NORTHWESTERN UNIVERSITY

Programming Protein Organization into Sequence-Encoded Architectures Using DNA

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILMENTS OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSPHY

Field of Chemistry

By

Peter Henry Winegar

EVANSTON, ILLINOIS

August 2022

ABSTRACT

Programming Protein Organization into Sequence-Encoded Architectures Using DNA

Peter Henry Winegar

Proteins are the nanoscale building blocks of life. Their sophisticated but well-defined architectures result in complex biological functions, including ones involved in metabolism, photosynthesis, transcription, translation, and immunity. To study and improve upon the natural functions of proteins, it is desirable to develop methodology for organizing proteins into targeted architectures. While methods exist for controlling proteins and even small collections of proteins into aggregate materials, it is still challenging to organize proteins reliably and quickly into predefined architectures. This is because the chemical structures of proteins and protein–protein interactions are highly complex. The work in this dissertation studies how complex protein–protein interactions can be replaced with well-defined DNA–DNA interactions to program the organization of proteins into sequence-encoded architectures, including single crystals and oligomers. In Chapter One, current approaches to synthesize protein architectures and their limitations are described. In Chapter Two, the impact of DNA on protein crystallization is investigated by replacing native protein–protein interactions with DNA–DNA interactions. In Chapter Three, a modular DNA scaffold is designed to explore the generalizable assembly of proteins into vast numbers of oligomers that contain exact numbers and orders of proteins. In Chapter Four, protein block co-oligomers are synthesized to study the impact of block design on assembly outcome. In Chapter Five, fundamental lessons learned from this work are summarized and future research directions are discussed. Collectively, this dissertation establishes methods to

program the organization of proteins into synthetic architectures and discusses the implications of these architectures on protein function, enabling future development of protein-based materials that mimic and surpass the natural functions of proteins.

_____

Peter H. Winegar

_____

Advisor: Chad A. Mirkin

ACKNOWLEDGEMENTS

My graduate research would not have been possible without extensive support from Northwestern University, mentors, collaborators, friends, and family. Pursuing a graduate degree during the last five years has been a turning point in my life because I have grown as a scientist, person, and leader. I thank my thesis advisor, Professor Chad Mirkin, who has established a dynamic research lab where graduate students have the freedom and resources to study fundamental scientific questions. I am exceedingly grateful for his support because he continually challenged me to improve as a scientist and communicator. I thank the other members of my thesis committee, Professor Amy Rosenzweig and Professor Milan Mrksich, as well as my qualifying exam committee chair, Professor Brian Hoffman, for many thoughtful discussions.

I have been blessed with outstanding graduate student and postdoctoral researcher mentors in the Mirkin Lab. Dr. Janet McMillan taught me how to pursue hypothesis-driven research and brought me into the lab's social groups. Dr. Oliver Hayes helped me troubleshoot experiments and chatted with me from his neighboring lab bench. Prof. Adrian Figg coached me in all aspects of scientific research, including writing and presenting. I thank Janet, Oliver, and Adrian for being role models for how to conduct myself personally and professionally. You are great friends!

During my graduate studies, I was fortunate to have the opportunity to serve as a graduate student mentor to several younger graduate students: Yinglun Ma, Catherine Davidson, Jennifer Delgado, and Rachel Chan. Each of you have so much enthusiasm for your research, and I have learned as much from you as I taught you.

I was incredibly fortunate to have pursued a Ph.D. in a research lab with an extensive and kind support staff. Thank you to the members of the office and research staff in the Mirkin Lab: Pam Watson, Elizabeth Forest, Bethany Hexom, Dr. Sara Rupich, Dr. Jenny Orbeck, Dr. Sarah Petrosko, and Dr. Tanushri Sengupta. I also acknowledge Robert Stawicki, Matthew Capek, Jennifer Delgado, and Vinzenz Mayer, who together synthesized hundreds of DNA strands for use in my research projects.

Dr. Pam Focia and Prof. Alfonso Mondragón were truly generous with their time to teach me skills related to protein crystallization, synchrotron X-ray diffraction, and protein structure determination. I thank the beamline scientists from Argonne National Laboratory's LS-CAT as well as staff scientists from Northwestern's IMSERC, NUANCE, SBF, and rPPC core facilities. These facilities are truly state-of-the-art and were essential for my Ph.D. research projects.

My scientific training did not begin in graduate school. I owe a debt of gratitude to all my undergraduate research mentors, including Professor Loredana Valenzano, Professor Laura Gagliardi, Professor João Marcelo Lopes, Professor Rigoberto Advincula, Professor Julia King, Professor Lujia (Luke) Liu, Dr. Dale Pahls, Dr. Varinia Bernales, and Dr. Brylee Tiu. Each of you spent many hours teaching me how to perform literature searches, generate hypotheses, design experiments, and present my findings. My success in graduate school can be attributed to your training.

Thank you to great friends and collaborators from the Mirkin Lab: Prof. Benjamin Partridge, Prof. Michelle Teplensky, Namrata Ramani, Katie Landy, Dr. Zachary Urbach, Max Distler, Dr. Alex Anderson, Prof. Devleena Samanta, Dr. Kyle Gibson, Zhenyu (Henry) Han, and many more. Thanks to my other friends at Northwestern and elsewhere: Kenton Hicks, Rueih-

PREFACE

This dissertation is submitted for the degree of Doctor of Philosophy at Northwestern University. The research conducted herein was conducted under the supervision of Professor Chad A. Mirkin in the Department of Chemistry, between November 2017 and August 2022. The following work is original.

Part of this work was presented in:

McMillan, J. R.; Hayes, O. G.; Winegar, P. H.; Mirkin, C. A. Protein Materials Engineering with DNA. *Acc. Chem. Res.* **2019**, *52*, 1939−1948.

Winegar, P. H.; Hayes, O. G.; McMillan, J. R.; Focia, P. J.; Figg, C. A.; Mirkin, C. A. DNA-Directed Protein Packing within Single Crystals. *Chem* **2020**, *6*, 1007–1017.

Figg, C. A.; Winegar, P. H.; Hayes, O. G.; Mirkin, C. A. Controlling the DNA Hybridization Chain Reaction. *J. Am. Chem. Soc.* **2020**, *142*, 8596–8601.

Winegar, P. H.*; Figg, C. A.*; Teplensky, M. H.; Ramani, N.; Mirkin, C. A. Modular Nucleic Acid Scaffolds for Synthesizing Sequence-Encoded Antibody Oligomers. *Chem* **2022**, https://doi.org/10.1016/j.chempr.2022.07.003.

Part of this work will be presented in:

Winegar, P. H.; Figg, C. A.; Ma, Y.; Han, Z.; Chan, R. R.; Landy, K. M.; Mirkin, C. A. Protein Block Co-Polymer Assembly Using DNA and Metal Coordination. *Manuscript in Preparation*.

*Equal Author Contribution

LIST OF ABBREVIATIONS AND NOMENCLATURE

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
A	Adenine
A	Anti-mouse-PD-1
AAO	L-amino acid oxidase
Ab	Antibody
B	Anti-mouse-TIGIT
Bp	Base pair
BSA	Bovine serum albumin
BSA(py)	Bovine serum albumin modified with pyridine groups
Bz	Benzoates
C	Cytosine
C	Anti-mouse-CTLA-4
cDNA	Complementary deoxyribonucleic acid
ConA	Concanavalin A
CPG	Controlled pore glass
Cryo EM	Cryogenic electron microscopy
Cy3	Cyanine 3
Cy5	Cyanine 5
Cys	Cysteine
D	Anti-human-PD-1
DBCO	Dibenzocyclooctyne
DLS	Dynamic light scattering
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ε	Extinction coefficient
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
Equiv.	Equivalents
Ex/em	Excitation/emission
FITC	Fluorescein
G	Guanine
GFP	Green fluorescent protein
GFP(py)	Green fluorescent protein modified with pyridine groups
$H1_4$	Hairpin sequence 1 with a mismatch at base 4
$H2_4$	Hairpin sequence 2 with a mismatch at base 4
HCR	Hybridization chain reaction
His	Histidine
HPLC	High-pressure liquid chromatography

I2$_4$	Initiator sequence 2 with a mismatch at base 4
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilodalton
Lys	Lysine
MALDI-TOF MS	Matrix-assisted laser desorption-ionization-time-of-flight mass spectrometry
MBP	Maltose-binding protein
MBP(py)	Maltose-binding protein modified with pyridine groups
mGFP	Mutant green fluorescent protein
M_n	Number-averaged molecular weight
M_w	Weight-averaged molecular weight
$MW_{exper.}$	Experimental molecular weight
$MW_{theor.}$	Theoretical molecular weight
ncDNA	Non-complementary deoxyribonucleic acid
NHS	*N*-hydroxysuccinimide
PAE	Programmable atom equivalent
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDI	Polydispersity index
PEG	Polyethylene glycol
PPIs	Protein–protein interactions
RL	*Renilla* luciferase
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RT	Room temperature
SA	Streptavidin
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
scDNA	Self-complementary deoxyribonucleic acid
SEC	Size-exclusion chromatography
SPDP	Succinimidyl 3-(2-pyridyldithio)propionate
SMCC	Succinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate
T	Thymine
T(DBCO)	DBCO-dT-CE phosphoramidite
TEM	Transmission electron microscopy
TEV	Tobacco etch virus
TOTO-3	Thiazole red homodimer
UV-vis	Ultraviolet-visible

This thesis is dedicated to my parents,

Keith and Denice Winegar

I am incredibly thankful for your unending love, support, and friendship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

CHAPTER ONE

Natural and Engineered Protein Architectures.

Material in this chapter is partly based upon published work:

McMillan, J. R.; Hayes, O. G.; Winegar, P. H.; Mirkin, C. A. Protein Materials Engineering with DNA. *Acc. Chem. Res.* **2019**, *52*, 1939−1948.

1.1. Proteins Structures, Interactions, and Functions.

1.1.1. Protein Structures and Interactions in Nature.

Proteins are an important class of natural biopolymers that function as nanoscale building blocks.[1] Proteins are linear and sequence-encoded polypeptides of the 20 canonical amino acids that fold into three-dimensional structures (Figure 1.1.A).[2-3] Interactions between the surfaces of proteins (protein–protein interactions (PPIs)) consist of complex combinations of electrostatic interactions, hydrophobic interactions, hydrogen bonds, and disulfide bonds (Figure 1.1.B).[4] Proteins assemble via PPIs into defined architectures, including discrete (*e.g.*, cyanobacteria photosystem I),[5] one-dimensional (1D, *e.g.*, actin and myosin filaments),[6] two-dimensional (2D, *e.g.*, bacterial S-layers),[7] and three-dimensional (3D, *e.g.*, collage networks)[8] architectures (Figure 1.1.C and D).[9-11] Many biological functions are dictated by these protein architectures, including the conversion of light energy into chemical energy,[5] the catalysis of chemical reactions,[12] the transcription and translation of genetic information,[13-14] the signaling of immune responses,[15] the directing of macroscopic movement,[6] and the transportation of molecules across membranes.[16] Therefore, it is desirable to prepare synthetic architectures of proteins to study and improve upon the natural functions of proteins.
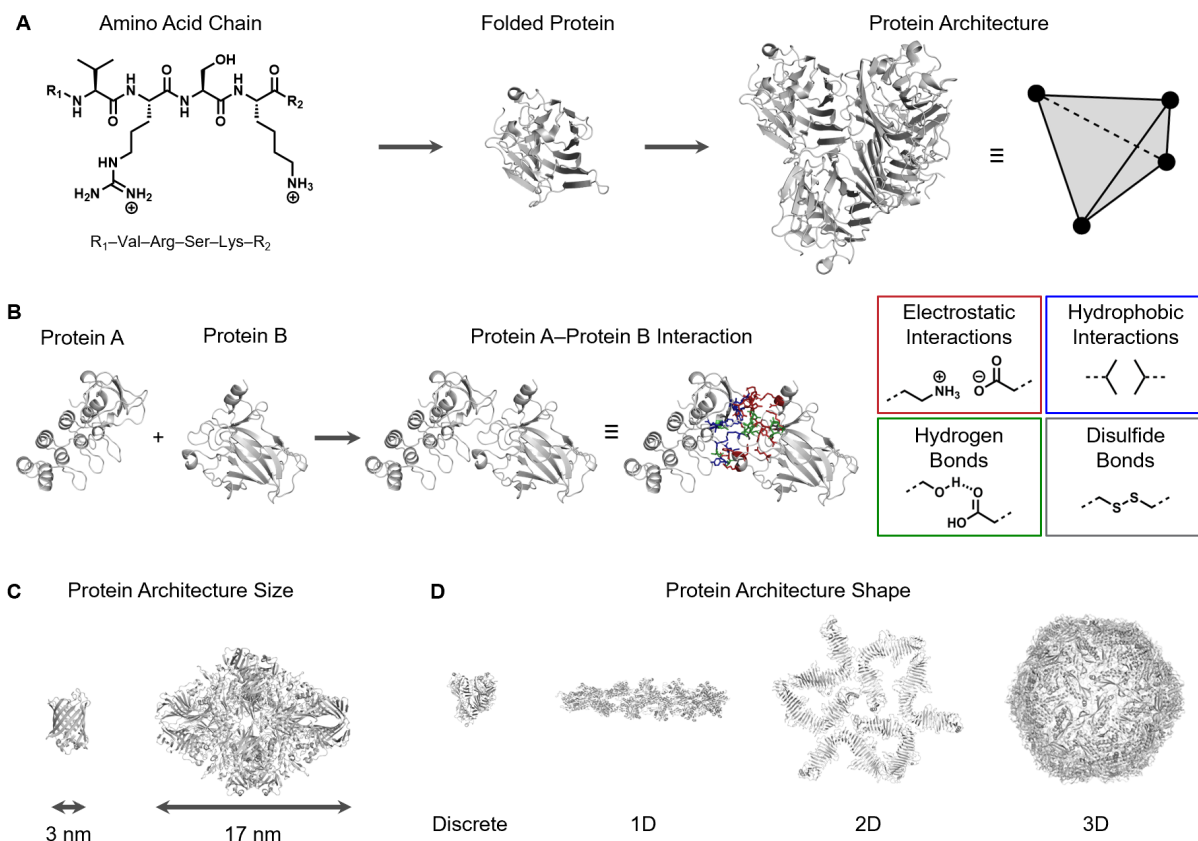
Figure 1.1. Structures of proteins and protein–protein interactions. (A) Proteins are linear chains of amino acids that fold into defined 3D structures. Folded proteins can further assemble into defined architectures. (B) An example protein–protein interaction (PPI) between the surfaces of the two model proteins, A and B, is depicted (PDB code: **1YCS**).[17] The A–B PPI contains electrostatic interactions (red), hydrophobic interactions (blue), and hydrogen bonds (green). Other PPIs contain disulfide bonds (dark gray). (C) Protein architectures have defined nanoscale sizes (PDB codes: **6UHJ** and **1JYX**)[18-19] as well as (D) discrete, 1D, 2D, or 3D shapes (PDB codes: **1YCS**, **3B63**, **5N8P**, and **6I9G**).[17, 20-22]

1.1.2. Methods to Prepare Synthetic Protein Architectures and Limitations.

Proteins and PPIs are chemically complex; therefore, it is challenging to design synthetic protein architectures. There are two main strategies that have been used to overcome this challenge: replacing native PPIs with synthetic PPIs and covalently linking proteins.[23-26] Proteins can be engineered to exhibit synthetic PPIs through the judicious design of protein amino acid sequences and the surface chemistry of proteins. Synthetic PPIs can involve metal coordination,[27-33] protein–ligand binding,[34-40] electrostatic interactions,[41-43] supramolecular host–guest binding,[35,

[38, 44-46] and hydrophobic interactions (Figure 1.2).[47-48] Recent advances in the *de novo* design of full proteins has enabled the design of synthetic PPIs that mimic the complexity of native PPIs.[49-58] In another approach, proteins can be covalently linked using approaches from chemistry or molecular biology. First, bioconjugation reactions can be used to covalently link proteins to each other[59-60] or to covalently link proteins to chemical scaffolds (*e.g.*, polymers).[61-65] Use of these approaches have greatly expanded following the introduction of bioorthogonal "click" reactions.[66-68] Second, multiple proteins can be fused and recombinantly expressed as a single polypeptide.[69-71] Proteins can also be fused to a subunit of a multimeric protein (*e.g.*, streptavidin) that assembles into supramolecular constructs[72-75] or to a subunit that can selectively form covalent bonds with a complementary group (*e.g.*, SpyTag/SpyCatcher,[76] SnoopTag/Snoopcatcher,[77] SNAPtag,[78] HALOtag,[79] or cutinase[80]).[77, 81-84] However, current methods to prepare synthetic protein architectures often require the introduction of recognition sites directly into the protein structure via site-directed mutagenesis and recombinant expression. Distinct protein mutants are often required to achieve each different synthetic protein architecture. Since many proteins can be challenging to mutate and recombinantly express (*e.g.*, proteins with post-translational modifications,[85] proteins with disulfide bonds,[86] toxic proteins,[87] or proteins that aggregate[88]), it can be challenging to generalize these methods to many proteins.
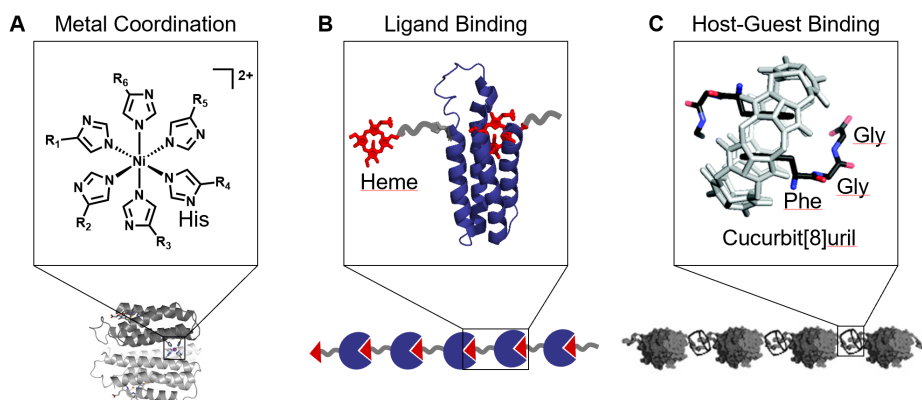
Figure 1.2. Methods to prepare synthetic protein architectures. (A) Metal coordination, (B) ligand binding, and (C) host-guest binding can be used to introduce synthetic PPIs between proteins and direct protein assembly into defined architectures. Adapted with permission from Kakkis, A. et al ref. [33]. Copyright 2020 Wiley. Adapted from Kitagishi, H. et al ref. [40]. Copyright 2007 American Chemical Society. Adapted with permission from Hou, C. et al ref. [46]. Copyright 2013 Wiley.

1.2. Programming Nanoscale Architectures Using DNA.

1.2.1. DNA Structures and Interactions.

Deoxyribonucleic acid (DNA) is a natural biopolymer that encodes genetic information in sequences of nucleobases: adenine (A), thymine (T), guanine (G), and cytosine (C).[89] Complementary sets of nucleobases (*i.e.*, A and T, G and C) interact via specific hydrogen bonding, termed Watson–Crick–Franklin base-pairing (Figure 1.3.A).[90] Pairs of DNA that contain complementary series of nucleobases hybridize and form a DNA double helix with predictable thermodynamics, specificity, and structure.[91] DNA can also form stimuli-responsive interactions; for example, C-rich DNA sequences form i-motifs at low pH[92] and G-rich sequences form G-quadruplexes in the presence of monovalent cations.[93] Pioneering breakthrough discoveries in solid-phase chemistry[94-95] and polymerase chain reaction[96] have enabled methods to prepare synthetic oligonucleotides with programmable sequences and lengths. These approaches are amenable to the synthesis of DNA strands that contain a myriad of chemical groups (*e.g.*, dyes,

amines, "click chemistry" groups, etc.). Together, these advances have led to the use of DNA as a synthetic ligand to program material structure at the nanoscale.
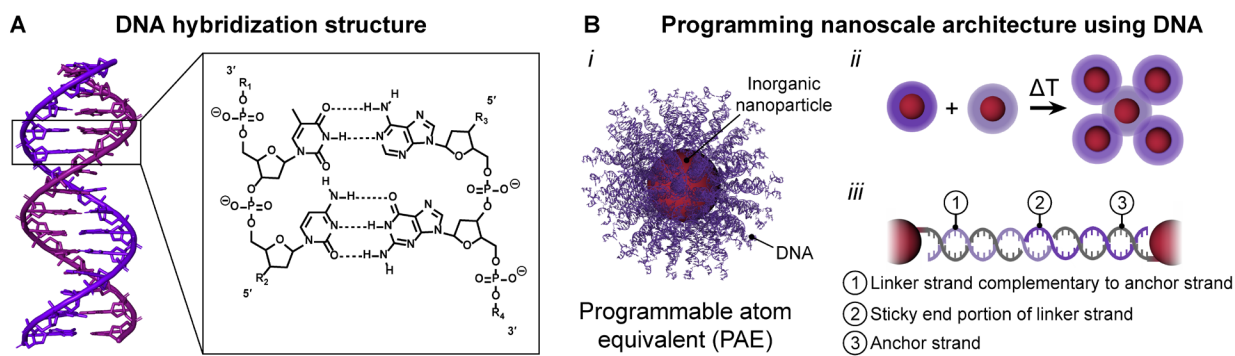
Figure 1.3. Programming nanoscale structure with DNA hybridization. (A) Model (left) and chemical structure (right) of the DNA double helix hybridization interaction. (B) DNA can be used to program nanoscale architecture. *i* For example, inorganic nanoparticles that are densely functionalized with a shell of radially-oriented DNA act as a programmable atom equivalent (PAE). *ii* PAEs with complementary sticky ends assemble into structures analogous to atomic lattices. *iii* Scheme of the DNA design that is used for PAE assembly.

1.2.2. Structural DNA Nanotechnology and Crystal Engineering Using DNA.

The use of DNA as a structural building block of materials was introduced by the Seeman Lab, who designed junctions of DNA[97-98] that can be combined into crystalline frameworks.[99] Similarly, DNA origami is a technique that uses many short DNA strands as structural "staples" between different sites on a long DNA template, directing its folding into arbitrary, predefined structures.[100-101] In a separate approach, the Alivisatos and Mirkin Labs discovered that DNA hybridization interactions between the surfaces of nanoparticles can control their assembly into discrete and extended structures, respectively.[102-103] An explosion of work has followed these original reports, where the programmable thermodynamics, specificities, and structures of DNA hybridization interactions are harnessed to program the assembly of nanoscale architectures.[104-108] DNA attached to the surface of nanoparticles are radially-oriented, therefore, the cooperative DNA interactions between two nanoparticles modified with complementary sequences are directional or

"bond-like."[109] These DNA-modified nanoparticles, termed programmable atom equivalents (PAEs), assemble via DNA "bonds" into structures analogous to atomic lattices that maximize the total number of DNA interactions (Figure 1.3.B).[104, 110]

1.3. Protein–DNA Materials.

1.3.1. Bioconjugation of DNA to Proteins.

A wide range of covalent bioconjugation methods have been developed to achieve the site-specific modification of proteins with synthetic oligonucleotides.[60, 111-112] Generally, these covalent bioconjugation reactions require low concentrations (nM–μM), mild temperatures (4–37 °C) and aqueous buffers (near physiological pH) to maintain the proper folding of proteins.[113] Common covalent bioconjugation reactions target reactive groups including amines (*N*-termini and lysine side chains),[114] thiols (cysteine side chains),[115] or carboxylic acids (*C*-termini, glutamic acid side chains, and aspartic acid side chains).[116] Proteins, on average, contain a higher abundance of amines and carboxylic acids as well as a lower abundance of thiols.[117] Therefore, bioconjugation methods can be selected to roughly match the desired number of DNA modifications per protein. The amino acid sequence of proteins can be modified using site-directed mutagenesis or genetic code expansion to control the number or location of reactive groups to be modified with DNA.[118] Alternatively, proteins can be attached to proteins noncovalently by taking advantage of native or introduced ligand- or DNA-binding sites on proteins.[111-112]

1.3.2. Materials Built from Protein–DNA Materials.

There are six key parameters of protein–DNA building blocks: (1) the amino acid sequences and folding structure of proteins, (2) the base sequences of DNA, (3) protein–DNA linkage structure, (4) DNA distribution on the protein, (5) number of DNA modifications per

protein, and (6) the shape and symmetry of the protein (Figure 1.4). By tuning these parameters, DNA-modified proteins have been organized via DNA–DNA interactions into a multitude of designed one-,[119-130] two-,[131-153] and three-dimensional[18, 154-167] protein architectures.
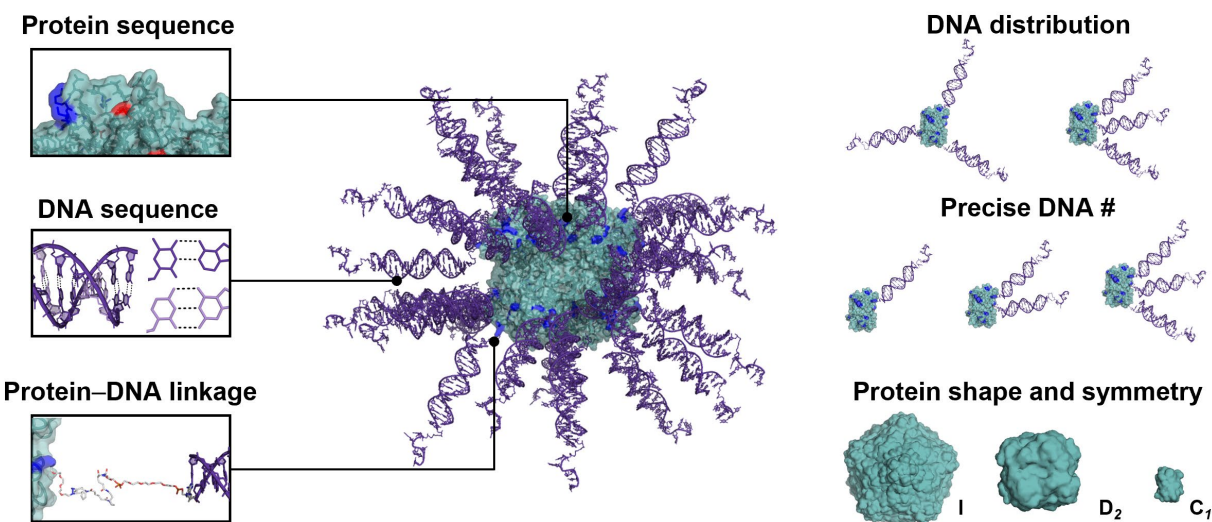
Figure 1.4. Design parameters of protein–DNA building blocks. Assembly architectures of proteins depend on many parameters, including amino acid sequence of proteins, base sequence of DNA, and protein–DNA attachment chemistry. Additional parameters include the distribution and number of DNA modifications on proteins as well as the inherent shape and symmetry of the protein.

Current approaches to build materials from protein–DNA building blocks can be classified as template-directed approaches (*i.e.*, proteins are tethered to a DNA structure)[168] or surface-directed approaches (*i.e.*, protein and DNA both contribute to the final structure).[106] In template driven approaches, the assembly architectures of proteins depend solely upon designed DNA interactions.[105, 168-170] DNA nanostructures such as tetrahedra,[131] junctions,[132-134] tiles,[141] and origami[135-136, 154] can be functionalized with proteins. Alternatively, DNA oligomerization methods including hybridization chain reaction (HCR) can organize protein-modified DNA monomers into discrete and extended architectures.[121-122] Recent work from the Mirkin Lab and others have designed directional DNA interactions between the surfaces of proteins, where

directionality is imposed by the protein–DNA attachment as well as steric and electrostatic crowding of DNA. By modifying opposite ends of proteins with as few as two DNA strands, proteins can be organized via surface-directed DNA interactions into one-[119-120] and three-dimensional structures.[157] The chemical heterogeneity, shape, and symmetry of protein surfaces enables their modification with an anisotropic distribution of DNA. These protein–DNA building blocks can exhibit one[158] or multiple (*e.g.*, Janus-type)[160, 163] directional interactions to program the complex or hierarchichal three-dimensional assembly of proteins.[156] Together, protein–DNA conjugates are highly exciting building blocks to create designer materials that harness the predictable interactions of DNA and the functions of proteins. However, one unmet challenge in this field is the ability to access a wide range of different sequence-encoded architectures without redesigning the protein, the DNA, or the protein–DNA linkage. This challenge is fully addressed by this dissertation.

