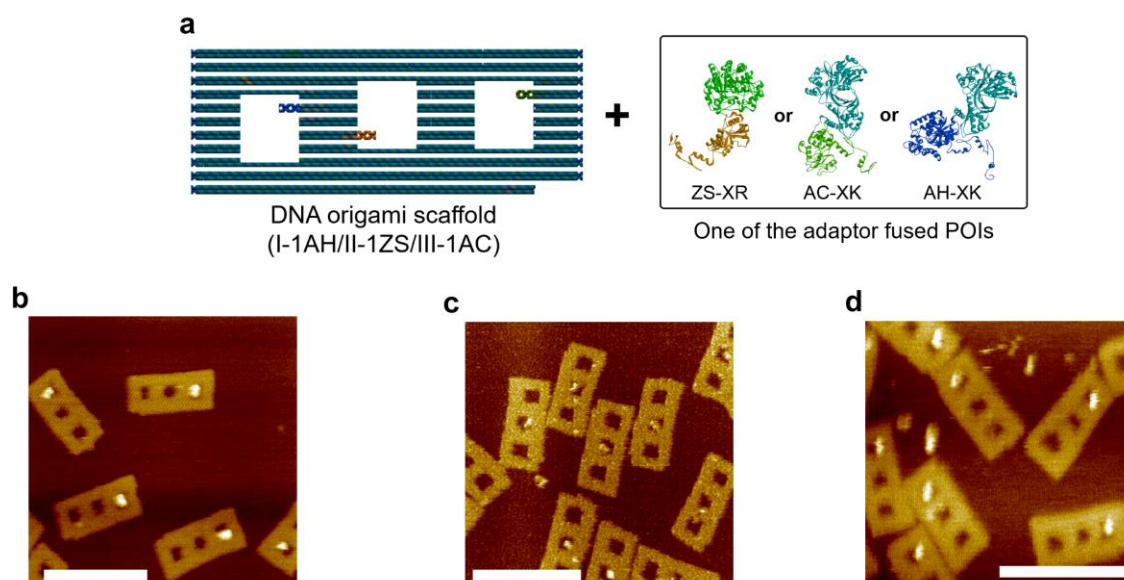


Figure 4.6. (a) An illustration for the setup of the orthogonal assembly of three adaptor fused POIs (ZS-XR, AC-XK, AH-XK) and DNA scaffold. (b-d) AFM images of (b) AH-XK, (c) ZS-XR, and (d) AC-XK on the DNA scaffold (I-1AH/II-1ZS/III-1AC), respectively. 5 nM of DNA scaffold was incubated with 5 molar equivalent of AH-XK, ZS-XR or AC-XK for 30 minutes on ice in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂ and 0.02% Tween20. Scale bars represent 200 nm.

Table 4.3. Total numbers and yields of the DNA scaffolds assembled with the modular adaptor fused enzymes analyzed by AFM.

DNA scaffold	Modular adaptor derivatives	Number of well-formed DNA scaffold	Numbers and yields of the modified cavities			AFM image
			Cavity I	Cavity II	Cavity III	
I-1AH/II-1ZS/III-1AC	AH-XK	142	132 [93%]	1 [1%]	3 [2%]	Figure 4.6b
	ZS-XR	174	2 [1%]	162 [93%]	1 [1%]	Figure 4.6c
	AC-XK	163	1 [1%]	3 [2%]	147 [90%]	Figure 4.6d

4.2.2 Construction of a sequential enzyme cascade and comparison of the effect of the assistance of the DNA scaffold

The set of orthogonal adaptors was applied for assembling three enzymes for a

part of the cascade reaction in the xylose metabolic pathway on the DNA scaffold (Figure 4.7). Previous report from the Morii laboratory demonstrated a two-step sequential enzymatic reaction system derived from the initial steps of xylose metabolic pathway (XR-XDH pathway).^{14,25,26} The first enzyme ZS-XR converted xylose into xylitol by consuming the cofactor NADH, then the second enzyme GCN4 adaptor fused XDH (G-XDH)²⁷ converts the resulting xylitol into xylulose by consuming NAD⁺ formed in the first step. This xylose metabolic pathway was extended to the third step of the reaction, where AC-XK converted xylulose to xylulose-5-phosphate by consuming ATP (Figure 4.8a).

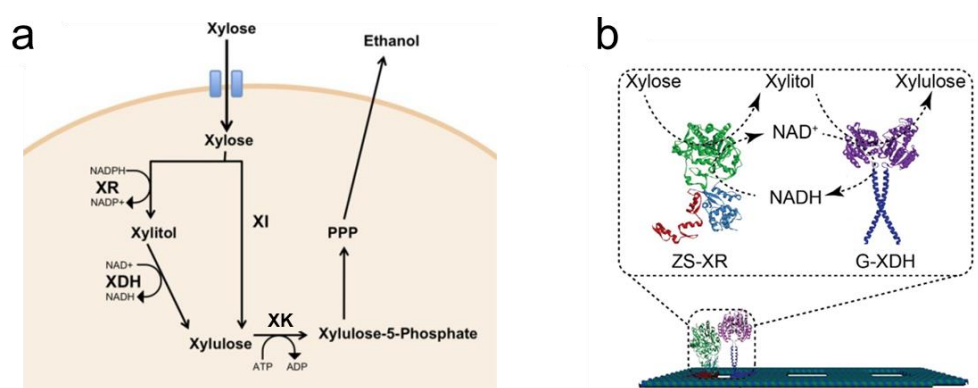


Figure 4.7 Xylose metabolic pathway. a) Xylose metabolism pathway from fungi, yeast, and bacteria.²⁰ XR: xylose reductase, XDH: xylitol dehydrogenase, XI: xylose isomerase, XK: xylose kinase, PPP: pentose phosphate pathway b) A schematic illustration of the artificial cascade reaction by two adaptor-fused enzymes (ZS-XR and G-XDH) coassembled on the DNA scaffold have been constructed in our previous study.¹⁴

