Protein adaptors assemble functional proteins on DNA scaffolds

Journal:	*ChemComm*
Manuscript ID	CC-FEA-06-2019-004661.R1
Article Type:	Feature Article

SCHOLARONE™
Manuscripts

Protein adaptors assemble functional proteins on DNA scaffolds

Tien Anh Ngo[1], Huyen Dinh[2], Thang Minh Nguyen[2], Fong Fong Liew[3], Eiji Nakata[2], and Takashi Morii[2*]

[1]Vinmec Biobank, Hi-tech Center, Vinmec Healthcare System, 458 Minh Khai, Ha Noi, Viet Nam

[2]Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan

[3]MAHSA University, Faculty of Dentistry, Bandar Saujana Putra, 42610 Jenjarom, Selangor, Malaysia.

[*]Corresponding Author: T. Morii: t-morii@iae.kyoto-u.ac.jp

Abstract: DNA is an attractive molecular building block to construct nanoscale structures for a variety of applications. In addition to their structure and function, modification the DNA nanostructures by other molecules opens almost unlimited possibilities for producing functional DNA-based architectures. Among the molecules to functionalize DNA nanostructures, proteins are one of the most attractive candidates due to their vast functional variations. DNA nanostructures loaded with various types of proteins hold promise for applications in the life and material sciences. When loading proteins of interest on DNA nanostructures, the nanostructures by themselves act as scaffolds to specifically control the location and number of protein molecules. The methods to arrange proteins of interest on DNA scaffolds at high yields while retaining their activity are still the most demanding task in constructing usable protein-modified DNA nanostructures. Here, we provide an overview of the existing methods applied for assembling proteins of interest on DNA scaffolds. The assembling methods were categorized into two main classes, noncovalent and covalent conjugation, with both showing pros and cons. The recent advance of DNA-binding adaptor mediated assembly of proteins on the DNA scaffolds is highlighted and discussed in connection with the future perspectives of protein assembled DNA nanoarchitectures.

1.   Introduction

Over the past thirty years, DNA has been considered as one of the most promising biomaterials to build programmable 2D and 3D nanostructures. DNA possesses unique properties, including the high specificity and predictable Watson-Crick base pairing (A-T and G-C), well defined conformations of the DNA double helix, e.g., a diameter of ~2 nm and ~3.4 nm per helical turn in B-DNA, and the convenient synthesis of oligodeoxynucleotides (ODN), that enable the design and construction of DNA nanostructures with nanometer precision. These characteristics have stimulated the continuous development of the field of structural DNA nanotechnology[1-10]. Particularly, in the early 1980s, Seeman used double helical DNA molecules to construct artificial 'Holiday' junction tiles[11] and double-crossovers (DX)[12]. Using these structural units, various periodic nanostructures with distinct topological and geometric features have been constructed[13]. In 2006, Rothemund reported DNA origami[14], which utilized hundreds of short ODN (staple strands) to fold a long single-stranded scaffold into a designed target shape by one-pot annealing. The DNA origami method has provided unlimited design and fabrication of spatially addressable 1D, 2D, and 3D nanostructures (**Figure 1**)[4-10]. DNA nanostructures have recently emerged as ideal scaffolds for spatially organizing functional molecules with nanometer precision by taking advantage of their sequence-driven programmability as comprehensively overviewed in some excellent reviews.[15-17] With almost unlimited functionality, such as molecular recognition, catalytic turnover, energy conversion, and translocation of ligands and ions across membranes, proteins are one of the most fascinating molecules that can be used to functionalize DNA nanostructures. Accurately controlling the positioning of proteins on DNA nanostructures has allowed us to investigate many processes, including the role of enzyme spatial organization in natural or artificial enzyme cascades[15-18].
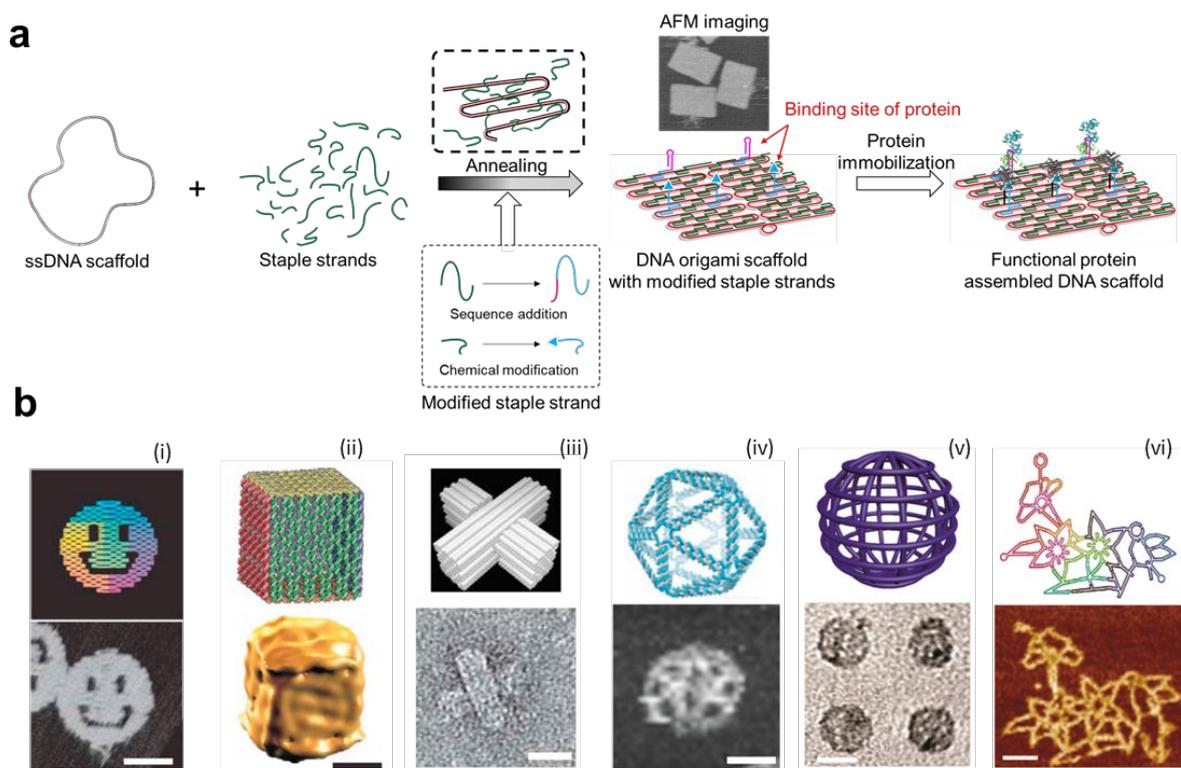
Figure 1. (a) A general schematic for the construction of a DNA origami scaffold containing protein binding sites, and the protein-assembled DNA scaffold. (b) Examples of constructed DNA origami with different patterns including: (i) DNA origami smiley face[14], (ii) cube DNA origami[19], (iii) a slotted cross constructed from a honeycomb DNA lattice[20], (iv-vi) wireframe DNA origami with different DNA rendering units[21–23]. Scale bars, (i-iii) 20 nm, (iv-vi) 50 nm. Reproduced from ref. 14, 19, 20, 22, 23 with permission from Springer Nature, copyright 2006, 2009, and 2015 and from ref. 21 with permission from AAAS, copyright 2016.

A number of methods have been used to arrange proteins of interest (POI) on DNA scaffolds[24-28]. Some of them show promising properties for functionalizing DNA scaffolds while retaining protein activity. The most widely used method for assembly of POI on DNA scaffolds relies on DNA-hybridization using POI covalently conjugated to ODN[29]. Other methods depend on noncovalent protein-DNA interactions, such as the biotin-avidin interaction[30–32], antibody-antigen interaction[33–36], Ni-NTA-hexahistidine interaction[37,38], reconstruction of apo-protein and DNA modified cofactors[39], aptamer binding[40–44], and

4

specific-sequence DNA-binding proteins[45-49], and are applicable depending on the required conditions. Crosslinking of genetically fused proteins and DNA, such as the reaction of a protein-tag and DNA modified with its substrate[50], is also applicable. From these methods, depending on the POI and on the particular DNA nanostructure, the most suitable is selected. Both the efficiency of POI loading and their activity once assembled on the DNA nanostructure significantly depend on the method that has been used to attach the POI on the DNA scaffold. In some cases, the direct conjugation of POI to the DNA scaffold is preferred (**Figure 2a**), although often the POI will have been modified with an ODN beforehand and hybridized to the complementary DNA sequence on the DNA scaffold (**Figure 2b**). However, methods to quantitatively arrange the POI in its functional form to DNA scaffolds remain to be established. In this article, we give an overview of the currently available methods to precisely arrange proteins on DNA scaffolds while retaining their native states. We discuss recent advances in the area of DNA-binding adaptors that specifically arrange proteins in their active forms on DNA nanostructures and their application for studying enzyme cascades.
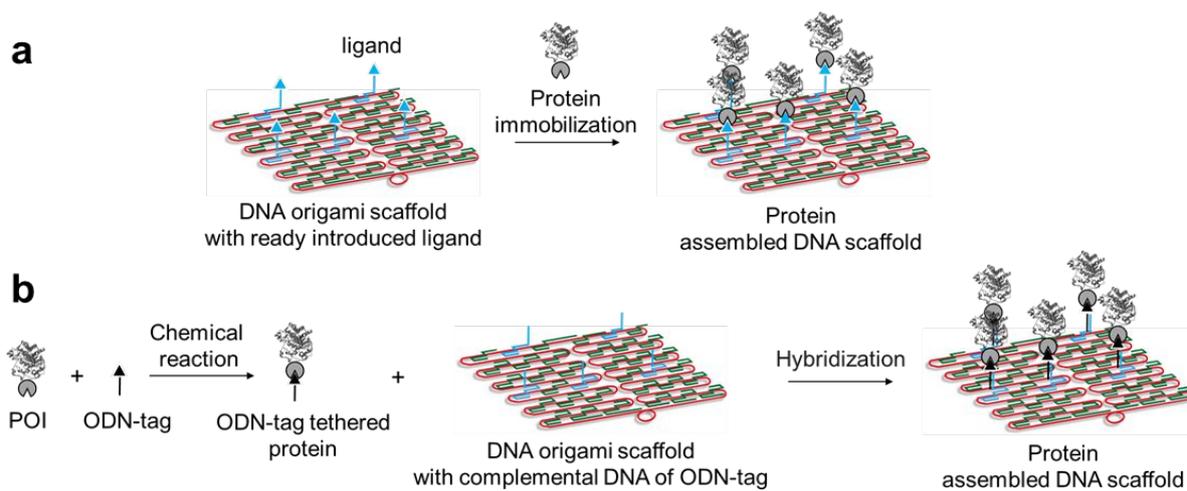
Figure 2. Classification of the protein immobilization processes. (a) Direct immobilization of the protein of interest (POI) at target positions on the DNA scaffold. (b) An oligodeoxynucleotide (ODN)-tag tethered POI assembled by hybridization to the complementary DNA sequence on the DNA scaffold.

2.   Common methods for arranging proteins on DNA scaffolds

To arrange a protein on a DNA scaffold, modification of the DNA or protein, or both of them, is required. In contrast to the chemical modification of DNA, such as the addition of a functional group at its terminals, the chemical modification of the POI often encounters difficulty because the modification processes have to be carried out in mild conditions to retain their activity. Alternatively, the POI is genetically fused to a handle module, such as a receptor protein. Until now, many methods for the conjugation of POI to DNA have been reported[24-28]. These methods are divided into two main classes, namely noncovalent (reversible) and covalent (irreversible) conjugations (**Table 1** and **2**). Each method has its own pros and cons, and thus, the choice of conjugation method depends on the intended use.

Table 1. Classification of protein-DNA conjugation methods (noncovalent (reversible) type)

Name of method	DNA modification	Protein modification	Dissociation constant (K_D)
Biotin-avidin[30–32]	biotin	No need (required to conjugate avidin and POI)	$\sim 10^{-15}$ M [30,31]
Antigen-antibody[33–36]	antigen	No need (required to conjugate antibody and POI)	10^{-11} to 10^{-4} M [34]
Ni-NTA-hexahistidine[37,38]	Ni-NTA	Genetic modification of POI with hexahistidine	10^{-9} to 10^{-7} M [38]
Reconstruction of apo protein and DNA modified cofactor[39]	cofactor	No need (limitation for cofactor dependent protein)	$\sim 10^{-9}$ M [39]
Aptamer for protein[40-44]	No need (hybridization)	No need (limited number of aptamers for protein)	10^{-12} to 10^{-7} M [40,42]
Specific-sequence DNA and DNA-binding protein[45-49]	Addition of the specific-sequence for DNA-binding protein	Genetic modification of POI with DNA-binding protein	10^{-15} to 10^{-7} M [47]