1.4. Thesis Overview.

In this dissertation, ill-defined protein–protein interactions (PPIs) are replaced with well-defined DNA hybridization interactions to program the assembly of proteins into sequence-encoded single crystals and oligomers using DNA. This dissertation teaches how the amino acid sequences of proteins, the base sequences of DNA, protein–DNA attachment sites, and protein–DNA linker structure of protein–DNA building blocks can regulate synthetic protein assembly architectures. Ultimately, these lessons are applied to unmet challenges in the field of protein assembly: designing protein packing within single crystals, organizing proteins into a wide variety of different oligomers that contain exact numbers and orders of proteins, and controlling the assembly morphologies of protein block co-oligomers.

In Chapter Two ("*DNA-Directed Protein Packing within Single Crystals.*"), the impact of replacing native PPIs with DNA hybridization interactions on protein crystallization is studied (Figure 1.5). Protein crystallization is directed by placing DNA hybridization interactions between the surfaces of proteins. The designs of DNA (*i.e.*, length, sequence, and complementarity) and DNA as protein–DNA linkage (*i.e.*, site and flexibility) are systematically varied to learn how these parameters impact crystallization. The important lesson learned is that decreasing flexibility of the introduced linkage between proteins is key for protein–DNA building blocks to crystallize. Together, this work is an essential step toward designing and engineering protein packing within single crystals and could lead to future applications in protein structure determination and functional protein crystal materials.
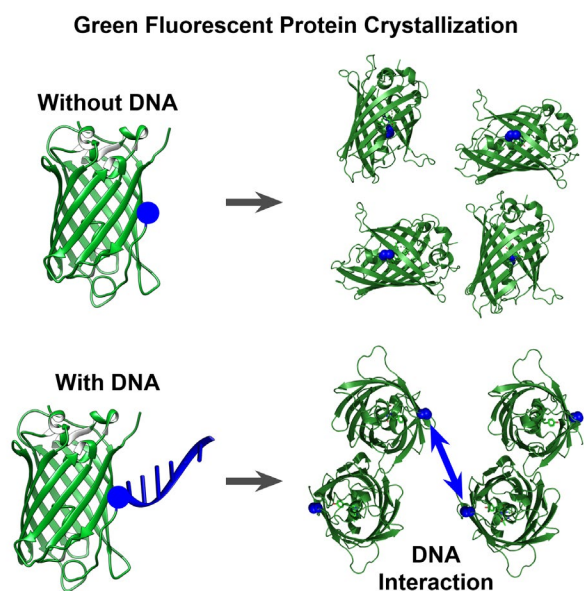
Figure 1.5. Studying the impact of DNA interactions on protein crystallization. In this work, the impact of DNA hybridization interactions on protein crystallization is investigated. The packing of proteins within crystals of mutant green fluorescent protein (mGFP, top) and DNA-modified mGFP (bottom) are distinctly different.

In Chapter Three ("*Modular Nucleic Acid Scaffolds for Synthesizing Monodisperse and Sequence-Encoded Antibody Oligomers.*"), a modular DNA scaffold is designed that can be used

to prepare a wide range of different monodisperse, sequence-encoded protein oligomers (Figure 1.6). Monomer building blocks are prepared by using generalizable bioconjugation to modify proteins with one synthetic oligonucleotide. Next, DNA base-pair recognition is used to organize monomers into oligomers with predefined numbers, stoichiometries, and oligomeric sequences of proteins. Importantly, nearly limitless numbers of different sequence-encoded oligomers can be synthesized without the need to redesign the proteins, the DNA scaffold, or the protein–DNA linkage. This synthetic advance enables subsequent studies to discover the fundamental relationships between protein oligomer structures and properties, which have significant implications for many fields (*e.g.*, therapeutics, catalysis, photosynthesis, and membrane transport).
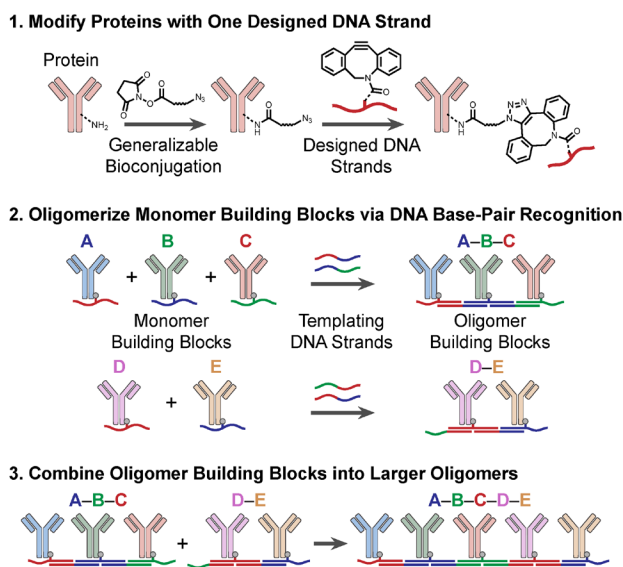
Figure 1.6. A generalizable approach to prepare sequence-encoded protein oligomers. A modular DNA scaffold is designed that can be used to organize proteins into nearly limitless numbers of different sequence-encoded oligomers. Proteins are modified with one designed DNA strand using generalizable bioconjugation chemistry. Next, monomers are oligomerized via DNA base-pairing into dimeric and trimeric building blocks that can be further combined into larger sequence-encoded oligomers.

In Chapter Four ("*Protein Block Co-Oligomer Assembly Using DNA and Metal Coordination.*"), protein block co-oligomers are synthesized using controlled DNA HCR to investigate the implications of block design on assembly morphology (Figure 1.7). A set of model proteins are functionalized with one DNA hairpin and proteins are modified with many coordinating ligands. This work will enable studies into how metal coordination between specific blocks can lead to predictable assembly morphologies of proteins. We will investigate the impact of the number of coordinating ligands per protein and the number of proteins with a block on assembly morphology. This work will ultimately lead to methods to achieve the self-segregating assembly of multiple proteins for uses as devices and synthetic organelles.
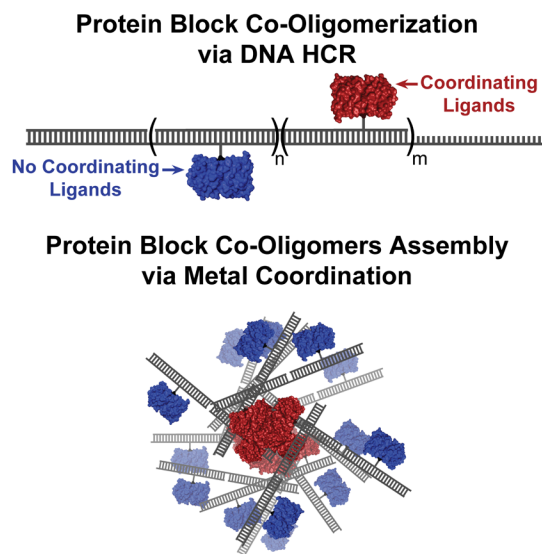
Figure 1.7. Investigating the impact block design on protein block co-oligomer assembly. Proteins can be assembled into protein block co-oligomers using controlled DNA HCR oligomerization. We will investigate how coordination between metal cations and proteins on specific blocks can direct the assembly of block co-oligomers.

In Chapter Five, fundamental lessons learned from this work are summarized and future research directions are discussed. As a whole, this dissertation establishes methods to program the

organization of proteins into predefined and sequence-encoded single crystals and oligomers. These protein architectures will enable the future development of protein-based materials that mimic and surpass the natural functions of proteins.

CHAPTER TWO

DNA-Directed Protein Packing within Single Crystals.

Material in this chapter is partly based upon published work:

2.1. Natural and Synthetic Protein Single Crystals.

In biology, periodic architectures of proteins (*i.e.*, bacterial S-layers,[7] actin and myosin,[6] and tubulin[171]) can dictate important biological functions (*i.e.*, glycan presentation,[7] muscle actuation,[6] and cargo transport,[171] respectively). The discovery of protein structure-function relationships requires the determination of the chemical and folding structures of proteins. Protein single crystals are synthetic protein architectures that have positional, orientational, and translational order; and X-ray crystallographic characterization of protein crystals can be used to determine protein structures.[172-176] In fact, over 160,000 structures of proteins have been solved using X-ray crystallography and deposited into the Protein Data Bank (PDB).[177] However, protein crystallization can be challenging for many proteins, because native interactions between proteins are complex. PPIs contain combinations of many weak noncovalent interactions (*e.g.*, hydrogen bonding, electrostatics, hydrophobics, and van der Waals interactions),[178-180] therefore, researchers interested in such structures have little control over crystallization and the type of crystals that form.[181-182]

Efforts to control protein crystallization have included modifications that affect charge,[41-43] hydrophobicity,[47-48] protein structure,[49-52] ligand binding,[35-38] and metal binding characteristics.[30-32] These methods often require proteins to be engineered (*i.e.*, site-directed mutagenesis and recombinant expression) to introduce recognition sites directly into the protein structure, potentially complicating the discovery of structure-function relationships[183-184]. In addition, different protein mutants are required for each modification to the synthetic PPI. Therefore, it is desirable to direct protein crystallization to different structural outcome without changes to the protein itself.

2.2. Challenges and Opportunities for Engineering Protein Crystals Using DNA.

In 2015, we introduced the concept of DNA-modification to control protein crystallization.[157] With isotropically and sometimes anisotropically functionalized structures, pseudo-crystalline materials could be realized, but to date, these techniques, in our hands or others, have not yielded structures suitable for single-crystal X-ray diffraction studies.[157-158, 160, 162] Based upon our previous work,[103, 106, 109, 119, 157-158, 160] we hypothesized that modifying proteins with single strands of DNA could be used to influence crystallization, and when combined with protein–protein interactions, could yield new crystal forms and atomic resolution structures. Herein, we explore this hypothesis with a model protein, mutant green fluorescent protein (mGFP). The effects of design parameters, including DNA sequence, DNA length, protein amino acid attachment position, and DNA base attachment position were systematically explored with respect to consequence on protein packing in the crystals (Figure 2.1). Importantly, for many of the systems studied, X-ray diffraction quality single crystals could be obtained. Elucidation of the resulting structures provided insight into the design parameters that control protein packing within such crystals. Taken together, the data demonstrate that a single DNA modification on the surface of a protein can be used to direct protein packing within a single crystal and, as such, is an important step forward in protein crystal engineering.
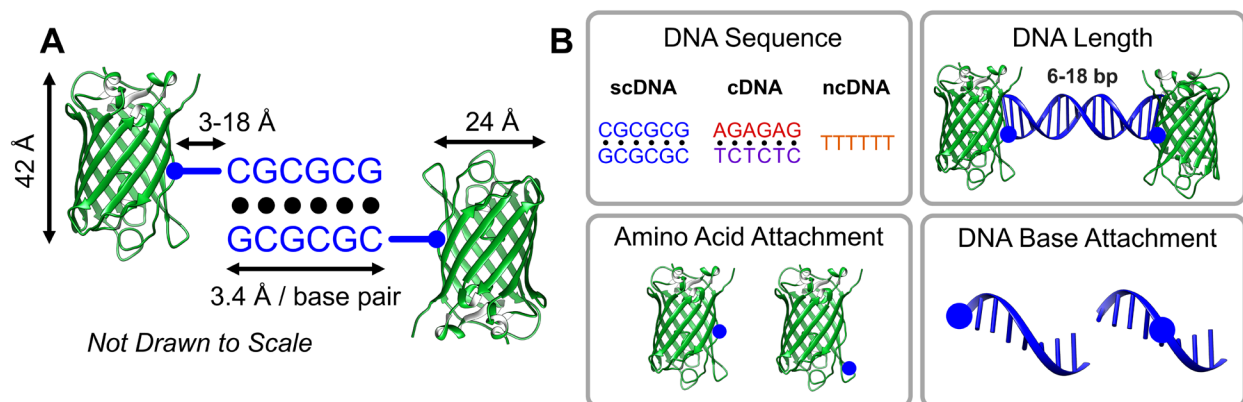
Figure 2.1. Design and parameter scope of mGFP–DNA conjugates studied. (A) Schematic of the DNA interaction between mGFP–DNA conjugates with dimensions for the mGFP, the DNA, and the mGFP–DNA linkage. (B) The design parameters explored include DNA sequence, DNA length, amino acid attachment position, and DNA base attachment position. DNA sequence was varied between self-complementary (scDNA), complementary (cDNA), and non-complementary (ncDNA) (upper left). DNA length was varied from 6–18 bases (upper right). DNA attachment positions were on the side (amino acid residue 148) or edge (amino acid residue 176 or 191) of the mGFP β-barrel (lower left). The sites within the DNA for attachment to the proteins were either internal or external (lower right).

2.2. Design and Synthesis of DNA-Modified Proteins.

2.2.1. Recombinant Expression of GFP Mutants.

To study how designed DNA interactions can influence the growth and packing in protein single crystals, we designed mutants of GFP that could be modified with one DNA strand using cysteine-specific bioconjugation chemistry. A single cysteine residue was positioned at a distinct surface location on different mutants, either on the side (C148 mGFP) or the edge (C176 mGFP or C191 mGFP) of the mGFP β-barrel (Table 2.1). Each mutant was recombinantly expressed in *E. coli* and isolated using standard protein purification techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and matrix-assisted laser desorption-ionization-time-of-flight mass spectrometry (MALDI-TOF MS) were used to characterize the C148 mGFP (Figure 2.2), C176 mGFP (Figure 6.1), and C191 mGFP (Figure 6.2).
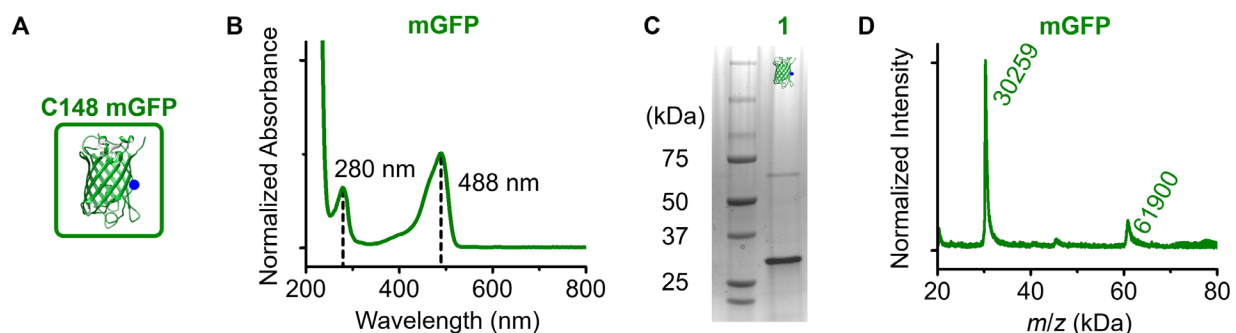
Figure 2.2. Characterization of C148 mGFP. (A) Schematic of C148 mGFP (green). (B) A UV-vis absorption spectrum that is normalized to the absorbance of the mGFP chromophore at 488 nm. The spectrum also contains a characteristic protein absorbance peak at 280 nm. (C) SDS-PAGE analysis shows C148 mGFP (lane 1) primarily in the thiol form (~30 kDa) with a small amount in the disulfide form (~60 kDa). (D) MALDI-TOF MS characterization shows an experimental C148 mGFP (green) mass of ~30.5 kDa.

Table 2.1. Amino acid sequences for GFP mutants.

Mutant	Amino Acid Sequence	Model
C148 mGFP	MRGSHHHHHHGMASMTGGQQMGRDLYENLYDDDDKMV SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGK LTLKFILTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQH DFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN RIELKGIDFKEDGNILGHKLEYNYN**C**HNVYIMADKQKNGI KVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYL STQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK	
C176 mGFP	MRGSHHHHHHGMASMTGGQQMGRDLYENLYDDDDKMV SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGK LTLKFILTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQH DFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI KVNFKIRHNIEDG**C**VQLADHYQQNTPIGDGPVLLPDNHYL STQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK	
C191 mGFP	MRGSHHHHHHGMASMTGGQQMGRDLYENLYDDDDKMV SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGK LTLKFILTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQH DFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI KVNFKIRHNIEDGSVQLADHYQQNTPIG**C**GPVLLPDNHYL STQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK	

The single surface-exposed cysteine residue on each GFP mutant is bolded and highlighted in blue.

2.2.2. DNA Design and Modification of GFP with One DNA Strand.

Next, we investigated the effect of introducing a designed DNA interaction between proteins on crystallization and protein packing within a single crystal. We varied design parameters including DNA sequence, DNA length, amino acid attachment position, and DNA base attachment position (Table 2.2). The wide range of DNA interactions that can be designed highlights the reason to direct protein crystallization using DNA.

Table 2.2. Oligonucleotide sequence design and characterization.

Name	Sequence (5′ → 3′)[a]	$\varepsilon_{260\ nm}$ (M^{-1}cm^{-1})[b]	MW$_{theor.}$ (Da)[b]	MW$_{exp.}$ (Da)
scDNA-1	*H₂N*-CGCGCG	51,400	1,930.2	1,960.3
cDNA-1	*H₂N*-GGCCGG	55,600	2,012.4	2,002.0
	H₂N-CCGGCC	48,600	1,932.3	1,919.3
cDNA-2	*H₂N*-AGAGAG	71,600	2,044.4	2,046.0
	H₂N-CTCTCT	45,800	1,897.3	1,898.3
ncDNA-1	*H₂N*-TTTTTT	49,200	1,942.4	1,929.9
cDNA-3	*H₂N*-AAGGAAGGA	106,200	3,000.1	3,005.9
	H₂N-TCCTTCCTT	69,900	2,794.9	2,797.4
cDNA-4	*H₂N*-AAGGAAGGAAGG	137,900	3,971.7	3,981.5
	H₂N-CCTTCCTTCCTT	91,700	3,677.5	3,677.5
cDNA-5	*H₂N*-AGTTAGGACTTACGCTAC	176,900	5,677.8	5,684.8
	H₂N-GTAGCGTAAGTCCTAACT	177,100	5,677.8	5,683.3
ncDNA-2	*H₂N*-TTTTTTTTT	73,500	2,855.0	2,889.1
scDNA-2	GCGC*T(NH₂)*AGC	80,600	2,508.8	2,510.2

[a]*H₂N*- represents a 5′ amino-modifier C6 phosphoramidite and *T(NH₂)* represents an amino-modifier C2 dT phosphoramidite (Scheme 6.1). [b]Extinction coefficients and theoretical molecular weights were calculated using the OligoAnalyzer Tool from IDT DNA.

Site-specific mono-functionalization of mGFP with DNA was performed according to a previously published procedure.[160] To install a single DNA strand onto mGFP, the amine on 10 equivalents (equiv.) of a DNA strand were reacted with the *N*-hydroxysuccinimide activated ester (NHS) on 300–500 equiv. of succinimidyl 3-(2-pyridyldithio)propionate (SPDP, Scheme 2.1). Excess unreacted SPDP was removed with two consecutive rounds of size-exclusion

chromatography (SEC). Next, the surface-exposed cysteine on mGFP was functionalized with pyridyl disulfide-modified DNA through a thiol-disulfide exchange reaction (Scheme 2.1). The mGFP–DNA conjugates were isolated from unreacted DNA and proteins using nickel affinity chromatography and anion exchange chromatography, respectively.

Scheme 2.1. Synthetic route to conjugate proteins to one DNA strand.
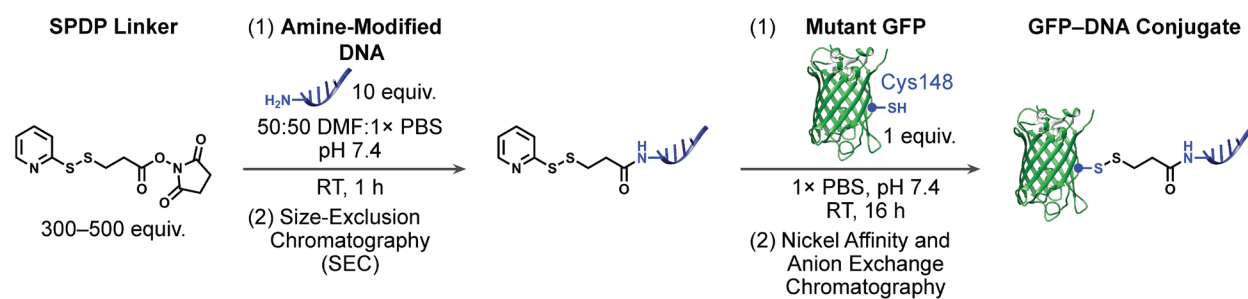
Following this approach, mGFP (Table 2.1) were conjugated to one copy of each designed DNA strand (Table 2.2). The mono-functionalization of mGFP with DNA and purification of mGFP–DNA conjugates were confirmed using UV-vis spectroscopy, SDS-PAGE, and MALDI-TOF MS. The successful synthesis of C148 mGFP–scDNA-1 (Figure 2.3), C148 mGFP–cDNA-1 (Figure 6.3), C148 mGFP–cDNA-2 (Figure 6.4), C148 mGFP–ncDNA-1 (Figure 6.5), C148 mGFP–cDNA-3 (Figure 6.6), C148 mGFP–cDNA-4 (Figure 6.7), C148 mGFP–cDNA-5 (Figure 6.8), C148 mGFP–ncDNA-2 (Figure 6.9), C176 mGFP–scDNA-1 (Figure 6.10), C191 mGFP–scDNA-1 (Figure 6.11), and C148 mGFP–scDNA-2 (Figure 6.12). These data conclusively demonstrate the attachment of one DNA strand per mGFP and the isolation of pure mGFP–DNA conjugates.
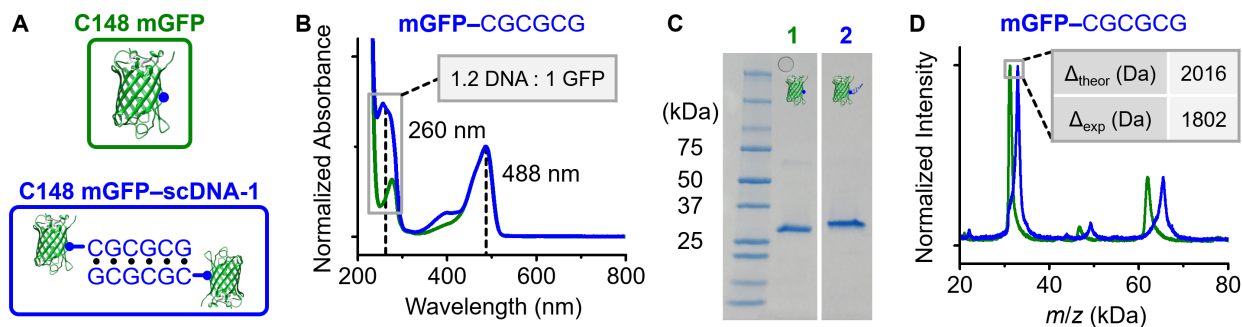
Figure 2.3. Characterization of C148 mGFP–scDNA-1. (A) Schematics of C148 mGFP (green) and C148 mGFP–scDNA-1 (blue). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-1 (blue) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–scDNA-1 compared to C148 mGFP corresponds to 1.2 scDNA-1 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–scDNA-1 (lane 2) show the conjugation of one scDNA-1 to C148 mGFP. Both images are from the same gel with intermediate lanes removed for clarity. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–scDNA-1 (blue) shows an experimental mass increase of 1802 Da that is consistent with the theoretical mass increase of mono-functionalization (2016 Da = 1930 Da (scDNA-1) + 86 Da (linker molecule)).

2.3. Studying the Impact of DNA on GFP Single Crystal Structure.

2.3.1. GFP and GFP–DNA Crystallization and X-ray Crystallography.

The mGFP proteins and mGFP–DNA conjugates were crystallized at a concentration of 5 mg/mL using vapor diffusion techniques and hundreds of crystallization conditions (*i.e.*, varying salt, precipitant, buffer, and temperature) were screened robotically in a high-throughput manner. Sitting drops of 1 µL of sample were mixed with 1 µL of a crystallization condition and placed next to a reservoir that contains of 70 µL of a crystallization condition. Crystallization experiments were incubated undisturbed for 2 weeks at both 4 and 22 °C. Obtained crystals (see Table 6.1 for complete information related to crystallization conditions) were characterized using X-ray crystallography at Argonne National Laboratory's Advanced Photon Source. Crystal structures were then solved for various mGFP proteins and mGFP–DNA conjugates (Table 2.3, see Tables 6.3–6.8 for complete information related crystallographic data and model refinement).

Table 2.3. Summary of mGFP and mGFP–DNA crystal properties.

Sample	GFP Mutant	PDB Code	Space Group	Cell Parameters (Å)	Cell Parameters (°)	Resolution (Å)
mGFP	C148	**6UHJ**	*P2₁2₁2₁*	51.51, 62.90, 69.40	90.00, 90.00, 90.00	1.50
mGFP	C176	**6UHK**	*I222*	88.93, 91.76, 151.71	90.00, 90.00, 90.00	1.90
mGFP–scDNA-1	C148	**6UHL**	*P12₁1*	64.87, 52.29, 86.80	90.00, 94.13, 90.00	1.91
mGFP and cDNA-1	C148	**6UHM**	*P2₁2₁2₁*	58.28, 61.76, 135.32	90.00, 90.00, 90.00	2.10
mGFP–cDNA-1	C148	**6UHN**	*P12₁1*	64.92, 52.18, 86.47	90.00, 94.24, 90.00	1.92
mGFP–cDNA-2	C148	**6UHO**	*P12₁1*	64.71, 52.22, 86.44	90.00, 94.23, 90.00	1.95
mGFP–ncDNA-1	C148	**6UHP**	*P12₁1*	59.05, 51.60, 100.41	90.00, 106.97, 90.00	2.90
mGFP–cDNA-3	C148	**6UHQ**	*C121*	106.63, 50.58, 56.69	90.00, 110.33, 90.00	2.85
mGFP–scDNA-2	C148	**6UHR**	*P2₁2₁2₁*	50.58, 50.89, 209.19	90.00, 90.00, 90.00	3.00

A structure of C191 mGFP is previously reported (PDB code: **4W6I**).[185] C148 mGFP–cDNA-4, C148 mGFP–cDNA-5, C148 mGFP–ncDNA-2, C176 mGFP–scDNA-1, and C191 mGFP–scDNA-1 did not crystallize.