Two types of DNA origami scaffolds which had the same number of binding sites for ZS-XR, G-XDH, and AC-XK but differed in the distance between the binding sites were prepared as shown in Figures 4.8b (I-4XR/II-4XDH/III-1XK) and 4.8c (I-4XR/I-4XDH/I-1XK). ZS-XR, G-XDH, and AC-XK were loaded on these DNA

scaffolds. The AFM images indicated that three enzymes were simultaneously located at the positions as designed (Figures 4.8d and 4.8e). A DNA scaffold assembled with only ZS-XR (I-4XR) and one coassembled with ZS-XR and G-XDH (I-4XR/II-4XDH and I-4XR/I-4XDH) were also prepared (Figure 4.9). The cascade reaction by three enzymes was analyzed by HPLC to quantitate the cofactors NADH, NAD⁺, ATP, and ADP, consumed or generated in the cascade reaction (Figures 4.8f and 4.8g) at steady-state conditions after 24 h (Figure 4.8, 4.9 and Table 4.4). The amount of ADP produced in the final step of the cascade reaction was used as a measure of the efficiency of the cascade reaction (Figures 4.8h and 4.8i). When the cascade reaction of three enzymes co-assembled system (I-4XR/II-4XDH/III-1XK or I-4XR/I-4XDH/I-1XK) was compared with that of two enzymes co-assembled system (I-4XR/II-4XDH or I-4XR/I-4XDH), the single enzyme loaded system (I-4XR), and the all free enzymes, the amount of ADP produced in the reaction with the three enzymes co-assembled system was higher than that with the two enzymes coassembled system and that with the single enzyme loaded system. In the cascade reaction by the three enzymes, the system with the first enzyme XR being loaded on the scaffold (I-4XR) was more efficient than the reaction system with all enzymes being free in the solution as shown in Figure 4.8 f and g, where "I-4XR" produced much more ADP than "All enzymes in bulk solution". The result indicated that XR could be more stable on the DNA scaffold than in the bulk solution because the consumption of NADH was also prevented in the case of "All enzymes in bulk solution" (Table 4.4). The result indicated the products of XR and XDH were efficiently transported within the co-assembled enzymes on the DNA scaffold. Moreover, the amount of ADP produced in the three enzymes co-assembled within 10 nm (Figure 4.8i: I-4XR/I-4XDH/I-1XK) was higher than that co-assembled with 50 nm distance (Figure 4.8h: I-4XR/II-4XDH/III-1XK).

These results were consistent with the previous results that the interenzyme distance is one of the important parameters to regulate the efficiency of the cascade reaction by facilitating transport of intermediates.^{3,11} The cascade reaction consisting of three enzymes studied here provides a useful model system to thoroughly investigate the effects of interenzyme distances, orientations, and ratios of the three enzymes.

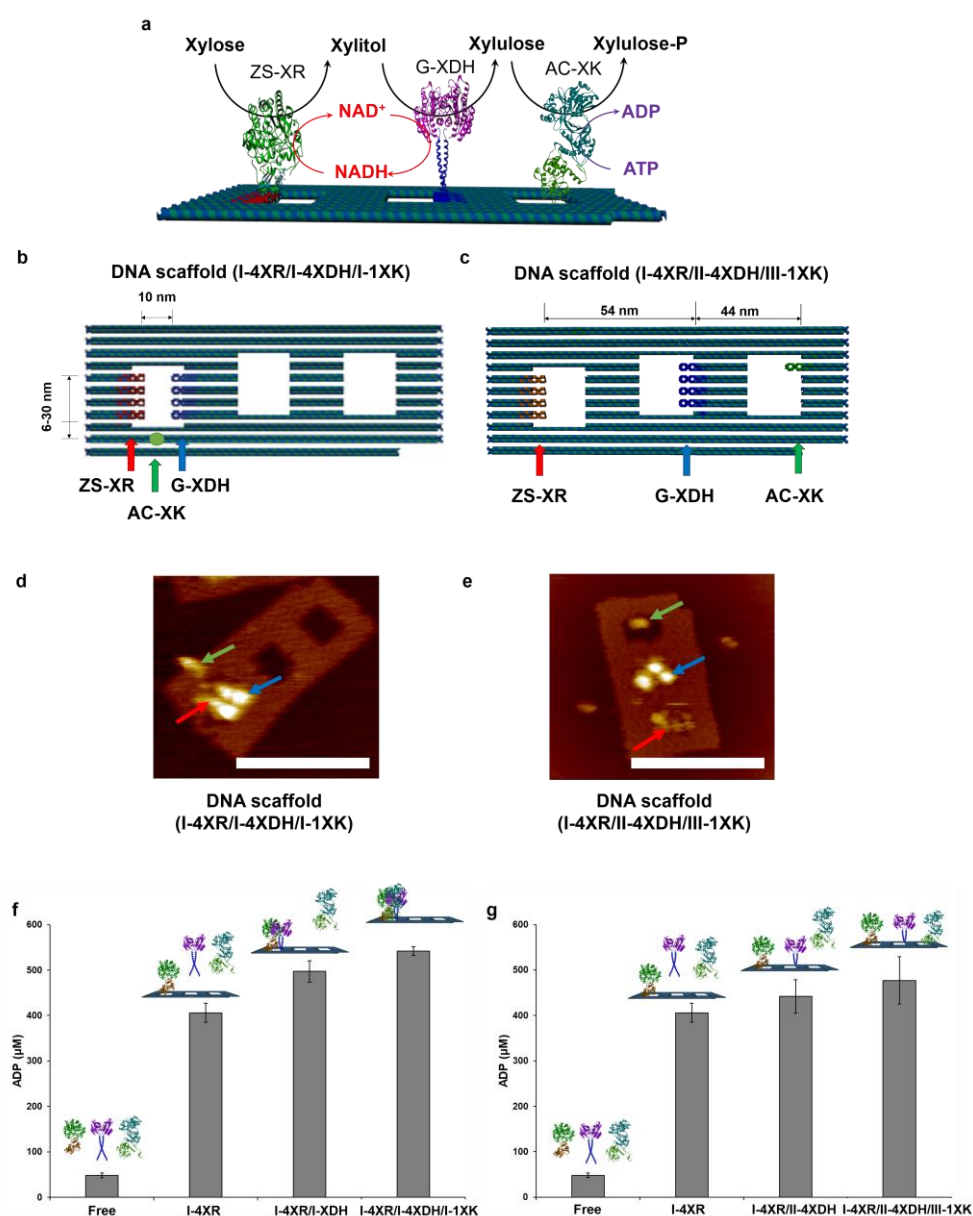


Figure 4.8. The three-step cascade reaction of xylose pathway activated on the DNA scaffold. (a) An illustration of the cascade reaction by three enzymes ZS-XR, G-XDH and AC-XK on the DNA scaffold. (b, c) Illustrations of DNA scaffolds (b) I-4XR/II-4XDH/III-1XK