2.1. Noncovalent protein-DNA interactions

The representative methods of noncovalent protein-DNA conjugation (**Figure 3**) include the specific binding of a ligand and protein, such as biotin-avidin[30–32], antigen-antibody[33–36], Ni-NTA-hexahistidine[37,38], reconstruction of apo-protein and DNA modified cofactor[39], aptamer for protein[40-44], and specific-sequence DNA and DNA-binding protein[45-49]. In general, these noncovalent conjugation methods depend on strong interactions and high selectivity between the ligand tethered to DNA and the protein receptor, or vice versa. Thus, the POI is directly assembled at a designed position on the DNA scaffold (**Figure 2a**). However, despite the strong binding affinities of these interactions, their reversible nature can cause difficulty in retaining the bound POI on the DNA scaffold as they are readily detached from their target positions under the equilibrium conditions.
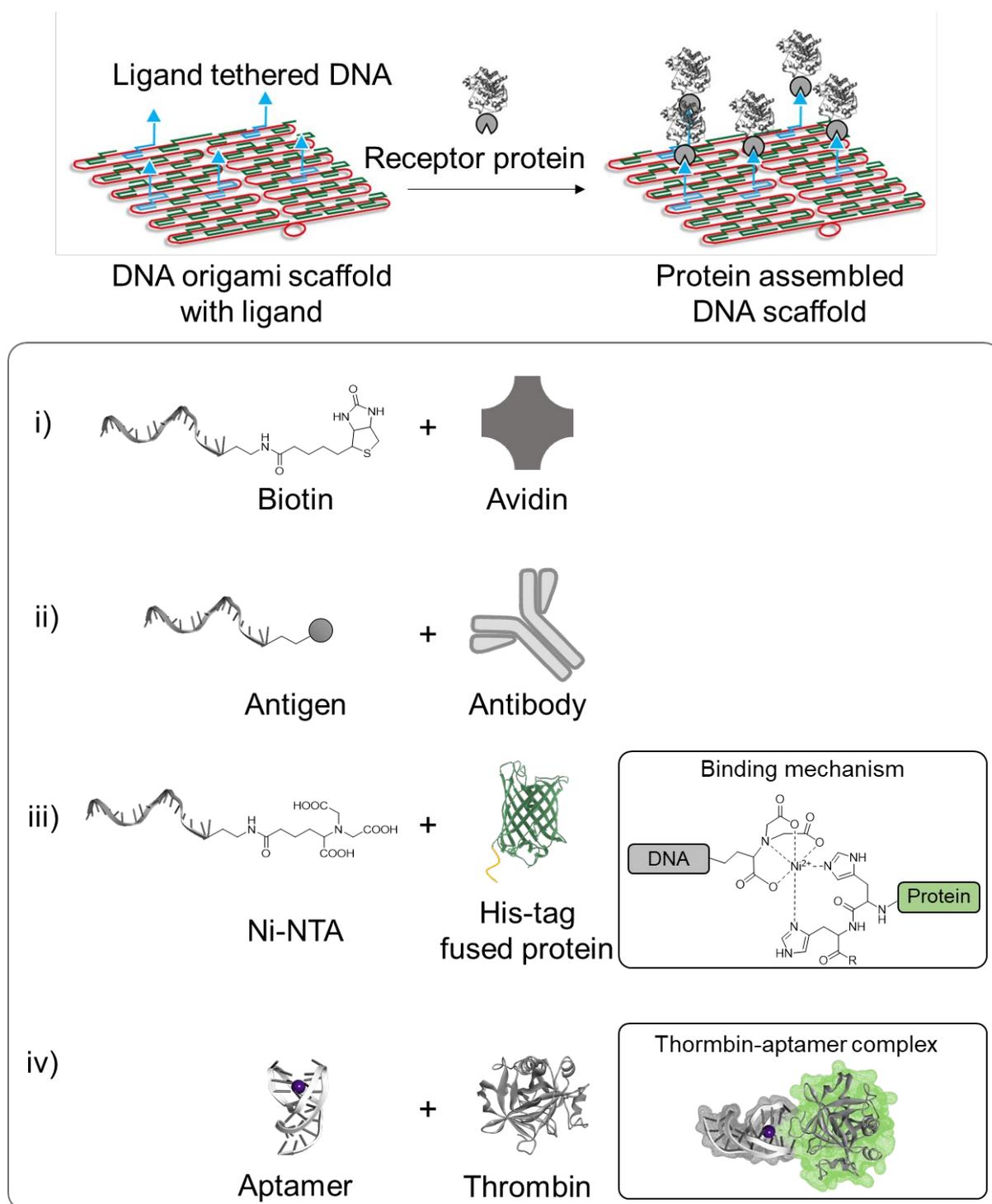
Figure 3. Representative noncovalent protein-DNA conjugation methods based on (i) biotin-avidin[30–32], (ii) antigen-antibody[33-36], (iii) Ni-NTA-hexahistidine[37,38], and (iv) aptamer-ligand[40-44] interactions (PDB ID: 5CMX[51]).

2.1.1.   *Biotin-avidin*

The biotin-avidin complex is formed by one of the strongest noncovalent interactions with an equilibrium dissociation constant in the femtomolar region[31]. Therefore, it is an attractive approach for assembling POI on DNA scaffolds[32]. Avidin is a tetrameric protein with four identical subunits that allows binding of four biotin derivatives. Biotinylated DNA strands can be used to target avidin and its derivatives to specific locations on a DNA scaffold. Streptavidin, an avidin derivative, was assembled in a linear array on DNA triple crossover tile[52] (**Figure 4a**), or on a 2D nanogrid array derived from a four-arm junction tile[32,53]. Similarly, a well-defined 2D array with controlled positioning of streptavidin was constructed on DNA origami[54]. Assembly of POI on DNA scaffolds using the biotin-avidin interaction was first reported by the Niemeyer group for the cascade reaction of NAD(P)H:FMN oxidoreductase (NFOR) and luciferase (LUC)[55]. NFOR and LUC were fused to the biotin carboxy carrier protein and biotinylated by biotin ligase when expressed in *E. coli*. Resulting mono biotinylated (30 ± 10%) NFOR and LUC were modified with streptavidin tethered ODN. The two enzyme-assembled ODN were allowed to hybridize to a biotinylated ODN that was introduced as a template to the surface of a streptavidin-coated microplate. When both enzyme-assembled ODN were hybridized to the same template ODN, the cascade reaction was enhanced more than two times than the reaction carried out with both enzyme-assembled ODN hybridized on separate templates. This method was also applied to assemble chemically biotinylated enzymes that are commercially available but the amount of modified biotin is not controlled on the 3D scaffold[56].

The simple assembly of POI utilizing the strong biotin-avidin interaction provides one of the highest assembly yields among the noncovalent conjugation methods[57]. However, this method is limited to the assembly of only one POI at a time and strict control of the number of POI is difficult because avidin derivatives exist as oligomer. Step-by-step modifications

are required for the assembly of two types of POI on the scaffold[56-58]. Multi-step assembly reactions require a great deal of caution to ensure the loading of proteins and enzymes in their functional forms. Another drawback is the possible reduction in the activity or deactivation of the enzyme during biotin conjugation, which is usually conducted in test tube. The biotin modification process could be carried out *in vivo* but the yield of biotinylated protein or enzyme was at most 30% depending on the culture conditions[55].
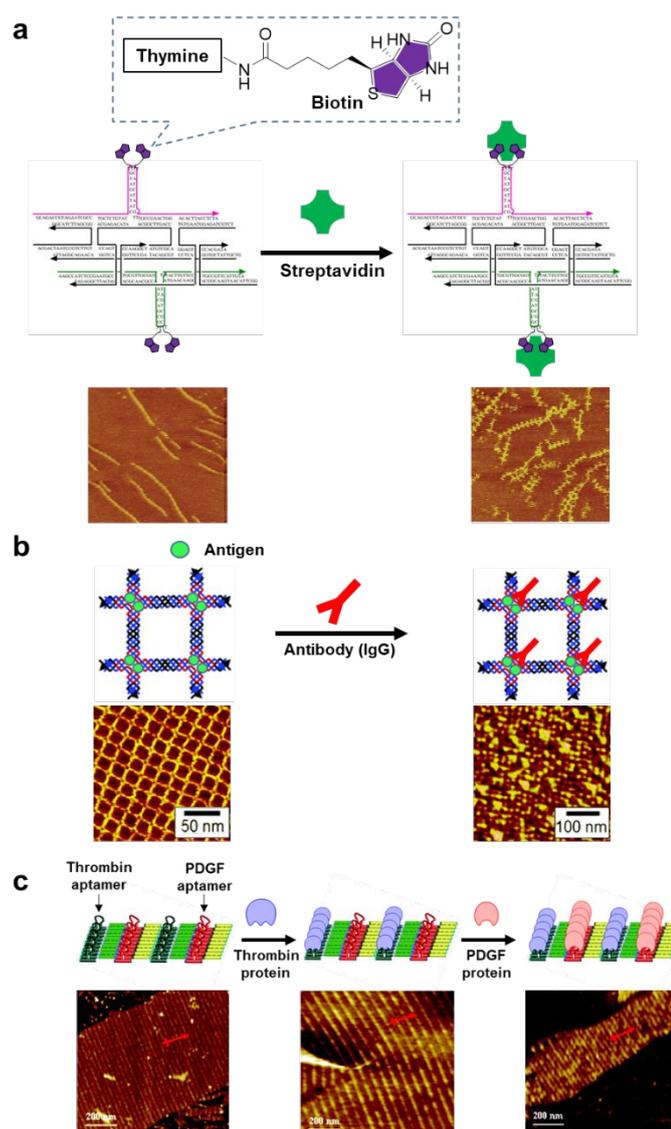
Figure 4. (a) Avidin-biotin interaction strategy for assembly of proteins on DNA nanostructures. Illustrations and AFM images of DNA triple crossover molecules (TX DNA tile) containing biotinylated oligonucleotides and assembled streptavidin[52]. (b) Immobilization of an antibody on a 2D DNA nanostructure through a specific antigen-antibody interaction[35]. (c) Immobilization of proteins on DNA nanostructures through specific aptamer

sequences. Representation of 2D DNA nanoarrays containing alternate thrombin and/or platelet-derived growth factor (PDGF) aptamers and the binding of their protein targets[62]. Reproduced from ref. 35, 52, and 62 with permission from ACS, copyright 2004, 2006, and 2007.

2.1.2.   *Antibody-antigen*

Antibodies are large proteins composed of two heavy and two light chains linked by disulfide bonds. Antibodies recognize unique molecules called antigens. Due to their high specificity and affinity for their antigens, antibodies are widely used as diagnostic tools. An antibody array on a DNA scaffold using the fluorescein (Fsc) – anti-fluorescein immunoglobulin antibody (IgG) interaction has been reported[33]. The Fsc antigen was tethered to ODN and assembled using these to form symmetrical cross-shaped antigen-DNA arrays (**Figure 4b**)[35]. Subsequent addition of IgG antibodies resulted in a well-defined periodic nanoarray with a pitch of approximately 20 nm. In 2012, the Mao group self-assembled IgG into a 3D hollow tetrahedron DNA nanostructure with Fsc antigens located at the edges of the tetrahedron[59]. Furthermore, the Fsc-IgG pair showed orthogonality to the biotin-streptavidin pair when different types of proteins were assembled on the DNA scaffold[36]. At least two antigens were located in close proximity for one antibody molecule, ensuring the effective binding of the "Y" shaped IgG molecule.

Another strategy constructed an ODN–peptide fusion, in which the peptide was the 'probe' for the specific antibody and the single strand ODN bound to the complementary single-strand DNA (capture probe) extended from the DNA surface[60]. A high-density myc-epitope peptide array was constructed from a DNA 'ABCD' tile array made by four DX motifs. The tile D was modified to contain a capture probe for positioning the myc-epitope peptide-ODN fusions via hybridization[60]. Addition of the anti-myc mouse antibody formed a series of parallel lines with a distance of approximately 64 nm and little or no nonspecific binding of the antibody. Even with the strong interaction between antibody and the antigen peptide, the antibody has to be equipped with another region to conjugate with POI or to fuse with POI to apply as the adaptor to locate POI on the DNA scaffold.

2.1.3.   Aptamers

Aptamers are short stretches of RNA or DNA, typically less than 100 nucleotides, that fold into three dimensional structures and display high affinities to target proteins or small molecules[44]. Having the same nucleic acid building blocks, it is straightforward to attach the aptamer sequence to the designed DNA tiles or staple strands in the DNA nanostructures, thereby providng the control on the location of the aptamer and the aptamer-bound POIs.

One of the first studies to use this strategy constructed a protein array on a linear DNA nanostructure assembled from periodic triple–crossover (TX) tiles[43]. Each TX tile contained a DNA hairpin loop with the aptamer sequence which bound specifically to thrombin. Addition of thrombin led to the formation of a periodic thrombin array with an interesting parallel pair of DNA arrays with the protein sandwiched inside. Extending this approach, a single chain variable fragment (scFv) was selected to bind to specific DNA aptamers on TX tile linear assemblies or cross-tile assemblies[61]. The scFv was proposed as a universal modular adaptor for the site-specific display of any POI.

With the purpose of arranging more than one type of protein on the DNA structure, different aptamer sequences were incorporated into a four-tile DNA 'ABCD' system in which tiles B and D contained aptamer sequences for thrombin and platelet-derived growth factor (PDGF), respectively (**Figure 4c**)[62]. The sequential addition of thrombin and PDGF resulted in a multi-protein 2D array with a periodic distance of approximately 64 and 32 nm, corresponding to the distances between two adjacent thrombin aptamers or thrombin–PDGF aptamers, respectively. The aptamers retained their functions after incorporation into the DNA origami scaffold in the presence of highly diverse sequences.

The fact that aptamers are formed by nucleic acids and need no further chemical modification makes the DNA aptamer a useful candidate for the construction of diagnostic

arrays. Godonoga *et al*., showed that the malaria protein biomarker from *Plasmodium falciparum*, lactate dehydrogenase (*Pf*LDH), was integrated on a rectangular DNA origami scaffold by binding its specific DNA aptamer[63]. *Pf*LDH bound specifically at aptamer loaded positions even in the presence of human blood plasma. The DNA scaffold bound *Pf*LDH retained its enzymatic activity, which indicated a promising malaria detection system that is constructed from DNA[63].

2.1.4.   Ni-NTA-hexahistidine

The bioorthorthogonal Ni-NTA interaction relies on complex formation of the divalent transition metal Ni^{2+} with the polyhistidine-tagged POI and a metal carrier, such as NTA, conjugated on the scaffold. This micromolar affinity interaction provides a moderately stable assembly of POI onto the scaffold. The polyhistidine-tagged protein binds to multiple NTA groups tethered to ODN with a sub-micromolar equilibrium dissociation constant, which is moderate stability as compared to the other combinations (**Table 1**). An ODN modified with three NTA groups (tris NTA ODN) binds tightly to the polyhistidine-tagged proteins with an equilibrium dissociation constant (K_D) as low as 6 nM[38]. Other divalent ions, such as Co^{2+}, Cu^{2+}, and Zn^{2+}, can also be used instead of Ni^{2+} in the Ni-NTA complex with a similar affinity, while Co^{3+} induces more stable complex formation[64].

The Ni-NTA interaction was used to assemble a DNA-templated protein array in a well-ordered form. The assembled proteins were used for single-molecule imaging of noncrystalline protein samples[65]. Unlike other methods where small domains are used as adaptors to assemble the POI on the scaffold, this method just requires a short peptide with six histidine residues appended at either the N- or C-terminal of the POI. This is the advantage of the method because the size and the shape of the assembled POIs will not be significantly disturbed when imaging them with atomic force microscopy (AFM) or electron microscopy. Several proteins and their complexes assembled on DNA-template protein arrays have been examined by transmission electron microscopy[65].

2.1.5. *Apo-protein reconstitution by the prosthetic group*

A semi-synthetic approach through reconstitution of apo-enzyme has been reported. For this method, the prosthetic group, such as porphyrin or flavin derivatives, was removed from the protein to yield the respective apo-enzyme, which was subsequently reconstituted using an artificial analogue[39]. This method offers a simple route to generate a hybrid ODN–enzyme conjugate. ODN is readily conjugated to the prosthetic group for reconstitution of the apo-enzyme[66]. There is potential to engineer tailor-made functional groups from the natural prosthetic groups by synthetic chemical methods and to further optimize the proteins function by means of site-directed mutagenesis or in vitro evolution. This method has been used to reconstitute apo-myoglobin[66,67] and apo-horseradish peroxidase[67,68]. Application of this method is limited for stable apo-enzymes, thus it has not been widely applied so far.

2.2.   Chemical modification of proteins and/or DNA

Formation of a covalent linkage between protein and DNA overcomes the instability of the reversible interaction between them. The protein is conjugated to an ODN before or after the preparation of the DNA scaffold. The *in-situ* POI-ODN conjugation on the DNA scaffold is preferable because it requires less or no complicated purification procedures before usage. The *in situ* conjugation is applicable depending on the reactivity and selectivity of the conjugation reaction.