2.3.2. Crystal Structures of GFP Mutants.

Crystal structures of C148 mGFP and C176 mGFP were determined (a structure of C191 mGFP is known, PDB code: **4W6I**)[185] prior to their functionalization with DNA as comparisons to structures obtained when DNA is present. While crystal structures of native GFP are well known,[186] the position of solvent-accessible cysteine residues on mGFP influences protein packing

through the formation of disulfide bonds.[185] The C148 mGFP was crystallized, and a 1.5 Å structure was determined in the space group $P2_12_12_1$ where C148 remains as a thiol (PDB code: **6UHJ**, Figure 2.4.A). The extended packing arrangement of mGFP within **6UHJ** is shown Figure 6.13. The structure is nearly identical to the majority of GFP structures in the PDB,[177] with nearly equivalent unit cell parameters and a root-mean-square deviation (RMSD) of 0.2 Å for all atoms from a previously reported GFP structure (PDB code: **4EUL**).[186] Crystals of C176 mGFP were characterized where C176 form disulfide bonds (product of oxidation) as a different structure in the space group $I222$ at 1.9 Å resolution (PDB code: **6UHK**, Figure 2.5). The extended packing arrangement of mGFP within **6UHK** is shown Figure 6.14.
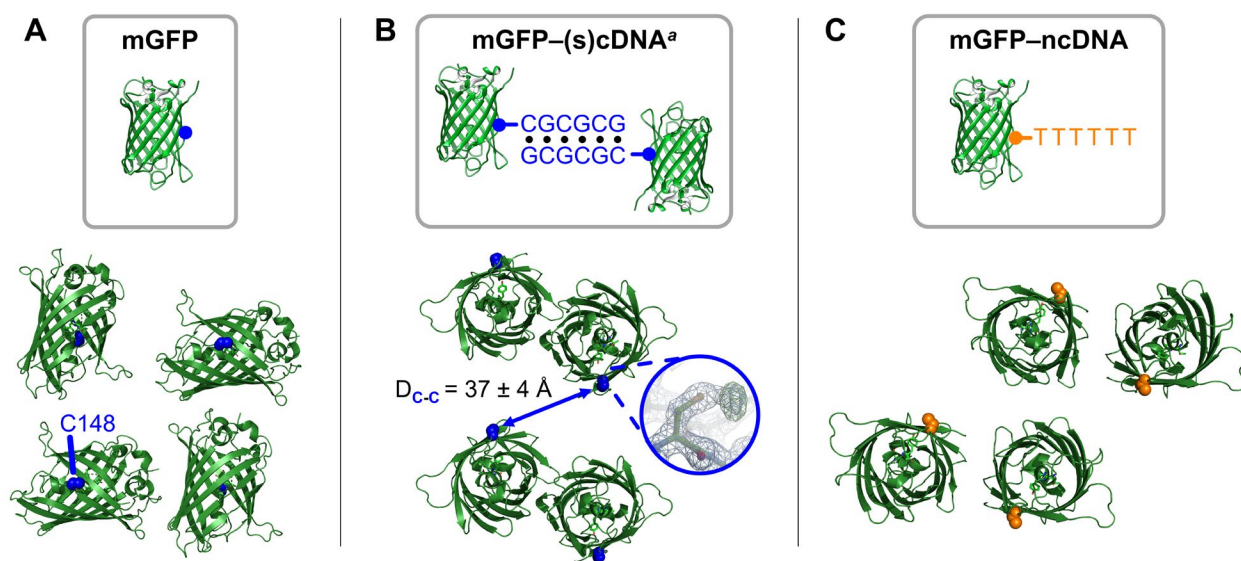
Figure 2.4. Single crystal structures of mGFP and mGFP–DNA. (A) A model of C148 mGFP (top). Four asymmetric units of the C148 mGFP crystal structure (PDB code: **6UHJ**, bottom), which is equivalent to previously reported GFP crystal structures. Proteins pack densely in this structure, and C148 (highlighted in blue) is involved in a protein–protein interaction. (B) A model of C148 mGFP–scDNA-1 (top). Two asymmetric units of the C148 mGFP–scDNA-1 crystal structure (PDB code: **6UHL**, bottom). In this structure, the DNA was not ordered past the disulfide mGFP–DNA attachment (inset). Pairs of C148 (blue) orient toward distinct regions of solvent space with a C148–C148 distance of 37 ± 4 Å that is within the theoretical distance for DNA hybridization (27–64 Å). (C) A model of the C148 mGFP–ncDNA-1 (top). Two asymmetric units from the C148 mGFP–ncDNA-1 crystal structure (PDB code: **6UHP**, bottom), where each C148 (orange) orients toward distinct regions of solvent space with no free path between C148 residues that would permit DNA hybridization.
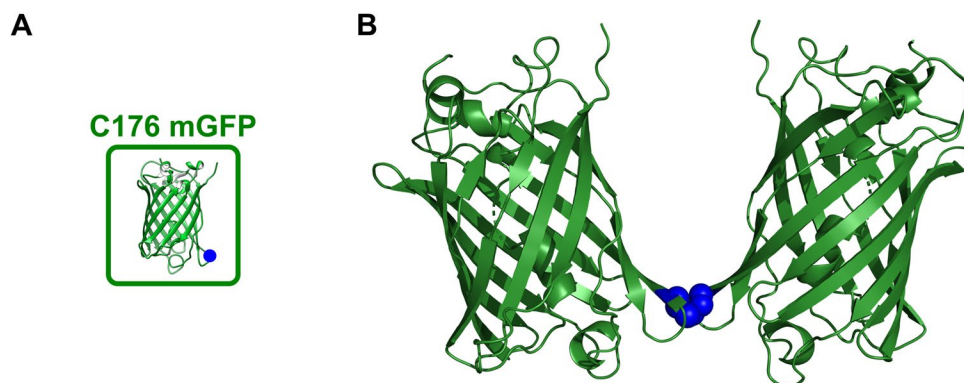
Figure 2.5. Crystal structure of C176 mGFP. (A) Model of C176 mGFP (green). (B) A subset of the C176 mGFP crystal structure (PDB code: **6UHK**) is depicted that highlights the C176–C176 disulfide bond (blue).

2.3.3. Crystal Structures of DNA-Modified GFP.

As a proof-of-concept that DNA interactions can modify the growth and packing of protein single crystals, we first studied the crystallization of mGFP modified with a 6-base self-complementary DNA strand (scDNA-1, Table 2.2) at the C148 position (mGFP–scDNA-1, Figure 2.3). DNA conjugation did not inhibit the protein's ability to crystallize, as the mGFP–scDNA-1 conjugate crystallized into thin plates (~100×200×10 µm). Significantly, a 1.9 Å resolution crystal structure was determined in the space group $P12_11$ (PDB code: **6UHL**, Figure 2.4.B). The extended packing arrangement of mGFP within **6UHL** is shown Figure 6.15. Furthermore, the structure has different unit cell parameters and protein packing with respect to the C148 mGFP crystal structure, indicating that the DNA modification plays a role in how the proteins are organized. In fact, the unit cell parameters and protein packing in the mGFP–scDNA-1 crystal are different relative to all previously reported GFP crystal structures. The crystal structure shows electron density for mGFP and the disulfide mGFP–scDNA-1 attachment, but not DNA. The flexibility of the linker used for protein conjugation (see Scheme 6.2.A for the chemical structures of the mGFP–DNA linkage) likely prevents DNA from ordering in the crystal. However, the

mGFP–scDNA-1 protein packing is consistent with the presence of hybridized DNA. Pairs of C148 residues orient towards distinct regions of solvent space and are separated by 37 ± 4 Å, a distance that corresponds well with the length of the duplexed DNA within the protein single crystals (theoretical distance for 6 bp duplex DNA is 27–64 Å, either in contracted/extended form with respect to the two alkyl linker molecules).[187] As an additional control experiment to confirm that covalent attachment of scDNA-1 to mGFP directs the mGFP–scDNA-1 crystal structure, a mixture of C148 mGFP and scDNA-1 was subjected to identical crystallization conditions as the conjugate. The crystals resulting from the mixture show a structure with a disulfide bond between surface-exposed cysteines (PDB code: **6UHM**, Figure 2.6), where mGFP packing is exclusively directed by inter-protein interactions. The extended packing arrangement of mGFP within **6UHM** is shown Figure 6.16. Taken together, these results show that the covalent attachment of a 6-base self-complementary DNA strand to mGFP leads to a change in protein–protein contacts during crystallization and, ultimately, different protein packing.
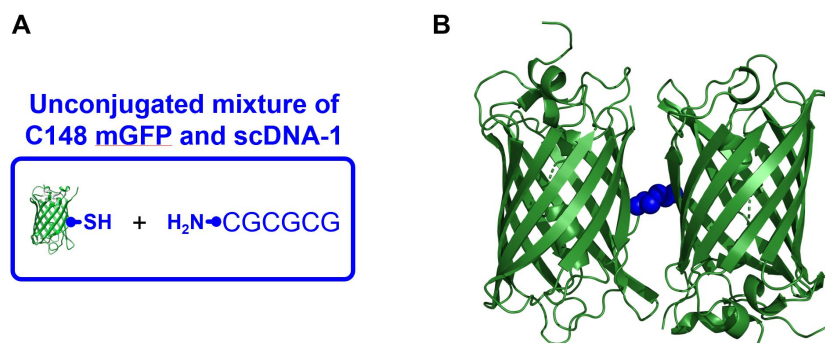
Figure 2.6. Crystal structure of unconjugated mixture of C148 mGFP and scDNA-1. (A) Model of the unconjugated mixture of C148 mGFP and scDNA-1 (blue). (B) A subset of the crystal structure of the unconjugated mixture of C148 mGFP and scDNA-1 (PDB code: **6UHM**) is depicted that highlights the C148–C148 disulfide bond (blue).

2.4. DNA Hybridization Can Direct mGFP Packing in Single Crystals.

2.4.1. Effect of DNA Sequence on Crystal Structure.

To explore whether DNA-directed protein packing using complementary strands is independent of specific sequence, two sets of 6-base complementary DNA were designed (cDNA-1 and cDNA-2, Table 2.2). The C148 mGFP was functionalized with the complementary DNA sequences separately, then corresponding mGFP–DNA conjugates (*i.e.*, mGFP–cDNA-1 and mGFP–cDNA-2, Figures 6.3 and 6.4, respectively) were mixed immediately prior to subjecting the mixture to crystallization experiments. Both mGFP–cDNA-1 and mGFP–cDNA-2 crystallized into thin plates, showing the same crystal morphology as mGFP–scDNA-1 crystals. Furthermore, 1.9 Å crystal structures for mGFP–cDNA-1 and mGFP–cDNA-2 have the same space group $P12_11$ and nearly equivalent unit cell parameters as the mGFP–scDNA-1 structure (PDB code: **6UHN** and **6UHO**, Figure 2.7). The extended packing arrangements of mGFP within **6UHN** and **6UHO** are shown Figures 6.17 and 6.18, respectively. The RMSD between mGFP–scDNA-1, mGFP–cDNA-1, and mGFP–cDNA-2 structures are less than 0.2 Å for all atoms, confirming that the protein packing of these structures is essentially equivalent. Therefore, (self-)complementary mGFP–DNA conjugates with a DNA length of 6-bases crystallize into practically identical single crystal forms, regardless of DNA sequence.
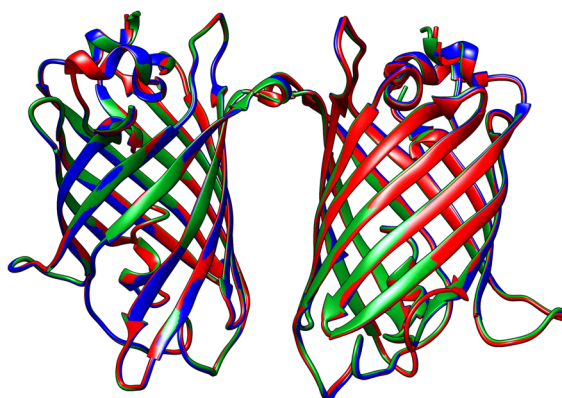
Figure 2.7. Comparison of crystal structures of with different complementary sequences. The asymmetric unit of the C148 mGFP–scDNA-1 (blue, PDB code: **6UHL**), C148 mGFP–cDNA-1 (red, PDB code: **6UHN**), and C148 mGFP–cDNA-2 (green, PDB code: **6UHO**) crystal structures are overlaid. Root-mean-square deviations (RMSDs) of 0.2 Å are calculated between pairs of these structures when comparing location of all atoms, indicating that the structures are nearly equivalent.

2.4.2. Effect of DNA Complementarity on Crystal Structure.

Next, we wanted to confirm the importance of DNA complementarity on the resulting crystal structure. A T6 non-complementary DNA strand (ncDNA-1, Table 2.2) was functionalized to C148 mGFP (mGFP–ncDNA-1, Figure 6.5) and crystallized. The mGFP–ncDNA-1 conjugates formed needle-like crystals, a distinct crystal morphology from mGFP and the three 6-base (self-) complementary mGFP–DNA conjugates. Moreover, a 2.9 Å resolution crystal structure was determined in the space group $P12_11$ for mGFP–ncDNA-1 with unit cell parameters and protein packing that are different from both those of mGFP and (self-)complementary mGFP–DNA conjugates (PDB code: **6UHP**, Figure 2.4.C). The extended packing arrangement of mGFP within **6UHP** is shown Figure 6.19. Clearly, the presence of non-complementary single stranded DNA still influences packing outcomes of mGFP, likely by filling space and altering the crystal contacts that may form between mGFP. However, the protein packing in the mGFP–ncDNA-1 structure is not consistent with DNA duplexing, as each C148 residue orients towards a different region of solvent space with no free path in solvent space between C148 residues that would permit DNA

hybridization (Figure 2.8). This result indicates the importance of DNA complementarity on protein packing outcomes in protein–DNA crystals and illustrates that protein packing within single crystals can be directed using programmable DNA interactions (*i.e.*, mGFP–scDNA-1, mGFP–cDNA-1, and mGFP–cDNA-2).
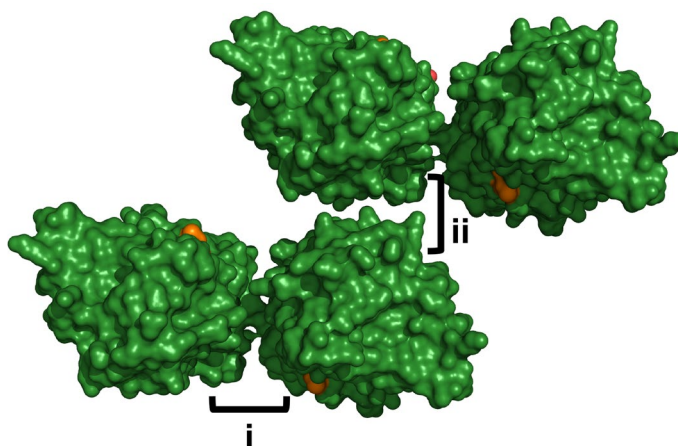
Figure 2.8. Crystal structure of C148 mGFP–ncDNA-1. Two asymmetric units from the C148 mGFP–ncDNA-1 crystal structure (PDB code: **6UHP**) with mGFP proteins depicted in a space-filling manner (green). Each C148 (orange) orients towards distinct regions of solvent space with no free path in solvent space between C148 residues that would permit DNA hybridization. Protein–protein interactions (i and ii) block the path between C148 residues.

2.4.3. Confocal Microscopy to Characterize DNA within Single Crystals.

Since no direct evidence of electron density for DNA was observed in the electron density maps for the mGFP–DNA crystals structures, to confirm the presence of the DNA, crystals were incubated with the DNA-intercalating dye thiazole red homodimer (TOTO-3) and imaged using confocal microscopy. TOTO-3 is a cationic, DNA duplex-sensitive dye that shows a several thousand-fold increase in fluorescence upon DNA intercalation due to decreased rotational freedom, which enforces a planar conformation.[188-189] Before dye addition, crystals of C148 mGFP, C148 mGFP–ncDNA-1, and C148 mGFP–cDNA-1 show mGFP fluorescence (485 nm excitation and 500–550 nm emission filter), but no TOTO-3 fluorescence (640 nm excitation and

663–738 nm emission filter). When TOTO-3 was added to crystals of C148 mGFP, as expected, no TOTO-3 fluorescence was observed because the mGFP crystals do not contain DNA (Figure 2.9.A). In contrast, a strong TOTO-3 fluorescence was observed for mGFP–ncDNA-1 (Figure 2.9.B) and mGFP–cDNA-1 crystals (Figure 2.9.C), providing evidence for the presence of DNA within the crystals of mGFP–ncDNA-1 and mGFP–cDNA-1. Surprisingly, no significant difference in the ratio of mGFP to TOTO-3 fluorescence was observed between mGFP–ncDNA-1 and mGFP–cDNA-1 crystals (Figure 2.9.D). While TOTO-3 is duplex-sensitive in solution, the behavior of TOTO-3 in the protein crystals is less understood. In our case, it is possible that TOTO-3 dye could interact with confined single stranded DNA in the protein crystals in a way that enforces planarity and induces fluorescence. Overall, this confocal microscopy experiment provides evidence for the presence of DNA in mGFP–ncDNA and mGFP–(s)cDNA crystals. Combining observations that (1) DNA is present in these crystals and (2) DNA complementarity determines crystallization outcomes leads to the conclusion that protein packing in single crystals can be modulated by DNA hybridization interactions.

Sample	Green:Far Red Signal Ratio
C148 mGFP	10.3 ± 4.7
C148 mGFP–ncDNA-1	1.8 ± 0.7
C148 mGFP–cDNA-1	1.4 ± 0.4
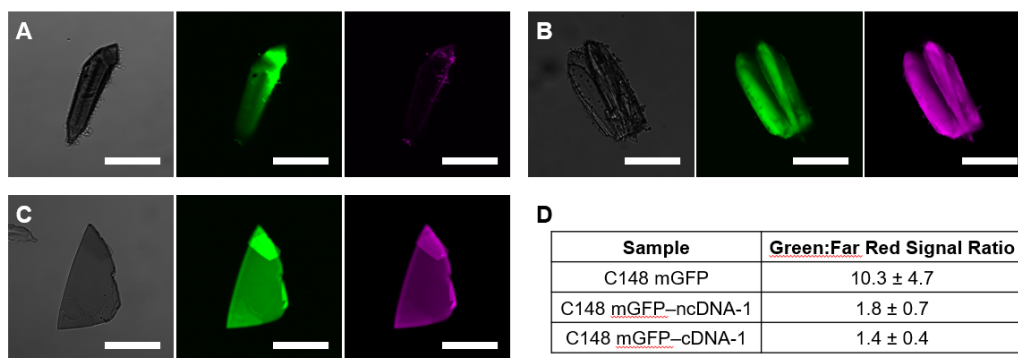
Figure 2.9. Confocal microscopy evidence for DNA in mGFP–DNA crystals. (A–C) Confocal microscopy images of (A) C148 mGFP, (B) C148 mGFP–ncDNA-1, and (C) C148 mGFP–cDNA-1 crystals after soaking the crystals for 30 min in the intercalating dye, TOTO-3. The crystals are imaged in a bright field (left), a green channel (middle, 485 nm excitation and 500–550 nm emission filter), and a far-red channel (right, 640 nm excitation and 663–738 nm emission filter). All scale bars are 50 μm. (D) The ratio of green:far-red fluorescence signals are compared across multiple crystals.

2.5. Studying the Effect of DNA Length on GFP Crystallization.

Since complementary DNA interactions can direct protein crystallization, we sought to determine if DNA length provides another parameter for affecting crystal packing arrangements. To investigate the effect of DNA interaction length on crystallization outcome, DNA interactions at various lengths (*i.e.*, 6, 9, 12, and 18 base pairs (bps)) were designed, and mGFP–DNA conjugates incorporating these interactions were synthesized. While a single crystal form of C148 mGFP was observed for three DNA duplexes of 6 base pairs (bp), an increase in DNA duplex length to 9 bp (mGFP–cDNA-3, Figure 6.6) led to a 2.9 Å structure in the space group *C121* (PDB code: **6UHQ**, Figure 2.10.A). The extended packing arrangement of mGFP within **6UHQ** is shown Figure 6.20. The protein packing within this structure is distinct from other mGFP–DNA structures and, importantly, pairs of C148 residues again orient towards distinct regions of solvent space, separated by 41 ± 6 Å, a distance that agrees with the length of the duplex DNA (theoretical distance for 9 bp duplex is 37–75 Å, either in the contracted/extended form with respect to the two alkyl linker molecules). However, when longer DNA ligands (12 bp: mGFP–cDNA-4 and 18 bp: mGFP–cDNA-5, Figures 6.7 and 6.8, respectively) were investigated, no crystallization was observed. This suggests that above an upper threshold for DNA duplex length, DNA prevents the formation of mGFP single crystals. Similarly, increasing the length of non-complementary DNA from 6 to 9 bases (mGFP–ncDNA-2, Figure 6.9) precluded crystallization. Taken together, mGFP–DNA crystallization and structural outcomes depend strongly on the length of designed DNA.
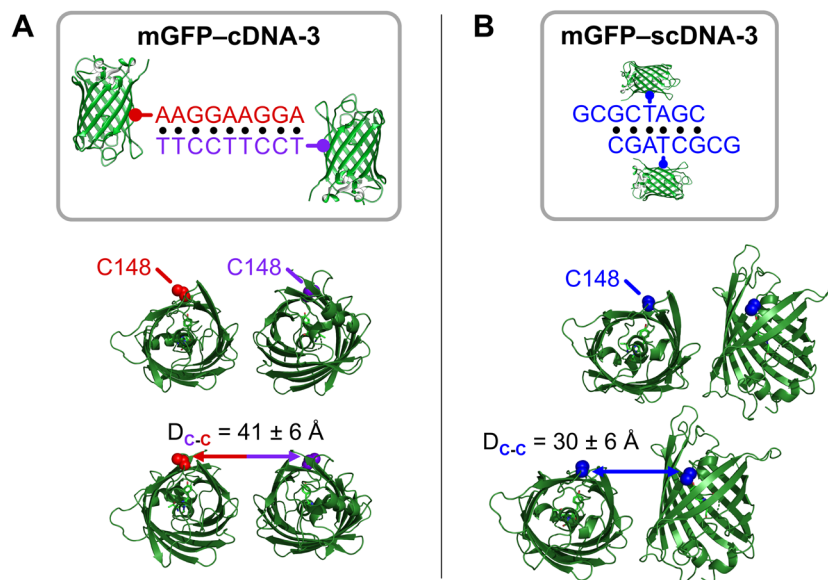
Figure 2.10. Crystal structures of C148 mGFP–cDNA-3 and mGFP–scDNA-3. A) A model of C148 mGFP–cDNA-3 (top). Four asymmetric units of the C148 mGFP–cDNA-3 crystal structure (PDB code: **6UHQ**, bottom). Pairs of C148 (red and purple) orient toward distinct regions of solvent space with a C148–C148 distance of 41 ± 6 Å that is within the theoretical distance for DNA hybridization (37–75 Å). (B) A model of C148 mGFP–scDNA-2 (top). Two asymmetric units of the C148 mGFP–scDNA-2 crystal structure (PDB code: **6UHR**, bottom). Pairs of C148 (blue) orient toward distinct regions of solvent space with a C148–C148 distance of 30 ± 6 Å that is within the theoretical distance for DNA hybridization (8–45 Å).

2.6. Investigating the Importance of GFP–DNA Linkage Chemistry.

2.6.1. Effect of Protein Attachment Site on Crystallization.

In addition to exploring how DNA design can influence crystal structures, protein–DNA attachment position represents another powerful design parameter, where changing attachment location can guide protein–protein interactions and therefore protein packing. The amino acid attachment position was varied by changing the location of the cysteine greater than 15 Å from the middle of the side of the mGFP β-barrel (C148 mGFP) to the edge of the mGFP β-barrel (C176 mGFP and C191 mGFP). The C176 mGFP and C191 mGFP were functionalized with scDNA-1 (C176 mGFP–scDNA-1 and C191 mGFP–scDNA-1, Figures 6.10 and 6.11, respectively), the same DNA which directed the crystallization and structure of C148 mGFP–scDNA-1. In contrast,

C176 mGFP–scDNA-1 and C191 mGFP–scDNA-1 conjugates did not crystallize, perhaps due to the high flexibility of loops at the edge of the mGFP β-barrel. Moreover, residues at position 176 and 191 are located at interfaces between GFP in other obtained crystal structures, potentially indicating their involvement in crystal packing contacts. These results exhibit the importance of amino acid attachment position on crystallization outcomes.

2.6.2. Effect of DNA Attachment Site on Crystallization.

Next, DNA base attachment position was changed from an external to an internal DNA base, which allows shorter inter-protein distances. Additionally, DNA strands with an internal base attachment position may be designed with short sticky end overhangs, which can lead to DNA ordering in single crystals.[125, 190] The C148 mGFP was functionalized with a 6 bp self-complementary DNA strand with a 2 base sticky end (C148 mGFP–scDNA-2, Figure 6.12). See Scheme 6.2.B for the linkage structure for this conjugate. The C148 mGFP–scDNA-2 conjugate was crystallized into a crystal form in the space group $P2_12_12_1$ (PDB code: **6UHR**, Figure 2.10.B). The extended packing arrangement of mGFP within **6UHR** is shown Figure 6.21. Similar to other mGFP–DNA crystal structures, DNA does not order, obscuring the effect of the sticky ends on crystal formation. That said, pairs of cysteines orient towards distinct regions of solvent space at a distance (30 ± 6 Å) that agrees with the length of the duplex DNA (theoretical distance for 8 bp duplex with internal attachment position is 8–45 Å), further confirming that DNA interactions can be extensively designed to influence the crystallization and packing of proteins. This structure suggests an additional layer of control provided by the DNA ligand and will be the subject of future investigations involving linker flexibility and sticky end design.

2.7. Analysis of Protein–Protein Interfaces in Crystal Structures.

Since different crystal structures were observed depending on the presence and design of DNA, we sought to understand the discrete changes in protein–protein interactions across observed single crystals. Towards this end, interfaces between mGFP in all obtained crystal structures were analyzed using PDBePISA[191] to study the effect of introduced DNA on protein packing. While some new interfaces arose, crystals of C148 mGFP–DNA show a protein–protein interface with 82–100 % amino acid residue similarity to the interface with the largest area observed in C148 mGFP. Importantly, this shows that despite dramatic changes in the overall protein packing, the largest protein–protein interface is conserved across all C148 mGFP and C148 mGFP–DNA crystals (Figure 2.11).
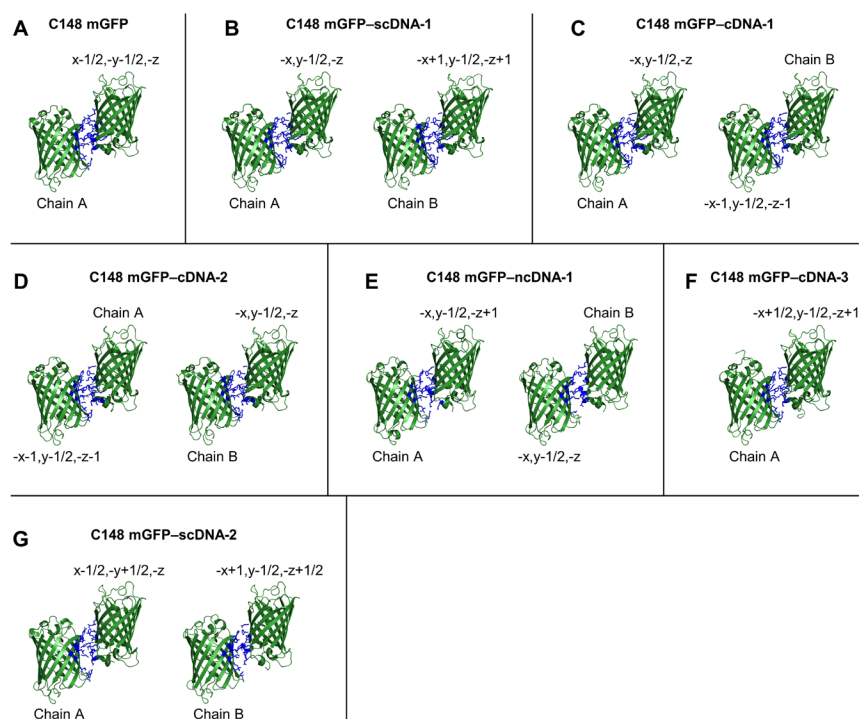
Figure 2.11. An interface that is conserved across all mGFP and mGFP–DNA crystals. A conserved interface (blue) between a protein in the asymmetric unit and a symmetry mate for (A) C148 mGFP, (B) C148 mGFP–scDNA-1, (C) C148 mGFP–cDNA-1, (D) C148 mGFP–cDNA-2, (E) mGFP–ncDNA-1, (F) C148 mGFP–cDNA-3, and (G) C148 mGFP–scDNA-2.