and (c) I-4XR/I-4XDH/I-1XK with 4 binding sites for ZS-ZR, 4 binding sites for G-XDH and 1 binding site for XK. (d, e) AFM images of the three enzymes bound on the DNA scaffolds (d) I-4XR/II-4XDH/III-1XK and (e) I-4XR/I-4XDH/I-1XK. The arrows in red, blue, and green indicate ZS-XR, G-XDH, and AC-XK, respectively. The scale bars represent 100 nm. (f, g) Production of ADP from the cascade reaction with one enzyme assembled on DNA scaffold (I-4XR: ZS-XR located on the scaffold and G-XDH and AC-XK in bulk solution), with the two enzymes assembled on DNA scaffold [(f) I-4XR/II-4XDH or (g) I-4XR/I-4XDH: ZS-XR and G-XDH located on the scaffold and AC-XK in the bulk solution], and with the three enzyme assembled on DNA scaffold [(f) I-4XR/II-4XDH/III-1XK and (g) I-4XR/I-4XDH/I-1XK: ZS-XR, G-XDH and AC-XK located on the scaffold]. The analysis was conducted with a reaction mixture incubated for 24 hours with ZS-XR, G-XDH and AC-XK located on the scaffold (d) I-4XR/II-4XDH/III-1XK or (e) I-4XR/I-4XDH/I-1XK in a buffer (pH 7.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 0.02% Tween20, 100 mM NaCl, 2 mM NADH, and 1 mM ATP. HPLC chromatograms (detected by UV at 260 nm) for determination of the amount of cofactors in the three enzymes cascade reactions were shown in Figure 4.9. The results were summarized in Table 4.4.

Table 4.4. Quantitation of the cofactors by HPLC

On the scaffold	Scaffold name	In bulk solution	ATP (μ M)	ADP (μ M)	NADH (μ M)	NAD ⁺ (μ M)
None	Free	ZS-XR G-XDH AC-XK	952 $\pm$ 5	48 $\pm$ 5	1311 $\pm$ 21	679 $\pm$ 21
ZS-XR	I-4XR	G-XDH AC-XK	587 $\pm$ 18	406 $\pm$ 21	73 $\pm$ 51	1910 $\pm$ 45
ZS-XR G-XDH	I-4XR/II-4XDH	AC-XK	557 $\pm$ 37	442 $\pm$ 37	132 $\pm$ 26	1867 $\pm$ 37
ZS-XR G-XDH AC-XK	I-4XR/II-4XDH/III-1XK	-	522 $\pm$ 53	477 $\pm$ 52	116 $\pm$ 15	1883 $\pm$ 11
ZS-XR G-XDH	I-4XR/I-4XDH	AC-XK	510 $\pm$ 24	497 $\pm$ 29	81 $\pm$ 5	1921 $\pm$ 12
ZS-XR G-XDH AC-XK	I-4XR/I-4XDH/I-1XK	-	459 $\pm$ 8	542 $\pm$ 10	102 $\pm$ 6	1897 $\pm$ 10

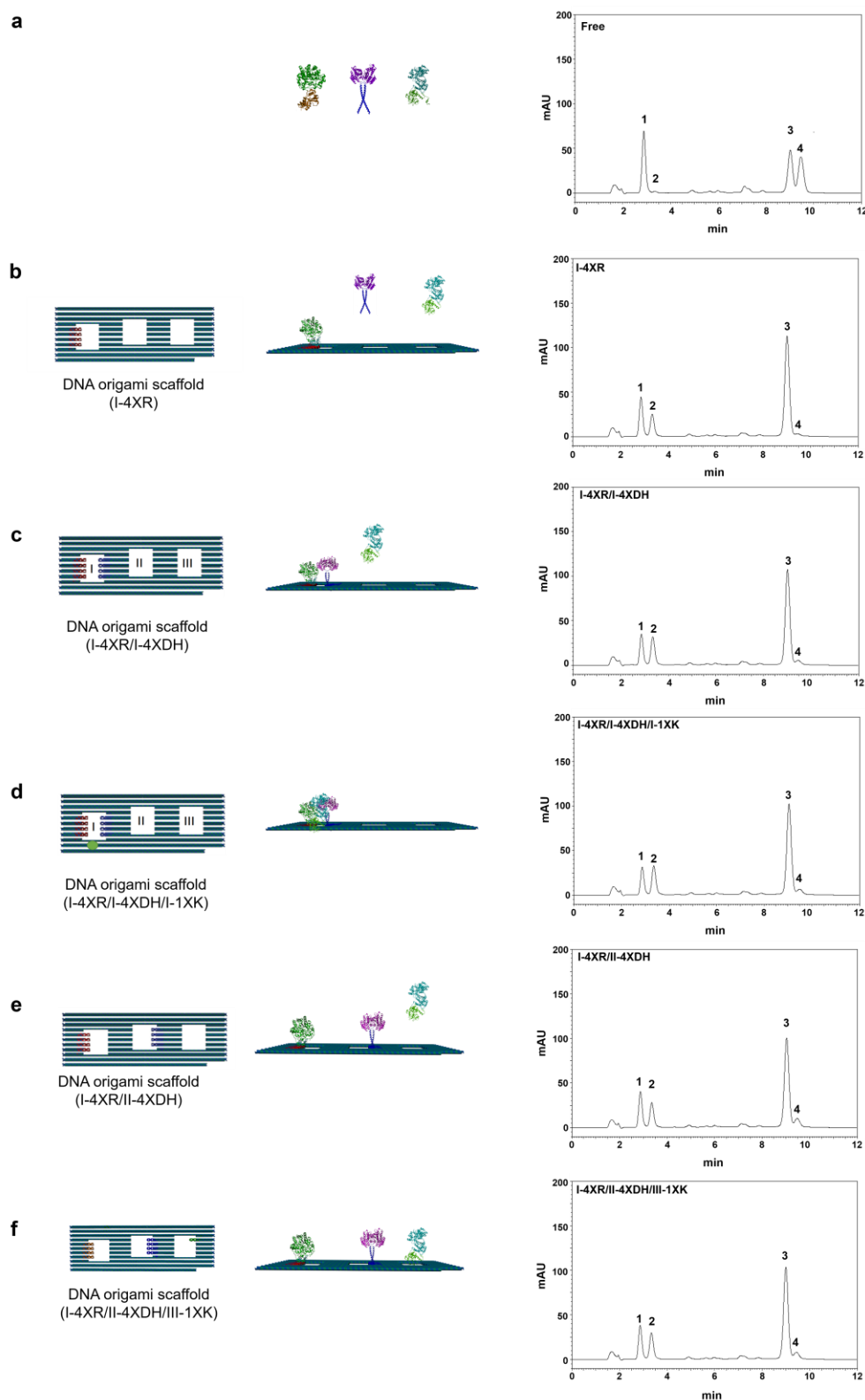


Figure 4.9. HPLC analyses (detected by UV at 260 nm) of cofactors in the three enzyme cascade reaction. Peaks 1, 2, 3 and 4 indicated ATP, ADP, NAD^+ and NADH, respectively. The analyses were carried out with the reaction mixture incubated for 24 hours (a) All enzyme

in bulk solution, (b) with ZS-XR located on the scaffold (I-4XR) and G-XDH and AC-XK in bulk solution, (c) with ZS-XR and G-XDH located on the scaffold (I-4XR/I-4XDH) and AC-XK in bulk solution, (d) ZS-XR, G-XDH and AC-XK located on the scaffold (I-4XR/I-4XDH/I-1XK), (e) with ZS-XR and G-XDH located on the scaffold (I-4XR/II-4XDH) and AC-XK in bulk solution, and (f) ZS-XR, G-XDH and AC-XK located on the scaffold (I-4XR/II-4XDH/III-1XK), respectively. Each reaction mixture for (a)-(f) contained 26 nM ZS-XR, 26 nM G-XDH and 6.5 nM AC-XK in a buffer (pH 7.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μM ZnCl₂, 0.02% Tween20, 2 mM NADH and 1 mM ATP. Reaction was started by addition of 200 mM xylose.