Table 2. Classification of protein-DNA conjugation methods (covalent conjugation)

Name of method	DNA modification	Protein modification	In situ conjugation of POI and DNA	k_{cat}/K_m (M⁻¹s⁻¹)
Hetero cross-linking of DNA and protein (random)[69]	Hetero cross-linker modified Protein	Hetero cross-linker modified DNA	No	n.d.
Hetero cross-linking of DNA and protein as DNA templated protein conjugation (DNA templated protein conjugation as a site-selective DNA-protein conjugation)[70]	Chemically activated oligonucleotide and a guiding oligonucleotide	Site-selectively conjugated by a chemically activated DNA and a guiding oligonucleotide	Possible	n.d.
Tag-protein[50,71]	Tag-substrate	Genetic modification of POI with a protein tag	Yes	10 to 10⁴
Modular adaptor[72-74]	Tag-substrate and specific sequence for DNA-binding module	Genetic modification of POI with modular adaptor	Yes	10⁵ to 10⁶
Relaxase (HUH-tag)[75-77]	Addition of the specific sequence for relaxase	Genetic modification of DNA-binding protein	Yes	n.d.

n.d.: no data

2.2.1. *Chemical crosslinking of protein and DNA via cross-linker*

The general approach to introduce an ODN-tag is to modify the purified (or commercially available) POI with an ODN using a heterobifunctional cross-linker (**Figure 5**). The cross-linking reaction between the POI and ODN is performed under physiological conditions. The ODN-tag modified POI is accurately attached to the predesigned complementary DNA sequence on the DNA scaffold through specific DNA hybridization. There are various chemical techniques[24-28] for introducing ODN-tags on the POI.

The maleimide and *N*-hydroxysuccinimide (NHS) ester derived heterobifunctional cross-linker was described as a typical example. The NHS ester-modified end of the heterobifunctional linker, such as sulfo-succinimidyl-4-(*N*-maleimidomethyl)-cyclohexane-1-carboxylate (Sulfo-SMCC), is covalently and randomly attached to a surface exposed lysine side chain or to the N-terminal amino group of the POI. The other maleimide-functionalized end is subsequently coupled to the thiol group incorporated in the ODN-tag. Other surface exposed side chains, such as cysteine or genetically introduced azide group, can also be used for covalent linkage by changing the functional group of cross-linkers with appropriate modification of ODN-tag (the variation of the previously used cross-linkers can be found in other reviews[24-28,78]). However, these methods suffer from several drawbacks, including a reduction in the activity of the POI. Because of the non-regioselective chemical modification using NHS cross-linkers, it is difficult to control the stoichiometry of the linkage and the assembly of the POI on the DNA scaffold through the DNA hybridization would not reach to a quantitative yield. Recently, a site-specific conjugation method using a heterobifunctional cross-linker, termed as a DNA-templated protein conjugation, was reported by Gothelf and co-workers[70], however, drawbacks, such as the requirement for redundant purification steps and the low conjugation yield, remain to be solved. More recently, this method was used to covalently attach antibodies on a pre-formed DNA nanostructure[79]. NTA modified ODNs

were inserted in the cavities of a DNA origami scaffold with NTA forming coordination metal complexes with histidine clusters on the Fc domain of the antibody. Subsequently, covalent linkages between the antibody and pre-introduced NHS groups in the cavity were formed at surface exposed lysine residues of the antibody. Instability of the pre-introduced NHS groups under the DNA annealing conditions should be compensated by increasing the number of NHS groups.
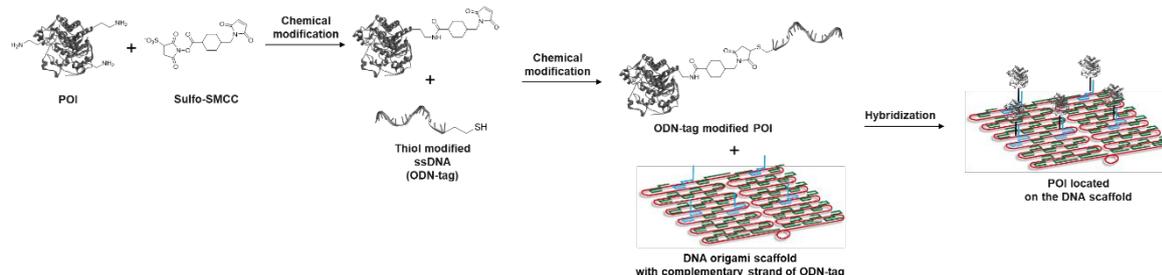
Figure 5. Schematic illustration of the residue-specific cross-linking reaction to modify a POI by ssDNA as an ODN-tag through covalent linkages and subsequent DNA/DNA hybridization to arrange the POI on the DNA scaffold.

2.2.2.   *Crosslinking of genetically fused protein with chemically modified DNA*

The self-ligating protein tag forms a covalent bond between the protein tag and its substrate. SNAP-tag[80] and Halo-tag[81] were applied to attach POI on the DNA scaffold[50]. The self-ligating protein tag was genetically fused to the POI, expressed in *E. coli*, and purified by the conventional method. Separately, ODN modified with the substrate of the self-ligating tag was prepared through the coupling reaction of the NHS modified substrate and the aminated ODN and assembled at specific positions on the DNA scaffold. The POI fused to the self-ligating protein tag reacted with the substrate tethered ODN on the DNA scaffold to form a covalent bond. Each self-ligating protein tag chemoselectively reacts with its respective substrate. Thus, POIs fused to different protein tags were orthogonally attached to their respective substrate modified ODNs. The self-ligating protein tags, SNAP-tag[80] and Halo-tag[81], showed high chemoselectivity to O^6-benzylguanine (BG) and 5-chlorohexane, respectively, with moderate reactivity. When SNAP-tag, Halo-tag, and monovalent streptavidin-fused POI were used for orthogonal assembly on DNA scaffolds, the reaction of the SNAP-tag and Halo-tag fused POI to their substrates on the DNA scaffolds proceeded very slowly resulting in low loading yields of 40–60%[50]. Therefore, the co-assembly yield of the three POI dropped to less than 10%. One of the reasons for such a low reactivity could be the electrostatic repulsion between the DNA scaffold and the protein-tag which could prevent the protein-tags from accessing their substrates on the DNA scaffold. The system conjugating a positively charged DNA binding protein and the self-ligating protein tag has overcome this drawback to accomplish almost quantitative assembling yields as described in the following section. The same group recently reported an improvement in the assembly yield (80%) by attaching a positively charged peptide, which was shorter than the DNA binding proteins, to the Halo-tag to improve the reaction constant from $\sim 10^2$ $M^{-1}s^{-1}$ to $\sim 10^4$ $M^{-1}s^{-1}$.[71] Though the SNAP-tag alone had moderate reactivity to BG modified ODN (10^2-10^3 $M^{-1}s^{-1}$)[72], a recent

study showed that fusing T7 RNA polymerase to the SNAP-tag significantly increased the assembly yield (95%)[82], possibly due to electrostatic interactions between the negatively charged ODN and the positively charged T7 RNA polymerase. In addition, there are several candidates which have the potential to be applied for attaching POI on the DNA scaffold. For example, CLIP-tag[83] is a protein tag derived from SNAP-tag and selectively reacts with benzylcytosine (BC), and SpyCatcher with SpyTag[84] and their derivatives[85-87] are the protein ligation approach using a short polypeptide SpyTag and its partner protein SpyCatcher. However, the reactivity of CLIP-tag with BC ($\sim 10^3$ $M^{-1}s^{-1}$)[83] and SpyCatcher with SpyTag ($\sim 10^3$ $M^{-1}s^{-1}$)[84] is lower than SNAP-tag with BG ($\sim 10^4$ $M^{-1}s^{-1}$)[80]. It might be one reason why CLIP-tag and SpyTag-SpyCatcher have not been applied for the purpose yet.

In some cases, POI activity was influenced by fusion with the self-ligating protein tag. The position at which the self-ligating protein tags are fused, i.e., the N or C-terminal, with or without a suitable linker, should be optimized. As described above, a serious problem in directly loading the self-ligating protein tag fused POI on the DNA scaffold is the requirement of a long incubation time and excess amounts of fusion protein due to the slow kinetics of covalent bond formation between the self-ligating protein tag and its substrate on the DNA scaffold[50]. This is not ideal to preserve POI activity. When these protein tags were used to construct an orthogonal assembly, the co-assembly yield was very low even though multiple redundant reaction sites for POI loading were introduced on the DNA scaffold to increase the loading yield. Such drawbacks caused difficulty in controlling the number of POI molecules or their stoichiometry on the DNA scaffold. It should be noted that care must be taken when introducing the substrates for self-ligating tag on the DNA scaffold. Because DNA origami is prepared through thermal denaturation and annealing processes, thermal stability of the substrates introduced on the DNA strand is highly required. In fact, thermally unstable groups, such as NHS[79], maleimide and thioester, are usually introduced on the DNA

scaffold after its completion of folding. Thus, available types of the pairs of self-ligating

protein tags and their substrates for assembling POIs on the DNA scaffold are very limited.

Until now, SNAP-tag[80], CLIP-tag[83], Halo-tag[81] and SpyCatcher[84-87] are the self-ligating

protein tags applied for assembling on DNA nanostructures.

2.3. *Applications of protein-assembled DNA scaffolds*

DNA nanostructures, including duplex DNA, are useful scaffolds to control the spatial organization of attached biomolecules, such as the interenzyme distance[88]. Willner and co-workers[89] successfully attached either two enzymes (GOx and HRP) or the enzyme-cofactor pair of NAD^+-dependent glucose dehydrogenase (GDH) and NAD^+ to predesigned positions on various DNA scaffolds through the hybridization of the enzyme conjugated DNA with the complementary DNA on the scaffold (**Figure 6a**). Yan and co-workers reported similar work whereby the same GOx/HRP pair was attached to a DNA nanostructure with systematic interenzyme distances (**Figure 6b**)[90]. They reported that yields of the enzyme cascade significantly increased at the interenzyme distance of 10 nm, suggesting that the transfer of the intermediate substrate (H_2O_2) between two enzymes followed limited surface diffusion for closely spaced enzymes. Furthermore, another group compared the cascade reaction of the GOx/HRP pair on 2D and 3D DNA origami templates using planar rectangular and tubular DNA origami structures[91]. The efficiency of the cascade reaction was notably higher for the enzyme pair encapsulated within the tubular structure than on the rectangular surface. To enhance the catalytic activity of the enzyme cascade and the stability of the enzyme, a DNA nanocage was constructed to accommodate two enzymes (GOx/HRP pair) with well controlled stoichiometry and spatial organization (**Figure 6c**)[92]. The DNA nanocage encapsulating the enzymes showed not only an increase in the activity of each enzyme and the efficiency of the cascade reaction but also increased the protection for the enzymes against proteases[92]. To extend the enzymatic cascade reaction to more than two enzymes, a three enzyme cascade reaction system of malic dehydrogenase (MDH), oxaloacetate decarboxylase, and lactate dehydrogenase (LDH) was constructed on a DNA scaffold[93]. In most of the cases described above, the substrate and/or the cofactor for the enzyme reactions were diffused in the bulk solution. In the following case, both the enzymes

and the cofactor were assembled on the DNA scaffold; the cofactor was attached to an arm swinging between the two enzymes[94]. The artificial swinging arm (**Figure 6d**) facilitated hydride transfer between the two enzymes, glucose-6-phosphate dehydrogenase (G6pDH) and MDH[94]. The group also demonstrated a model for regulating a two enzyme pathway, G6pDH–MDH and G6pDH–LDH, on a rectangular DNA origami platform by controlling the location of NAD^+ between the two enzyme pairs (**Figure 6e**)[95].

The DNA nanostructures for studying the enzyme cascades described above were all static in nature. Applying a dynamic DNA nanostructure for such an enzyme cascade would facilitate regulation of the efficiency of the cascade reaction and accessibility of the substrate to the enzyme. DNA tweezers were used as dynamic and switchable DNA nanostructures to investigate GOx and HRP enzyme cascade reactions. Two enzymes were attached to each end of the tweezer and the interenzyme distance was tuned by the open and close states which were switched by strand-displacement[96]. In another system, ODN-tag conjugated enzymes were loaded through DNA hybridization into a 3D DNA nanostructure that opened or closed using DNA strand displacement as a lock mechanism[97]. Estimation of the number of enzymes inside the 3D DNA nanostructure is one of the more difficult tasks in the characterization and quantitative analysis of this construct. Direct quantitation of the loaded enzyme inside the 3D DNA nanostructure tends to be difficult. In order to estimate the amount of enzyme loaded, a gold nanoparticle conjugated ODN-tag was reacted with the 3D DNA nanostructure in its open form. Transmission electron microscopy images of the sample revealed a single gold nanoparticle within the cavity for 10–15% of the DNA nanostructures[97].