The second largest interface in crystals of C148 mGFP is also conserved for crystals of C148 mGFP–scDNA-1, C148 mGFP–cDNA-1, and C148 mGFP–cDNA-2 (89, 100, and 73 % similar, respectively) and partially conserved for crystals of C148 mGFP–ncDNA-1 (50 % similar, Figures 2.12). Interestingly, this interface is not observed with increases in DNA length or changes in DNA base attachment position, indicating that the nature of protein packing is dictated by an interplay between protein–protein and DNA hybridization interactions. This suggests we could use existing protein crystal structures to inform the placement of DNA ligands to engineer protein packing within single crystals.
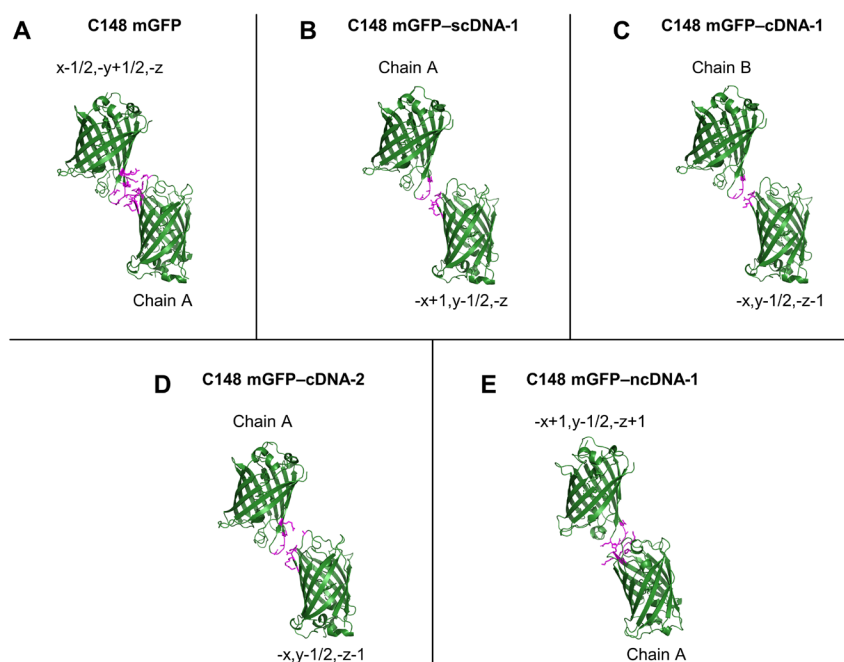
Figure 2.12. An interface that is conserved across some mGFP and mGFP–DNA crystals. A conserved interface (magenta) between a protein in the asymmetric unit and a symmetry mate for (A) C148 mGFP, (B) C148 mGFP–scDNA-1, (C) C148 mGFP–cDNA-1, (D) C148 mGFP–cDNA-2, and (E) C148 mGFP–ncDNA-1.

2.8. Conclusion.

Growth of protein single crystals involves complex protein–protein interactions which are challenging to design and predict. This work demonstrates how replacing such interactions with

highly programmable DNA interactions could enable structural control over protein packing within single crystals. We report the first protein single crystal structure where DNA hybridization interactions between the surfaces of proteins direct the packing of proteins within the crystal. An analysis of protein–protein interfaces within mGFP and mGFP–DNA crystals reveals that interfaces may be conserved or disrupted, regardless of overall crystal structure.

Furthermore, we demonstrate that DNA complementarity, DNA length, and protein–DNA attachment position all influence crystallization and protein packing structural outcomes. Although the resulting crystal structure is shown to be independent of DNA sequence (while maintaining complementarity), crystallization only occurred for mGFP–DNA conjugates when DNA duplexes were less than or equal to 9 bp. Interestingly, changing the DNA length or the attachment of DNA to the protein through an internal base modification afforded more crystal structures that further demonstrate the versatility of this approach and the large design space to be explored. Importantly, minimizing the flexibility of the DNA linkage between protein surfaces (*i.e.*, decreasing DNA length, attaching DNA to a β-strand instead of a loop, or decreasing length of the protein–DNA linker) is important to obtain crystals suitable for X-ray crystallographic structure determination.

Furthermore, we have recently reported that DNA can eliminate, augment, and replace native protein interfaces in single crystals of concanavalin A (ConA).[161] In ConA–DNA crystals, crystallographic analysis led to structure determination of proteins, DNA, and the protein–DNA linker. This unambiguously indicates that DNA hybridization can control the packing of ConA into single crystals. In addition, fine changes to the DNA design (*i.e.*, 0 bases to 4 bases to 6 bases) lead to predictable expansion of the crystal unit cell along one crystallographic axis. Together, these works are essential steps toward designing and engineering protein packing within single

crystals and could lead to future applications in protein structure determination and functional protein crystal materials.

2.9. Materials and Methods.

See Appendix A (Section 6.1).

CHAPTER THREE

Modular Nucleic Acid Scaffolds for Synthesizing Monodisperse and Sequence-Encoded Antibody Oligomers.

3.1. Natural Oligomers of Proteins with Encoded Sequences.

In biology, many proteins assemble into defined oligomeric architectures that contain exact numbers and oligomeric sequences of multiple different proteins.[9-11] Herein, oligomeric sequence of proteins and oligomeric protein sequence are defined as the order of proteins within an oligomeric architecture. The oligomeric architectures of proteins can dictate the biological (*e.g.*, human *IgM* antibodies contain five protein subunits),[15] catalytic (*e.g.*, eukaryotic ribonucleic acid (RNA) polymerase II contains twelve protein subunits),[13] photophysical (*e.g.*, cyanobacteria photosystem I contains twelve protein subunits),[5] and membrane transport (*e.g.*, *Streptomyces lividans* potassium channel contains four protein subunits)[192] properties of proteins.

3.2. Opportunities and Challenges in Synthetic Protein Oligomers

To mimic and potentially surpass the natural properties of protein oligomers, the modular synthesis of different oligomers is needed. A versatile synthetic protein oligomerization method would: (1) provide access to a large number of proteins per oligomer, (2) provide access to any oligomeric sequence of the same or different proteins, (3) be generalizable with regard to proteins, and (4) not require mutations of the amino acid sequence of proteins and recombinant protein expression. A method that meets all four criteria would enable the study of how the identity, number, stoichiometry, oligomeric sequence, and architecture of proteins affects the emergent properties of oligomers. While strategies have been developed to prepare synthetic protein oligomers[23-26] and study how oligomerization affects protein properties,[193-196] there is no current method that meets all four listed criteria. In this work, we explored the design and synthesis of a single modular nucleic acid scaffold that can be used to organize proteins into a near limitless array of monodisperse and sequence-encoded protein oligomers (Figure 3.1).

Figure 3.1. Modular DNA scaffold to prepare sequence-encoded oligomers. In this work, we design a set of six DNA strands that can be used as a modular DNA scaffold to organize proteins into an expansive array of monodisperse and sequence-encoded oligomers. This generalizable method will enable the synthesis of different oligomeric sequences of proteins.

3.2.1. Molecular Biology Approaches to Oligomerize Proteins.

Protein oligomers are frequently prepared using techniques from molecular biology, including genetic engineering and recombinant expression of mutated proteins (Figure 3.2). Generally, fusion proteins are designed via genetic engineering and recombinantly expressed to achieve the desired oligomerization structure. Three common methods for oligomerization using fusion proteins include: direct expression of protein oligomers as a single polypeptide[69-71]; fusion of proteins to a subunit of a multimeric protein (*e.g.*, streptavidin) that assembles into supramolecular constructs (Figure 3.2.A)[72-75]; or fusion of proteins to a subunit that can selectively form covalent bonds with a complementary group (*e.g.*, SpyTag/SpyCatcher,[76]

SnoopTag/Snoopcatcher,[77] SNAPtag,[78] HALOtag,[79] or cutinase,[80] Figure 3.2.B).[77, 81-84] Furthermore, recent advances in the design of protein–protein interfaces enable sophisticated control over synthetic protein architectures (Figure 3.2.C).[53-58] Each of these methods requires mutations of the amino acid sequence of proteins and recombinant protein expression. However, many proteins are challenging to prepare via recombinant expression (*e.g.*, proteins with post-translational modifications,[85] proteins with disulfide bonds,[86] toxic proteins,[87] or proteins that aggregate[88]), potentially limiting the scope of proteins that can be oligomerized through these methods.

A **Protein fusion to assembling units**
(*e.g.*, streptavidin)

B **Protein fusion to conjugation units**
(*e.g.*, SNAPtag, HALOtag)

C **Designed protein–protein interfaces**

Protein **A**   Protein **A**   Protein **A**

✓ Large number of proteins per oligomer
✗ Any possible oligomeric protein sequence
✓ Generalizable with regard to proteins
✗ Does not require protein mutations

✓ Large number of proteins per oligomer
✓ Any possible oligomeric protein sequence
✓ Generalizable with regard to proteins
✗ Does not require protein mutations

✓ Large number of proteins per oligomer
✗ Any possible oligomeric protein sequence
✓ Generalizable with regard to proteins
✗ Does not require protein mutations

Figure 3.2. Molecular biology approaches to oligomerize proteins. Protein oligomers can be synthesized using techniques from molecular biology, including (A) protein fusion to assembling units, (B) protein fusion to conjugation units, and (C) design of protein-protein interfaces.

3.2.2. Chemical Approaches to Oligomerize Proteins.

Chemical approaches to assemble proteins are another powerful method to control oligomerization (Figure 3.3). The amino acid sequence of proteins can be mutated to incorporate (un)natural amino acids at defined positions for interactions such as electrostatic,[42] supramolecular host–guest binding,[35, 38, 44-45] metal coordination,[27-29] or covalent linking.[59-60] However, without extensive chemical design, modification of protein amino acid sequences and/or recombinant protein expression, it is challenging to access monodisperse and sequence-encoded oligomers that are larger than dimers or trimers. Chemical approaches to directly oligomerize proteins post-

expression without modifying the amino acid sequence of proteins or using an assembly template (*vide infra*) were expanded with the introduction of bioorthogonal "click" reactions (Figure 3.3.A).[66-68] While these reactions have been used to oligomerize therapeutically relevant proteins (*e.g.*, antibodies), it is challenging to achieve oligomers larger than dimers or trimers.[197]
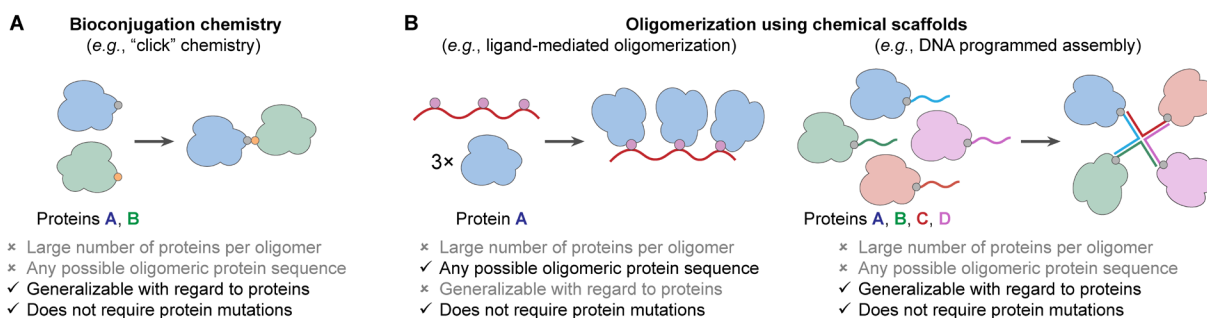
Figure 3.3. Chemistry approaches to oligomerize proteins. Protein oligomers can be synthesized using techniques from chemistry, such as (A) bioconjugation chemistry and (B) scaffold-directed oligomerization of proteins.

Proteins can be oligomerized using attachment to chemical scaffolds (*e.g.*, polymers[61-65] or DNA[105-106, 168-170], Figure 3.3.B). Templated assembly of proteins using DNA is one of the most promising and versatile approaches to organize proteins into oligomers larger than dimers or trimers. The utility of this approach is a result of the programmability of nucleic acids where specific, defined assemblies can be accessed solely based on DNA sequence design.[105-106, 168-170] For example, proteins have been covalently[111, 113, 198] or noncovalently[111-112] modified with oligonucleotides and the resulting constructs have been organized via DNA–DNA interactions into a multitude of protein oligomers with one-,[119-130] two-,[131-153] and three-dimensional[18, 154-167] architectures. However, in each of these systems, DNA design must be changed to synthesize defined protein oligomers that contain different numbers or oligomeric sequences of proteins. For example, a protein tetramer can be readily synthesized using a DNA tetrahedron scaffold[131] or a

four-arm Holliday DNA junction scaffold.[132-134] However, to synthesize a protein pentamer, every DNA sequence must be re-designed to account for the one additional protein. Modular multi-protein constructs can be realized on large DNA origami structures,[135-136, 154] but most of the composition (*e.g.*, >80%) of these constructs is DNA instead of protein. Molecular constructs that minimize the required amount of DNA such that most of the chemical properties are dictated by the identity and organization of proteins are inaccessible using these techniques.[169] Together, these limitations significantly hamper any studies where access to libraries of different protein oligomers with discrete stoichiometries and oligomeric sequences of proteins could provide insight into how protein–protein interactions and cooperativity can be exploited for enhanced properties of oligomers.

3.3. Designing a Modular DNA Scaffold to Organize Proteins into Predefined Sequences.

We hypothesized that a single set of designed DNA strands could be used as a modular scaffold to organize proteins into oligomers with exact stoichiometries and oligomeric sequences (Figure 3.1). This novel DNA design would enable different proteins to be precisely organized into an expansive array of monodisperse, sequence-encoded oligomers. Herein, we tested our hypothesis by designing a modular six-strand DNA scaffold (Table 3.1) and using it to oligomerize commercially available and therapeutically relevant proteins (*i.e.*, antibodies). The designed DNA scaffold consists of three distinct DNA strands that can be conjugated to proteins and three distinct DNA strands that template the assembly of DNA-modified proteins into oligomers via DNA–DNA interactions (Scheme 3.1). Importantly, each of the six DNA strands contains two distinct binding domains and the sites for attachment to proteins can be located anywhere on the DNA strands. Using the designed DNA scaffold, monodisperse, sequence-encoded monomer, dimer, and trimer

building blocks are synthesized. Next, these building blocks are used to access a larger oligomer

(*i.e.*, pentamer) that contains a defined number and oligomeric sequence of proteins.

Scheme 3.1. Design of the modular DNA scaffold.

A DNA strands for protein conjugation: two 20-base nucleic acid sequences (*i.e.*, S2 and S3, S4 and S5, and S6 and S1) joined by a phosphoramidite containing dibenzocyclooctyne (DBCO) and a dye (*i.e.*, Cy3, Cy5, and FITC)
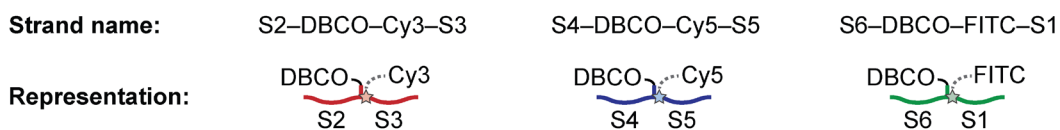
B Template DNA strands: two 20-base nucleic acid sequences (*i.e.*, S1′ and S2′, S3′ and S4′, and S5′ and S6′) complementary to different strands used for protein conjugation

C
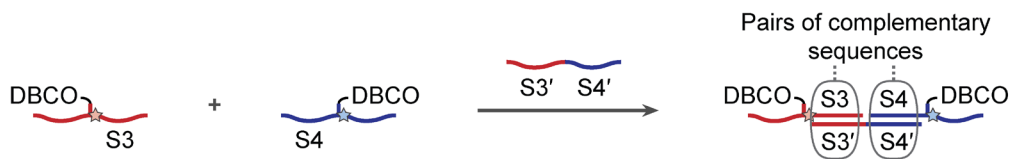
Labels represent distinct sequences within DNA strands, where prime symbols indicate sequence complementarity (*e.g.*, S3 is a complementary sequence to S3′ and S4 is a complementary sequence to S4′)

Table 3.1. DNA Sequences and Characterization.

Strand Name[a]	DNA Sequence $(5' \rightarrow 3')$[b]	$\varepsilon_{260 \text{ nm}}$ $(M^{-1}cm^{-1})$[c]	$MW_{theor.}$ (Da)[c]	$MW_{exp.}$ (Da)
S2–DBCO–Cy3–S3	AACATCTTGTGCTCAATATC–T(DBCO)–Cy3–TAGTCGTCTACGTAACAGTC	406,100	13,485	13,479
S4–DBCO–Cy5–S5	TTAGGCTGGATCTCGCGTTC–T(DBCO)–Cy5–TGCACAGACCCATGTACTCG	394,400	13,545	13,563
S6–DBCO–FITC–S1	TAACTCGTGAACGTATGCTC–T(DBCO)–FITC–AATCATCAGTACTCACCTAG	407,900	13,571	13,545
S1'–S2'	GTAACGAT[d]–GATATTGAGCACAAGATGTT–CTAGGTGAGTACTGATGATT	488,000	14,925	14,900
S3'–S4'	CATTCAGA[d]–GAACGCGAGATCCAGCCTAA–GACTGTTACGTAGACGACTA	479,800	14,779	14,764
S5'–S6'	GTCATGTT[d]–GAGCATACGTTCACGAGTTA–CGAGTACATGGGTCTGTGCA	469,700	14,854	14,832

[a]The names of DNA strands for protein conjugation are listed in the 5′ to 3′ direction, while the names of template DNA strands are listed in the 3′ to 5′ direction. [b]DBCO-dT-CE, cyanine 3, cyanine 5, and fluoresceine phosphoramidites are represented by T(DBCO), Cy3, Cy5, and FITC, respectively (see Scheme 6.3). [c]Extinction coefficients and theoretical molecular weights were calculated using the OligoAnalyzer Tool from IDT DNA. [d]These DNA sequences are toeholds that were included in synthesized template strands but were not utilized in this work.

Importantly, the foundational examples shown herein are a fraction of the possible oligomeric sequences that are accessible using the modular six-strand DNA scaffold (Table 3.2, Figure 3.4). For example, if five different proteins are used, there are, in principle, 3,125 different accessible pentameric sequences. Overall, this generalizable synthetic route will enable future investigations into how the identity, stoichiometry, oligomeric sequence, and architecture of proteins in oligomers affect the properties of these constructs.

Example: 27 trimer sequences of 3 different proteins
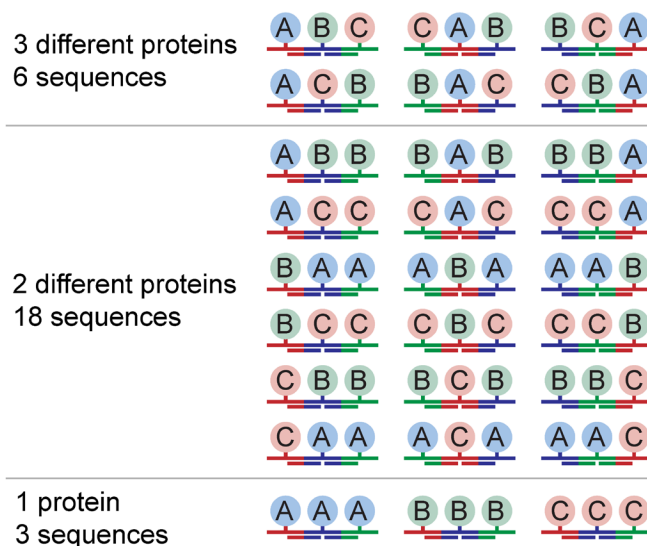
Figure 3.4. Accessible oligomeric trimer sequences of three proteins.

Table 3.2. Number of accessible oligomeric protein sequences.

		Number of different oligomeric proteins sequences			
		Up to 2 different proteins	Up to 3 different proteins	Up to 4 different proteins	Up to 5 different proteins
Oligomer size	Dimers	4	9	16	25
	Trimers	8	27	64	125
	Tetramers	16	81	256	625
	Pentamers	32	243	1024	3125

3.4. Synthesis of Oligomeric Protein–DNA Building Blocks.

3.4.1. Generalizable Approach to Modify Proteins with One DNA Strand.

Three commercially available IgG antibodies commonly used as checkpoint inhibitors (*i.e.*, **anti-mouse-PD-1** (**A**), **anti-mouse-TIGIT** (**B**), and **anti-mouse-CTLA-4** (**C**)) were chosen for

the sequence-encoded oligomerization of proteins using a modular DNA scaffold. To install a single DNA strand onto either **A**, **B**, or **C**, each antibody was reacted with 2 equiv. of an oligo(ethylene glycol) molecule containing an *N*-hydroxysuccinimide activated ester and an azide (NHS–PEG$_{12}$–N$_3$) for 45 min (Figure 3.5.A, left). This chemistry targets the primary amines (*e.g.*, ε-amines on lysines and α-amines on N-termini)[199] on both the Fc and Fab regions of the antibody and was chosen because it is generalizable with regard to proteins. Although the exact location of conjugation cannot be controlled, we expected that the number of azide modifications per antibody would be controlled by low numbers of equivalents of NHS–PEG$_{12}$–N$_3$ added. We hypothesized that the low number of primary amines modified would not inhibit the target binding characteristics of antibodies. After purification by SEC, the azide on the surface of each antibody underwent a strain-promoted azide-alkyne cycloaddition reaction with 5 equiv. of DNA strands containing dibenzocyclooctyne (DBCO) and a fluorophore (*i.e.*, Cyanine 3 (Cy3), Cyanine 5 (Cy5), or fluorescein (FITC)) and two distinct 20 base nucleic acid sequences (Table 3.1, Scheme 3.1.A, *i.e.*, S2–DBCO–Cy3–S3, S4–DBCO–Cy5–S5, or S6–DBCO–FITC–S1). After 16 h, roughly 25–30% of antibodies were modified with one DNA strand (Figure 3.6). Next, unreacted DNA was removed from the reaction mixture using SEC. Anion exchange chromatography was used to isolate antibodies that were functionalized with a single DNA strand from unreacted antibodies and antibodies that were functionalized with multiple DNA strands (Figure 3.2.A, right). Three different protein–DNA conjugates (*i.e.*, S2–**A**–Cy3–S3, S4–**B**–Cy5–S5, and S6–**C**–FITC–S1) were prepared and confirmed to contain a single DNA functionalization via SDS-PAGE (Figure 3.5.B) and SEC (Figure 3.5.C).
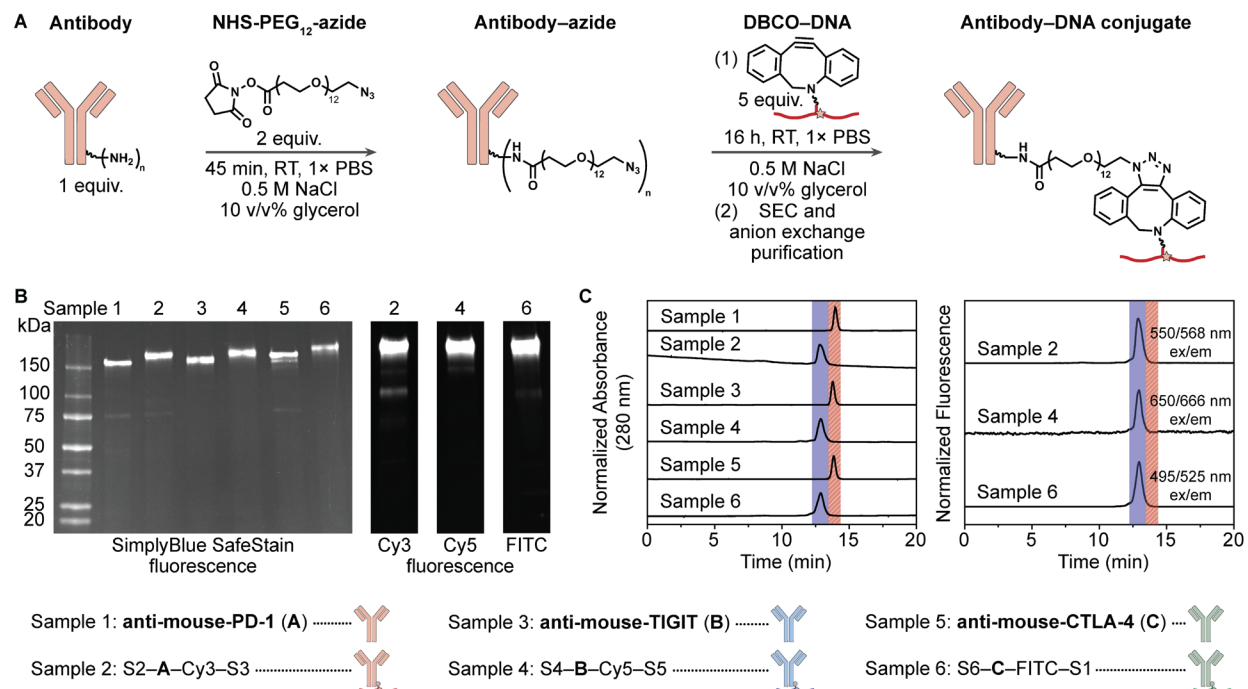
Figure 3.5. Conjugating antibodies to one DNA strand. (A) Primary amines on the surface of antibodies were functionalized with azides using an NHS–PEG$_{12}$–N$_3$ linker. Next, dibenzocyclooctyne (DBCO)-modified DNA was conjugated to azide-modified antibodies via a strain-promoted azide-alkyne cycloaddition (SPAAC). (B) Mouse antibodies (*i.e.*, sample 1: **anti-mouse-PD-1** (**A**), sample 3: **anti-mouse-TIGIT** (**B**), and lane 5: **anti-mouse-CTLA-4** (**C**)) and antibody–DNA conjugates (*i.e.*, lane 2: S2–**A**–Cy3–S3, sample 4: S4–**B**–Cy5–S5, and sample 6: S6–**C**–FITC–S1) were characterized using SDS-PAGE. A single gel was imaged for SimplyBlue SafeStain, Cy3, Cy5, and FITC fluorescence. The same samples were characterized using analytical SEC, where the elution times of samples were tracked using the absorbance at 280 nm (left) and fluorescence of DNA-conjugated fluorophores (right). In both graphs, the elution times from 12.25–13.50 min are highlighted in blue (shaded area) and the elution times from 13.50–14.30 min are highlighted in red (shaded area with diagonal lines).

3.5. Synthesis and Characterization of Sequence-Encoded Protein Oligomers.

3.5.1. Monodisperse and Sequence-Encoded Protein Dimers.

Protein oligomers were synthesized by mixing the purified protein–DNA conjugates (Figure 3.6.A and E: lanes 1–3) with template DNA strands (Table 3.1, Scheme 3.1.B, *i.e.*, S1'–S2', S3'–S4', or S5'–S6'). The template strands were designed as complements to two 20 base nucleic acid sequences on different antibody–DNA conjugates (Table 3.3, Scheme 3.1.C). For example, the S5' DNA sequence on the template strand S5'–S6' is complementary to the S5 DNA sequence on S4–**B**–Cy5–S5 and the S6' DNA sequence is complementary to the S6 DNA sequence

on S6–**C**–FITC–S1. Equal amounts of the **B**–DNA conjugate, **C**–DNA conjugate, S5'–S6' template strand, and S1'–S2' template strand were mixed to synthesize a protein dimer with the oligomeric sequence S4–**B**–**C**–S2' (Figure 3.6.B, Scheme 3.2.A) at an assembly yield of 68% (Figure 6.24.A and D). Oligomers that contain greater than two antibodies were not observed in the assembly mixture because there are no DNA sequences that are complementary to either the S4 or S2' DNA sequences. Protein dimers were isolated from unreacted monomers and template strands in the assembly mixture using SEC purification and characterized with agarose gel electrophoresis (Figure 3.6.E: lane 4). Importantly, the agarose gel showed a single band for dimers with only the expected Cy5 and FITC fluorescence and lower electrophoretic mobility than either antibody–DNA conjugate alone. Therefore, monodisperse and sequence-encoded protein dimers with the oligomeric sequence S4–**B**–**C**–S2' were successfully synthesized.

Table 3.3. Calculated melting temperatures of designed DNA–DNA Interactions.