4.3 Discussion

Xylose is a major product of the hydrolysis of lignocellulosic biomass, which can be used by a variety of organisms to produce bioethanol (Figure 4.7a).²⁸ Attempt to industrializing the production of bioethanol by fermenting xylose was studied but still at the primary stage.^{20,29} The study in Morii laboratory have utilized the DNA nanotechnology to co-assemble the first two enzymes (xylose reductase and xylitol dehydrogenase) of the xylose pathway at close proximity to enhance the efficiency of cascade reaction. Xylose reductase and xylitol dehydrogenase were fused to the modular adaptors ZF-SNAP and GCN4 adaptor to form ZS-XR and G-XDH, respectively. Interenzyme distances of ZS-XR and G-XDH were accurately controlled on DNA scaffold for 10 nm, 54 nm, 98 nm, and 249 nm (G-XDH in free form). Investigation on the biomolecular transport of substrates to XDH from XR provides important insights that the interenzyme distance contribute more to the efficiency of consecutive reactions in the biomolecular transport system than that of unimolecular transporting enzyme cascade

In this chapter, I co-assembled the three enzymes (ZS-XR, G-XDH, and

AC-XK) derived from xylose pathway on the DNA scaffold. Several configurations of the assembly of three enzymes ZS-XR, G-XDH, and AC-XK on the 3-well DNA scaffold were investigated. The interenzyme distances were set approximately 50 nm (Figure 4.8d) or 10 nm (Figure 4.8e) for three enzymes or with one or two downstream enzymes in bulk solution (Figure 4.8, 4.9). The efficiency of the three enzymes cascade reaction is enhanced when these three enzymes were co-assembled on the DNA scaffold, and further enhanced when placed at close proximity. These results are consistent with the previous results of the two enzymes system that the interenzyme distance is one of the important parameters to regulate the efficiency of the cascade reaction by facilitating transport of intermediates. This study is the first investigation of the cascades reaction of three enzymes from the natural metabolic pathway on a DNA scaffold.

It is worth mentioning that the three enzymes system on DNA scaffold could be further optimized by changing the stoichiometry, orientation or placed in the space-confined scaffold for enhancing the efficiency of the production of the final product.

4.4 Conclusion

The modular adaptor was successfully applied to assemble xylulose kinase on the DNA scaffold. The cascade reaction of three enzymes of the xylose pathway was activated on a DNA scaffold. The efficiency of the cascade reaction was enhanced when the three enzymes were located at the close proximity, relevant to the previous study.^{15,18}

4.5 Materials and methods

4.5.1 Materials

The single-stranded M13mp18, restriction enzymes (NdeI and HindIII), pSNAP-tag (T7)-2 Vector (N9181S), BG-GLA-NHS (S9151S), BC-GLA-NHS (S9237S) were purchased from New England Biolabs. pCLIPf(T7)-2 Vector was generously donated from New England Biolabs. HaloTag 7 vector (pFN18K : G2681) and Halo-Tag Succinimidyl Ester (O2) Ligand (P1691) were purchased from Promega. Purified oligonucleotides as the staple strands for DNA origami, oligonucleotide primers for gene construction and all other oligonucleotides were obtained from Sigma-Aldrich (St. Louis, MO), Thermo Fisher Scientific Inc. (Waltham, MA, USA) or Gene Design Inc. (Osaka, Japan). *Escherichia coli* BL21 (DE3) competent cells were purchased from Invitrogen (Carlsbad, CA). Mini Elute Gel Extraction Kit was prepared from QIAGEN (Tokyo, Japan). HiTrap SP XL cation exchange column (5 ml), HisTrap HP column (5 ml), and Sephacryl S-400 were from GE Healthcare Japan Inc. (Tokyo, Japan). PrimeSTAR HS DNA polymerase, T4 DNA ligase, and E. Coli DH5 α competent cells were obtained from TaKaRa Bio Inc. (Shiga, Japan). Ultrafree-MC-DV was obtained from Merck Millipore (Darmstadt, Germany). COSMOSIL packed column PBr (4.6 i.d. $\times$ 150 mm), Cosmosil 5C18-MS II column (4.6 i.d. $\times$ 150 mm) and HPLC-grade acetonitrile were purchased from Nacalai Tesque (Kyoto, Japan). β -Nicotinamide adenine dinucleotide in reduced (NADH) and oxidized (NAD⁺) forms were obtained from Oriental Yeast (Tokyo, Japan). Adenosine triphosphate (ATP) was obtained from Sigma-Aldrich (St. Louis, MO). D-xylulose from Santa Cruz Biotechnology (Santa Cruz, LA). Xylose, gel electrophoresis grade acrylamide, bis(acrylamide), phenol, and all other chemicals and reagents were purchased from Wako Chemicals (Tokyo, Japan) or Nacalai Tesque (Kyoto, Japan). Phosphate buffer (PB) was prepared as 20 mM Na₂HPO₄ and 20 mM NaH₂PO₄. DNA origami buffer (pH 8.0) contained 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂.

4.5.2 Preparation of the modular adaptor fused enzymes

Overexpression and purification of the ZS-XR and G-XDH were described in the previous report.¹⁴ As a typical example, preparation of AC-XK is shown below. All of the vectors encoding protein-tag derivatives (pET-30a-XK, pET-30a-AC-XK, pET-30a-AH-XK) were prepared in a similar manner as described previously¹⁴ and the protein-tag derivatives (AC-XK, AH-XK) were overexpressed and purified in the same manner.

To construct the vector for AC-XK (pET-30a-AC-XK), AC-XK coding gene in pAC-XK (T7)-2 vector was amplified by PCR using the following primer pairs.

Forward primer (F_EcoRI_AC-XK):

TAATAAGAATTCGGCGGCTCCGGCGGCTCCGACAAAGATTGCGAAA

Reverse primer (R_AC-XK_HindIII):

TTATTAAAGCTTTTAATGATGGTGATGATGATGGTGATGATGGTGGGTACCATT
AACCTCGAGCCCGGGG.