Despite the fact that ODN-conjugated enzymes often show significant reductions in enzyme activity as a result of the chemical modification process,[89-91,94,96,98,99] the method of chemical modification of the POI by ODN and successive hybridization to the target DNA

sequence of the DNA scaffold is still useful to arrange the POI on the DNA scaffold in the fields of protein/enzyme chemistry due to its simplicity and adaptability[98-100]. Ideally, the methods to conjugate POI to DNA scaffolds should ensure a high assembly yield at the desired position without cumbersome handling or long preparation time. In particular, a high co-assembly yield of each POI is important for the quantitative evaluation of the subsequent application. Application of DNA-binding proteins[101,102] and their derivatives[72-74] as genetically fused adaptors as reported by our group overcomes the drawbacks described above and improves the assembling yield of POI on DNA scaffolds. The details of this method are described in the following sections.
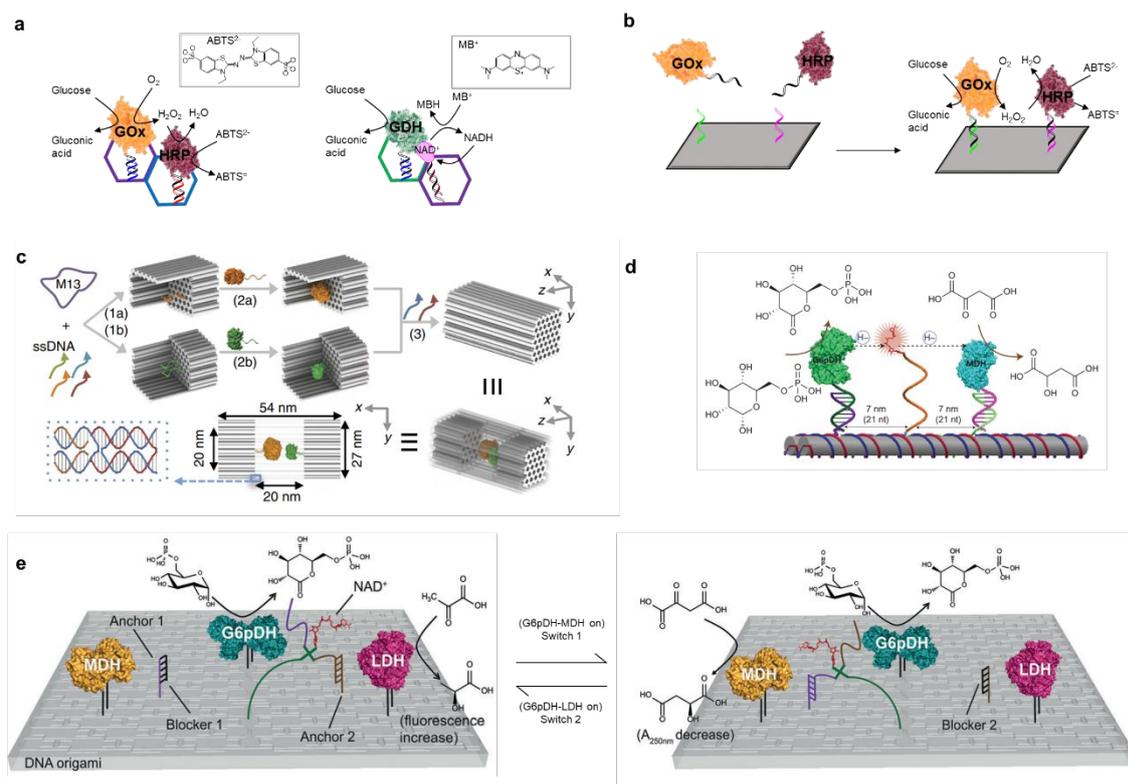
Figure 6. Application of protein-assembled DNA scaffold. **(a)** Assembly of the GOx and HRP enzymes (*left*) and the NAD$^+$/GDH system (*right*) on the two-hexagon scaffold using different lengths of tethers linking to the scaffold[89]. (b) Assembly of GOx and HRP enzymes on DNA nanostructures for an enzymatic cascade investigation[90]. (c) Design and characterization of a DNA nanocage encapsulating a pair of GOx (orange) and HRP (green) enzymes with well-controlled stoichiometry and spatial organization for the enhancement of

catalytic activity and stability[92]. (d) The organization of G6pDH and MDH enzymes on a DNA DX tile and the NAD$^+$-modified single-stranded poly(T)$_{20}$ positioning between the two enzymes[94]. (e) Illustration of the enzyme pathway regulation system on DNA origami[95]. Reproduced from ref. 89, 92, and 94 with permission from Springer Nature, copyright 2009, 2014, and 2016, from ref. 90 with permission from ACS, copyright 2012, and from ref. 95 with permission from John Wiley and Son, copyright 2016.

3.   DNA-binding proteins as adaptors for assembling proteins on DNA scaffolds

3.1.   *DNA-binding adaptors for reversible assembly of proteins via noncovalent protein-DNA interactions*

3.1.1.   *Zinc finger proteins*

As described above, most of the methods employed to arrange ODN-modified POI on DNA nanostructures through the DNA hybridization require harsh conditions during the chemical modification procedure of ODN to POI. In order to develop a method that is free from the chemical modification and fully based on the protein components, we chose the sequence-specific DNA-binding proteins as adaptors to arrange proteins in their functional forms at target sites on the DNA scaffold.

A large amount of structural information is available for DNA-protein complexes, facilitating the structure-based design of fusion proteins of the DNA-binding domain and POI[46]. Zinc-finger proteins (ZFP) are one of the well-studied classes of DNA-binding proteins, and the artificially designed ZFPs, according to the recognition rule[47,103], have been shown to bind a wide variety of DNA sequences[45,47,49]. Each zinc-finger domain is capable of recognizing a tract of four base pairs in the major groove of a DNA duplex. A three-fingered protein recognizes a tract of ten base pairs with an affinity of nanomolar equilibrium dissociation constant[104,105]. ZFP have been utilized to re-assemble split fragments of green fluorescent protein (GFP) by bringing them into close proximity on the DNA duplex in a sequence enabled reassembly strategy[106]. ZFP were genetically fused to each of the GFP split fragments which reassembled only in the presence of duplex DNA that contained adjacent ZFP binding sites (**Figure 7a**)[106].
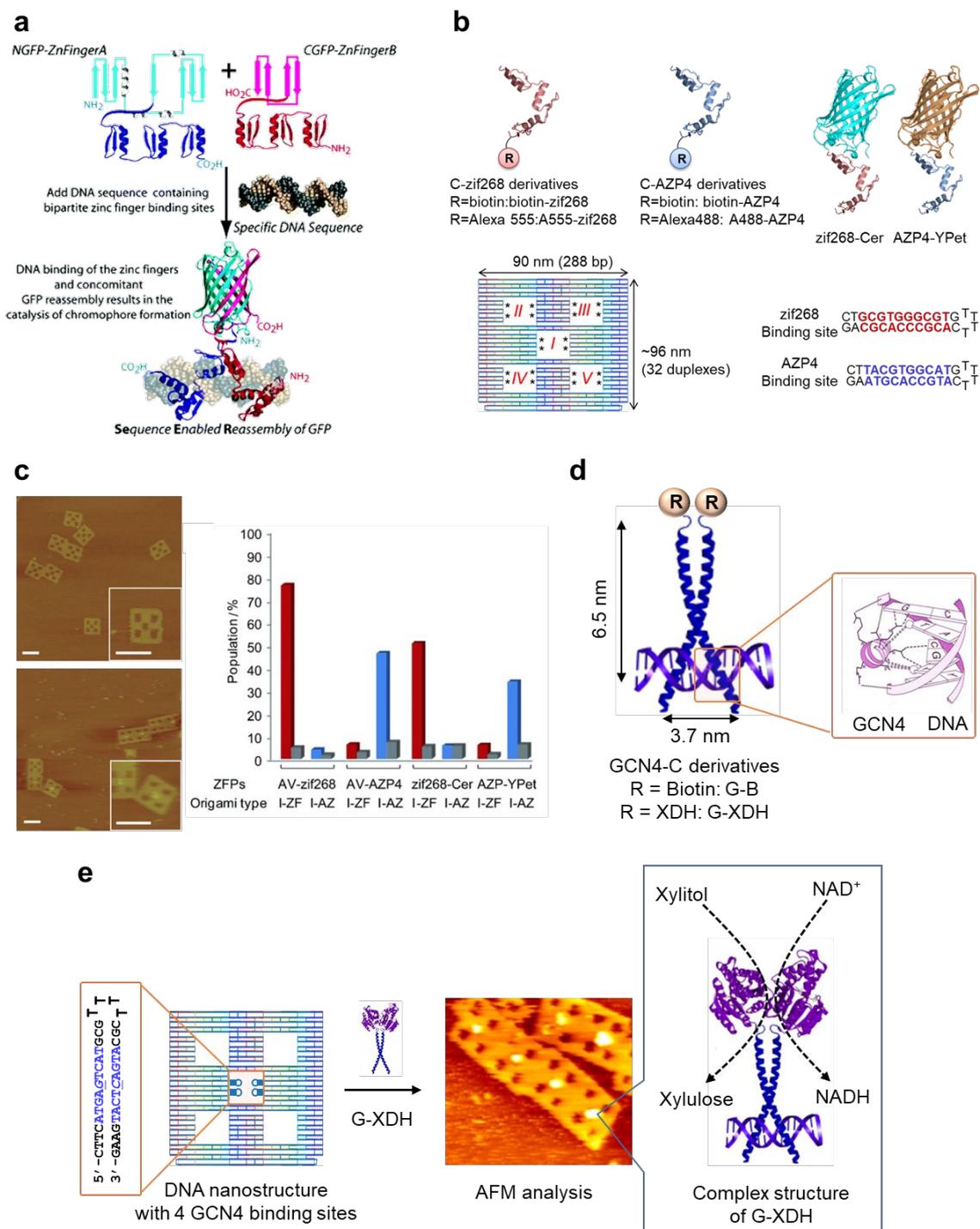
Figure 7. (a-c) Zinc-finger proteins (ZFP) as protein adaptors to arrange protein on a DNA scaffold. (a) The sequence enabled reassembly strategy. NGFP-ZnFingerA (cyan and blue) comprises residues 1-157 of GFP fused by a 15-residue linker to the ZFP zif268. CGFP-ZnFingerB (pink and red) comprises residues 158-238 of GFP fused by a 15-residue linker to the ZFP PBSII[106]. (b) (*top*) Structures of the ZFP adaptors (zif268 and AZP4) and ZFP adaptor-fused proteins (zif268-Cer and AZP4-YPet). (*bottom*) An illustration of DNA origami that contains five cavities (I, II, III, IV, and V) where positions of the binding sequences for zinc-finger adaptors

are indicated by asterisks (*)[101]. (c) AFM images of unmodified I-ZF (*left*, *top*), which is DNA origami with four zif268 adaptor binding sites at the central cavity I, and I-ZF modified with avidin-attached biotin-zif268 (AV-zif268) (*left*, *bottom*). Insets: magnified images (scale bars = 100 nm). (*right*) Selective binding of ZFP adaptors to the target site was estimated by counting the number of ZFP-bound origami structures in the AFM images[101]. (d, e) The basic leucine zipper protein (GCN4) as a homodimeric protein adaptor to arrange POI on the DNA scaffold. (d) Structure of the GCN4-DNA complex[108]. e) A scheme showing the procedure to arrange GCN4 adaptor-fused enzymes (G-XDH) at specific positions on the DNA scaffold with yields of over 85%[102]. Reproduced from ref. 101 with permission from John Wiley and Son, copyright 2016, from ref. 102 and 108 with permission from Elsevier, copyright 1992 and 2014, and from ref. 106 with permission from ACS, copyright 2005.

We have reported the first application of ZFP as adaptors to arrange ZFP-fused proteins on DNA origami scaffolds[101]. Two types of well-characterized ZFP, zif268[104] and AZP4[105], each recognizing its unique target DNA sequence, were chosen as orthogonal adaptors for the zinc finger binding sequences on the DNA structure (**Figure 7b**)[101]. Both adaptors were specifically and orthogonally located within the cavities of a rectangular DNA origami structure with binding yields of up to 70% for zif268 and 45% for AZP4. These yields were consistent with the affinity between ZFP and their target DNA, and a nonspecific binding yield was less than 10% based on the statistical analyses of AFM images (**Figure 7c**) [101].

3.1.2. Basic leucine zipper proteins

The basic leucine zipper (bZIP) protein GCN4 was used as an adaptor for dimeric POI[102]. The yeast transcription factor GCN4 is a bZIP class protein that forms a parallel coiled-coil and targets DNA sequences of nine to ten base pairs with an equilibrium dissociation constant (K_D) in the low nanomolar range (**Figure 7d**)[107-110]. With its simple helical structure, the bZIP monomer is easily fused to the N- or C-terminal of a POI without

having to consider specific conditions for its folding (**Figure 7d**). Taking advantage of these characters, we used GCN4 as a homodimeric adaptor to arrange dimeric proteins, a common active enzyme form[111], on DNA nanostructures (**Figure 7e**)[73,102]. Each monomer of a dimeric enzyme xylitol dehydrogenase (XDH) from *Pichia stipitis* that oxidizes xylitol to xylulose in an NAD[+] dependent manner[112] was fused to the C-terminal of GCN4 monomer through a Gly-Gly-Ser linker to yield an active adaptor-fused enzyme G-XDH (**Figure 7e**, *right*). Interestingly, the activity of homodimeric XDH slightly increased in the form of G-XDH as compared to the wild-type[102]. Homodimer formation of GCN4 likely reinforced the stability of the XDH dimer. This is a rare case when compared to the modifications of enzymes by ODN that usually result in a decrease in enzyme activity[89-91,94,96,98,99].

Both GCN4 and G-XDH were confirmed to accurately bind at positions containing specific DNA-binding sequence on DNA scaffolds with high yields (over 80%)[102]. Importantly, G-XDH exhibited high enzymatic activity when specifically assembled on the DNA scaffold. Thus, GCN4 serves as a useful homodimeric adaptor to arrange homodimeric proteins in their functional forms on DNA nanostructures. Specific and orthogonal targeting of GCN4 and ZFP adaptors to their respective addresses on DNA nanostructure[102] enabled us to precisely arrange two different enzymes, XDH and xylose reductase (XR), at predesigned locations with variations in the interenzyme distances on a DNA scaffold[73].

3.2.   *Modular adaptors for covalent conjugation of genetically modified proteins to chemically modified DNA*

For functional analyses and for practical applications, the prerequisites for protein-assembled DNA scaffolds include (1) high loading yields of the POI to the designed positions in a short incubation time, (2) stable DNA-POI conjugation, and (3) retaining the activity of assembled POI, especially for enzymes. By considering the stability of POI assembly, the

major drawback for the use of adaptors described in the precedent section is the reversible nature of their DNA-binding complexes. Stable conjugation of POI to DNA scaffolds through covalent linkages ensures long term stability of the assembly[50,71,75-77]. In order to form a covalent bond between the POI and the DNA scaffold, post-translational chemical modifications of the POI are not desirable for unstable enzymes as mentioned above. A possible method to tether POI to DNA scaffolds is to fuse POI with a self-ligating protein tag[50], which has shown several drawbacks due to the slow kinetics of covalent bond formation (10^2-10^3 $M^{-1}s^{-1}$)[72] as described in section 2.2.2.

The modular adaptor consisting of a DNA binding protein and a self-ligating protein tag compensates the drawbacks associated with the DNA-binding adaptor and the self-ligating protein tag (**Figure 8**). POI fused with modular adaptors were assembled through a covalent linkage at a defined position on the DNA scaffold in almost quantitative yields with fast reaction kinetics under mild conditions[72-74]. SNAP-tag was fused to the C-terminal of the DNA-binding ZFP zif268[104] to form a modular adaptor ZF-SNAP that has the characteristics of both modules to rapidly form a stable covalent bond in a chemoselective manner at the designated DNA sequence on the DNA scaffold. In fact, the second-order rate constant for ZF-SNAP and the substrate tethered target DNA (10^5~10^6 $M^{-1}s^{-1}$) was determined to be almost 1000 times higher than that for the SNAP-tag alone (10^2~10^3 $M^{-1}s^{-1}$) **(Figure 8a)**[72]. This is due to the fast binding kinetics of zif268 with its target DNA and the successive increase in the effective concentration of substrate[74]. Monomeric XR fused to ZF-SNAP (ZS-XR) was specifically loaded on the DNA scaffold with fast reaction kinetics and a high reaction yield while fully retaining enzymatic activity **(Figure 8b)**[73]. In addition, the modular adaptor is useful for tightly controlling the number of enzyme molecules by simply varying the number of modular adaptor-binding sites on the DNA scaffold **(Figure 8c, d, e)**[73].