DNA–DNA Interaction	T_m (°C)[a]
S1 and S1'	60
S2 and S2'	60
S3 and S3'	62
S4 and S4'	68
S5 and S5'	68
S6 and S6'	64

[a]Melting temperatures were calculated using the OligoAnalyzer Tool from IDT DNA with DNA concentrations of 5 μM and salt concentrations from 1× PBS.
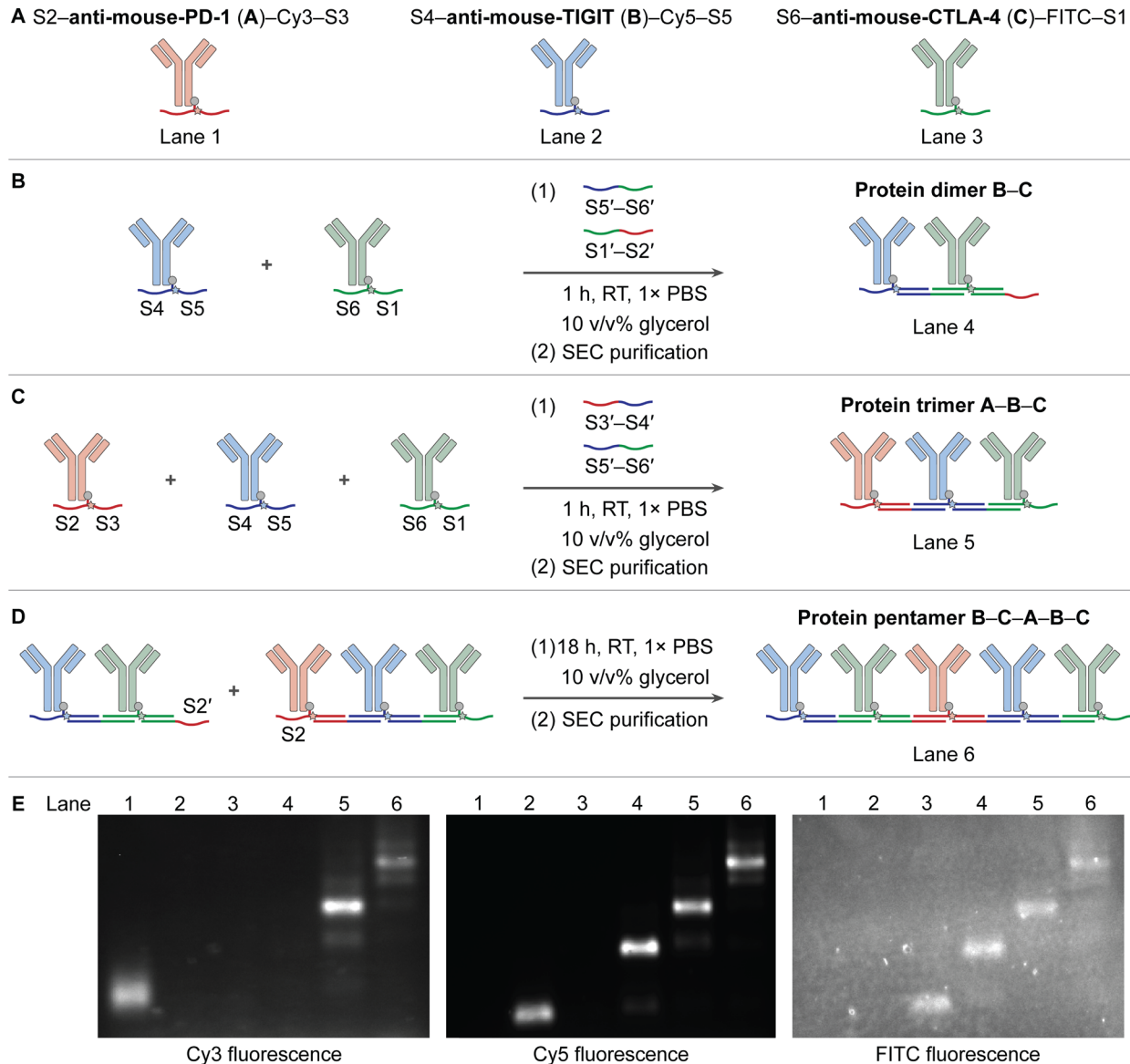
Figure 3.6. Sequence-encoded antibody oligomerization using DNA–DNA interactions. (A) Antibody–DNA conjugates and template DNA strands were assembled using DNA–DNA interactions into sequence-encoded protein (B) dimers and (C) trimers. Dimers and trimers were subsequently assembled using DNA–DNA interactions into sequence-encoded protein (D) pentamers. (E) Agarose gel characterization of antibody–DNA conjugates (*i.e.*, lane 1: S2–**anti-mouse-PD-1** (**A**)–Cy3–S3, lane 2: S4–**anti-mouse-TIGIT** (**B**)–Cy5–S5, and lane 3: S6–**anti-mouse-CTLA-4** (**C**)–FITC–S1) along with sequence-encoded antibody dimers (*i.e.*, lane 4: S4–**B**–**C**–S2'), trimers (*i.e.*, lane 5: S2–**A**–**B**–**C**–S1), and pentamers (*i.e.*, lane 6: S4–**B**–**C**–**A**–**B**–**C**–S1). A single gel was imaged for Cy3, Cy5, and FITC fluorescence and these images are merged into one composite image in Figure 6.28.

Scheme 3.2. Model structures of antibody oligomers.

A Antibody dimer model

Drawn to scale

B Antibody trimer model

Drawn to scale

C Antibody pentamer model

Drawn to scale
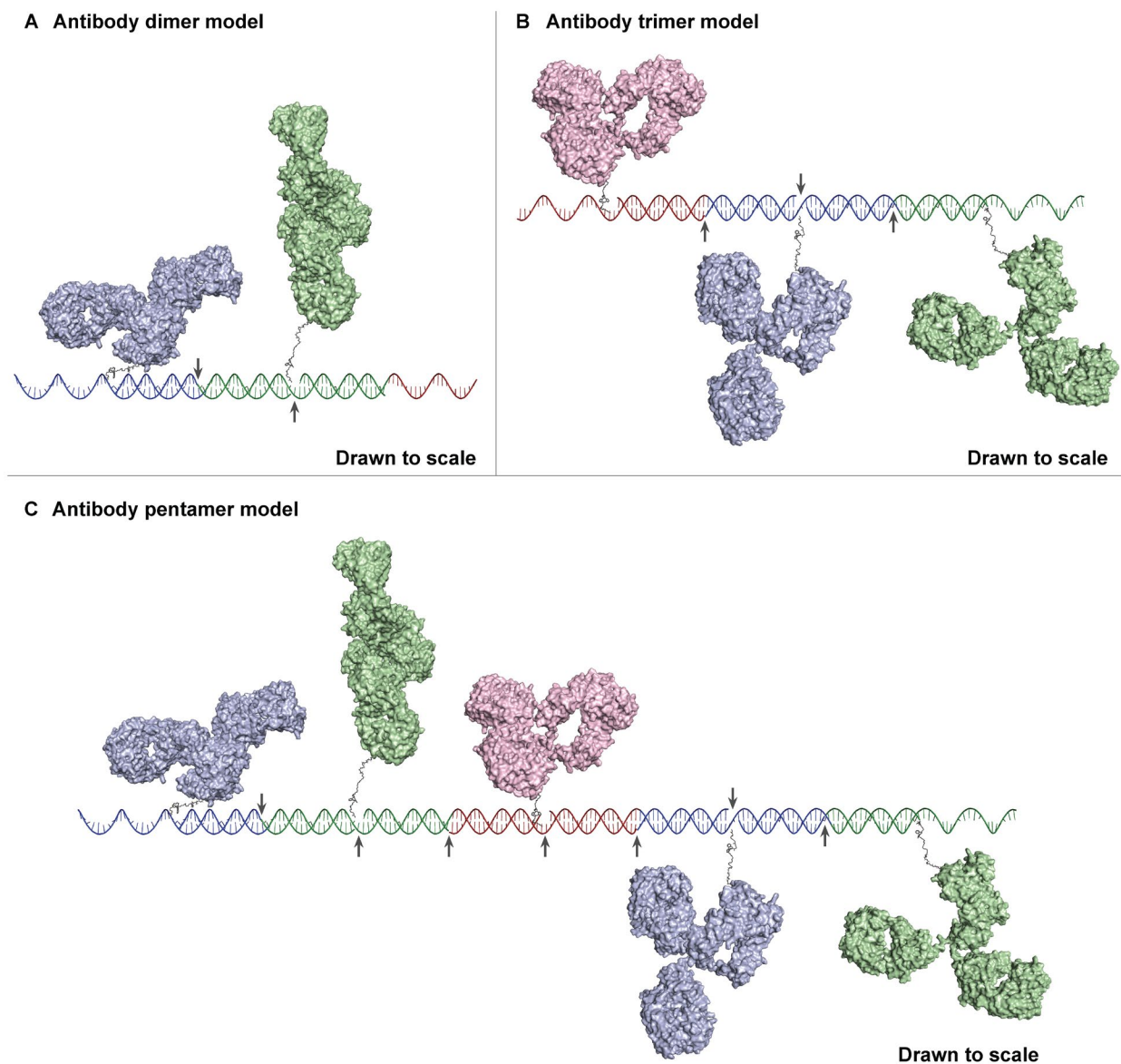
Model structures were created in PyMol of the antibody (A) dimers, (B) trimers, and (C) pentamers. An IgG antibody structure (PDB code: **1IGT**)[200] was used to model **anti-mouse-PD-1** (light red), **anti-mouse-TIGIT** (light blue), and **anti-mouse-CTLA-4** (light green). The ideal B-form DNA structure was used for the DNA scaffold. Nicks in the DNA scaffold are potential sites of DNA bending and are marked on the models with gray arrows. The model structures are drawn to scale.

3.5.2. Monodisperse and Sequence-Encoded Protein Trimers.

Next, the synthesis of a sequence-encoded protein trimer was targeted. Equal amounts of the **A**–DNA conjugate, **B**–DNA conjugate, **C**–DNA conjugate, S3'–S4' template strand, and S5'–S6' template strand were mixed to synthesize a protein trimer with the oligomeric sequence S2–**A**–**B**–**C**–S1 (Figure 3.6.C, Scheme 3.2.B) at an assembly yield of 27% (Figure 6.24.B and D). Oligomers that contain greater than three antibodies or trimers containing alternative oligomeric antibody sequences were not observed in the assembly mixture because there are no DNA sequences that are complementary to either the S2 or S1 DNA sequences. Protein trimers were isolated from the assembly mixture using SEC purification and characterized with agarose gel electrophoresis (Figure 3.6.E: lane 5). The agarose gel showed a single band for the trimers with the expected Cy3, Cy5, and FITC dye fluorescence as well as lower electrophoretic mobility on an agarose gel than the dimers. Therefore, these results indicate that monodisperse and sequence-encoded protein trimers with the oligomeric sequence S2–**A**–**B**–**C**–S1 were successfully synthesized. Importantly, no disassembly of S4–**B**–**C**–S2' or S2–**A**–**B**–**C**–S1 oligomers were observed over 10 days of storage at 4 °C.

3.6. Studying Antibody Properties After Oligomerization.

To ensure that this synthetic technique is generalizable, different protein oligomers were synthesized, including a protein dimer with the oligomeric sequence **A**–**B** (Figure 3.10) and a protein trimer with the oligomeric sequence, **A**–**B**–**B** (Figure 6.25). Furthermore, another antibody, **anti-human-PD-1 (D)**, was functionalized with a single DNA strand of S2–DBCO–Cy3–S3, S4–DBCO–Cy5–S5, or S6–DBCO–FITC–S1 (Figure 6.23), and the resulting constructs were organized into protein dimers with the oligomeric sequence **D**–**D** (Figures 6.26) and trimers with

the oligomeric sequence **D**–**D**–**D** (Figure 6.27). Analytical SEC analysis of antibody–DNA conjugates, dimers, and trimers show a single peak for each sample with decreases in retention time as degree of oligomerization increases (Figures 6.25.C and 6.26.C).

The target binding characteristics of human antibodies after functionalization with DNA and oligomerization with the modular DNA scaffold were investigated using antigen binding and checkpoint inhibitor activity cellular assays. We studied **D**, **D**–DNA conjugates, **D**–**D** dimers, and **D**–**D**–**D** trimers using these assays and confirmed that checkpoint inhibitor activity and antigen binding were retained in each sample (Figures 3.7 and 3.8). Importantly, an antibody dimer, **A**–**B**, exhibited minimal degradation in the cellular media used in these experiments (Figure 6.30).
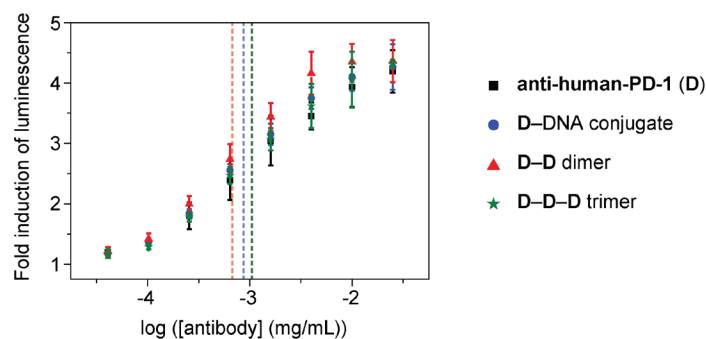
Sample	EC$_{50}$ (µg/mL)	95% Confidence Interval for EC$_{50}$ (µg/mL)
anti-human-PD-1 (**D**)	1.04	0.866–1.21
D–DNA conjugate	0.868	0.761–0.991
D–**D** dimer	0.683	0.591–0.789
D–**D**–**D** trimer	1.04	0.868–1.24

Figure 3.7. Cellular assay to characterize checkpoint inhibitor activity. The PD-1 blockade bioassay was run with treatments of **anti-human-PD-1** (**D**) (black square), **D**–DNA conjugate (blue circle), **D**–**D** dimer (red triangle), and **D**–**D**–**D** trimer (green star). EC50 values and 95% confidence intervals for EC50 values were calculated for each treatment (see table) and are represented on the plot with colored, dashed lines (lines for **D**–DNA conjugates and **D**–**D**–**D** trimers are overlaid).
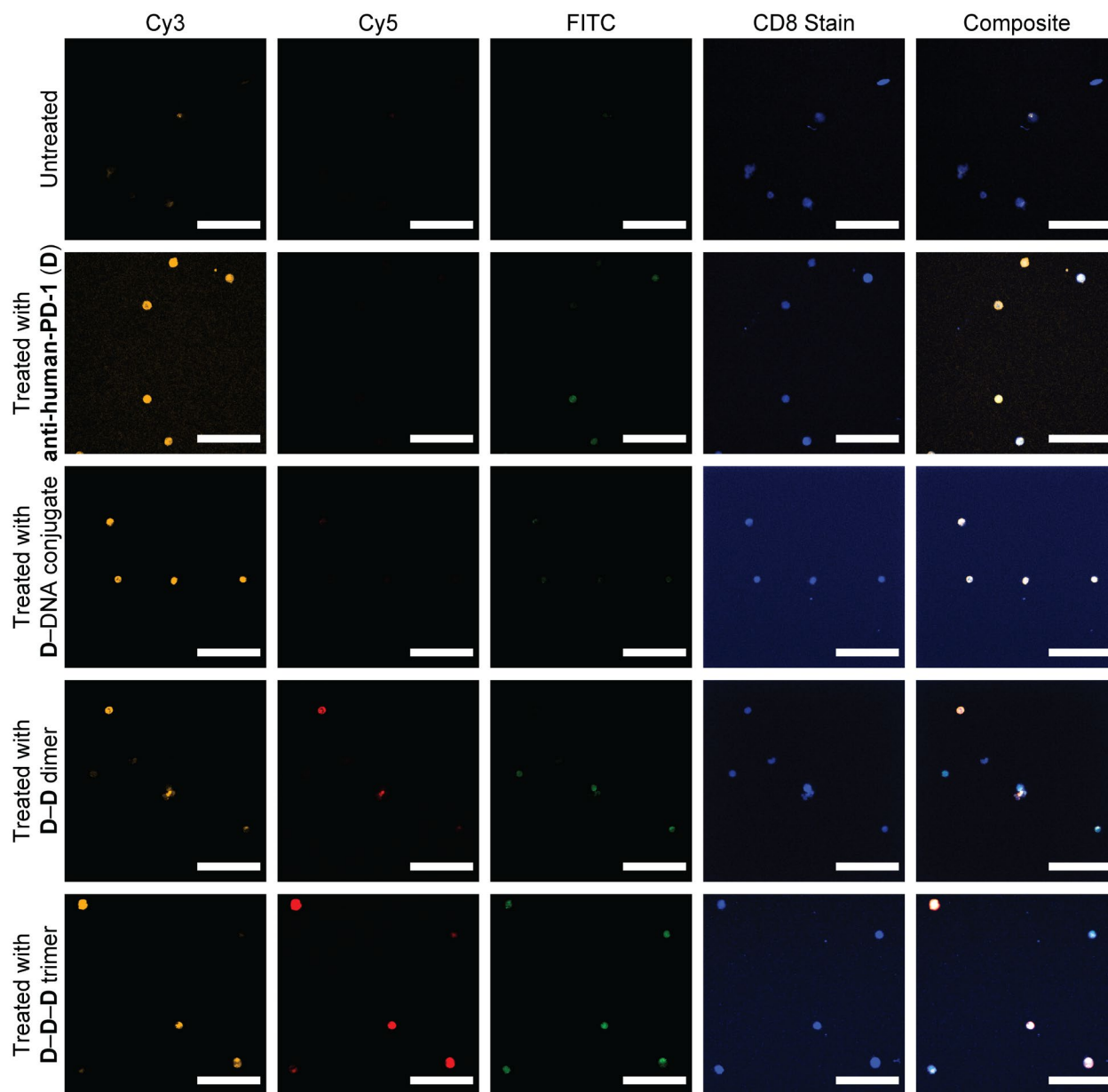
Figure 3.8. Confocal microscopy characterization of antibody binding. Representative confocal microscopy images of untreated human peripheral blood mononuclear cells (hPMBCs) with upregulated PD-1 and the same hPMBCs treated with **anti-human-PD-1 (D)** (modified with ~2.6 AF546 dye modifications per antibody), **D**–DNA conjugate, **D**–**D** dimer, and **D**–**D**–**D** trimer. Cells were imaged for Cy3, Cy5, FITC, and CD8. The Cy3, Cy5, FITC, and CD8 stain images were also merged into one composite image. Images for each sample were taken using the same laser settings and were processed identically. All scale bars are 50 μm.

3.7. Combining Protein Oligomer Building Blocks into Larger Oligomers.

Finally, a protein dimer and trimer were used as building blocks to synthesize a monodisperse and sequence-encoded protein pentamer where three different antibodies are organized into a precise oligomeric sequence. The S4–**B**–**C**–S2' protein dimer and S2–**A**–**B**–**C**–S1 protein trimer were mixed together at a 1:1 ratio and the specific binding between the S2' DNA sequence on the dimer and the S2 DNA sequence on the trimer leads to the synthesis of a sequence-encoded protein pentamer with the oligomeric sequence S4–**B**–**C**–**A**–**B**–**C**–S1 (Figure 3.6.D, Scheme 3.2.C) at an assembly yield of 58% (Figure 6.24.C and D). Oligomers that contain greater than five antibodies were not observed in the assembly mixture. Protein pentamers were isolated from other unreacted dimers, trimers, and template strands in the assembly mixture using SEC purification and characterized with agarose gel electrophoresis (Figure 3.6.E: lane 6). The pentamers showed a single band with the expected Cy3, Cy5, and FITC dye fluorescence as well as lower electrophoretic mobility on an agarose gel than the trimers. Therefore, monodisperse and sequence-encoded protein pentamers with the oligomeric sequence S4–**B**–**C**–**A**–**B**–**C**–S1 were successfully synthesized. This is the first reported monodisperse antibody pentamer that contains different antibodies in a predefined oligomeric sequence.

3.8. Discussion.

The designed set of six DNA strands was used as a modular scaffold to organize proteins into oligomers with programmed identity, stoichiometry, and oligomeric sequence. This scaffold provided access to monomer, dimer, and trimer building blocks that could be modularly combined independent from the identity of proteins. Therefore, all of the criteria for a versatile synthetic protein oligomerization method were met: (1) providing access to a large number of proteins per

oligomer, (2) providing access to any oligomeric sequence of the same or different proteins, (3) being generalizable with regard to proteins, and (4) not requiring mutations of the amino acid sequence of proteins and recombinant protein expression.

Established chemistry was used to functionalize a primary amine (*e.g.*, ε-amine on a lysine or α-amine on a N-terminus) on proteins with a single DNA strand.[199] Nearly all proteins contain primary amines, so this approach is generalizable with regard to proteins, including proteins that are commercially available, isolated from natural sources, or recombinantly expressed. Many other covalent[111, 113] and noncovalent[111-112] bioconjugation methods could also be used to modify proteins with one of the DNA strands reported here. By taking advantage of these approaches, nearly any protein can be modified with a single DNA strand and organized into monodisperse, sequence-encoded oligomers using the designed DNA scaffold.

While modular multi-protein constructs can be prepared using large DNA origami scaffolds, proteins comprise less than 20% of the mass of these constructs.[135-136, 154] This large amount of DNA compared to protein means that most of the solution properties and interactions of these constructs are dictated by the DNA scaffold instead of by the proteins. In cases where protein binding interactions are integral to function (*e.g.*, antibody–antigen binding), this large amount of DNA may affect target recognition and accessibility. In contrast, using the modular DNA scaffold reported in this work, proteins make up most of the mass of oligomeric constructs. For example, proteins comprise 84%, 86%, and 85% of the mass of the sequence-encoded **B**–**C** dimer, **A**–**B**–**C** trimer, and **B**–**C**–**A**–**B**–**C** pentamer, respectively. While other protein assembly techniques using nucleic acids also provide access to oligomers mostly comprised of proteins (*e.g.*,

oligomerization using a DNA tetrahedron scaffold[131] or a four-arm Holliday DNA junction scaffold[135-136, 154]), they lack modularity to access different numbers of proteins per construct.

In principle, the modular DNA scaffold described herein provides access to vast numbers of different oligomeric sequences and sizes. For example, 3 different proteins could be oligomerized into trimeric constructs with 27 different oligomeric sequences, including homo-oligomers of one protein, and hetero-oligomers of 2 or 3 proteins (Figure 3.1.C, *ii*). Oligonucleotides in the DNA scaffold interact through Watson–Crick–Franklin base pairing to form a right-handed double helix. Therefore, two oligomers with reversed oligomeric sequences of proteins (*e.g.*, **A**–**A**–**B** and **B**–**A**–**A**) form different structures and are considered as different sequences. Likewise, 2 different proteins could be oligomerized into dimeric constructs with 4 different oligomeric sequences, 4 different proteins could be oligomerized into tetrameric constructs with 256 different oligomeric sequences, and 5 different proteins could be oligomerized into pentameric constructs with 3,125 different oligomeric sequences (Table 3.2). Furthermore, each oligomer building block synthesized using this method inherently contains living chain ends where more units could be added to access larger oligomers (*e.g.*, hexamers, heptamers, and octamers, Scheme 3.3). Considering the growing number of discovered proteins, the foundational oligomers synthesized in this work illustrate the unlimited number of protein oligomers that could be accessed via a single modular DNA scaffold.

Scheme 3.3. Proposed routes to access various sequence-encoded oligomers.
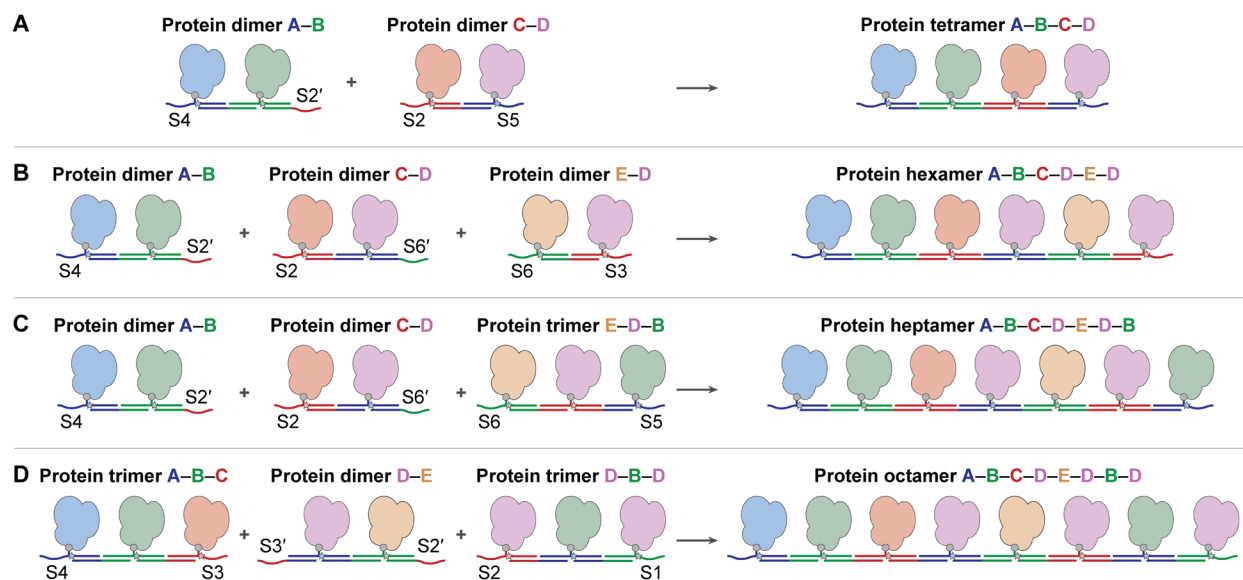
(A) Tetramer, (B) hexamer, (C) heptamer, and (D) octamer can potentially be accessed by combining dimer and trimer building blocks.

3.9. Conclusion.

In conclusion, this work shows how monodisperse, sequence-encoded protein oligomers can be synthesized using generalizable bioconjugation chemistry with a judiciously designed DNA scaffold. This versatile protein oligomerization approach is powerful and useful because oligomers with different stoichiometries and oligomeric sequences can be synthesized without the need to redesign the proteins or the DNA scaffold. Importantly, this synthetic advance will enable subsequent studies to understand the fundamental relationships between protein oligomer structures and properties, which have significant implications for many fields (*e.g.*, therapeutics, catalysis, photosynthesis, and membrane transport).

3.10. Materials and Methods.

See Appendix B (Section 6.2).

CHAPTER FOUR

Protein Block Co-Oligomer Assembly Using DNA and Metal Coordination.

4.1. Assembly of Block Co-Polymers.

Block co-polymers are molecules that contain multiple polymeric sections with different constituent monomers. The size and chemical composition of blocks within the polymers regulate their assembly into defined morphologies, including micelles, worms, branched worms, and vesicles.[201-203] These assembly structures have widespread applications in lithography, photonics and controlled drug delivery.[204] Recent investigations into sequence-encoded DNA block co-oligomers have discovered that assembly morphology depends on the sequence of DNA.[205-206] The modification of block co-polymers assembly structures with proteins would imbue these materials with the functions of proteins.

4.2. Block Co-Oligomerization of Proteins.

Proteins are exciting monomers for block co-oligomers because they contain exact ratios of different functional groups (*i.e.*, amines, carboxylic acids, thiols) and defined functions (*e.g.*, catalysis). Proteins have been covalently linked to polymers and in these structures, the polymer directs the assembly morphology of proteins.[69, 207-210] However, the assembly of protein block co-oligomers where the number and identities of proteins dictate assembly morphology is less well understood. It remains challenging to prepare protein block co-oligomers (current methods to synthesize protein oligomers are discussed in Chapter Three).

DNA hybridization chain reaction (HCR) has been used to oligomerize DNA-modified proteins, where living chain ends on the DNA HCR scaffold enable the chain extension of oligomers (Figure 4.1.A).[121] Recently, we reported that the introduction of a base-pair mismatch in the duplex of HCR hairpins can be used to differentiate the rates of initiation and propagation in DNA HCR oligomerization (Figure 4.1.B).[122] This advance enabled the controlled

oligomerization of DNA-modified GFP and the synthesis of GFP block co-oligomers (GFP_2-*b*-GFP_2, Figure 4.1.C).[122]
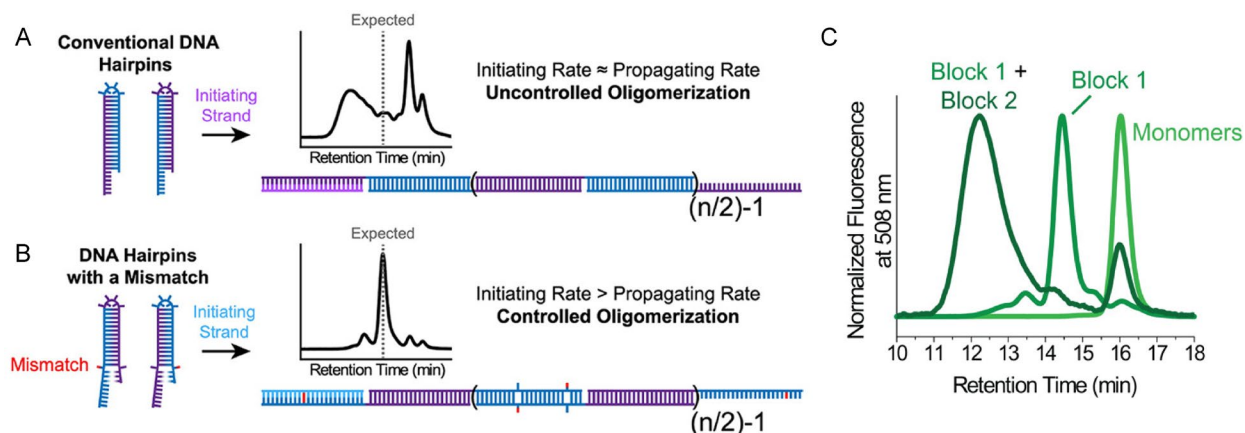
Figure 4.1. Protein block co-oligomer characterization. (A) In conventional hybridization chain reaction (HCR), the rates of initiation and propagation are similar, leading to uncontrolled oligomerization. (B) When DNA hairpins are modified with a mismatch, the rate of initiation is increased and the rate of propagation is decreased, resulting in controlled oligomerization. (C) Size-exclusion chromatogram of monomers (GFP–DNA, light green), oligomers (GFP_2, green), and block co-oligomers (GFP_2–*b*–GFP_2, dark green).