The PCR products were run on a 1% agarose gel (TAE) and were purified by Mini Elute Gel Extraction Kit. The PCR products and pET-30a-AC-XK were digested with EcoRI and HindIII and were purified in the same manner. These products were incubated with T4-DNA-ligase. The mixture was transformed into *E. coli* DH5 α competent cells for amplification. The vector encoding AC-XK (termed as pET-30a-AC-XK) was purified, sequence checked, then transformed into *E. coli* BL21(DE3) competent cells.

4.5.3 Overexpression and purification of adaptor fused enzymes (ZS-XR, G-XDH, AC-XK, and AH-XK)

The transformed cells were grown at 37 °C until OD₆₀₀ reached 0.5, and protein expression was induced with 1 mM IPTG for 24 h at 25 °C. The soluble fraction of the cell lysate containing recombinant protein was loaded to HisTrap HP column equilibrated with a 50 mM phosphate buffer (pH 8.0) containing 200 mM NaCl and 1 mM DTT, and eluted by imidazole gradient. The main fractions containing the target protein were collected, loaded to HiTrap SP HP column equilibrated with a 50 mM phosphate buffer (pH 7.0) containing 1 mM DTT and eluted by NaCl gradient. The purified protein was dialyzed by using a 50 mM phosphate buffer (pH 8.0) containing 1 mM DTT, 50 µM ZnCl₂, and 50% glycerol and stocked at -20°C. The purity of target protein was checked by SDS-PAGE. The major band in SDS-PAGE corresponded to the calculated molecular weight with estimated purity of over 95% (Figure 4.3). The modular adaptors were characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, SA matrix). AC-XK: m/z calcd 100901, found 100933; AH-XK: m/z calcd 116473, found 116435. Amino acid sequences and calculated molecular weights of recombinant proteins used in this study are shown in Table 4.1.

4.5.4 Preparation of substrate modified ODNs

A coupling reaction between amino modified ODNs (100 µM) and the succinimidyl derivative of protein-tag substrate (1 mM) was carried out in a 100 mM phosphate buffer (pH 8.0) for 8 hours at ambient temperature. The substrate modified ODN was purified by reversed-phase HPLC on a Cosmosil 5C18-MS II column (4.6 x 150 mm, elution with 100 mM triethylammonium acetate buffer, pH 7.0, liner gradient over 30 min from 2.5 to 30% acetonitrile at flow rate of 1.0 mL.min⁻¹), and

characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu, HPA matrix). ODN-5j-AZ-BC: m/z calcd 23632, found 23624; Preparation of ODN-8e-AZ-CH and ODN-24D-AZ-BC were described in Chapter 2. The preparation of ODNs containing the binding sites for assembly of ZS-XR and G-XDH on DNA scaffolds showed in Figure 4.8, 4.9 was described in ref. 14.

Table 4.5. The staple strands including zif268 and AZP4 binding sites with substrates modified T (T^{BG} , T^{CH} , T^{BC} for T modified with BG, BC, and CH, respectively). The zif268 and AZP4 binding sites on the staple strands were colored in red and blue, respectively.

Oligo name	Sequence (from 5' to 3')
5j-AZ-BC	GCTGAGAGACAGACAACTTATGCCACGTAGCGTT ^{BC} TTCGCTACGTGGC ATAAG TATTTTAAACGCTCATGGAAATA
8E-AZ-CH	CTACTAACTTATGCCACGTAGCGTT ^{CH} TTCGCTACGTGGCATAAG TGACCATTAGATACAACGAGTAGA
11g-ZF-BG	AACAGGTCCTTACGCCACGTAGCGTT ^{BG} TTCGCGCGTGGCGTAAG GAACCAGACCGGAAGATTCGAGC
24D-AZ-BC	GGACAGATCTTATGCCACGTAGCGTT ^{BC} TTCGCTACGTGGCATAAG AAATTGTGTCGAAATCTGTATCAT

4.5.5 Preparation of the DNA origami scaffold with three cavities

A solution (50 μ L) containing M13mp18 single-stranded DNA (New England Biolabs, 10 nM), staple DNA strands (5 equiv, 50 nM, the nucleotide sequences of all the staple DNA strands were shown in Chapter 2 and Table 3.4) in a buffer (40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, pH 8.0) was heated at 95 $^{\circ}C$ for 1 min, annealed at 53 $^{\circ}C$ for 30 min, and kept at 4 $^{\circ}C$ by using a thermal cycler. The samples were purified by size-exclusion chromatography (400 μ L volume of Sephacryl S-400, GE Healthcare) equilibrated with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM $MgCl_2$ in Ultrafree-MC-DV (Millipore).

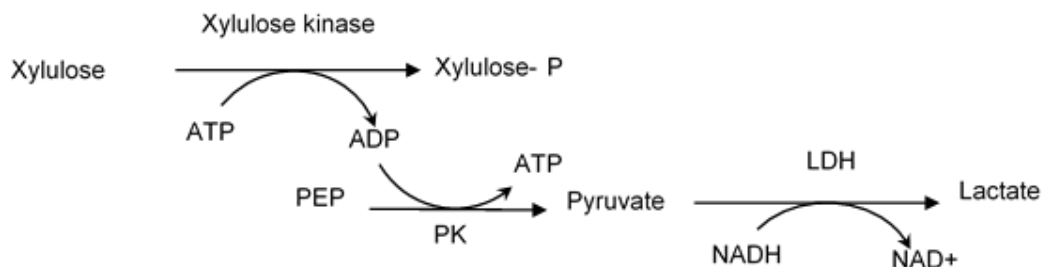
4.5.6 Preparation of DNA origami scaffold assembled with modular adaptor fused enzymes.

A DNA origami scaffold was incubated with AC-XK, AH-XK, ZS-XR, and G-XDH under the conditions shown in the caption of figures or tables. For example, the crosslinking reaction of each of modular adaptor was carried out for 30 minutes at ambient temperature with 5 nM DNA scaffold and 100 nM AC-XK in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, and 0.02% Tween20. The mixture was purified by size-exclusion chromatography (400 μ L volume of Sephacryl S-400 in Ultrafree-MC-DV) equilibrated with the buffer used for the reaction to remove the unbound modular adaptors. The fractions containing DNA origami scaffold were utilized for AFM analyses.