When assembling several types of POI at defined positions on the DNA scaffold in their functional forms, a series of modular adaptors with orthogonality and fast reaction kinetics under mild conditions are required, especially for thermally unstable POI (**Figure 9, 10**)[74]. A series of modular adaptors consisting of a DNA-binding domain and a self-ligating protein tag were systematically constructed to evaluate orthogonal covalent bond formation at specific positions on a DNA scaffold. Three DNA-binding domains (zif268, AZP4, and GCN4) and three protein tags (SNAP-tag, CLIP-tag, and Halo-tag) were adapted to construct orthogonal modular adaptor candidates (**Figure 9c**)[74]. Among the nine possible modular adaptor candidates, a set of three modular adaptors (ZF-SNAP, AZ-CLIP, and AZ-Halo) (**Figure 9d**) were chosen to orthogonally react at their respective target sites with quantitative yields (**Figure 10a**). It should be noted that the one-pot co-assembly yield of three different modular adaptors was 90% at ambient temperatures within 5 min or on ice for 20 min (**Figure 10b**). The apparent rate constants for the cross-linking reactions of the three different modular adaptors are within the same order of magnitude ($10^5 \sim 10^6$ $M^{-1}s^{-1}$). Such characteristics for multiple cross-linking reactions are extremely useful to simultaneously assemble various POI on the DNA scaffold using a one-pot reaction. The short reaction time is suitable for loading thermally unstable enzymes on DNA scaffolds while maintaining their activity. In fact, the enzyme xylulose kinase (XK), which was reported to be thermally unstable[113], was successfully loaded on a DNA scaffold with its full enzymatic activity retained. The high loading yields achieved by modular adaptor-fused enzymes to a single binding site on a DNA scaffold allows us to control the position and number of enzymes on the DNA scaffold through covalent linkages at different DNA sequences.

Our study suggested a principle for a design that expands the modular adaptor system to orthogonally assemble a number of POI on a DNA scaffold based on available DNA-binding proteins and self-ligating protein tags. One of the drawbacks of the modular adaptor

approach is a few choices of the protein tags because thermally unstable chemical substrates could not be used as mentioned above (see section 2.2.2.). In order to overcome the drawback, we proposed a general strategy to construct a set of modular adaptors with complete orthogonality and retaining high reactivity based on the kinetic parameters for the cross-linking reaction and the association and dissociation rate constants for the noncovalent complex formation between the modular adaptor and the substrate modified ODN[74]. When the rate constant for crosslinking reaction is much smaller than the dissociation rate constant of the complex between the target DNA sequence and modular adaptor, the apparent rate constant for the formation of crosslink between the modular adaptor and the substrate modified ODN is governed by the equilibrium dissociation constant of the DNA-modular adaptor complex. Based on this theory, we have recently provided a set of orthogonal modular adaptors, in which the same chemoselective cross-linking domain (CLIP-tag) is shared by different types of DNA binding domains. These modular adaptors in fact undergo the cross-linking reaction at different DNA sequences exclusively governed by the sequence specific recognition by the DNA binding domain[114]. A possible drawback in the use of modular adaptors is that the size of the modular adaptors, *e.g.,* 48 kDa for AZ-Halo, 33 kDa for ZF-SNAP, and 32 kDa for AZ-CLIP. Genetic fusion of the modular adaptor to the POI could be an issue depending on the intended application of POI.
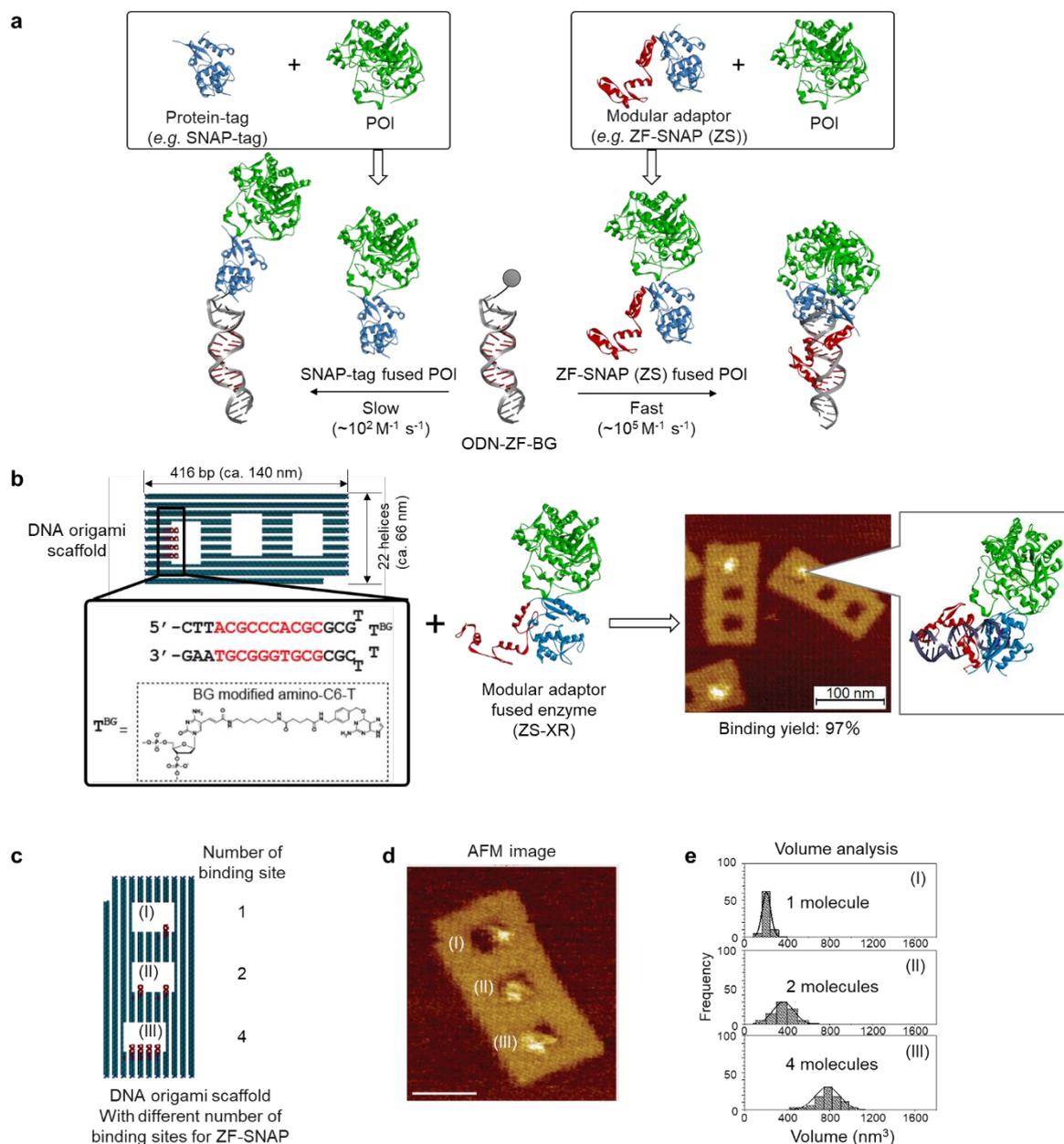
Figure 8. Modular adaptor for assembling POI to a specific binding site on a DNA origami scaffold. (a) A scheme represents the covalent bond formation between the substrate-modified oligonucleotide (ODN-ZF-BG) and protein-of-interest (POI) fused modular adaptor ZF-SNAP (ZS) with fast reaction kinetics (*right*) or POI fused protein tag (SNAP-tag) with slow reaction kinetics (*left*). (b) An illustration of complex formation of a DNA origami scaffold with binding sites for a modular adaptor fused enzyme (ZS-XR). The target sequence for ZF-SNAP (or ZS-XR), the structure of benzylguanine (BG) modified amino-C6-T (T^{BG}), and the location of T^{BG} adjacent to the target sequence are shown (*left*). An AFM image of ZS-XR bound to the DNA origami scaffold is shown (*right*). (c) Illustration of the DNA origami scaffold with different numbers of binding sites for ZF-SNAP. (d) An AFM image of ZS-XR bound to the DNA origami scaffold. The scale bar

represents 100 nm. (e) Frequency distributions of molecular volumes of ZS-XR for each cavity (I, II, III) with

different numbers of binding sites. Reproduced from ref. 73 with permission from ACS, copyright 2016.
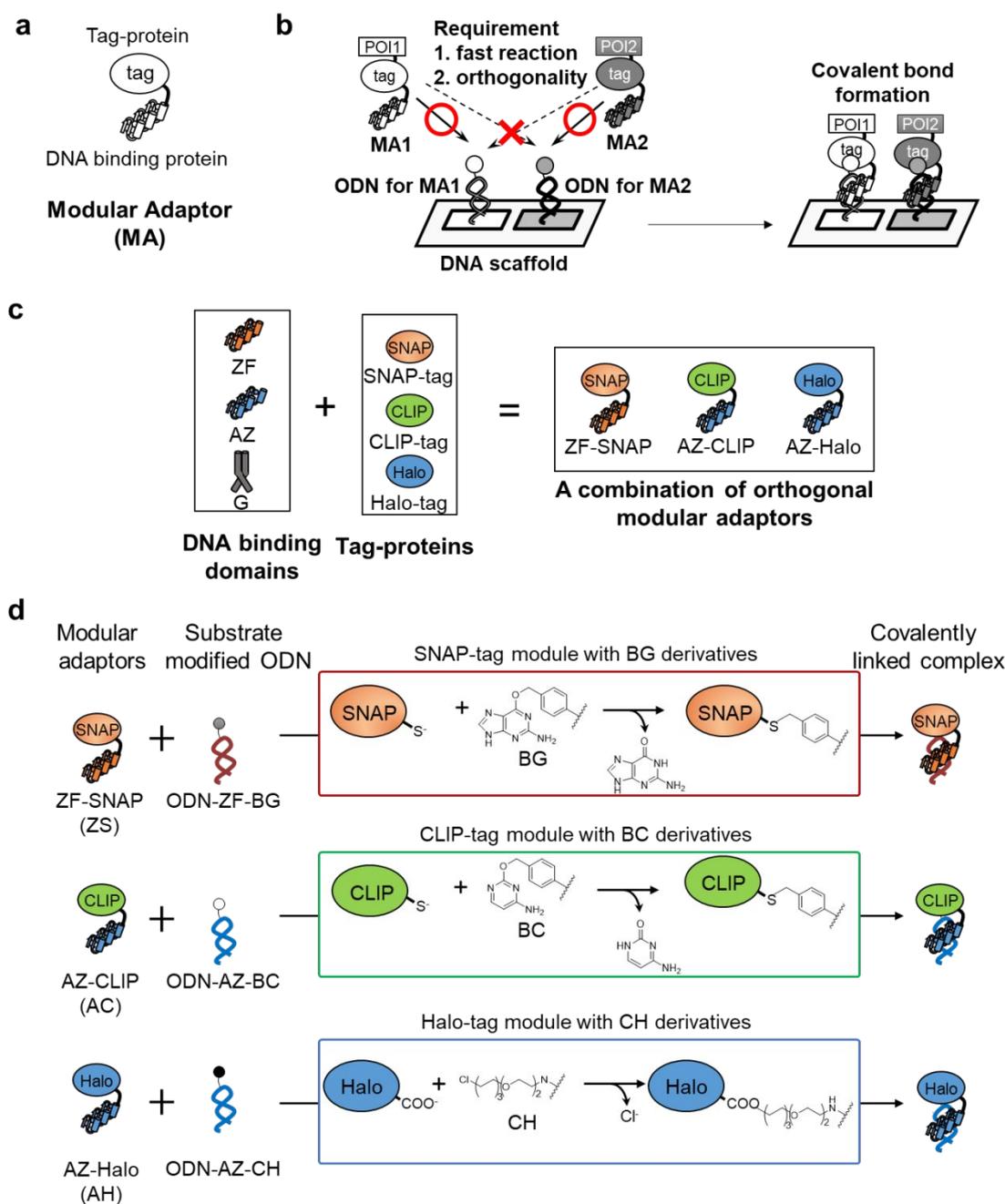
Figure 9. (a) An illustration of a modular adaptor. (b) A scheme representing the fast and orthogonal loading of proteins of interest (POI) fused modular adaptors at defined DNA sites on the DNA scaffold. (c) Combination of the DNA-binding domains and protein tags for constructing an orthogonal combination of modular adaptors. (d) Reaction schemes representing the cross-linking reactions between the modular adaptors and the substrates incorporated ODN. Reproduced from ref. 74 with permission from ACS, copyright 2017.
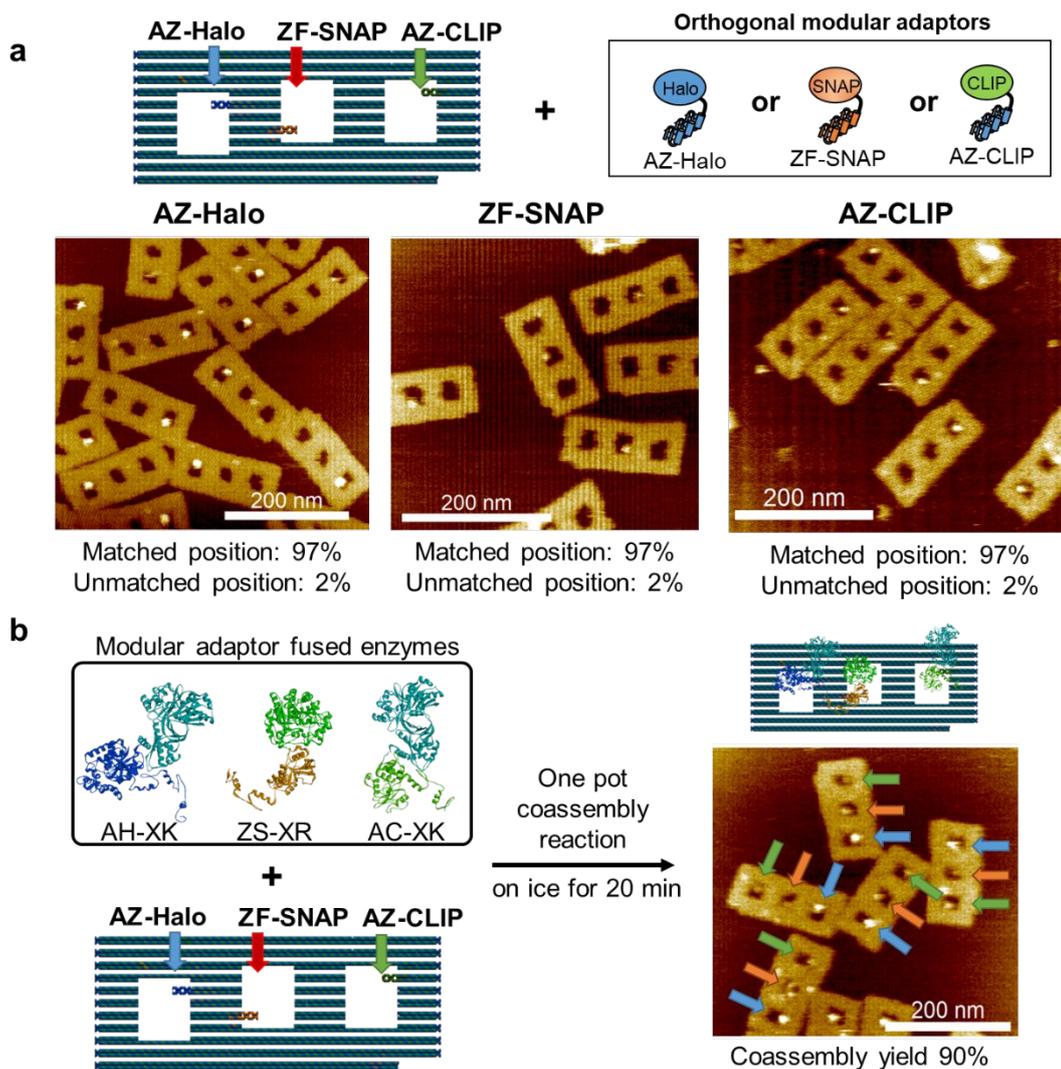
Figure 10. (a) An illustration of a DNA origami scaffold in which each cavity contains a single site for one of the modular adaptors (t*op*). AFM images of the DNA scaffold with the modular adaptor (AZ-Halo, ZF-SNAP, or AZ-CLIP) at its predesigned position were shown (*bottom*). (b) An illustration and AFM image of one-pot co-assembly of three adaptor fused enzymes (ZS-XR, AC-XK, and AH-XK) on a DNA scaffold. The loading and co-assembly yields are shown below each image. Reproduced from ref. 74 with permission from ACS, copyright 2017.