We hypothesize that the assembly morphology of protein block co-oligomers depends on the number and identity of proteins per block. A wide range of protein block co-oligomers and their assembly properties will be studied to test this hypothesis. First, model proteins GFP, MBP, and BSA will be prepared that contain one surface-exposed thiol and varying numbers of surface-exposed amines. Each protein will be modified with one DNA HCR hairpin strand at the thiol and varying numbers of pyridine coordinating ligands at the amines. Next, proteins will be assembled using DNA HCR into a diverse set of different block co-oligomers with varying numbers and identities of proteins (Scheme 4.1.A). After adding metal cations to the block co-oligomers, the impact of block design on resulting assembly morphologies will be studied (Scheme 4.1.B).

Scheme 4.1. Protein block co-oligomer synthesis.
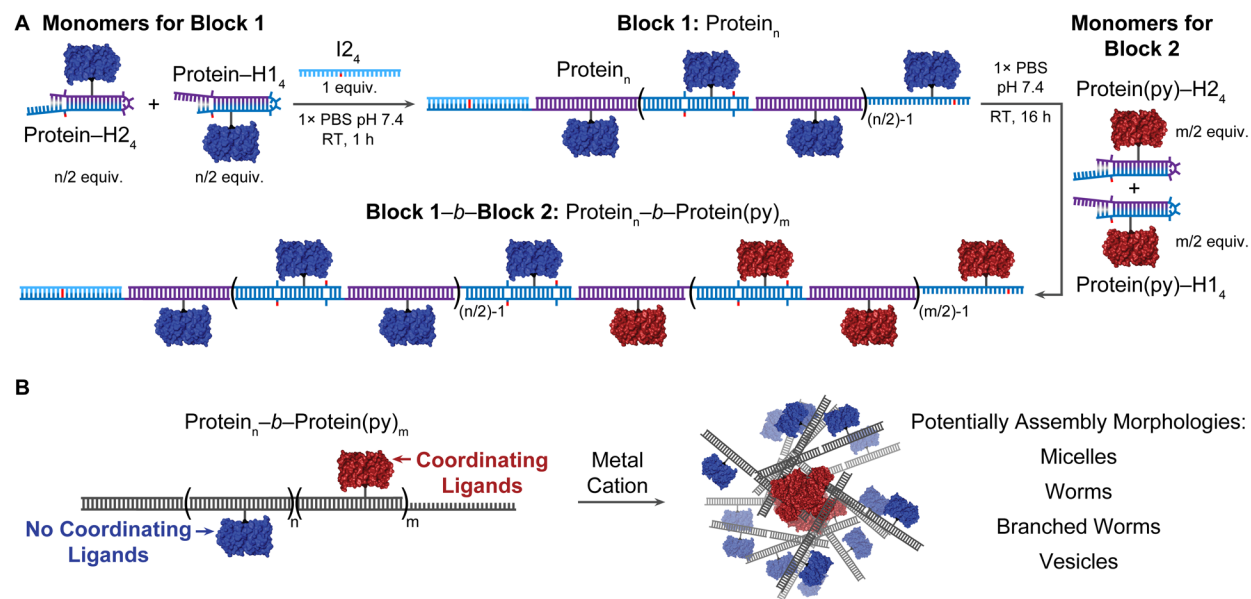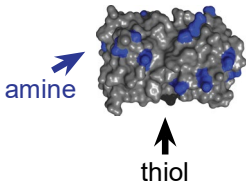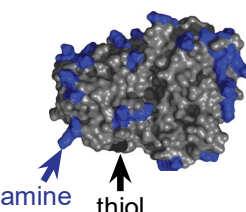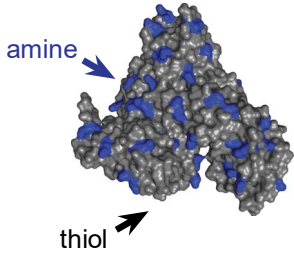
(A) Monomers for Block 1 are protein–H1$_4$ and protein–H2$_4$. When the initiator strand is added to this mixture, these proteins assemble via controlled HCR oligomerization. The overhanging sticky end acts as an initiator for the HCR oligomerization after a second set of monomers, protein(py)–H1$_4$ and protein(py)–H2$_4$, are added. (B) The assembly morphologies of block co-oligomers upon the addition of metal cations will be studied. The impact of the number of proteins and protein identity in each block on assembly morphology will be investigated.

4.2.1. Preparation of Proteins with Different Ratios of Reactive Groups.

Model monomeric proteins were selected that contain one surface-exposed thiol and varying numbers of surface-exposed amines (Tables 4.1 and 4.2): mutant GFP (1 thiol and 19 amines), mutant maltose-binding protein (MBP, 1 thiol and 36 amines), and native bovine serum albumin (BSA, 1 thiol and 59 amines). While BSA contains 35 cysteine residues, 34 form 17 disulfide bonds, so only one cysteine is surface-exposed. Mutant GFP and mutant MBP were recombinantly expressed in *E. coli* and isolated using standard protein purification techniques. Proteins were characterized using SDS-PAGE and MALDI-TOF MS (Figure 4.2.B, lanes 1, 3, and 5; Figure 4.3).

Table 4.1. DNA Sequences and Characterization.

Protein	Model Structure	Number of Surface-Exposed Thiols	Number of Surface-Exposed Amines
Mutant GFP	amine thiol	1 (1 surface-exposed cysteine)	19 (1 *N*-termini and 18 lysines)
Mutant MBP	amine thiol	1 (1 surface-exposed cysteine)	36 (1 *N*-termini and 35 lysines)
Native BSA	amine thiol	1 (1 surface-exposed cysteine not involved in disulfide bonds)	59 (1 *N*-termini and 58 lysines)

The surface-exposed cysteine on the model of BSA is behind the front face of the protein. and is highlighted with an arrow. The relative sizes of the model structures are drawn to scale.

Table 4.2. Amino acid sequences of mutant GFP, mutant MBP, and native BSA.

Protein	Amino Acid Sequence
Mutant GFP (thiol on C191)	**M**HHHHHHGGGGENLYFQSGGGGGMVS**K**GEELFTGVVPILVELDGDV NGH**K**FSVSGEGEGDATYG**K**LTL**K**FILTTG**K**LPVPWPTLVTTLTYGV QCFSRYPDHM**K**QHDFF**K**SAMPEGYVQERTIFF**K**DDGNY**K**TRAEV**K** FEGDTLVNRIEL**K**GIDF**K**EDGNILGH**K**LEYNYNSHNVYIMAD**K**Q**K**N GI**K**VNF**K**IRHNIEDGSVQLADHYQQNTPIG**C**GPVLLPDNHYLSTQSA LS**K**DPNE**K**RDHMVLLEFVTAAGITLGMDELY**K**
Mutant MBP (thiol on C75)	**M**HHHHHHGGGGENLYFQSGGGGG**K**IEEG**K**LVIWINGD**K**GYNGLAEV G**KK**FE**K**DTGI**K**VTVEHPD**K**LEE**K**FPQVAATGDGPDIIFWAHDRFGG YAQ**C**GLLAEITPD**K**AFQD**K**LYPFTWDAVRYNG**K**LIAYPIAVEALSLI YN**K**DLLPNPP**K**TWEEIPALD**K**EL**K**A**K**G**K**SALMFNLQEPYFTWPLIA ADGGYAF**K**YENG**K**YDI**K**DVGVDNAGA**K**AGLTFLVDLI**K**N**K**HMNA DTDYSIAEAAFN**K**GETAMTINGPWAWSNIDTS**K**VNYGVTVLPTF**K**G QPS**K**PFVGVLSAGINAASPN**K**ELA**K**EFLENYLLTDEGLEAVN**K**D**K**PL GAVAL**K**SYEEELA**K**DPRIAATMENAQ**K**GEIMPNIPQMSAFWYAVRT AVINAASGRQTVDEAL**K**DAQT
Native BSA (thiol on C34)	**D**TH**K**SEIAHRF**K**DLGEEHF**K**GLVLIAFSQYLQQ**C**PFDEHV**K**LVNELT EFA**K**TCVADESHAGCE**K**SLHTLFGDELC**K**VASLRETYGDMADCCE **K**QEPERNECFLSH**K**DDSPDLP**K**L**K**PDPNTLCDEF**K**ADE**KK**FWG**K**YL YEIARRHPYFYAPELLYYAN**K**YNGVFQECCQAED**K**GACLLP**K**IETM RE**K**VLTSSARQRLRCASIQ**K**FGERAL**K**AWSVARLSQ**K**FP**K**AEFVEV T**K**LVTDLT**K**VH**K**ECCHGDLLECADDRADLA**K**YICDNQDTISS**K**L**K**E CCD**K**PLLE**K**SHCIAEVE**K**DAIPENLPPLTADFAED**K**DVC**K**NYQEA**K** DAFLGSFLYEYSRRHPEYAVSVLLRLA**K**EYEATLEECCA**K**DDPHAC YSTVFD**K**L**K**HLVDEPQNLI**K**QNCDQFE**K**LGEYGFQNALIVRYTR**K**V PQVSTPTLVEVSRSLG**K**VGTRCCT**K**PESERMPCTEDYLSLILNRLCVL HE**K**TPVSE**K**VT**K**CCTESLVNRRPCFSALTPDETYVP**K**AFDE**K**LFTFH ADICTLPDTE**K**QI**KK**QTALVELL**K**H**K**P**K**ATEEQL**K**TVMENFVAFVD **K**CCAADD**K**EACFAVEGP**K**LVVSTQTALA

Surface-exposed thiols from cysteines are highlighted (bolded and underlined) in black and surface-exposed amines from lysines and N-termini are highlighted (bolded) in blue.

4.2.2. Modification of Proteins with One DNA Strand for Oligomerization.

To organize proteins into linear block co-oligomers using DNA HCR, proteins must be modified with one copy of a hairpin DNA (Table 4.3, *i.e.*, H1$_4$–NH$_2$ or H2$_4$–NH$_2$). The single surface-exposed thiol on each protein was functionalized with one amine-modified DNA strand (Figure 4.2.A). To achieve this, DNA was first snap-cooled in 1× PBS by heating to 95 °C for 4

min, cooled to 4 °C over 1 minute, held at 4 °C for 4 min, then brought to room temperature. Next, 1 equiv. of DNA was reacted with 50 equiv. of succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) in 50:50 DMF:1× PBS, pH 7.4 for 1 h at RT. DNA was purified from excess SMCC with SEC. Finally, 3–5 equiv. of purified DNA was then reacted with 1 equiv. of GFP, MBP, or BSA overnight at RT. Anion exchange chromatography was used to isolate proteins that were functionalized with a single DNA strand from unreacted proteins and DNA. Six different protein–DNA conjugates (*i.e.*, GFP–H1$_4$, GFP–H2$_4$, MBP–H1$_4$ MBP–H2$_4$, BSA–H1$_4$, and BSA–H2$_4$) were prepared and confirmed to contain a single DNA functionalization via SDS-PAGE (Figure 4.2.B).

Table 4.3. DNA Sequences and Characterization.

Strand Name	DNA Sequence (5′ → 3′)[a]	$\varepsilon_{260\ nm}$ (M^{-1}cm^{-1})[b]	MW$_{theor.}$ (Da)[b]	MW$_{exp.}$ (Da)
I2$_4$	CAA AGT **G**TA GGA TTC GGC GTG	242,200	7,473	7,468
H1$_4$–NH$_2$	TTA ACC CAC GCC GAA TCC TAG ACT CAA AGT CTA GGA *T*(**NH₂**)TC GG**G** GTG	471,200	14,786	14,706
H2$_4$–NH$_2$	AGT CTA GGA T*T*(**NH₂**)C GGC GTG GGT TAA CAC GCC GAA TCC TA**C** ACT TTG	457,800	14,759	14,945

[a]**X** (bolded and underlined base) denotes the position of mismatch in DNA sequences. *T*(**NH₂**) denotes the thymine base that is modified with a primary amine from the phosphoramidite amino-modifier C6 dT (see Scheme 6.4). [b] Extinction coefficients and theoretical molecular weights were calculated using the OligoAnalyzer Tool from IDT DNA.
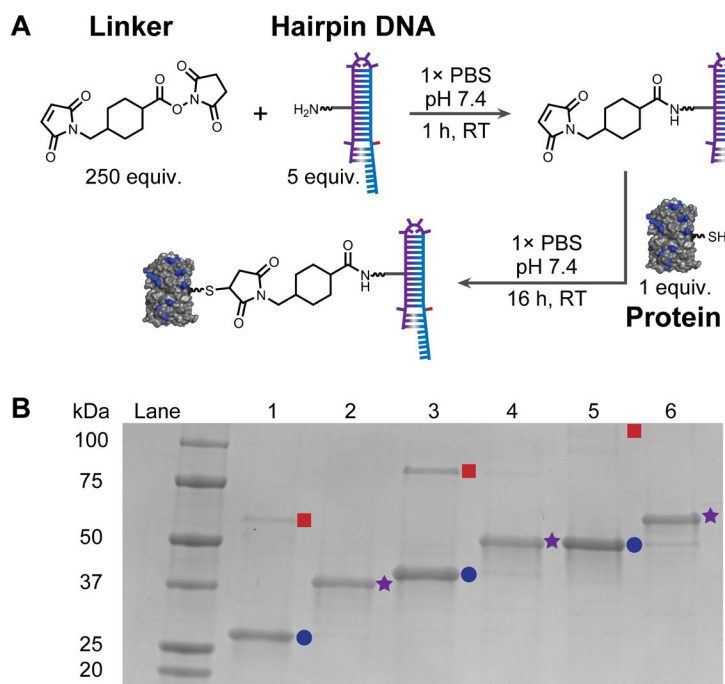
Figure 4.2. Protein–DNA building block synthesis. (A) Reaction scheme of protein functionalization with DNA. (B) SDS-PAGE characterization of proteins (*i.e.*, lane 1: GFP, lane 3: MBP, lane 5: BSA) and protein–DNA conjugates (*i.e.*, lane 2: GFP–DNA, lane 4: MBP–DNA, lane 6: BSA–DNA). The blue circles highlight bands identified as protein monomers, the red squares highlight bands identified as protein disulfide-mediated dimers, and the purple star highlight bands identified as proteins modified with one DNA strand.

4.3. Protein Modification with Coordinating Ligands.

4.3.1. Modification of Proteins with Coordinating Ligands.

Amines on GFP, MBP, and BSA were modified with pyridine groups via a reductive amination reaction to prepare GFP(py), MBP(py), and BSA(py). Each protein was reacted with 20,000 equiv. 4-pyridine carboxaldehyde in the presence of sodium cyanoborohydride at 4 ° C in 1× PBS (Figure 4.3.A). The number of pyridine modifications per protein was determined using MALDI-TOF MS to be 16 per GFP(py), 35 per MBP(py), and 56 per BSA(py), translating to yields of 84%, 97%, and 95%, respectively. Differences in yields between proteins could result from differences in surface accessibility or local chemical environment of specific amines on each protein.
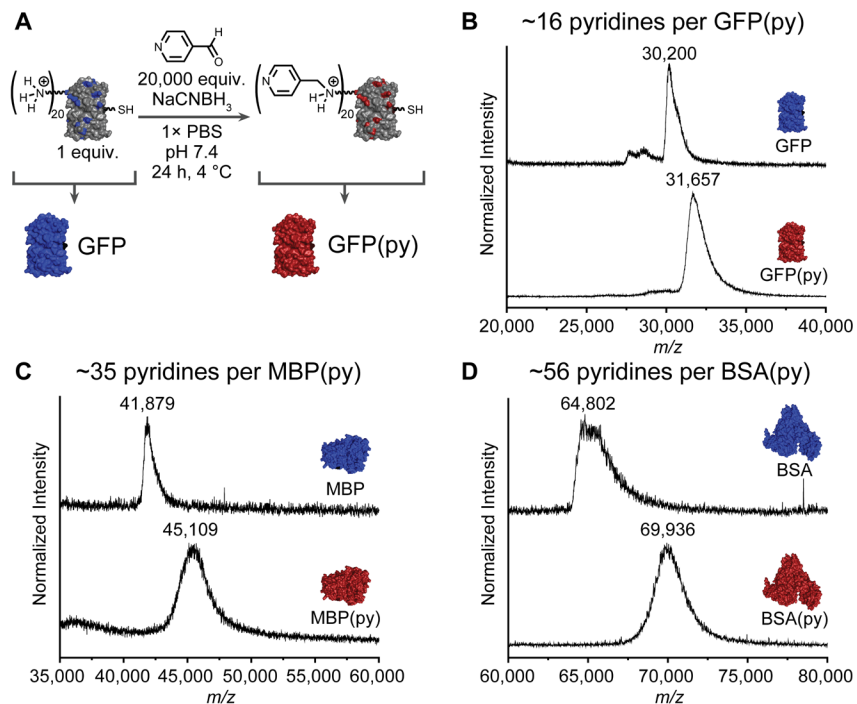
Figure 4.3. Modification of proteins with coordinating pyridine ligands. (A) Scheme of the reductive amination reaction used to modify amines on proteins with pyridine groups. MALDI-TOF MS was used to characterize the number of pyridine modification modifications on (A) GFP, (B) MBP, and (C) BSA. To calculate the number of pyridine modifications, the observed mass-to-charge ratios proteins (top) and pyridine-modified proteins (bottom) were compared.

4.3.2. Protein Assembly Via Metal Coordination.

Next, the assembly of GFP(py) and BSA(py) via metal coordination was studied. GFP, GFP(py), BSA, and BSA(py) were added to 125 equiv. of nickel (II) chloride and aggregation was observed for GFP(py) and BSA(py), but not GFP and BSA (Figure 4.4.A and D, *i* and *ii*). Next, the addition of 500 equiv. of ethylenediaminetetraacetic acid (EDTA) led to the disappearance of the aggregates. Dynamic light scattering was used to characterize the sizes of protein assemblies. The sizes of GFP and BSA show little change upon addition of nickel cations and EDTA (Figure 4.4.B and E). In contrast, GFP(py) and BSA(py) show dramatic increases in size when nickel cations are added and complete reversal to their original size upon addition of EDTA (Figure 4.4.C

and F). Together, these results indicate that proteins that are modified with pyridine ligands can be reversibly assembled into aggregate structures via metal coordination with nickel cations.
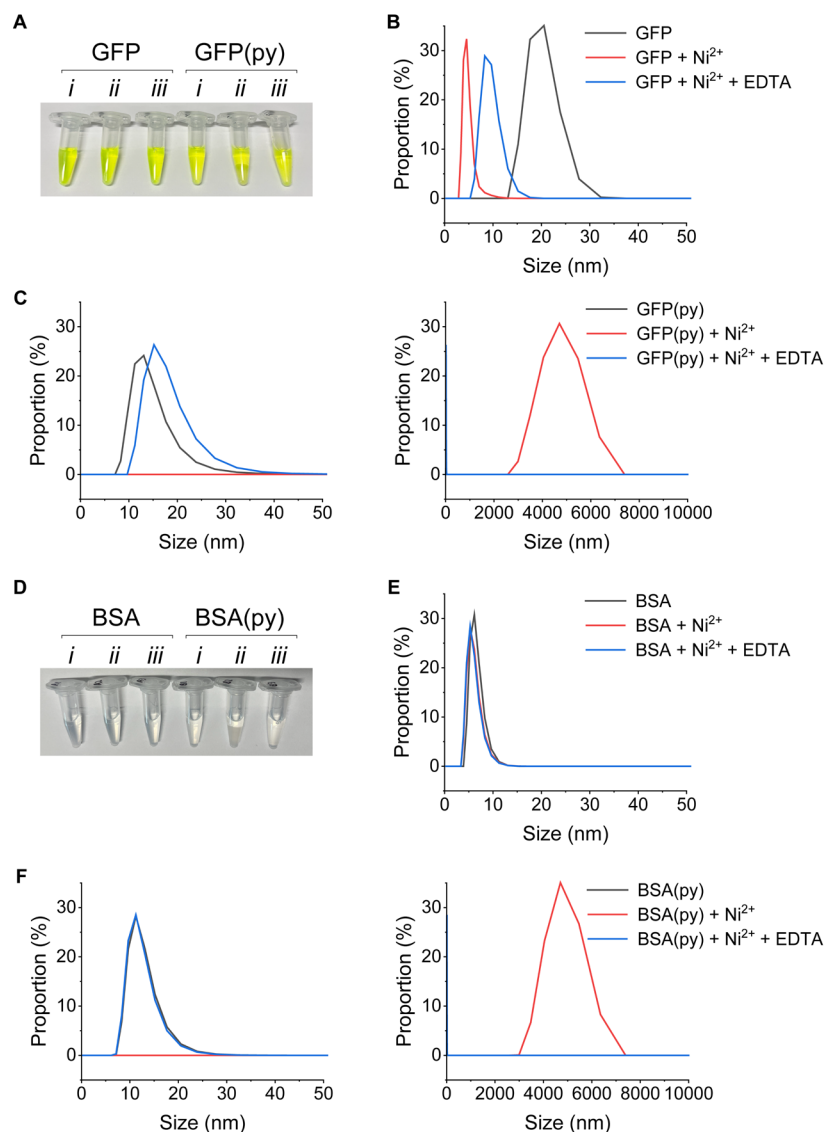
Figure 4.4. Characterization of protein assembly via metal coordination. (A) Optical images of *i* GFP (left) and GFP(py) (right) and the same proteins added to *ii* nickel (II) cations and, subsequently, *iii* EDTA. Characterization of (B) GFP and (C) GFP(py) assembly state via dynamic light scattering (DLS). (D) Optical images of *i* BSA (left) and BSA(py) (right) and the same proteins added to *ii* nickel (II) cations and, subsequently, *iii* EDTA. Characterization of (E) BSA and (F) BSA(py) assembly state via DLS.

4.4. Conclusion.

In this work, we synthesized and purified a wide variety of protein–DNA monomers for controlled DNA HCR oligomerization, including GFP–H1$_4$, GFP–H2$_4$, MBP–H1$_4$, MBP–H2$_4$, BSA–H1$_4$, and BSA–H2$_4$. The surface-exposed amines on each of these proteins were modified with pyridine ligands and coordination between metal cations and these coordinating ligands was found to direct the assembly of proteins. In ongoing experiments, these protein–DNA and protein(py)–DNA (*e.g.*, GFP(py)–H1$_4$, GFP(py)–H2$_4$, MBP(py)–H1$_4$, MBP(py)–H2$_4$, BSA(py)–H1$_4$, and BSA(py)–H2$_4$) building blocks will be oligomerized using controlled HCR into a wide number of different oligomers, including GFP$_2$–*b*–GFP(py)$_2$, GFP$_2$–*b*–GFP(py)$_4$, GFP$_2$–*b*–GFP(py)$_6$, GFP$_2$–*b*–GFP(py)$_8$, GFP$_2$–*b*–MBP(py)$_2$, GFP$_2$–*b*–BSA(py)$_2$, MBP$_2$–*b*–GFP(py)$_2$, MBP$_2$–*b*–MBP(py)$_2$, MBP$_2$–*b*–BSA(py)$_2$, BSA$_2$–*b*–GFP(py)$_2$, BSA$_2$–*b*–MBP(py)$_2$, and BSA$_2$–*b*–BSA(py)$_2$. Next, these block co-oligomers will be assembled using metal coordination and study under electron microscopy (EM) to learn how the design of blocks (*i.e.*, block size or protein identity) affect assembly morphology. Ultimately, this work will enable the generalizable self-segregating assembly of multiple proteins for applications as devices.

4.5. Materials and Methods.

See Appendix C (Section 6.3).

CHAPTER FIVE

Conclusion and Outlook.

5.1. Conclusion.

This dissertation establishes generalizable approaches to program the organization of proteins into sequence-encoded architectures using judiciously designed DNA. Importantly, vast numbers of different protein architectures can be accessed by tuning four parameters of protein–DNA building blocks: (1) the amino acid sequences of proteins, (2) the base sequences of DNA, (3) protein–DNA attachment sites, and (4) protein–DNA linker structure. We have discovered design principles that enable the preparation of synthetic protein single crystals and sequence-encoded oligomers.

In Chapter One ("*Natural and Engineered Protein Architectures.*"), the importance of proteins and protein architectures in biology is described. Next, existing approaches to prepare synthetic protein architectures are described along with corresponding limitations. Finally, we motivate the use of designed DNA interactions to program protein assembly into synthetic architectures.

In Chapter Two ("*DNA-Directed Protein Packing within Single Crystals.*"), the impact of replacing native PPIs with DNA hybridization interactions on protein crystallization is studied. The first single crystal structures are reported where DNA hybridization interactions between the surfaces of proteins direct the packing of proteins. We analyzed how these structures change according to the design of DNA length, sequence, and complementarity as well as protein–DNA attachment sites and linker structure. One lesson learned is that crystallization is observed when the flexibility of the introduced DNA linkage between proteins is minimized. This is achieved by decreasing DNA length, attaching DNA to a defined secondary structure instead of a loop, and decreasing length of the protein–DNA linker. Together, this work is an essential step toward

designing and engineering protein packing within single crystals and could lead to future applications in protein structure determination and functional protein crystal materials.

In Chapter Three ("*Modular Nucleic Acid Scaffolds for Synthesizing Monodisperse and Sequence-Encoded Antibody Oligomers.*"), a modular DNA scaffold is designed that can be used to prepare monodisperse, sequence-encoded protein oligomers in a generalizable manner. This versatile protein oligomerization approach is powerful and useful because oligomers that contain different numbers, stoichiometries, and oligomeric sequences of proteins can be synthesized without the need to redesign the proteins, DNA, or the protein–DNA linkage. Importantly, this synthetic advance will enable subsequent studies to discover the fundamental relationships between protein oligomer structures and properties, which have significant implications for many fields (*e.g.*, therapeutics, catalysis, photosynthesis, and membrane transport).

In Chapter Four ("*Protein Block Co-Oligomer Assembly Using DNA and Metal Coordination.*"), protein block co-oligomers are synthesized to investigate the implications of block design on assembly morphology. DNA HCR is used to organize DNA-modified proteins into sequence-encoded block co-oligomers. Next, metal coordination between proteins that are modified with coordinating ligands and metal cations lead to the assembly of proteins. This work enables future investigations into how the design of protein block co-oligomers (*i.e.*, block size, protein identity) impact the assembly morphology of protein block co-oligomers. This work will ultimately enable the generalizable self-segregating assembly of multiple proteins for applications as devices.

This dissertation establishes method to program the organization of proteins into sequence-encoded protein single crystals and oligomers. These methods are generalizable to any protein of

interest and accessible to any chemist. Furthermore, nearly limitless different structures can be prepared without any redesign of proteins, DNA, or the protein–DNA linkage. Overall, this work will enable future development of designer protein-based materials that mimic and even surpass the natural functions of proteins.

5.2. Outlook.

Life depends upon the dynamic regulation of complex structures of proteins. Protein conformation and, therefore, protein function is controlled by environmental cues. However, it remains challenging to engineer proteins to reversibly switch between different structural states in response to stimuli. In contrast, DNA hybridization interactions can be readily designed to change structure in response to a wide range of stimuli, including pH, light, and complementary DNA strands. The lessons learned in this thesis will be harnessed to design synthetic protein–DNA architectures (*e.g.*, oligomers and hydrogels) that exhibit programmed and dynamic functions.