4.5.7 AFM imaging and statistical analysis

The sample was deposited on freshly cleaved mica surface (1.5 mm ϕ) and adsorbed for 5 min at ambient temperature, then washed three times with a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, and 12.5 mM MgCl₂. The sample was scanned in tapping mode using a fast-scanning AFM system (Nano Live Vision, RIBM Co. Ltd., Tsukuba, Japan) with a silicon nitride cantilever (Olympus BL-AC10DS-A2). At least three independent preparations of each sample were analyzed by AFM and several images were acquired from different regions of the mica surface. The total number of DNA scaffolds corresponds to the number of expected rectangular shapes possessing three cavities observed by AFM. The specific and nonspecific binding of modular adaptors were counted for only the modular adaptor bound to the perfectly folded DNA scaffold. The yield of DNA-scaffold-assembled modular adaptor fused enzyme was calculated as in Chapter 2.

4.5.8 Xylulose kinase assay



Scheme 4.1. Coupled enzyme assay for studying the activity of XK derivatives

The catalytic activity of xylulose kinase derivatives (wild-type XK, AC-XK, and AH-XK) were analyzed by the coupled enzyme assay (Scheme 4.1) with pyruvate kinase (PK) and lactate dehydrogenase (LDH) as described in previous report with slight modification.^{21,30,31} The XK activity was measured by monitoring NADH oxidation by LDH at 340 nm with Shimadzu UV-vis spectrophotometer UV-1700. The assays were performed in a buffer (pH 7.6) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl₂, 1 mM DTT, 1 μ M ZnCl₂, 0.02% Tween20, 100 mM NaCl, 1.1 mM ATP, 2.3 mM phosphoenolpyruvate (PEP), 0.2 mM NADH, 4.8 U/ml PK, 4.5 U/ml LDH and the indicated concentration of xylulose (0.03 to 2 mM). The reaction was started by an addition of xylulose and carried out at 25 °C. The kinetic parameters K_m and k_{cat} were determined by fitting to the Michaelis-Menten plot (Figure 4.4 and Table 4.2).

4.5.9 Analysis of the enzyme cascade reaction

An enzyme cascade reaction catalyzed by three enzymes (ZS-XR, G-XDH, and AC-XK) was designed and analyzed on the three cavities DNA scaffold (Figure 4.8b, 4.8c and Figure 4.9) with or without the binding sites for ZS-XR at cavity I (4 binding sites), G-XDH at cavity I or II (4 binding sites) and AC-XK at cavity I or III (1 binding

site). The calculated distance between the binding site of ZS-XR and G-XDH represented as $d(\text{ZS-XR/G-XDH})$ and G-XDH and AC-XK represented as $d(\text{G-XDH/AC-XK})$ are determined as remarked below (Figure 4.8b, 4.8c). DNA scaffold (I-4XR/II-4XDH/III-1XK): $d(\text{ZS-XR/G-XDH}) = 54 \text{ nm}$, $d(\text{G-XDH/AC-XK}) = 44 \text{ nm}$; DNA scaffold (I-4XR/I-4XDH/I-1XK): $d(\text{ZS-XR/G-XDH}) = 10 \text{ nm}$, $d(\text{G-XDH/AC-XK}) = 6\text{--}30 \text{ nm}$. The reaction was started with an addition of xylose (200 mM) to a mixture of ZS-XR, G-XDH, and/or AC-XK located on the DNA scaffold (6.5 nM) (the total concentration of ZS-XR, G-XDH and AC-XK were set to 26 nM, 26 nM and 6.5 nM, respectively) in a buffer (pH7.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM MgCl_2 , 1 mM DTT, 1 μM ZnCl_2 , 0.02% Tween20, 100 mM NaCl, 2 mM NADH, and 1 mM ATP. The amounts of cofactors (ATP, ADP, NADH, and NAD^+) were monitored by HPLC at 260 nm. HPLC conditions: COSMOSIL packed column PBr (4.6 i.d. $\times$ 150 mm) eluent A: 10% methanol in 20 mM phosphate buffer (pH 7.0); eluent B: 50% methanol in 20 mM phosphate buffer (pH 7.0); gradient of eluent B increased from 0% to 20% in 1 min, and to 30% in 10 min; flow rate of 1 mL/min.

4.6 References

- (1) Pavletich, N. P.; Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science*. **1991**, 252, 809–817.
- (2) Sera, T.; Uranga, C. Rational Design of Artificial Zinc-Finger Proteins Using a Nondegenerate Recognition Code Table. *Biochemistry* **2002**, 41, 7074–7081.
- (3) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A General Method for the Covalent Labeling of Fusion Proteins with Small Molecules in Vivo. *Nat. Biotechnol.* **2003**, 21, 86–89.

- (4) Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **2008**, *15*, 128–136.
- (5) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; et al. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.
- (6) Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, *99*, 237–247.
- (7) Li, J.; Chao, J.; Shi, J.; Fan, C. Cotranscriptionally Folded RNA Nanostructures Pave the Way to Intracellular Nanofabrication. *ChemBioChem* **2015**, *16*, 39–41.
- (8) Zhang, F.; Jiang, S.; Wu, S.; Li, Y.; Mao, C.; Liu, Y.; Yan, H. Complex Wireframe DNA Origami Nanostructures with Multi-Arm Junction Vertices. *Nat. Nanotechnol.* **2015**, *10*, 779–784.
- (9) Benson, E.; Mohammed, A.; Gardell, J.; Masich, S.; Czeizler, E.; Orponen, P.; Högberg, B. DNA Rendering of Polyhedral Meshes at the Nanoscale. *Nature* **2015**, *523*, 441–444.
- (10) Dunn, K. E.; Dannenberg, F.; Ouldrige, T. E.; Kwiatkowska, M.; Turberfield, A. J.; Bath, J. Guiding the Folding Pathway of DNA Origami. *Nature* **2015**, *525*, 82–86.
- (11) Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **2012**, *134*, 5516–5519.
- (12) Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter, N. G.; Woodbury, N. W.; Yan, H. Multi-Enzyme Complexes on DNA Scaffolds Capable of