3.3. HUH-tag for covalent conjugation of genetically modified proteins to unmodified DNA

The HUH-tag derived from HUH-endonuclease[75-77] provides an alternative enzymatic approach for covalent bond formation between the protein adaptor and the target DNA through a mechanism that involves the processing of a range of mobile genetic elements by catalyzing the cleavage and ligation of DNA, such as Rep proteins, relaxases and transposases. HUH-endonuclease contains a conserved pair of metal-coordinating histidine residues separated by a hydrophobic residue. The reaction starts with the nicking of single-stranded DNA at a specific sequence in its origin of replication followed by the formation of a covalent phosphotyrosine intermediate, in which the 5'-end of DNA is linked to a specific tyrosine residue in the HUH protein[115]. Purified HUH-proteins form stable covalent bonds with ODN bearing the origin of replication sequence *in vitro*. The reaction proceeds under a variety of conditions, including *in vitro*, standard cell culture media, cellular lysates, and cell fixation. Therefore, HUH-tag is applicable to immobilize POI to DNA nanostructures *in vitro* and in cultured cells without functional disruption when fused to several nuclear, cytoplasmic, and cell-surface target proteins[76,77]. The relaxase domain, a type of HUH-endonuclease, is a small monomer (20–30 kDa) with a low K_D for its target DNA sequence and is easily fused to POI by genetic modification. The advantage of using relaxase is that under physiological conditions unmodified DNA serves as its target site on the DNA scaffold.

Several kinds of relaxase domains with different sequence selectivity's were used to arrange multiple POI on DNA scaffolds[75,76]. The loading yields of these relaxase-fused POI ranged from 40 to 50% for a single binding site which was comparable to those using other chemical modification methods. This method using relaxase could be used to develop a new class of orthogonal, sequence-selective protein adaptors for DNA nanotechnology[75,76].

4. Application of DNA-binding adaptors for assembling proteins on DNA scaffolds

4.1. Assembling proteins of interest on DNA scaffolds using cell lysates

To arrange POI on DNA origami scaffolds, zif268 and AZP4 were fused to a cyan fluorescent protein variant (Cerulean) and a yellow fluorescent protein variant (YPet), respectively[101]. The adaptor derivatives zif268-Cerulean and AZP4-YPet bound to the expected locations through their DNA-binding sequences on the DNA scaffold as confirmed by gel-electrophoretic analysis and AFM imaging (**Figure 11a**). An advantage of using a DNA-binding protein as the adaptor to arrange POI onto DNA origami scaffolds is that the adaptor fused POI can be expressed in *E. coli* with no further modifications. In fact, zif268-Cerulean selectively bound the target DNA sequence even when the cell lysate containing the overexpressed zif268-Cerulean was applied directly to the DNA scaffold, indicating a potential application for the adaptor system *in vivo*.

4.2. Assembling proteins of interest on DNA scaffold in cells

The DNA-binding adaptors, zif268 and AZP4, were utilized to assemble the transmembrane protein complex, G-protein-gated inwardly rectifying Kir3 K^+ channel, on DNA origami scaffolds (**Figure 11b**)[116]. The DNA scaffold cavities differed in size but contained the same number of binding sites in order to serve as templates to accommodate the Kir3 subunits in an optimal arrangement. Formation of the hetero-tetrameric assembly of Kir3 channel subunits was controlled by the orthogonal binding of zif268- and AZP4-conjugated to Kir3.1 and Kir3.4 subunits, respectively, to the DNA scaffold. The oligomerization states and spatial arrangements of the Kir3 subunits using adaptors and the DNA scaffold system were also controlled in living cells to reveal an enhancement in whole-cell current by the formation of a hetero-tetrameric K^+ channel (**Figure 11b**).
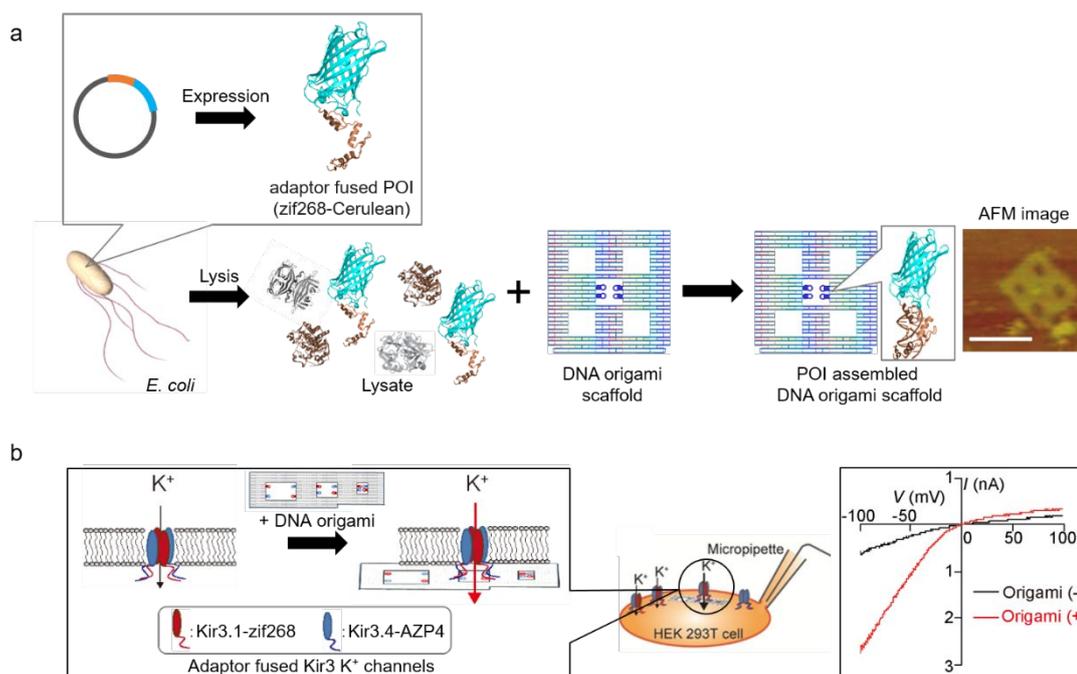
Figure 11. Application of zinc finger protein adaptors to arrange functional proteins on DNA scaffolds in cell lysates or in the cell. (a) An illustration showing the direct assembly of the adaptor-fused protein expressed by *E. coli* on a DNA scaffold using *E. coli* lysate. (*right*) An AFM image of a DNA scaffold treated with *E. coli* lysate containing zif268-Cerulean[101]. (b) A schematic illustrating the functional enhancement of Kir3.1/3.4 channels using a DNA scaffold in living cells. (r*ight*) Representative curves of whole-cell currents in Kir3.1-zif268/Kir3.4-AZP4-expressing HEK293T cells with and without DNA scaffold[116]. Reproduced from ref. 101 and 116 with permission from John Wiley and Sons, copyright 2012 and 2018.

4.3.   *Two-step sequential enzymatic reaction systems on a DNA scaffold*

The ability of DNA-binding adaptors to arrange the POI at a specifically designed position on a DNA scaffold with high orthogonality enabled the construction of an artificial enzyme cascade. We focused on the XR-XDH pathway within the ᴅ-xylose metabolic pathway[117-119]. In this pathway, the first enzyme of the cascade, XR, converts xylose to xylitol using the cofactor NADH. The resulting products, xylitol and NAD^+, simultaneously react with the second enzyme XDH, which converts xylitol to xylulose by recycling the NAD^+ to NADH (**Figure 12a**). A 2D DNA origami structure with three cavities containing specific binding sequences for the adaptors ZF-SNAP (ZS) and GCN4 (**Figure 12b, c**) was prepared as a

scaffold to co-assemble the XR/XDH pair. For XDH attachment, XDH was fused to the C-terminal of the GCN4 adaptor to give an active adaptor fused enzyme G-XDH (**Figure 7e, 12a**). XR was fused to the modular adaptor ZS[73], which consisted of both the DNA-binding protein zif268 (**Figure 7a**)[104] and the chemoselective SNAP-tag module[80], yielding an adaptor fused enzyme ZS-XR (**Figure 12a**). The ZFP zif268 bound to its specific DNA sequence on the DNA scaffold, while the SNAP-tag formed a covalent linkage with BG which had been incorporated adjacently to the target DNA sequence (**Figure 12b**). Interestingly, the adaptor fused enzyme, ZS-XR, exhibited a higher enzymatic activity than the parent XR[73]. Using both GCN4 and ZS adaptors, the positions and the numbers of two adaptor-fused enzymes ZS-XR and G-XDH were accurately controlled on a DNA scaffold through the location and number of adaptor binding sequences (**Figure 12d**). These characteristics enabled us to investigate the sequential reactions of XR and XDH. The efficiency of the two-step reaction was highly dependent on the interenzyme distances (**Figure 12f**). The interenzyme distance contributed more to the efficiency of consecutive reactions in the biomolecular transport system (xylitol and NAD^+) than that of a unimolecular transporting enzyme cascade.
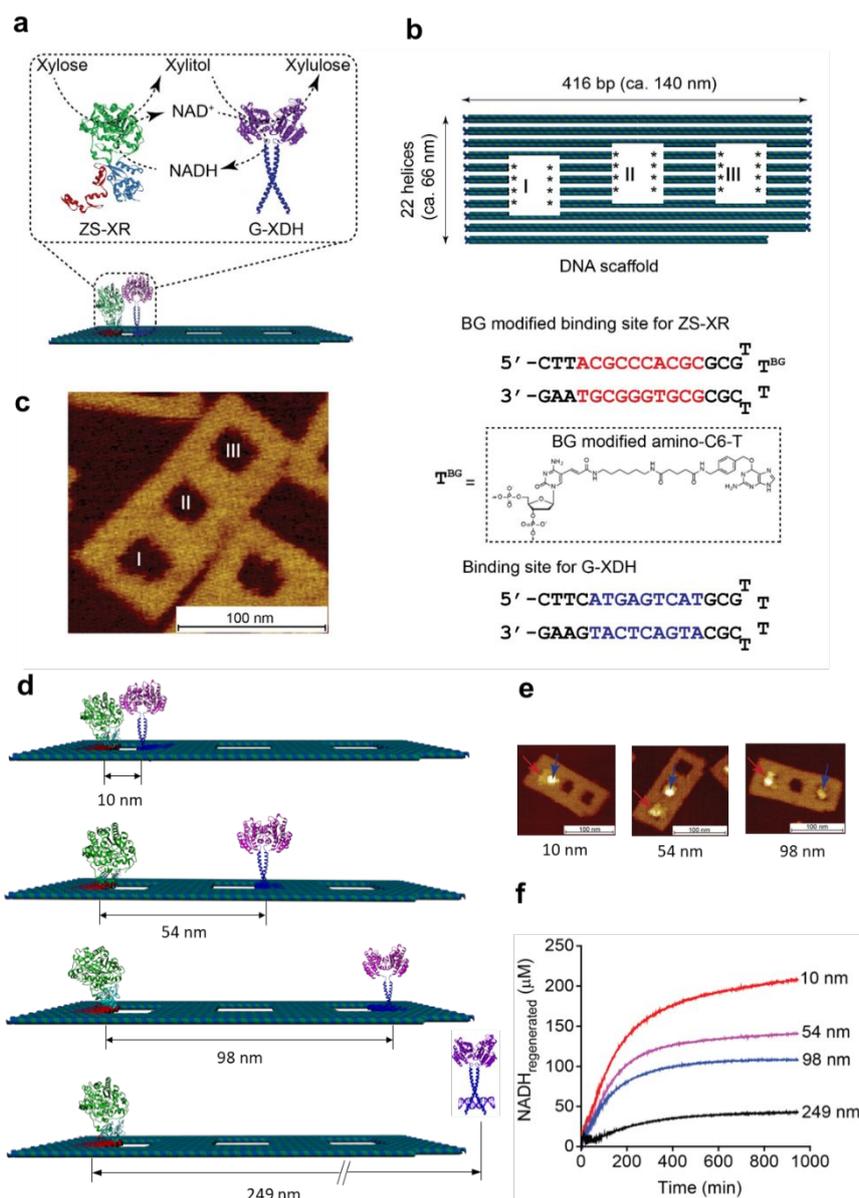
Figure 12. Assembling two types of adaptor-fused enzymes on a DNA scaffold. (a) A schematic illustration of the cascade reaction by two adaptor-fused enzymes (ZS-XR and G-XDH) co-assembled on the DNA scaffold. (b) An illustration of the DNA scaffold (*top*) showing three cavities each of which holds up to eight DNA hairpins containing the specific binding sequences for either ZS-XR or G-XDH at predesigned positions as indicated by asterisks (*). Also shown is the chemical structure of BG-modified thymidine, denoted as "T^{BG}" (*bottom*). (c) An AFM image of the DNA scaffold before assembly of the adaptor-fused enzyme. (d) Illustration of assembled ZS-XR with or without G-XDH on the DNA scaffold with different interenzyme distances. (e) AFM images indicate the co-assembly of ZS-XR and G-XDH on DNA scaffolds with different interenzyme distances (Scale bar represents 100 nm). (f) Time-course reaction profiles for NADH when two enzymes were

co-assembled 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 distance of 249 nm. Reproduced from ref. 73 with permission from ACS, copyright 2016.