5.2.1. Dynamic Protein Oligomers to Regulate Protein Functions.

Proteins are poised to revolutionize chemical synthesis and materials science through the discovery, design, and synthesis of functional protein-based materials that display complex yet well-coordinated collective behavior. However, it remains synthetically challenging to reliably disrupt and override PPIs and protein functions. We hypothesize that DNA interactions can reversibly regulate PPIs and protein functions (Scheme 5.1). This hypothesis will be tested by judiciously designing DNA to regulate the PPIs and functions of split fluorescent proteins (FPs) and a multisubunit enzyme as model systems.

Scheme 5.1. Controlling the interactions between proteins using DNA.
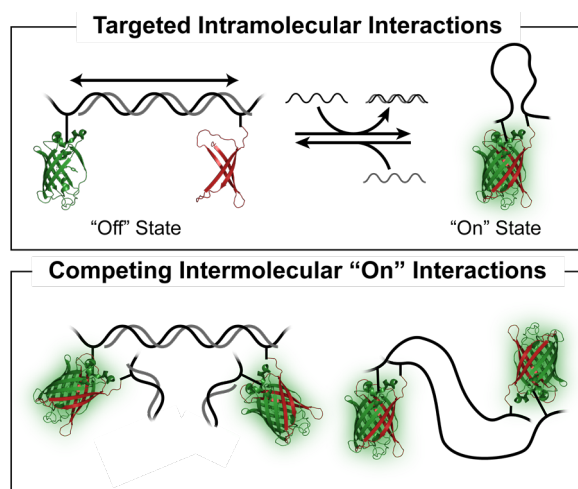
A split protein is a protein that has been split into two separate polypeptide chains. For example, the GFP β-barrel contains 11 β-strands (denoted hereafter as GFP_{1-11}) that can be split into two sections (*i.e.*, GFP_{1-7} and GFP_{8-11}; GFP_{1-9} and GFP_{10-11}; GFP_{1-10} and GFP_{11}). Corresponding pairs of split GFP are nonfunctional when separate and assemble reversibly with distinct binding thermodynamics to exhibit fluorescence.[211] Therefore, split GFP is an ideal model system because its protein–protein interactions result in a simple *in situ* fluorescence readout. Here, we will recombinantly express GFP_{1-7}, GFP_{8-11}, GFP_{1-9}, GFP_{10-11}, GFP_{1-10}, and GFP_{11} that contain a C-terminal reactive amino acid (*i.e.*, cysteine or an unnatural amino acid that contains an azide). If it is challenging to express a particular split GFP, we can express fusions between the split GFP and a highly soluble protein, MBP joined by a protease cleave site (*e.g.*, MBP–tobacco etch virus (TEV) protease site–GFP_{11}).[212-213] Next, we will functionalize the reactive C-terminal amino acid on each split GFP with a single DNA strand. We will use DNA–DNA interactions and DNA ligation to prepare molecular dimers that contain one copy of each split GFP fragment (Figure 5.1.A, *i.e.*, GFP_{1-7}–GFP_{8-11}, GFP_{1-9}–GFP_{10-11}, GFP_{1-10}–GFP_{11}). We hypothesize that the addition of a complementary template DNA strand to the split GFP oligomer will lead to a rigid DNA

duplex that spatially isolates each split GFP fragment. This spatial separation will prevent fluorescence (the "off" state) because the thermodynamic gain of DNA–DNA interactions is larger than the thermodynamic gain of protein–protein interactions ($\Delta G_{DNA} > \Delta G_{PPI}$, Figure 5.1.B). By removing the complementary template DNA strand using a strand displacement reaction, the split protein fragments will be separated by a flexible, single-stranded DNA region, leading to protein assembly and fluorescence turn-on (the "on" state, $\Delta G_{DNA} < \Delta G_{PPI}$, Figure 5.1.C). The proportion of split GFP oligomers in "on" and "off" states will be quantified using bulk fluorescence measurements on a plate reader. Together, this work will result in the ability to use designed DNA–DNA interactions to modulate the interactions and fluorescence of split GFP fragments.
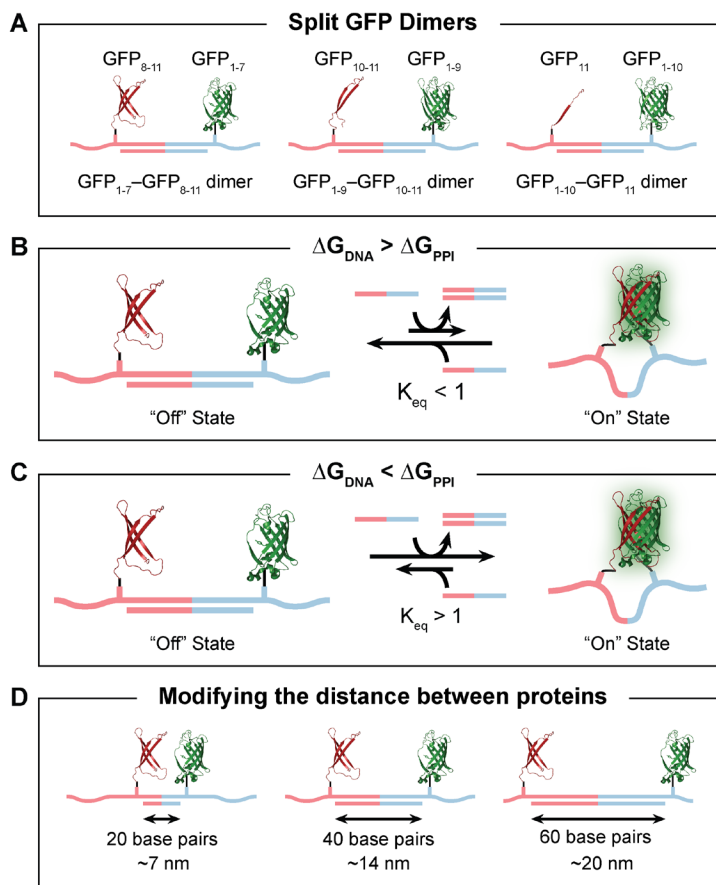
Figure 5.1. Regulating protein fluorescence using DNA. (A) Split GFP oligomers will be prepared using DNA. DNA–DNA interactions will compete with the protein–protein interaction in split GFP oligomers. (B) If the free energy of the DNA interaction is greater than the free energy of the protein–protein interaction, then the equilibrium will be biased towards the "off" state. (C) If the free energy of the DNA interaction is less than the free energy of the protein–protein interaction, then the equilibrium will be biased towards the "on" state. (D) The number of base pairs between protein fragments will be varied to study how the local concentration of split GFP fragments affects inter- and intramolecular protein–protein interactions.

Understanding intramolecular protein–protein interactions is important because this will allow us to develop explicit correlations between protein function (*i.e.*, split GFP fluorescence) and DNA design. Due to the propensity of split GFP to reassemble, we expect that there will be competing metastable inter- and intramolecular interactions (Scheme 5.1). To learn how to favor intramolecular interactions over intermolecular interactions, we will study how global and local

concentrations of split GFP fragments affects assembly outcomes. We will vary global split protein concentration by changing oligomer concentrations in solution. In addition, we will vary local split protein concentration by changing the number of bases between the two fragments to place split protein fragments at different positions along the DNA strands (Figure 5.1.D). Subsequently, using gel electrophoresis and SEC, we can determine conditions where intramolecular interactions dominate over intermolecular interactions. If achieving exclusively intramolecular interactions using the described methods proves challenging, we can alternatively vary binding strength of the template DNA strand to the split GFP dimer (*e.g.*, increase G-C content, introduce locked nucleic acids), the rigidity of DNA design (*e.g.*, parallel duplexes, six-helix duplexes), or the stoichiometry of split GFP fragments in the protein oligomer. By designing split protein oligomers that exclusively exhibit intramolecular protein–protein interactions, we can investigate how DNA–DNA interaction thermodynamics can be programmed to override the protein–protein interaction thermodynamics.

To quantify the thermodynamics of protein–protein interactions between split GFP fragments on oligomers, we will measure the decrease in bulk fluorescence intensity that is caused by DNA–DNA interactions with known thermodynamics. We will specifically design template DNA strands with binding energies that are predefined by DNA sequence design. The measured protein–protein interaction thermodynamics will then be compared to that of unmodified split GFP fragments as characterized by traditional techniques (*i.e.*, isothermal or differential scanning calorimetry, analytical ultracentrifugation, surface plasmon resonance)[214] to understand how the incorporation of split GFP fragments into DNA-based materials affects protein–protein interactions. The robustness of this method will be determined by studying the thermodynamics

of GFP$_{1-7}$–GFP$_{8-11}$, GFP$_{1-9}$–GFP$_{10-11}$, and GFP$_{1-10}$–GFP$_{11}$ dimers. While the DNA conjugation sites on these pairs of split GFP are located on the same end of the GFP β-barrel, DNA conjugation sites may be located on opposite sides for other pairs of proteins. Therefore, we will express mutants of split GFP proteins with reactive handles for DNA functionalization at different locations (*i.e.*, N-terminus) and explore how the relative location of DNA functionalization affects protein–protein interaction thermodynamics within the dimers. Learning how different oligonucleotide designs affect the assembly thermodynamics of protein–DNA oligomers using a system with a simple fluorescent readout will lay the foundation for understanding how protein functions can be controlled using DNA–DNA interactions to disrupt and override protein–protein interactions. Overall, the lessons learned from this work will enable the design and synthesis of functional protein-based materials where protein functions are reversibly regulated using DNA.

Studying how to modulate the properties of fluorescent proteins is an important first step in understanding how to use DNA–DNA interactions to control protein function. However, controlling enzymatic reactions presents a significant and necessary challenge to address. Specifically, affecting enzymes through the introduction of reversible DNA–DNA interactions will allow us to regulate the chemical transformations of nearly any enzyme-based chemical syntheses (*e.g.*, cascade reactions). Additionally, spatially organizing enzymes within a DNA scaffold could lead to new reaction pathways where a variety of stimuli (*e.g.*, light, chemical) can be used to modulate complex syntheses. The model enzyme, split *Renilla* luciferase (RL), is an ideal candidate to begin studying how we can regulate enzyme functions using DNA because the split RL enzyme catalyzes the oxidation of coelenterazine to an excited state oxyluciferin product that then emits a photon, providing a simple *in situ* chemiluminescence readout. While separate

split RLs are nonfunctional, split RL fragments interact reversibly to restore chemiluminescence because enzymatic catalysis is restored.[215] First, we will express split RL fragments with a surface reactive amino acid (*i.e.*, cysteine or an unnatural amino acid that contains an azide) on either the solvent-exposed C- or N-terminus. Subsequently, we will functionalize the reactive amino acid on each split RL fragment with a single DNA strand and prepare dimers that contain one copy of each fragment. We hypothesize that the addition of a complementary template DNA strand to the split RL oligomer will lead to a rigid DNA duplex that spatially isolates each split RL fragment and prevents enzymatic catalysis (the "off" state). In contrast, after removal of the complementary template DNA strand using a strand displacement reaction, the split protein fragments will be separated by a flexible, single-stranded DNA region, leading to protein assembly and catalysis turn-on (the "on" state). The "on" and "off" states of enzymatic catalysis will be measured by bulk chemiluminescence on a plate reader. Gel electrophoresis and SEC will be used to determine conditions that favor intramolecular interactions over intermolecular interactions. While the distance-dependencies of split RL enzymatic catalysis[216] and other split enzymes are not well-understood,[217] such insight could be important for analytical techniques (*e.g.*, *in vitro* protein–protein interaction characterization) that rely on protein fragment-assisted complementation.[218] By controlling the length of the rigid, duplexed region between split RL fragments or varying the distance that spilt RL fragments are placed from each other along an oligonucleotide sequence, we can control the distance between these fragments. This will allow us to study the dependence of enzymatic activity on distance. By learning the thermodynamics of how DNA design affects protein–protein interactions, we can design systems to switch between protein "on" and "off" states, creating pathways to control new biomaterials. Overall, these fundamental concepts will be

important as we transition from studying model enzymes with defined and understood split protein designs to enzymes whose properties are less understood.

While the model systems of fluorescent proteins and RL enzymes have split protein analogs, the design of new split protein analogs of enzymes is nontrivial. Conversely, the majority of proteins assemble into oligomeric complexes where protein assembly dictates function.[219] Therefore, modulating the assembly of nonfunctional protein subunits assembly into functional protein oligomeric complexes is a more broadly applicable approach to control protein function. Using the tetrameric enzyme catalase, we will investigate how the function of a multimeric protein can be reversibly switched between "on" and "off" states. Catalase is a tetrameric protein with prosthetic heme groups that catalyze the degradation of hydrogen peroxide into water and oxygen.[220-221] While tetrameric catalase exhibits turnover rates that are orders of magnitude higher than the prosthetic group alone, single heme-containing catalase subunits show no turnover.[222] Therefore, we hypothesize that DNA–DNA interactions can be used to override catalase subunit interactions, dictating the assembly and disassembly of subunits and leading to controlled "on" and "off" states (Figure 5.2).
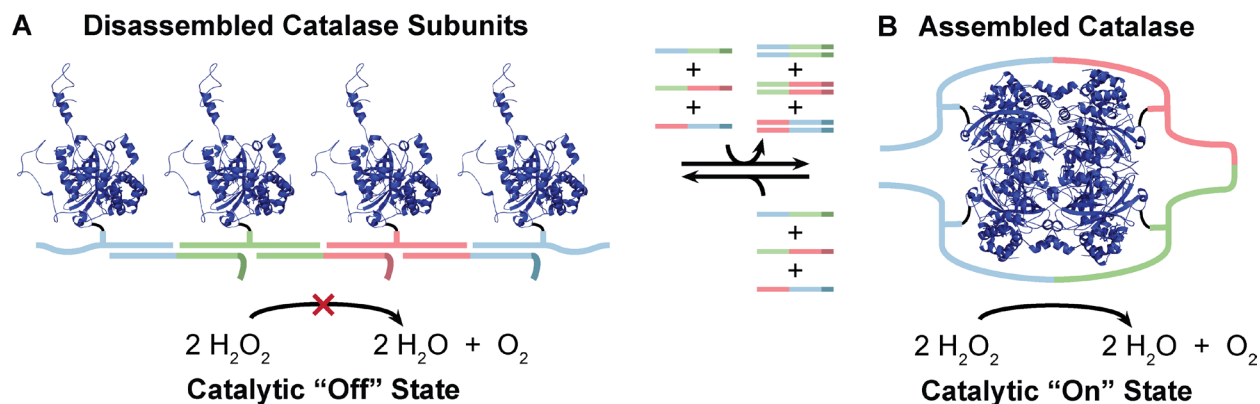
Figure 5.2. Regulating enzyme function using DNA. Oligomers containing four catalase subunits will be prepared. (A) When template DNA strands duplex with the oligomer, the catalase subunits will be spatially separated, resulting in a catalytic "off" state. (B) When the template DNA strands are removed via a toehold-mediated strand displacement reaction, the catalase subunits will assemble into tetrameric catalase, resulting in a catalytic "on" state.

To synthesize catalase oligomers, we will initially target the C-termini on tetrameric catalase for functionalization with DNA because the N-terminal region of catalase is involved in protein–protein interactions during the assembly of catalase subunits into tetramers. First, we will express catalase with a C-terminal cysteine residue. Next, we will disassemble the tetramer into single catalase subunits in basic buffers (pH > 10) and functionalize the C-terminal cysteine of the subunits with a single DNA strand. Alternatively, we could functionalize the C-terminal cysteines of tetrameric catalase with DNA strands and then disassemble the catalase into its subunits. Next, standard protein purification techniques (*i.e.*, SEC, anion exchange chromatography) will be used to isolate the single catalase subunits that are conjugated to DNA. After functionalizing the catalase termini with DNA, we will prepare molecular oligomers that contain four catalase subunits. DNA–DNA interactions will then be used to direct the assembly of catalase subunits into either an assembled tetrameric "on" state or a disassembled monomeric "off" state. Catalase assembly and its resulting restoration of enzymatic function will be characterized using commercially available colorimetric activity assays. In addition, we will use DNA–DNA interactions to sequentially

assemble catalase subunits into dimers, trimers, and/or tetramers to understand how oligomeric state affects catalytic activity and determine whether the assembly pathway affects function. Overall, by using catalase as a model system, we will discover how to override protein subunit assembly to control the functions of multimeric protein complexes using DNA.

Protein–protein interactions are complex and difficult to control, and therefore the design of materials that leverage the inherent functions of proteins remains challenging. Objective 1 will address this challenge by developing a general platform to understand and exploit protein–protein interactions and protein functions to realize reversible activity according to DNA inputs. Additionally, Objective 1 will overcome challenges in translating protein functions into materials by developing a method to oligomerize proteins with predefined sequence and architecture. Together, the proposed studies center on learning how to leverage DNA–DNA binding thermodynamics to override protein–protein interactions and protein assembly, leading to protein function that can be controlled between "on" and "off" states. Looking forward, this platform will enable independent DNA-mediated control over multiple steps in catalytic pathways to enable clean energy by mimicking and surpassing the efficiency of solar energy conversion into chemical fuel by photosystem proteins or nitrogen reduction to ammonia by nitrogenase proteins.

5.2.2. Protein–DNA Hydrogels that Exhibit Dissipative Assembly.

Hydrogels are a class of materials consisting of 3D crosslinked networks.[223] Traditionally, hydrogels rely on polymeric scaffolds (*e.g.,* polyethylene glycol, polyacrylamide*)* to form extended networks. However, the capabilities of synthetic polymers are limited relative to biological materials. In contrast, hydrogels contain protein–protein and protein–ligand interactions can exhibit dynamic and responsive properties, such as light and enzyme responsive hydrogel

network structure.[224] These novel biomaterials have proven to be transformative in fields such as sensing, therapeutics, and catalysis.[225-230] Conversely, DNA hydrogels are promising materials because their properties (*e.g.*, softness, pore size, stimuli-responsiveness) are highly programmable and dynamic.[231] Through DNA design, hydrogels can be synthesized with molecularly-defined network structures[232] that exhibit tunable and stimuli-responsive properties (*e.g.*, oligonucleotides, pH, and light- and temperature-triggered chemical changes).[233-236]

Efforts to combine the benefits of protein and DNA hydrogels have led to the preparation of dynamic peptide– and protein–DNA hydrogels; however, the peptide or protein is typically conjugated to a pre-existing DNA network and, resultantly, DNA predominantly defines hydrogel properties.[169, 237-239] We hypothesize that hydrogels with judiciously designed DNA backbones and protein–ligand/enzyme–substrate crosslinks will enable independent and reversible control over backbone rigidity and crosslink density (Figure 5.3), a feat that is difficult, if not impossible to obtain using any other synthetic method. We will also investigate how these materials can be designed to exhibit dissipative material properties (Figure 5.4, *e.g.*, modulus, toughness, elasticity, shear-thinning).
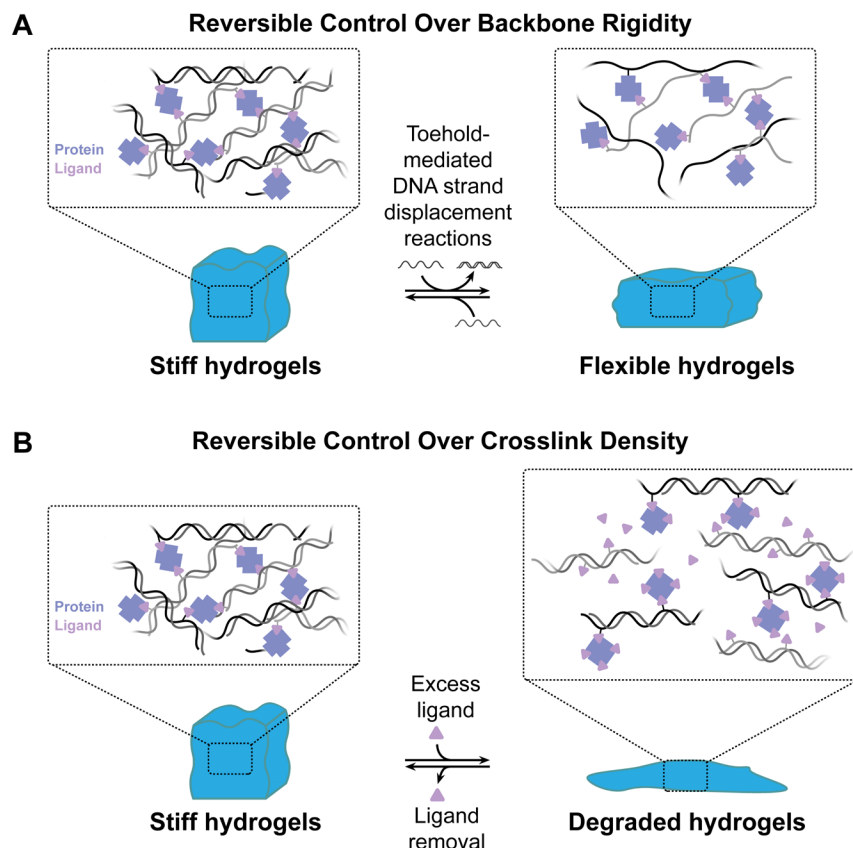
Figure 5.3. Synthesis of a dynamic protein–DNA hydrogel. Hydrogels will form by mixing proteins and ligand–DNA oligomers. (A) The backbone rigidity in these hydrogels can be reversibly modulated between stiff and flexible states via DNA toehold-mediated strand displacement reactions. These reactions will switch DNA between rigid double-stranded states and flexible single-stranded states. (B) The crosslink density in these hydrogels can be reversibly modulated by introducing and removing excess protein ligand.

The hypothesis will be tested with responsive protein–DNA hydrogels that comprised of DNA backbones and protein–ligand/enzyme–inhibitor crosslinking points. Hydrogels will be synthesized by mixing ligand/inhibitor–DNA oligomers with corresponding multivalent proteins. Specifically, we will study a strong protein–ligand interaction pair (streptavidin (SA) and biotin, where $K_a = 2.5 \times 10^{13}$ M^{-1}),[240] a weak protein–ligand interaction pair (concanavalin A (ConA) and mannose, where $K_a = 3.0 \times 10^4$ M^{-1}),[241] and an enzyme–inhibitor ligand pair (L-amino acid oxidase (AAO) and *m*-chlorobenzoate (Bz), where the inhibitor ligand can be displaced by substrate (*i.e.*,

an L-amino acid)) will be investigated as potential crosslinking points. DNA hairpin strands will be prepared using solid-phase DNA synthesis techniques and ligand/inhibitor modifications will be incorporated at specific locations in the oligonucleotide sequence. The ligand- and inhibitor ligand-containing DNA strands will be analyzed via MALDI-TOF MS to confirm the number of ligands per monomer. These ligand/inhibitor hairpins will be assembled into oligomers using DNA HCR, which will provide control over the number of ligand molecules per oligomer.[122] Subsequently, the ligand/inhibitor–DNA oligomers will be ligated with T4 DNA ligase, yielding oligomers where ligand molecules are conjugated to a single DNA strand. Overall, this oligomerization approach will allow us to modulate the stoichiometry of ligand and the relative number of different ligands (*e.g.*, the ratio of strong-binding biotin to weak-binding mannose) per DNA scaffold to determine how protein–ligand crosslink design affects the hydrogel rheological characteristics.

The dynamic characteristics of the hydrogels will be studied in response to different chemical stimuli that affect the DNA–DNA interactions, protein–ligand interactions, or both. First, we hypothesize that modulating the duplexing state of the DNA scaffolds will lead to changes in hydrogel backbone rigidity and macroscopic stiffness changes without affecting the crosslink density of the material. This property is difficult to achieve with traditional hydrogels (Figure 5.3.A). The stiffness of the hydrogel will be modified by performing toehold-mediated strand displacement reactions. During this transition, we expect that the hydrogel network will remain intact, but the stiffness (*i.e.*, magnitude of the storage modulus) will decrease because the rigidity and mobility of the DNA scaffold is changing. Furthermore, we will investigate the reversibility of this DNA-mediated transformation. In hydrogels with identical DNA designs, properties are

expected to be defined by the protein–ligand interaction affinity. Next, the hydrogels will be subjected to different chemical stimuli to induce structural transformations. For example, high concentrations of ligand (1–10 mM) should lead to a transition in the hydrogel from a stiff state to a degraded hydrogel state (Figure 5.3.B, *i.e.*, gel to sol transition), which can be quantitatively determined using rheological characterization and monitoring the tan(δ), the ratio between G″ and G′. We will study the impact of protein equilibrium states by varying the free ligand concentration (*e.g.*, 0.1 µM – 10 mM). We expect that hydrogels in equilibrium with increasing amounts of free ligand will have decreased crosslinking, which can be detected by monitoring gel stiffness. These experiments will allow us to understand how protein–ligand binding equilibrium impacts hydrogel properties. We will investigate the reversibility of this hydrogel transformation by characterizing hydrogel properties as we cycle between the addition and removal of chemical stimuli. Overall, through this research, we will establish synthetic routes to achieve protein–DNA hydrogels with designed and adaptable properties.

Studying the hydrogels containing SA–biotin and ConA–mannose crosslinks will provide the fundamental understanding necessary to pursue more complex, dissipative systems where enzymatic activity modulates hydrogel properties. Inhibitor–DNA oligomers will be synthesized with designs similar to those described above. Hydrogels formed by mixing multivalent enzymes with inhibitor–DNA oligomers through this approach are expected to be reconfigurable materials that can respond, transform, and adapt to chemical potentials. Specifically, the goal of this work will be to design hydrogels that display dissipative, self-regulating properties upon the introduction of specific chemical reagents and then, through the catalytic consumption of that reagent, restore the hydrogel to its initial state (Figure 5.4).
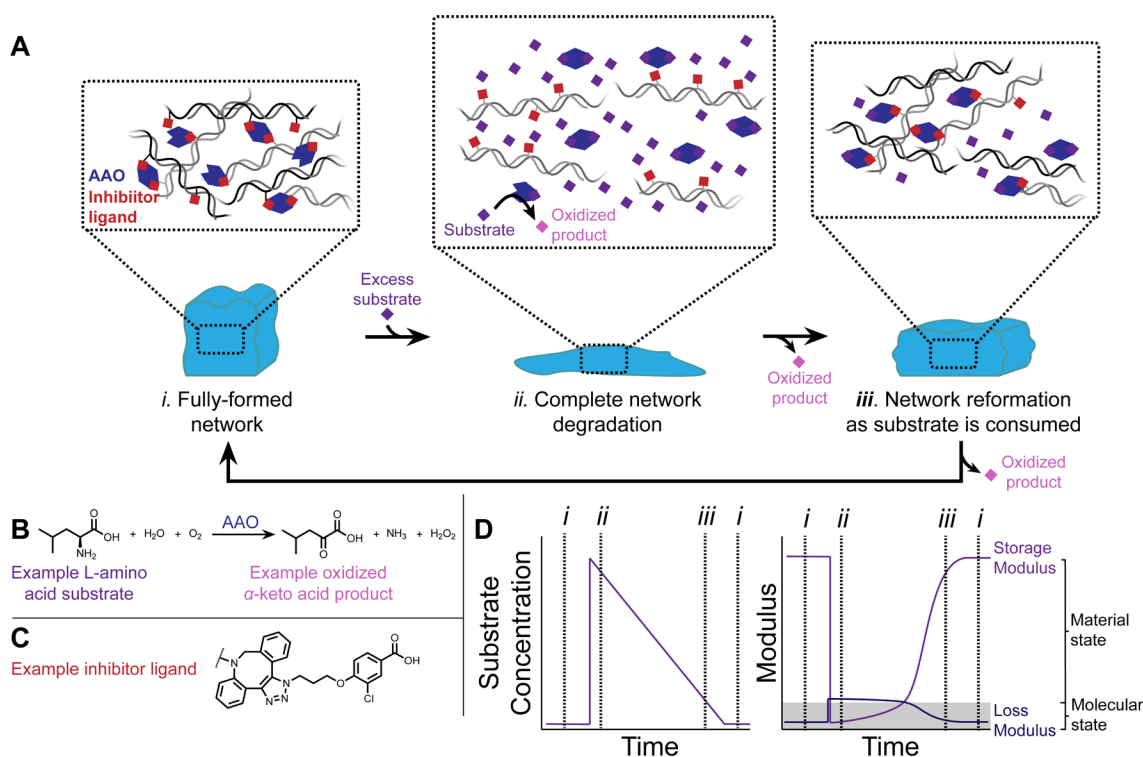
Figure 5.4. Dissipative assembly properties of protein–DNA hydrogels. (A) Hydrogels formed by supramolecular crosslinks of L-amino acid oxidase (AAO) and an inhibitor ligand (*i*) can be transiently reconfigured by introducing excess AAO substrate, which will degrade the network by saturating the active site of AAO (*ii*). As the substrate is consumed, the hydrogel network will reform (*iii*). (B) AAO catalyzes the oxidation of L-amino acids. (C) Benzoates will be conjugated to DNA to act as competitive inhibitors for AAO. (D) As substrate concentration is increased in the fully formed network (*i*), the storage modulus of the material will decrease upon network degradation (*ii*) forming a liquid. As the substrate is consumed and its concentration decreases, the storage modulus will increase as the hydrogel network reforms (*iii*).