- Substrate Channelling with an Artificial Swinging Arm. *Nat. Nanotechnol.* **2014**, 9, 531–536.
- (13) Liu, M.; Fu, J.; Qi, X.; Wootten, S.; Woodbury, N. W.; Liu, Y.; Yan, H. A Three-Enzyme Pathway with an Optimised Geometric Arrangement to Facilitate Substrate Transfer. *ChemBioChem* **2016**, 17, 1097–1101.
- (14) Ngo, T. A.; Nakata, E.; Saimura, M.; Morii, T. Spatially Organized Enzymes Drive Cofactor-Coupled Cascade Reactions. *J. Am. Chem. Soc.* **2016**, 138, 3012–3021.
- (15) Rajendran, A.; Nakata, E.; Nakano, S.; Morii, T. Nucleic-Acid-Templated Enzyme Cascades. *ChemBioChem* **2017**, 18, 696–716.
- (16) Niemeyer, C. M.; Koehler, J.; Wuerdemann, C. DNA-Directed Assembly of Bienzymic Complexes from in Vivo Biotinylated NAD(P)H:FMN Oxidoreductase and Luciferase. *ChemBioChem* **2002**, 3, 242–245.
- (17) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, 4, 249–254.
- (18) Linko, V.; Nummelin, S.; Aarnos, L.; Tapio, K.; Toppari, J.; Kostiainen, M. DNA-Based Enzyme Reactors and Systems. *Nanomaterials* **2016**, 6, 139.
- (19) Zhao, Z.; Fu, J.; Dhakal, S.; Johnson-Buck, A.; Liu, M.; Zhang, T.; Woodbury, N. W.; Liu, Y.; Walter, N. G.; Yan, H. Nanocaged Enzymes with Enhanced Catalytic Activity and Increased Stability against Protease Digestion. *Nat. Commun.* **2016**, 7, 10619.
- (20) Kricka, W.; Fitzpatrick, J.; Bond, U. Metabolic Engineering of Yeasts by Heterologous Enzyme Production for Degradation of Cellulose and Hemicellulose from Biomass: A Perspective. *Front. Microbiol.* **2014**, 5, 574.

- (21) Eliasson, A.; Hofmeyr, J. H. S.; Pedler, S.; Hahn-Hägerdal, B. The Xylose Reductase/Xylitol Dehydrogenase/Xylulokinase Ratio Affects Product Formation in Recombinant Xylose-Utilising *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* **2001**, 29, 288–297.
- (22) Fujiwara, K.; Adachi, T.; Doi, N. Artificial Cell Fermentation as a Platform for Highly Efficient Cascade Conversion. *ACS Synth. Biol.* **2018**, 7, 363–370.
- (23) Huang, X.; Li, M.; Mann, S. Membrane-Mediated Cascade Reactions by Enzyme-Polymer Proteinosomes. *Chem. Commun.* **2014**, 50, 6278–6280.
- (24) Pival, S. L.; Birner-Gruenberger, R.; Krump, C.; Nidetzky, B. D-Xylulose Kinase from *Saccharomyces cerevisiae*: Isolation and Characterization of the Highly Unstable Enzyme, Recombinantly Produced in *Escherichia coli*. *Protein Expr. Purif.* **2011**, 79, 223–230.
- (25) Jeffries, T. W. Engineering Yeasts for Xylose Metabolism. *Curr. Opin. Biotechnol.* **2006**, 17, 320–326.
- (26) Matsushika, A.; Inoue, H.; Kodaki, T.; Sawayama, S. Ethanol Production from Xylose in Engineered *Saccharomyces cerevisiae* Strains: Current State and Perspectives. *Appl. Microbiol. Biotechnol.* **2009**, 84, 37–53.
- (27) Ngo, T. A.; Nakata, E.; Saimura, M.; Kodaki, T.; Morii, T. A Protein Adaptor to Locate a Functional Protein Dimer on Molecular Switchboard. *Methods* **2014**, 67, 142–150.
- (28) Winkelhausen, E.; Kuzmanova, S. Microbial Conversion of D-Xylose to Xylitol. *J. Ferment. Bioeng.* **1998**, 86, 1–14.
- (29) Wei, N.; Oh, E. J.; Million, G.; Cate, J. H. D.; Jin, Y. S. Simultaneous Utilization of Cellobiose, Xylose, and Acetic Acid from Lignocellulosic Biomass for Biofuel Production by an Engineered Yeast Platform. *ACS Synth. Biol.* **2015**, 4, 707–713.

- (30) Shamanna, D. K. K.; Sanderson, K. E. E. Uptake and Catabolism of D-Xylose in *Salmonella Typhimurium* LT2. *J. Bacteriol.* **1979**, *139*, 64–70.
- (31) Katahira, S.; Mizuike, A.; Fukuda, H.; Kondo, A. Ethanol Fermentation from Lignocellulosic Hydrolysate by a Recombinant Xylose- and Cellooligosaccharide-Assimilating Yeast Strain. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 1136–1143.

CHAPTER 5

General conclusion

In this thesis, a new method for orthogonally and simultaneously assembling multiple enzymes through modular adaptors on DNA scaffold was developed. By applying the modular adaptors, an assembly of three enzymes driving the metabolic cascade reaction was studied. This method is based on the modularization of the protein adaptor to contain two functional modules: the position targeting module (using the sequence-specific DNA binding protein) and the covalent bond formation module (using self-ligating protein tag).

In Chapter 2, I first expanded the combinations of DNA binding protein (ZF, AZ) and the protein-tag (SNAP-tag, CLIP-tag, and Halo-tag) to select an orthogonal set of modular adaptors to locate more than two types of enzymes on a DNA scaffold. The denaturing PAGE and AFM imaging analyses were used to study the orthogonal reaction by four types of modular adaptors (ZF-SNAP, AZ-SNAP, AZ-CLIP, and AZ-Halo). The study indicated the modular adaptors with the same DNA binding protein (AZ-SNAP, AZ-CLIP, and AZ-Halo) showed good orthogonality; the better result was obtained by the combination of ZF-SNAP, AZ-CLIP, and AZ-Halo, where the orthogonality was governed by both the substrate selectivity of protein-tags and the sequence specificity of the DNA binding modules. AFM imaging showed the co-assembly of ZF-SNAP, AZ-CLIP, and AZ-Halo were more than 90% in a very short incubation time (5 min incubation), which is the highest yield among the currently available methods for co-assembling three proteins on a DNA scaffold.

Chapter 3 focused on a more general approach for developing the method to assemble a number of POIs orthogonally on a DNA scaffold. Kinetic aspects of covalent bond formation between the modular adaptor and its substrate modified oligodeoxynucleotide (ODN) showed a new strategy for expanding the number of orthogonal modular adaptors by combining the same protein-tag with difference

sequence-specific DNA binding proteins. This next generation of modular adaptors was termed as “sequence-specific modular adaptors”, in which the covalent bond formation occurred exclusively at the substrate modified ODN containing the specific sequence of the DNA binding module. The sequence-specific reaction of modular adaptors depends on the balance between the binding rate constant of the DNA binding protein and the rate constant for the covalent bond formation of the protein-tag module. In Chapter 3, I succeed to construct three orthogonal modular adaptors (ZF-CLIP, AZ-CLIP, and CLIP-G) that were derived from three DNA binding proteins (ZF, AZ, and GCN4) and a single protein-tag (CLIP-tag).