4.3. Three-step sequential enzymatic reaction systems on a DNA scaffold

We have developed three types of modular adaptors with orthogonal coassembly yields reaching 90% in a short incubation time as described in the previous section (section 3.2).[73] To demonstrate the robustness of our assembly system, modular adaptors were used to assemble three types of enzymes at defined positions on a DNA scaffold to drive consecutive enzymatic reactions (**Figure 13**). The XR-XDH pathway described in the previous section was extended to the third step to convert xylulose to xylulose-5-phosphate by xylulose kinase (XK)[113,117-119]. A modular adaptor consisting of AZP4 and a CLIP-tag (AC) showed orthogonality to both the modular adaptor ZS and the GCN4 adaptor when fused to XK (AC-XK). The enzyme cascade of XR/XDH/XK (**Figure 13a**) was constructed on a DNA scaffold with varying interenzyme distances, (**Figure 13b**) as confirmed by AFM imaging (**Figure 13c**). The cascade reaction for the three enzyme co-assembled system (**Figure 13c**, I-4XR/I-4XDH/I-1XK or I-4XR/II-4XDH/III-1XK) was investigated and compared with that of the two and one enzyme co-assembled systems. The amount of ADP produced by the three step reaction, the measure of three step reaction, was higher than that for the two and one enzyme loaded systems. ADP production for the three enzyme co-assembled system with a distance of 10 nm (**Figure 13c**: I-4XR/I-4XDH/I-1XK) was higher than those co-assembled with a distance of 50 nm (**Figure 13c**: I-4XR/II-4XDH/III-1XK). Consistent with previous results, the interenzyme distance is an important parameter for the efficiency of the cascade reaction for the effective transport of intermediates. However, the three enzyme co-assembled system showed a marginal increase in ADP production over the two enzyme co-assembled system. Products of XR and XDH diffuse three-dimensionally while the enzymes are assembled on a

2D scaffold. In increasing the number of enzyme reaction steps, the spatial arrangement of

the enzyme becomes an important factor for the design of the DNA scaffold.
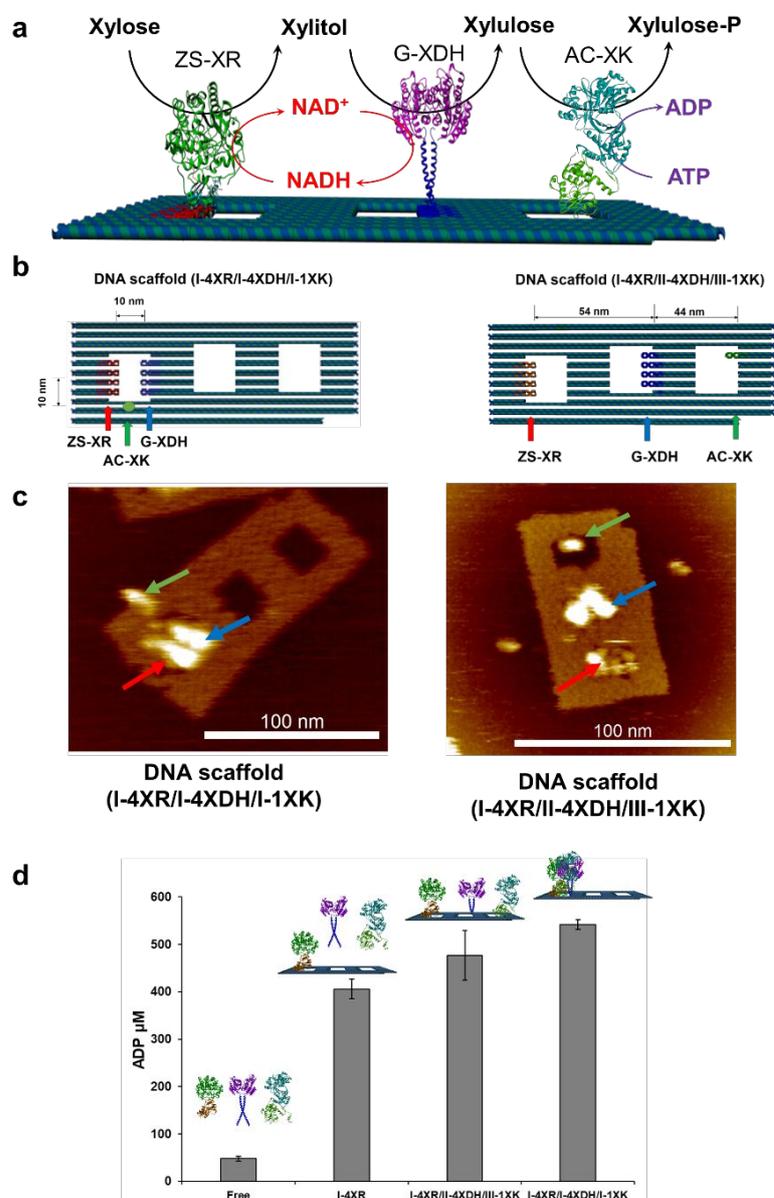
Figure 13. (a) An illustration showing the three-step cascade reaction adapted from the xylose pathway

assembled on the DNA scaffold using three types of adaptor-fused enzymes, xylose reductase (ZS-XR), xylitol

dehydrogenase (G-XDH), and xylulose kinase (AC-XK). (b) Design of DNA scaffolds with different

interenzyme distances for co-assembled ZS-XR (red arrow), G-XDH (blue arrow), and AC-XK (green arrow).

(c) AFM images of three enzymes bound on the DNA scaffold, ZS-XR, G-XDH and AC-XK. (d) Efficiencies of

the three enzymes cascade reaction on the DNA scaffold or in bulk solution. Reproduced from ref. 74 with

permission from ACS, copyright 2017.

5. Conclusion and future perspective

The power of DNA nanotechnology stimulated the rapid development of not only novel DNA nanostructures but also a variety of applications for the spatial arrangements of molecules and materials, such as nanoparticles, fluorophores, and proteins[120]. As proteins have a variety of functions, their ordered assembly is quite attractive in many fields. However, only a limited number of methods are available for specifically arranging proteins in their functional forms on DNA nanostructures[121]. The most commonly used method to arrange POI on DNA scaffolds relies on the hybridization of DNA-protein conjugates to its complementary DNA sequence on the scaffolds[24-28,122]. Though convenient, this method often results in a reduction or loss of activity upon chemical modification of the POI by the ODN[89-91,94,96,98,99]. Other methods based on the noncovalent interactions are available to arrange POI at specific locations on DNA nanostructures. However, each of these methods also suffers from drawbacks, such as the incomplete assembly of the POI, time consuming reactions, and reduced activity of the assembled POI.

These issues were challenged by protein adaptors that arrange POI at predesigned positions on the DNA scaffold through the specific complex formation of DNA-binding proteins (zif268, AZP4, and GCN4)[101,102]. The POI is genetically fused either to a monomeric or dimeric adaptor, and expressed and purified as an all protein-based adaptor-fused POI under similar conditions as for the POI without any modification. The adaptor-fused enzymes were found to maintain or even enhance the activity of the parent enzymes[73,102]. By using the adaptor-based method, POIs were efficiently loaded on DNA scaffolds at predesigned positions with fast binding kinetics and high loading yields. A drawback to the use of DNA-binding proteins as adaptors is that the attachment of the POI to the DNA scaffold relies on the noncovalent adaptor-DNA complex. Even though many DNA-binding proteins have high

affinities to their cognate DNA sequences, the adaptor-DNA complex could dissociate at high temperatures or at low concentrations of adaptor and/or DNA scaffold. The stability of DNA-adaptor complexes and the loading yield of adaptor-fused POI was improved using modular adaptors, a recombinant DNA-binding adaptor and a chemoselective protein-tag. The modular adaptor fused POI binds to predesigned positions on the DNA scaffold with higher loading yields and stability by covalent bond formation of the protein-tag and its substrate on the DNA scaffold, thereby maintaining the fast binding kinetics and sequence selectivity of parent DNA-binding adaptors[72-74]. As described in section 3.3, three modular adaptors have been used to orthogonally arrange multiple types of enzymes on a DNA scaffold[74].

There are still many issues which need to be resolved before establishing a general method to arrange POI on DNA scaffolds. As the number of POI to be arranged on the DNA scaffold increases, further orthogonal series of modular adaptors will be required. A possible approach for increasing the number of orthogonal modular adaptors is to design a modular adaptor that forms a covalent bond between the adaptor and the tag substrate exclusively depending on specific DNA sequence recognition by the DNA adaptor binding module[74]. By this way, a number of orthogonal modular adaptors could be available even from a single type of chemoselective protein tag, as demonstrated recently[114]. Alternatively, relaxases has potential as a new class of orthogonal, sequence-selective protein adaptors[75-77]. The size of adaptors could be an issue when assembling small POI or when attaching many POIs that need to be in contact with each other. For such assemblies, the steric hindrance of adaptors could reduce the loading yield of POI and limit the distance between them.

Association of multiple proteins is often crucial in driving metabolic pathways and signal transduction, spatial arrangement of proteins, especially enzymes and receptors, and is the key in understanding the mechanisms behind the extraordinary efficiency and specificity of metabolic reactions and signal transduction in cellular activities[123,124]. For example,

enzymes for sequential metabolic reactions are confined and/or aligned in close proximity in the compartment to regulate their spatial organization, preventing the loss of reaction intermediates by random diffusion, thereby accelerating transfer between enzymes[125]. Formation of higher order protein and nucleic acid complexes during transcription and translation also confines the spatial orientation of the active sites of proteins. Therefore, spatial organization of biological molecules could be a dynamic process that enables them to exert profound effects on cellular function. How do we explore the chemistry of enzymes and receptors in defined spatial orientations? DNA nanotechnology provides 2D or 3D DNA nanoarchitectures that can adopt an almost limitless variety of shapes[1-10] that serve as scaffolds to arrange proteins and other molecules in almost every possible spatial configuration. Our approach assembles protein molecules one by one at defined locations on DNA scaffolds at high yields with near nanometer precision to mimic the possible organization of biological molecules during dynamic processes in the cell. Though stable attachment of proteins through modular adaptors dismisses the dynamic equilibria or movement seen inside the cell, such a snap-shot of spatially organized enzymes and receptors is suitable for exploring their chemistry in far greater detail than is possible within cells. A combination of DNA scaffolds and appropriate methods to spatially orientate specific proteins opens an avenue to examine spatial chemistry for biological molecules.

6.   References

1   N. C. Seeman, *Mol. Biotechnol.*, 2007, **37**, 246–257.

2   N. C. Seeman, *Annu. Rev. Biochem.*, 2010, **79**, 65–87.

3   F. A. Aldaye, A. L. Palmer and H. F. Sleiman, *Science*, 2008, **321**, 1795–1799.

4   A. V. Pinheiro, D. Han, W. M. Shih and H. Yan, *Nat. Nanotechnol.*, 2011, **6**, 763–772.

5   V. Linko and H. Dietz, *Curr. Opin. Biotechnol.*, 2013, **24**, 555–561.

6   F. Zhang, J. Nangreave, Y. Liu and H. Yan, *J. Am. Chem. Soc.*, 2014, **136**, 11198–
     11211.

7   Y. J. Chen, B. Groves, R. A. Muscat and G. Seelig, *Nat. Nanotechnol.*, 2015, **10**, 748–
     760.

8   F. Hong, F. Zhang, Y. Liu and H. Yan, *Chem. Rev.* 2017, **117**, 12584-12640.

9   S. Nummelin, J. Kommeri, M. A. Kostiainen and V. Linko, *Adv. Mater.*, 2018, **30**, 1–
     12.

10   P. Wang, T. A. Meyer, V. Pan, P. K. Dutta and Y. Ke, *Chem*, 2017, **2**, 359–382.

11   N. C. Seeman, *J. Theor. Biol.*, 1982, **99**, 237–247.

12   F. Tsu-Ju and N. C. Seeman, *Biochemistry*, 1993, **32**, 3211–3220.

13   N. C. Seeman, *Nature*, 2003, **421**, 427–431.

14   P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.

15   V. Linko, S. Nummelin, L. Aarnos, K. Tapio, J. Toppari and M. Kostiainen,
     Nanomaterials, 2016, **6**, 139.

16   A. Rajendran, M. Endo and H. Sugiyama, *Angew. Chemie - Int. Ed.*, 2012, **51**, 874–
     890.

17   A. Rajendran, E. Nakata, S. Nakano and T. Morii, *ChemBioChem*, 2017, **18**, 696–716.

18   I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov and M. Sigman, *Nat. Chem.*, 2016, **8**, 299–309.

19   E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf and J. Kjems, *Nature*, 2009, **459**, 73–76.

20   S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf and W. M. Shih, *Nature*, 2009, **459**, 414–418.

21   R. Veneziano, S. Ratanalert, K. Zhang, F. Zhang, H. Yan, W. Chiu and M. Bathe, *Science*, 2016, **352**, 1534.

22   F. Zhang, S. Jiang, S. Wu, Y. Li, C. Mao, Y. Liu and H. Yan, *Nat. Nanotechnol.*, 2015, **10**, 779–784.

23   E. Benson, A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen and B. Högberg, *Nature*, 2015, **523**, 441–444.

24   C. M. Niemeyer, *Angew. Chemie - Int. Ed.*, 2010, **49**, 1200–1216.

25   B. Saccà and C. M. Niemeyer, *Chem. Soc. Rev.*, 2011, **40**, 5910–5921.

26   J. B. Trads, T. Tørring and K. V. Gothelf, *Acc. Chem. Res.*, 2017, **50**, 1367–1374.

27   Y. R. Yang, Y. Liu and H. Yan, *Bioconjug. Chem.*, 2015, **26**, 1381–1395.

28   M. Madsen and K. V. Gothelf, *Chem. Rev.*, 2019. ASAP DOI: 10.1021/acs.chemrev.8b00570

29   C. M. Niemeyer, T. Sano, C. L. Smith and C. R. Cantor, *Nucleic Acids Res.*, 1994, **22**, 5530–5539.

30   N. M. Green, *Biochem. J.*, 1963, **89**, 599–609.

31   M. Wilchek and E. A. Bayer, *Anal. Biochem.*, 1988, **171**, 1–32.

32   H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, and T. H. LaBean, *Science*, 2003, **301**, 1882–1884.

33   E. W. Voss and D. E. Lopatin, *Biochemistry*, 1971, **10**, 208–213.

34   B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter in *Molecular biology of the cell*, Garland Science,4th edition, 2002, B Cells and Antibodies.