I next tested the robustness of the orthogonal modular adaptors by using AZ-CLIP (AC) and AZ-Halo (AH) to assemble the unstable enzyme xylulose kinase (XK) on DNA scaffold (Chapter 4). Indeed, I succeed to assemble and to characterize XK in the form of adaptor fused enzyme (AC-XK and AH-XK). I further constructed an artificial xylose metabolism pathway of three adaptor fused enzymes: ZS-XR (ZF-SNAP fused xylose reductase), G-XDH (GCN4 fused xylitol dehydrogenase), and AC-XK on a DNA scaffold. The efficiency of the three enzyme cascade reaction was significantly enhanced when these three enzymes were placed at close proximity. This study is the first study to investigate the cascade reaction of three enzymes derived from the nature metabolism pathway on a DNA scaffold.

The method developed in this thesis offers an efficient technique to assemble multiple enzymes on a DNA scaffold. This method facilitates the study of the artificial cascade metabolism pathway at the close proximity on a DNA nanoscaffold. The number of modular adaptors library is easily expanded to construct more complicated artificial metabolic pathways on a DNA scaffold, such as a surface-confined or a three-dimension DNA scaffold. The modular adaptors are also useful for assembling protein complexes

on a design DNA scaffold mimicking the *in vivo* system to study the spatial and stoichiometric effect of protein compartmentalization that occurred inside the cell.

List of publications

1. Design of modular protein-tags for the orthogonal covalent bond formation at specific DNA sequences

Thang Minh Nguyen, Eiji Nakata, Masayuki Saimura, Huyen Dinh, and Takashi Morii

J. Am. Chem. Soc. **2017**, *139*, 8487

(Chapter 2, 4)

2. Development of sequence specific modular adaptors

Thang Minh Nguyen, Eiji Nakata, Masayuki Saimura, and Takashi Morii

To be submitted

(Chapter 3)

Other Publication

3. DNA binding adaptors to assemble protein of interest on a DNA scaffold.

Eiji Nakata, Huyen Dinh, Thang Minh Nguyen, and Takashi Morii.

Methods Enzymol. vol. 617, **2019** (*in press*)

4. Review article: DNA-binding proteins as the potential adaptors to locate functional protein on DNA nanoscaffold

Tien Anh Ngo, Huyen Dinh, Thang Minh Nguyen, Fong Fong Liew, Eiji Nakata, and Takashi Morii

Manuscript in preparation

List of presentations

International conference

1. Expansion of modular adaptors for covalently locating multiple enzymes on DNA nanoscaffold

Thang Minh Nguyen, Eiji Nakata, Masayuki Saimura, and Takashi Morii

The 42nd International Symposium on Nucleic Acids Chemistry (ISNAC2015),
2015.9.24, Himeji, Japan

2. Site-specific covalent modification of individual protein molecule on DNA nanoscaffold

Eiji Nakata, Huyen Dinh, Thang Minh Nguyen, Tien Anh Ngo, Masayuki Saimura,
Takashi Morii The 42nd International Symposium on Nucleic Acids Chemistry
(ISNAC2015), 2015.9.24, Himeji, Japan

3. Development of modular adaptors to covalently locate multiple enzymes on DNA nanoscaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo,
Masayuki Saimura, Takashi Morii

2015 International Chemical Congress of Pacific Basin Societies (Pacifichem2015),
Hawaii, USA

4. Development of various types of protein adaptors to locate a single molecule of functional protein on molecular switchboard

Tien Anh Ngo, Eiji Nakata, Huyen Dinh, Thang Minh Nguyen, Masayuki Saimura,
Takashi Morii

2015 International Chemical Congress of Pacific Basin Societies (Pacifichem2015),
Hawaii, USA

5. Evaluation of orthogonal modular adaptors for assembling proteins on DNA scaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo, Masayuki Saimura,

Takashi Morii

The 43rd International Symposium on Nucleic Acids Chemistry (ISNAC2016),
2016.9.27-29, Kumamoto, Japan

6. Expansion of modular adaptors for covalently locating a number of proteins on DNA nanostructure

Eiji Nakata, Thang Minh Nguyen, Masayuki Saimura, Takashi Morii

Institute for Chemical Research International Symposium 2016 (ICRIS'16),
2016.3.7-8. Uji, Japan.

7. Development of modular adaptors to covalently locate multiple enzymes on DNA nanoscaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo, Masayuki Saimura,
Takashi Morii

Kyoto - Ajou - Zhejiang Joint Symposium on Energy Science 2017.1, Uji, Japan

8. Assembly of multiple enzymes on DNA scaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo, Masayuki Saimura,
Takashi Morii

The Second International Symposium on Biofunctional Chemistry (ISBC2017),
2017.12.14-16, Uji, Japan

9. Orthogonal modular adaptors to locate three enzymes for cascade reaction on DNA scaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Masayuki Saimura, Takashi Morii

The 44rd International Symposium on Nucleic Acids Chemistry (ISNAC2017),
2017.11.14-16, Tokyo, Japan

Domestic conference

10. Development of orthogonal modular adaptors for assembling multiple proteins

on DNA nanostructure

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo, Masayuki Saimura, Takashi Morii

日本化学会第 96 春季年会, 2016.3.24-27., Kyoto, Japan

11. Orthogonal modular adaptors for assembling multiple enzymes on DNA scaffold

Thang Minh Nguyen, Eiji Nakata, Huyen Dinh, Tien Anh Ngo, Masayuki Saimura, Takashi Morii

日本化学会第 97 春季年会, 2017-3.16-20, Yokohama, Japan

12. DNA ナノ構造体に酵素を配置した分子スイッチボード

中田栄司, Ngo Anh Tien, Dinh Huyen, Nguyen Minh Thang, 才村正幸, 森井 孝

第 10 回バイオ関連化学シンポジウム, 2016.9.7-9, Kanazawa, Japan

13. Enzymatic cascade reactions on the DNA scaffold

Eiji Nakata, Huyen Dinh, Thang Minh Nguyen, Masayuki Saimura, Takashi Morii

日本化学会第 97 春季年会, 2017-3.16-20, Yokohama, Japan

14. Sequence specific modular adaptors for orthogonal location of multiple enzymes on a DNA scaffold

Thang Nguyen, Eiji Nakata, Masayuki Saimura, Takashi Morii

日本化学会第 98 春季年会, 2017-3.20-23, Chiba, Japan

15. efficiency of the enzyme cascade reaction on a DNA scaffold

Lin Peng, Thang Nguyen, Huyen Dinh, Eiji Nakata, Masayuki Saimura, Takashi Morii

日本化学会第 98 春季年会, 2017-3.20-23, Chiba, Japan

List of awards

1. Student Award 2017

Institute of Advanced Energy Science, Kyoto University.

2. CSJ Student Presentation Award 2017

The 97th Annual Meeting of the Chemical Society for Chemistry (CSJ) at Hiyoshi Campus, Keio University, Yokohama, Japan, March 2017.