35   Y. He, Y. Tian, Y. A. E. Ribbe and C. Mao, *J. Am. Chem. Soc.* 2006, **128**, 12664–12665.

36   T. Yamazaki, J. G. Heddle, A. Kuzuya, M. Komiyama, *Nanoscale*, 2014, **6**, 9122–9126.

37   W. Shen, H. Zhong, D. Neff and M. L. Norton, *J. Am. Chem. Soc.*, 2009, **131**, 6660–6661.

38   R. P. Goodman, C. M. Erben, J. Malo, W. M. Ho, M. L. Mckee, A. N. Kapanidis and A. J. Turberfield, *ChemBioChem*, 2009, **10**, 1551–1557.

39   L. Fruk, C. H. Kuo, E. Torres and C. M. Niemeyer, *Angew. Chemie - Int. Ed.*, 2009, **48**, 1550–1574.

40   E. N. Brody, M. C. Willis, J. D. Smith, S. Jayasena, D. Zichi and L. Gold, *Mol. Diagnosis*, 1999, **4**, 381–388.

41   C. Lin, E. Katilius, Y. Liu, J. Zhang, and H. Yan, S *Angew. Chemie - Int. Ed.* 2006, **45**, 5296–5301.

42   S. Rinker, Y. Ke, Y. Liu, R. Chhabra and H. Yan, *Nat. Nanotechnol.*, 2008, **3**, 418–422.

43   Y. Liu, C. Lin, H. Li and H. Yan, *Angew. Chemie - Int. Ed.*, 2005, **44**, 4333–4338.

44   L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564–566.

45   A. Klug and J. W. Schwabe, *FASEB J.*, 1995, **9**, 597–604.

46   N. M. Luscombe, S. E. Austin, H. M. Berman and J. M. Thornton, *Genome Biol.*, 2000, **1**, reviews001.1–001.37.

47   C. O. Pabo, E. Peisach and R. A. Grant, *Annu. Rev. Biochem*, 2001, **70**, 313–340.

48   M. Papworth, P. Kolasinska and M. Minczuk, *Gene*, 2006, **366**, 27–38.

49   K. J. Brayer and D. J. Segal, *Cell Biochem. Biophys.*, 2008, 50, 111–131.

50   B. Saccà, R. Meyer, M. Erkelenz, K. Kiko, A. Arndt, H. Schroeder, K. S. Rabe and C. M. Niemeyer, *Angew. Chemie - Int. Ed.*, 2010, **49**, 9378–9383.

51   I. R. Krauss, V. Spiridonova, A. Pica, V. Napolitano, F. Sica, *Nucleic Acids Res.* 2016, **44**, 983-991.

52   H. Li, S. H. Park, J. H. Reif, T. H. LaBean and H. Yan, *J. Am. Chem. Soc.*, 2004, **126**, 418–419.

53   K. Lund, Y. Liu, S. Lindsay and H. Yan, *J. Am. Chem. Soc.*, 2005, **127**, 17606–17607.

54   A. Kuzuya, M. Kimura, K. Numajiri, N. Koshi, T. Ohnishi, F. Okada and M. Komiyama, *ChemBioChem*, 2009, **10**, 1811–1815.

55   C. M. Niemeyer, J. Koehler and C. Wuerdemann, *ChemBioChem*, 2002, **3**, 242–245.

56   V. Linko, M. Eerikäinen and M. A. Kostiainen, *Chem. Commun.*, 2015, **51**, 5351–5354.

57   N. V. Voigt, T. Tørring, A. Rotaru, M. F. Jacobsen, J. B. Ravnsbaek, R. Subramani,

W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher and K. V. Gothelf, *Nat. Nanotechnol.*, 2010, **5**, 200–203.

58   K. Numajiri, M. Kimura, A. Kuzuya and M. Komiyama, *Chem. Commun.*, 2010, **46**, 5127–5129.

59   C. Zhang, C. Tian, F. Guo, Z. Liu, W. Jiang and C. Mao, *Angew. Chem. Int. Ed.*, 2012, **51**, 3382–3385.

60   B. A. R. Williams, K. Lund, Y. Liu, H. Yan and J. C. Chaput, *Angew. Chem. Int. Ed.*, 2007, **46**, 3051–3054.

61.   H. Li, T. H. LaBean and D. J. Kenan, *Org. Biomol. Chem.*, 2006, **4**, 3420-3426

62   R. Chhabra, J. Sharma, Y. Ke, Y. Liu, S. Rinker, S. Lindsay and H. Yan, *J. Am. Chem. Soc.*, 2007, **129**, 10304–10305.

63   M. Godonoga, T.-Y. Lin, A. Oshima, K. Sumitomo, M. S. L. Tang, Y.-W. Cheung, A. B. Kinghorn, R. M. Dirkzwager, C. Zhou, A. Kuzuya, J. A. Tanner and J. G. Heddle, *Sci. Rep.*, 2016, **6**, 21266.

64   S. V. Wegner and J. P. Spatz, *Angew. Chemie - Int. Ed.*, 2013, **52**, 7593–7596.

65   D. N. Selmi, R. J. Adamson, H. Attrill, A. D. Goddard, R. J. C. Gilbert, A. Watts and A. J. Turberfield, *Nano Lett.*, 2011, **11**, 657–660.

66   L. Fruk and C. M. Niemeyer, *Angew. Chemie - Int. Ed.*, 2005, **44**, 2603–2606.

67   L. Fruk, J. Müller and C. M. Niemeyer, *Chem. - A Eur. J.*, 2006, **12**, 7448–7457.

68   L. Fruk, J. Müller, G. Weber, A. Narváez, E. Domínguez and C. M. Niemeyer, *Chem. - A Eur. J.*, 2007, **13**, 5223–5231.

69   C. M. Niemeyer, W. Bürger and R. M. J. Hoedemakers, *Bioconjug. Chem.*, 1998, **9**, 168–175.

70   C. B. Rosen, A. L. B. Kodal, J. S. Nielsen, D. H. Schaffert, C. Scavenius, A. H. Okholm, N. V. Voigt, J. J. Enghild, J. Kjems, T. Tørring and K. V. Gothelf, *Nat. Chem.*, 2014, **6**, 8040–809.

71   K. J. Koßmann, C. Ziegler, A. Angelin, R. Meyer, M. Skoupi, K. S. Rabe and C. M. Niemeyer, *ChemBioChem*, 2016, **17**, 1102–1106.

72   E. Nakata, H. Dinh, T. A. Ngo, M. Saimura and T. Morii, *Chem. Commun.*, 2015, **51**, 1016–1019.

73   T. A. Ngo, E. Nakata, M. Saimura and T. Morii, *J. Am. Chem. Soc.*, 2016, **138**, 3012–3021.

74   T. M. Nguyen, E. Nakata, M. Saimura, H. Dinh and T. Morii, *J. Am. Chem. Soc.*, 2017, **139**, 8487–8496.

75   S. Sagredo, T. Pirzer, A. Aghebat Rafat, M. A. Goetzfried, G. Moncalian, F. C. Simmel and F. de la Cruz, *Angew. Chemie Int. Ed.*, 2016, **55**, 4348–4352.

76   K. N. Lovendahl, A. N. Hayward and W. R. Gordon, *J. Am. Chem. Soc.*, 2017, **139**, 7030–7035.

77   G. Bernardinelli and B. Högberg, *Nucleic Acids Res.*, 2017, **45**, 1–9.

78   R. Meyer, S. Giselbrecht B. E. Rapp, M. Hirtz, C. M. Niemeyer, *Curr. Opin. Chem. Biol.* 2014, **18**, 8-15.

79   X. Ouyang, M. De Stefano, A. Krissanaprasit, A. L. Bank Kodal, C. Bech Rosen, T. Liu, S. Helmig, C. Fan and K. V. Gothelf, *Angew. Chemie - Int. Ed.*, 2017, **56**, 14423–14427.

80   A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nat. Biotechnol.*, 2003, **21**, 86–89.

81   G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood, *ACS Chem. Biol.*, 2008, **3**, 373–382.

82   T. Masubuchi, M. Endo, R. Iizuka, A. Iguchi, D. H. Yoon, T. Sekiguchi, H. Qi, R. Iinuma, Y. Miyazono, S. Shoji, T. Funatsu, H. Sugiyama, Y. Harada, T. Ueda and H. Tadakuma, *Nat. Nanotechnol.*, 2018, **13**, 933–940.

83   A. Gautier, A. Juillerat, C. Heinis, I. R. Corrêa Jr., M. Kindermann, F. Beaufils and K. Johnsson, *Chem. Biol.*, 2008, **15**, 128–136.

84   B. Zakeri, J.O. Fierer, E. Celik, E.C. Chittock, U. Schwarz-Linek, V.T. Moy, M. Howarth, *Proc. Natl. Acad. Sci. USA* 2012,**109**, E690–E697.

85   S. C. Reddington, M. Howarth, *Curr. Opin. Chem. Biol.* 2015, **29**, 94-99.

86   A. Banerjee, M. Howarth, *Curr. Opin. Biotech.* 2018, **51**, 16-23.

87   D. Hatlen, T. T. D. Linke, J. C. Leo, *Int. J. Mol. Sci.* 2019, **20**, 2129.

88   J. Müller and C. M. Niemeyer, *Biochem. Biophys. Res. Commun.*, 2008, **377**, 62–67.

89   O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman and I. Willner, *Nat. Nanotechnol.*, 2009, **4**, 249–254.

90.   J. Fu, M. Liu, Y. Liu, N. W. Woodbury and H. Yan, *J. Am. Chem. Soc.*, 2012, **134**, 5516–5519.

91   Y. Fu, D. Zeng, J. Chao, Y. Jin, Z. Zhang, H. Liu, D. Li, H. Ma, Q. Huang, K. V. Gothelf and C. Fan, *J. Am. Chem. Soc.*, 2013, **135**, 696–702.

92   Z. Zhao, J. Fu, S. Dhakal, A. Johnson-Buck, M. Liu, T. Zhang, N. W. Woodbury, Y. Liu, N. G. Walter and H. Yan, *Nat. Commun.*, 2016, **7**, 10619.

93 M. Liu, J. Fu, X. Qi, S. Wootten, N. W. Woodbury, Y. Liu and H. Yan, *ChemBioChem*, 2016, **17**, 1097–1101.

94 J. Fu, Y. R. Yang, A. Johnson-Buck, M. Liu, Y. Liu, N. G. Walter, N. W. Woodbury and H. Yan, *Nat. Nanotechnol.*, 2014, **9**, 531–536.

95 G. Ke, M. Liu, S. Jiang, X. Qi, Y. R. Yang, S. Wootten, F. Zhang, Z. Zhu, Y. Liu, C. J. Yang and H. Yan, *Angew. Chemie - Int. Ed.*, 2016, **55**, 7483–7486.

96 L. Xin, C. Zhou, Z. Yang and D. Liu, *Small*, 2013, **9**, 3088–3091.

97 S. M. Douglas, I. Bachelet and G. M. Church, *Science*, 2012, **335**, 831.

98 G. Grossi, M. Dalgaard Ebbesen Jepsen, J. Kjems and E. S. Andersen, *Nat. Commun.*, 2017, **8**, 1–8.

99 J. Fu, Y. R. Yang, S. Dhakal, Z. Zhao, M. Liu, T. Zhang, N. G. Walter and H. Yan, *Nat. Protoc.*, 2016, **11**, 2243–2273.

100 G. Grossi, A. Jaekel, E. S. Andersen and B. Saccà, *MRS Bull.*, 2017, **42**, 920–924.

101 E. Nakata, F. F. Liew, C. Uwatoko, S. Kiyonaka, Y. Mori, Y. Katsuda, M. Endo, H. Sugiyama and T. Morii, *Angew. Chemie - Int. Ed.*, 2012, **51**, 2421–2424.

102 T. A. Ngo, E. Nakata, M. Saimura, T. Kodaki and T. Morii, *Methods*, 2014, **67**, 142–150.

103 T. Sera and C. Uranga, *Biochemistry*, 2002, **41**, 7074–7081.

104 N. P. Pavletich and C. O. Pabo, *Science*, 1991, **252**, 809–817.

105 H. A. Greisman and C. O. Pabo, *Science*, 1997, **275**, 657–661.

106 C. I. Stains, J. R. Porter, A. T. Ooi, D. J. Segal and I. Ghosh, *J. Am. Chem. Soc.*, 2005, **127**, 10782–10783.

107   E. K. O'Shea, R. Rutkowski and P. S. Kim, *Science*, 1989, **243**, 538–542.

108   T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–1237.

109   P. König and T. J. Richmond, *J. Mol. Biol.*, 1993, 233, 139–154.

110   W. Keller, P. König and T. J. Richmond, *J. Mol. Biol.*, 1995, **254**, 657–667.

111   G. Mei, A. Di Venere, N. Rosato and A. Finazzi-Agrò, *FEBS J.*, 2005, **272**, 16–27.

112   S. Watanabe, T. Kodaki and K. Makino, *J. Biol. Chem.*, 2005, **280**, 10340–10349.

113   S. L. Pival, R. Birner-Gruenberger, C. Krump and B. Nidetzky, *Protein Expr. Purif.*, 2011, **79**, 223–230.

114   T. M. Nguyen, E. Nakata, Z. Zhang, M. Saimura, H. Dinh and T. Morii, *Chem. Sci.*, 2019, DOI: 10.1039/C9SC02990G.

115   M. Chandler, F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalian and B. Ton-Hoang, *Nat. Rev. Microbiol.*, 2013, **11**, 525–538.

116   T. Kurokawa, S. Kiyonaka, E. Nakata, M. Endo, S. Koyama, E. Mori, N. H. Tran, H. Dinh, Y. Suzuki, K. Hidaka, M. Kawata, C. Sato, H. Sugiyama, T. Morii and Y. Mori, *Angew. Chemie - Int. Ed.*, 2018, **57**, 2586–2591.

117   E. Winkelhausen and S. Kuzmanova, *J. Ferment. Bioeng.*, 1998, **86**, 1–14.

118   T. W. Jeffries, *Curr. Opin. Biotechnol.*, 2006, **17**, 320–326.

119   A. Matsushika, H. Inoue, T. Kodaki and S. Sawayama, *Appl. Microbiol. Biotechnol.*, 2009, **84**, 34-53.

120   A. F. Jorge and R. Eritja, *Pharmaceutics*, 2018, **10**, 268.

121   S. L. Kuan, F. R. G. Bergamini and T. Weil, *Chem. Soc. Rev.*, 2018, **47**, 9069–9105.

122 Y. Sakai, M. S. Islam, M. Adamiak, S. C.-C. Shiu, J. A. Tanner and J. G. Heddle, *Genes (Basel).*, 2018, **9**, 571.

123 C. M. Agapakis, P. M. Boyle and P. A. Silver, *Nat. Chem. Biol.*, 2012, **8**, 527–535.

124 R. J. Conrado, J. D. Varner and M. P. DeLisa, *Curr. Opin. Biotechnol.*, 2008, **19**, 492–499.

125 L. J. Sweetlove and A. R. Fernie, *Nat. Commun.*, 2018, **9**, 2136.