To design dissipative networks, we will employ the dimeric L-amino acid oxidase (AAO) as a model enzyme. This enzyme catalyzes the reaction of L-amino acids (e.g., leucine) into oxidized α-keto acids and hydrogen peroxide (Figure 5.4.B).[242] However, the enzyme can be competitively inhibited using benzoates (Figure 5.4.C, Bz), which block the active site through a combination of hydrophobic and electrostatic interactions.[243] Hydrogels will be synthesized by mixing the AAO with the Bz–DNA oligomers developed in Objective 2.2 and characterized using rheological techniques to establish hydrogel properties in the absence of AAO substrates. Next,

amino acid substrates will be introduced to the system and the catalytic activity of AAO will be monitored by measuring the concentration of hydrogen peroxide generated using fluorescein bis(benzyl boronic ester) (FBBBE), a dye that becomes fluorescent upon reaction with hydrogen peroxide. We expect that the substrate, in large excess, will displace Bz, resulting in the degradation of AAO–Bz crosslinks, which will cause a gel-sol transition (Figure 5.4.D, *ii*). As the amino acid is consumed and the concentration of the substrate decreases, AAO–Bz crosslinks will reform. Re-establishment of these crosslinks will reform the hydrogel (Figure 5.4.D, *iii*) into a material with the same properties as the original network (Figure 5.4.D, *i*). We will analyze the viscoelastic property changes of the hydrogel during enzymatic reactions using time sweep rheological characterization to quantitatively determine how the hydrogels change with substrate conversion. If measuring gel properties upon the introduction of a chemical stimulus proves to be too challenging on an oscillatory rheometer, we plan to utilize in situ techniques such as particle tracking microrheology which tracks the Brownian motion of beads through complex, viscoelastic materials.[244-245] In microrheology, tracking algorithms are used to trace the motion of particles and mathematical relationships are used to extract values such as viscosity, storage modulus and loss modulus. By measuring the storage modulus over time and across cycles, we can correlate changes in hydrogel properties to enzyme turnover rates obtained in solution using the FBBBE assay. By tuning when and how substrate is added to the hydrogel, we can develop quantitative relationships between the amount of substrate added and the rate at which hydrogel properties change and recover. Furthermore, we will add substrate to specific regions of a hydrogel to obtain materials with a spatial gradient of dynamic properties. Overall, we will learn how enzymatic reactivity can be exploited to synthesize hydrogels that display dissipative and self-regulating properties

according to chemical stimuli. This insight can potentially be translated and applied to many of the known enzyme–substrate pairs with known competitive-binding inhibitors. Realizing transient hydrogel property transformations will be a significant step towards synthesizing materials that exhibit dissipative material properties.

This proposed research will generate a modular approach to synthesize protein–DNA hydrogels where every aspect of the material plays an integral and reversible role in defining material properties. By leveraging the programmability of DNA and the wide range of protein–ligand binding affinities, these materials will enable the independent study of how DNA hybridization interactions in the hydrogel backbone and protein–ligand interactions in crosslinks affect hydrogel materials properties. These hydrogels will be able to undergo transient and reversible changes between many states (i.e., stiff hydrogels to flexible hydrogels to degraded hydrogels) according to stimuli inputs or local chemical fluxes. Such materials could have important applications in sensing, therapeutics, and catalysis.

REFERENCES

1. Kessel, A.; Ben-Tal, N., *Introduction to Proteins: Structure, Function, and Motion*, 2nd ed.; Chapman and Hall/CRC, New York, **2018**.

2. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.A.A.; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021**, *596*, 583–589.

3. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; Millán, C.; Park, H.; Adams, C.; Glassman, C.R.; DeGiovanni, A.; Pereira, J.H.; Rodrigues, A.V.; Dijk, A.A.v.; Ebrecht, A.C.; Opperman, D.J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M.K.; Dalwadi, U.; Yip, C.K.; Burke, J.E.; Garcia, K.C.; Grishin, N.V.; Adams, P.D.; Read, R.J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876.

4. Schreiber, G., Protein–Protein Interaction Interfaces and Their Functional Implications. In *Protein–Protein Interaction Regulators*, 2020; pp 1–24.

5. Jordan, P.; Fromme, P.; Witt, H.T.; Klukas, O.; Saenger, W.; Krauß, N. Three-Dimensional Structure of Cyanobacterial Photosystem I at 2.5 Å Resolution. *Nature* **2001**, *411*, 909–917.

6. Gotti, C.; Sensini, A.; Zucchelli, A.; Carloni, R.; Focarete, M.L. Hierarchical Fibrous Structures for Muscle-Inspired Soft-Actuators: A Review. *Appl. Mater. Today* **2020**, *20*, 100772.

7. Sleytr, U.B.; Schuster, B.; Egelseer, E.-M.; Pum, D. S-Layers: Principles and Applications. *FEMS Microbiol. Rev.* **2014**, *38*, 823–864.

8. Shoulders, M.D.; Raines, R.T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958.

9. Ahnert, S.E.; Marsh, J.A.; Hernández, H.; Robinson, C.V.; Teichmann, S.A. Principles of Assembly Reveal a Periodic Table of Protein Complexes. *Science* **2015**, *350*, aaa2245.

10. Marsh, J.A.; Teichmann, S.A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem. 84*, 551–575.

11. Pieters, B.J.G.E.; van Eldijk, M.B.; Nolte, R.J.M.; Mecinović, J. Natural Supramolecular Protein Assemblies. *Chem. Soc. Rev.* **2016**, *45*, 24–39.

12. Nivina, A.; Yuet, K.P.; Hsu, J.; Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **2019**, *119*, 12524–12547.

13. Bushnell, D.A.; Kornberg, R.D. Complete, 12-Subunit Rna Polymerase Ii at 4.1-Å Resolution: Implications for the Initiation of Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6969–6973.

14. Cate, J.H.; Yusupov, M.M.; Yusupova, G.Z.; Earnest, T.N.; Noller, H.F. X-Ray Crystal Structures of 70s Ribosome Functional Complexes. *Science* **1999**, *285*, 2095–2104.

15. Perkins, S.J.; Nealis, A.S.; Sutton, B.J.; Feinstein, A. Solution Structure of Human and Mouse Immunoglobulin M by Synchrotron X-Ray Scattering and Molecular Graphics Modelling: A Possible Mechanism for Complement Activation. *J. Mol. Biol.* **1991**, *221*, 1345–1366.

16. Cherezov, V.; Rosenbaum, D.M.; Hanson, M.A.; Rasmussen, S.G.F.; Thian, F.S.; Kobilka, T.S.; Choi, H.-J.; Kuhn, P.; Weis, W.I.; Kobilka, B.K.; Stevens, R.C. High-Resolution Crystal Structure of an Engineered Human B2-Adrenergic G Protein–Coupled Receptor. *Science* **2007**, *318*, 1258-1265.

17. Gorina, S.; Pavletich, N.P. Structure of the P53 Tumor Suppressor Bound to the Ankyrin and Sh3 Domains of 53bp2. *Science* **1996**, *274*, 1001–1005.

18. Winegar, P.H.; Hayes, O.G.; McMillan, J.R.; Figg, C.A.; Focia, P.J.; Mirkin, C.A. DNA-Directed Protein Packing within Single Crystals. *Chem* **2020**, *6*, 1007–1017.

19.   Juers, D.H.; Heightman, T.D.; Vasella, A.; McCarter, J.D.; Mackenzie, L.; Withers, S.G.; Matthews, B.W. A Structural View of the Action of Escherichia Coli (Lacz) B-Galactosidase. *Biochemistry* **2001**, *40*, 14781–14794.

20.   Cong, Y.; Topf, M.; Sali, A.; Matsudaira, P.; Dougherty, M.; Chiu, W.; Schmid, M.F. Crystallographic Conformers of Actin in a Biologically Active Bundle of Filaments. *J. Mol. Biol.* **2008**, *375*, 331–336.

21.   Bharat, T.A.M.; Kureisaite-Ciziene, D.; Hardy, G.G.; Yu, E.W.; Devant, J.M.; Hagen, W.J.H.; Brun, Y.V.; Briggs, J.A.G.; Löwe, J. Structure of the Hexagonal Surface Layer on Caulobacter Crescentus Cells. *Nat. Microbiol.* **2017**, *2*, 17059.

22.   Lončar, N.; Rozeboom, H.J.; Franken, L.E.; Stuart, M.C.A.; Fraaije, M.W. Structure of a Robust Bacterial Protein Cage and Its Application as a Versatile Biocatalytic Platform through Enzyme Encapsulation. *Biochem. Biophys. Res. Commun.* **2020**, *529*, 548–553.

23.   Zhu, J.; Avakyan, N.; Kakkis, A.; Hoffnagle, A.M.; Han, K.; Li, Y.; Zhang, Z.; Choi, T.S.; Na, Y.; Yu, C.-J.; Tezcan, F.A. Protein Assembly by Design. *Chem. Rev.* **2021**, *121*, 13701–13796.

24.   Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem. Rev.* **2016**, *116*, 13571–13632.

25.   King, N.P.; Lai, Y.-T. Practical Approaches to Designing Novel Protein Assemblies. *Curr. Opin. Struct. Biol.* **2013**, *23*, 632–638.

26.   Permana, D.; Putra, H.E.; Djaenudin, D. Designed Protein Multimerization and Polymerization for Functionalization of Proteins. *Biotechnol. Lett.* **2022**, *44*, 341–365.

27.   Churchfield, L.A.; Tezcan, F.A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **2019**, *52*, 345–355.

28.   Golub, E.; Subramanian, R.H.; Esselborn, J.; Alberstein, R.G.; Bailey, J.B.; Chiong, J.A.; Yan, X.; Booth, T.; Baker, T.S.; Tezcan, F.A. Constructing Protein Polyhedra Via Orthogonal Chemical Interactions. *Nature* **2020**, *578*, 172–176.

29. Zhang, L.; Bailey, J.B.; Subramanian, R.H.; Groisman, A.; Tezcan, F.A. Hyperexpandable, Self-Healing Macromolecular Crystals with Integrated Polymer Networks. *Nature* **2018**, *557*, 86–91.

30. Lawson, D.M.; Artymiuk, P.J.; Yewdall, S.J.; Smith, J.M.A.; Livingstone, J.C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; Thomas, C.D.; Shaw, W.V.; Harrison, P.M. Solving the Structure of Human H Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* **1991**, *349*, 541–544.

31. Brodin, J.D.; Ambroggio, X.I.; Tang, C.; Parent, K.N.; Baker, T.S.; Tezcan, F.A. Metal-Directed, Chemically Tunable Assembly of One-, Two- and Three-Dimensional Crystalline Protein Arrays. *Nat. Chem.* **2012**, *4*, 375–382.

32. Sontz, P.A.; Bailey, J.B.; Ahn, S.; Tezcan, F.A. A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3d Protein Crystals. *J. Am. Chem. Soc.* **2015**, *137*, 11598–11601.

33. Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R.D.; Tezcan, F.A. Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites. *Angew. Chem. Int. Ed.* **2020**, *59*, 21940–21944.

34. Ringler, P.; Schulz, G.E. Self-Assembly of Proteins into Designed Networks. **2003**, *302*, 106-109.

35. Engilberge, S.; Rennie, M.L.; Dumont, E.; Crowley, P.B. Tuning Protein Frameworks Via Auxiliary Supramolecular Interactions. *ACS Nano* **2019**, *13*, 10343–10350.

36. Alex, J.M.; Rennie, M.L.; Volpi, S.; Sansone, F.; Casnati, A.; Crowley, P.B. Phosphonated Calixarene as a "Molecular Glue" for Protein Crystallization. *Cryst. Growth Des.* **2018**, *18*, 2467–2473.

37. Sakai, F.; Yang, G.; Weiss, M.S.; Liu, Y.; Chen, G.; Jiang, M. Protein Crystalline Frameworks with Controllable Interpenetration Directed by Dual Supramolecular Interactions. *Nat. Commun.* **2014**, *5*, 4634.

38. Rennie, M.L.; Fox, G.C.; Pérez, J.; Crowley, P.B. Auto-Regulated Protein Assembly on a Supramolecular Scaffold. *Angew. Chem. Int. Ed.* **2018**, *130*, 13960–13965.

39.   Dotan, N.; Arad, D.; Frolow, F.; Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals. *Angew. Chem. Int. Ed.* **1999**, *38*, 2363–2366.

40.   Kitagishi, H.; Oohora, K.; Yamaguchi, H.; Sato, H.; Matsuo, T.; Harada, A.; Hayashi, T. Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme−Heme Pocket Interactions. *J. Am. Chem. Soc.* **2007**, *129*, 10326-10327.

41.   Cohen-Hadar, N.; Lagziel-Simis, S.; Wine, Y.; Frolow, F.; Freeman, A. Re-Structuring Protein Crystals Porosity for Biotemplating by Chemical Modification of Lysine Residues. *Biotechnol. Bioeng.* **2011**, *108*, 1–11.

42.   Simon, A.J.; Zhou, Y.; Ramasubramani, V.; Glaser, J.; Pothukuchy, A.; Gollihar, J.; Gerberich, J.C.; Leggere, J.C.; Morrow, B.R.; Jung, C.; Glotzer, S.C.; Taylor, D.W.; Ellington, A.D. Supercharging Enables Organized Assembly of Synthetic Biomolecules. *Nat. Chem.* **2019**, *11*, 204–212.

43.   Künzle, M.; Eckert, T.; Beck, T. Binary Protein Crystals for the Assembly of Inorganic Nanoparticle Superlattices. *J. Am. Chem. Soc.* **2016**, *138*, 12731–12734.

44.   McGovern, R.E.; McCarthy, A.A.; Crowley, P.B. Protein Assembly Mediated by Sulfonatocalix[4]Arene. *Chem. Commun.* **2014**, *50*, 10412–10415.

45.   Bai, Y.; Luo, Q.; Liu, J. Protein Self-Assembly Via Supramolecular Strategies. *Chem. Soc. Rev.* **2016**, *45*, 2756–2767.

46.   Hou, C.; Li, J.; Zhao, L.; Zhang, W.; Luo, Q.; Dong, Z.; Xu, J.; Liu, J. Construction of Protein Nanowires through Cucurbit[8]Uril-Based Highly Specific Host–Guest Interactions: An Approach to the Assembly of Functional Proteins. **2013**, *52*, 5590-5593.

47.   Yamada, H.; Tamada, T.; Kosaka, M.; Miyata, K.; Fujiki, S.; Tano, M.; Moriya, M.; Yamanishi, M.; Honjo, E.; Tada, H.; Ino, T.; Yamaguchi, H.; Futami, J.; Seno, M.; Nomoto, T.; Hirata, T.; Yoshimura, M.; Kuroki, R. 'Crystal Lattice Engineering,' an Approach to Engineer Protein Crystal Contacts by Creating Intermolecular Symmetry: Crystallization and Structure Determination of a Mutant Human Rnase 1 with a Hydrophobic Interface of Leucines. *Protein Sci.* **2007**, *16*, 1389–1397.

48.   Cattani, G.; Vogeley, L.; Crowley, P.B. Structure of a Pegylated Protein Reveals a Highly Porous Double-Helical Assembly. *Nat. Chem.* **2015**, *7*, 823–828.

49. King, N.P.; Bale, J.B.; Sheffler, W.; McNamara, D.E.; Gonen, S.; Gonen, T.; Yeates, T.O.; Baker, D. Accurate Design of Co-Assembling Multi-Component Protein Nanomaterials. *Nature* **2014**, *510*, 103–108.

50. Brunette, T.J.; Parmeggiani, F.; Huang, P.-S.; Bhabha, G.; Ekiert, D.C.; Tsutakawa, S.E.; Hura, G.L.; Tainer, J.A.; Baker, D. Exploring the Repeat Protein Universe through Computational Protein Design. *Nature* **2015**, *528*, 580–584.

51. Doyle, L.; Hallinan, J.; Bolduc, J.; Parmeggiani, F.; Baker, D.; Stoddard, B.L.; Bradley, P. Rational Design of A-Helical Tandem Repeat Proteins with Closed Architectures. *Nature* **2015**, *528*, 585–588.

52. Laganowsky, A.; Zhao, M.; Soriaga, A.B.; Sawaya, M.R.; Cascio, D.; Yeates, T.O. An Approach to Crystallizing Proteins by Metal-Mediated Synthetic Symmetrization. *Protein Sci.* **2011**, *20*, 1876–1890.

53. Lai, Y.-T.; Reading, E.; Hura, G.L.; Tsai, K.-L.; Laganowsky, A.; Asturias, F.J.; Tainer, J.A.; Robinson, C.V.; Yeates, T.O. Structure of a Designed Protein Cage That Self-Assembles into a Highly Porous Cube. *Nat. Chem.* **2014**, *6*, 1065–1071.

54. Divine, R.; Dang, H.V.; Ueda, G.; Fallas, J.A.; Vulovic, I.; Sheffler, W.; Saini, S.; Zhao, Y.T.; Raj, I.X.; Morawski, P.A.; Jennewein, M.F.; Homad, L.J.; Wan, Y.-H.; Tooley, M.R.; Seeger, F.; Etemadi, A.; Fahning, M.L.; Lazarovits, J.; Roederer, A.; Walls, A.C.; Stewart, L.; Mazloomi, M.; King, N.P.; Campbell, D.J.; McGuire, A.T.; Stamatatos, L.; Ruohola-Baker, H.; Mathieu, J.; Veesler, D.; Baker, D. Designed Proteins Assemble Antibodies into Modular Nanocages. *Science* **2021**, *372*, eabd9994.

55. Boyken Scott, E.; Chen, Z.; Groves, B.; Langan Robert, A.; Oberdorfer, G.; Ford, A.; Gilmore Jason, M.; Xu, C.; DiMaio, F.; Pereira Jose, H.; Sankaran, B.; Seelig, G.; Zwart Peter, H.; Baker, D. De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network–Mediated Specificity. *Science* **2016**, *352*, 680–687.

56. Bale, J.B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T.O.; Gonen, T.; King, N.P.; Baker, D. Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. *Science* **2016**, *353*, 389–394.

57. Chen, Z.; Johnson, M.C.; Chen, J.; Bick, M.J.; Boyken, S.E.; Lin, B.; De Yoreo, J.J.; Kollman, J.M.; Baker, D.; DiMaio, F. Self-Assembling 2d Arrays with De Novo Protein Building Blocks. *J. Am. Chem. Soc.* **2019**, *141*, 8891–8895.

58.   Sahtoe, D.D.; Praetorius, F.; Courbet, A.; Hsia, Y.; Wicky, B.I.M.; Edman, N.I.; Miller, L.M.; Timmermans, B.J.R.; Decarreau, J.; Morris, H.M.; Kang, A.; Bera, A.K.; Baker, D. Reconfigurable Asymmetric Protein Assemblies through Implicit Negative Design. *Science* **2022**, *375*, eabj7662.

59.   Mogilevsky, C.S.; Lobba, M.J.; Brauer, D.D.; Marmelstein, A.M.; Maza, J.C.; Gleason, J.M.; Doudna, J.A.; Francis, M.B. Synthesis of Multi-Protein Complexes through Charge-Directed Sequential Activation of Tyrosine Residues. *J. Am. Chem. Soc.* **2021**, *143*, 13538–13547.

60.   Koniev, O.; Wagner, A. Developments and Recent Advancements in the Field of Endogenous Amino Acid Selective Bond Forming Reactions for Bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495–5551.

61.   Griffith, B.R.; Allen, B.L.; Rapraeger, A.C.; Kiessling, L.L. A Polymer Scaffold for Protein Oligomerization. *J. Am. Chem. Soc.* **2004**, *126*, 1608–1609.

62.   Broyer, R.M.; Grover, G.N.; Maynard, H.D. Emerging Synthetic Approaches for Protein–Polymer Conjugations. *Chem. Commun.* **2011**, *47*, 2212–2226.

63.   Tao, L.; Kaddis, C.S.; Loo, R.R.O.; Grover, G.N.; Loo, J.A.; Maynard, H.D. Synthesis of Maleimide-End-Functionalized Star Polymers and Multimeric Protein−Polymer Conjugates. *Macromolecules* **2009**, *42*, 8028–8033.

64.   Tao, L.; Kaddis, C.S.; Ogorzalek Loo, R.R.; Grover, G.N.; Loo, J.A.; Maynard, H.D. Synthetic Approach to Homodimeric Protein–Polymer Conjugates. *Chem. Commun.* **2009**, 2148–2150.

65.   Heredia, K.L.; Grover, G.N.; Tao, L.; Maynard, H.D. Synthesis of Heterotelechelic Polymers for Conjugation of Two Different Proteins. *Macromolecules* **2009**, *42*, 2360–2367.

66.   Sletten, E.M.; Bertozzi, C.R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998.

67.   Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

68. Devaraj, N.K. The Future of Bioorthogonal Chemistry. *ACS Cent. Sci.* **2018**, *4*, 952–959.

69. Paloni, J.M.; Miller, E.A.; Sikes, H.D.; Olsen, B.D. Improved Ordering in Low Molecular Weight Protein–Polymer Conjugates through Oligomerization of the Protein Block. *Biomacromolecules* **2018**, *19*, 3814–3824.

70. Structural Genomics Consortium, A.e.F.d.M.; Biologiques, B.S.G.C., China Structural Genomics; Consortium, I.C.f.S.a.F.I., Israel Structural; Proteomics Center, J.C.f.S.G., Midwest Center for; Structural Genomics, N.Y.S.G.R.C.f.S.; Genomics, N.S.G.C., Oxford Protein; Production Facility, P.S.P.F., Max Delbrück Center for; Molecular Medicine, R.S.G.P.I., and SPINE2-; Complexes Protein Production and Purification. *Nat. Methods* **2008**, *5*, 135–146.

71. Lapenta, F.; Aupič, J.; Strmšek, Ž.; Jerala, R. Coiled Coil Protein Origami: From Modular Design Principles Towards Biotechnological Applications. *Chem. Soc. Rev.* **2018**, *47*, 3530–3542.

72. Halin, C.; Gafner, V.; Villani, M.E.; Borsi, L.; Berndt, A.; Kosmehl, H.; Zardi, L.; Neri, D. Synergistic Therapeutic Effects of a Tumor Targeting Antibody Fragment, Fused to Interleukin 12 and to Tumor Necrosis Factor A. *Cancer Res.* **2003**, *63*, 3202–3210.

73. Kipriyanov, S.M.; Breitling, F.; Little, M.; Dübel, S. Single-Chain Antibody Streptavidin Fusions: Tetrameric Bifunctional Scfv-Complexes with Biotin Binding Activity and Enhanced Affinity to Antigen. *Hum. Antibodies* **1995**, *6*, 93–101.

74. Zhang, J.; Tanha, J.; Hirama, T.; Khieu, N.H.; To, R.; Tong-Sevinc, H.; Stone, E.; Brisson, J.-R.; Roger MacKenzie, C. Pentamerization of Single-Domain Antibodies from Phage Libraries: A Novel Strategy for the Rapid Generation of High-Avidity Antibody Reagents. *J. Mol. Biol.* **2004**, *335*, 49–56.

75. Kim, Y.E.; Kim, Y.-n.; Kim, J.A.; Kim, H.M.; Jung, Y. Green Fluorescent Protein Nanopolygons as Monodisperse Supramolecular Assemblies of Functional Proteins with Defined Valency. *Nat. Commun.* **2015**, *6*, 7134.

76. Zakeri, B.; Fierer, J.O.; Celik, E.; Chittock, E.C.; Schwarz-Linek, U.; Moy, V.T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E690–E697.

77.   Veggiani, G.; Nakamura, T.; Brenner, M.D.; Gayet, R.V.; Yan, J.; Robinson, C.V.; Howarth, M. Programmable Polyproteams Built Using Twin Peptide Superglues. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1202–1207.

78.   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.

79.   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.

80.   Hodneland, C.D.; Lee, Y.-S.; Min, D.-H.; Mrksich, M. Selective Immobilization of Proteins to Self-Assembled Monolayers Presenting Active Site-Directed Capture Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5048–5052.

81.   Modica, J.A.; Iderzorig, T.; Mrksich, M. Design and Synthesis of Megamolecule Mimics of a Therapeutic Antibody. *J. Am. Chem. Soc.* **2020**, *142*, 13657–13661.

82.   Modica, J.A.; Skarpathiotis, S.; Mrksich, M. Modular Assembly of Protein Building Blocks to Create Precisely Defined Megamolecules. *ChemBioChem* **2012**, *13*, 2331–2334.

83.   Modica, J.A.; Lin, Y.; Mrksich, M. Synthesis of Cyclic Megamolecules. *J. Am. Chem. Soc.* **2018**, *140*, 6391–6399.

84.   Kimmel, B.R.; Modica, J.A.; Parker, K.; Dravid, V.; Mrksich, M. Solid-Phase Synthesis of Megamolecules. *J. Am. Chem. Soc.* **2020**, *142*, 4534–4538.

85.   Conibear, A.C. Deciphering Protein Post-Translational Modifications Using Chemical Biology Tools. *Nat. Rev. Chem.* **2020**, *4*, 674–695.

86.   de Marco, A. Strategies for Successful Recombinant Expression of Disulfide Bond-Dependent Proteins in Escherichia Coli. *Microb. Cell Fact.* **2009**, *8*, 26.

87. Rosano, G.L.; Ceccarelli, E.A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, *5*, 411–421.

88. Singh, A.; Upadhyay, V.; Upadhyay, A.K.; Singh, S.M.; Panda, A.K. Protein Recovery from Inclusion Bodies of Escherichia Coli Using Mild Solubilization Process. *Microb. Cell Fact.* **2015**, *14*, 41.

89. Cooper, G.M.; Hausman, R.E.; Hausman, R.E., *The Cell: A Molecular Approach*, ASM press Washington, DC, **2007**; Vol. 4.

90. Watson, J.D.; Crick, F.H. In *The Structure of DNA*, Cold Spring Harbor Symp. Quant. Biol., Cold Spring Harbor Laboratory Press: 1953; pp 123–131.

91. John SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.

92. Day, H.A.; Pavlou, P.; Waller, Z.A.E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.

93. Lipps, H.J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends Cell Biol.* **2009**, *19*, 414–422.

94. Letsinger, R.L.; Mahadevan, V. Oligonucleotide Synthesis on a Polymer Support. *J. Am. Chem. Soc.* **1965**, *87*, 3526–3527.

95. Caruthers, M.H.; Barone, A.D.; Beaucage, S.L.; Dodds, D.R.; Fisher, E.F.; McBride, L.J.; Matteucci, M.; Stabinsky, Z.; Tang, J.Y., Chemical Synthesis of Deoxyoligonucleotides by the Phosphoramidite Method. In *Methods Enzymol.*, Academic Press: 1987; Vol. 154, pp 287–313.

96. Saiki, R.K.; Scharf, S.; Faloona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A.; Arnheim, N. Enzymatic Amplification of B-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **1985**, *230*, 1350–1354.

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

98. Kallenbach, N.R.; Ma, R.-I.; Seeman, N.C. An Immobile Nucleic Acid Junction Constructed from Oligonucleotides. *Nature* **1983**, *305*, 829–831.

99. Zheng, J.; Birktoft, J.J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P.E.; Ginell, S.L.; Mao, C.; Seeman, N.C. From Molecular to Macroscopic Via the Rational Design of a Self-Assembled 3d DNA Crystal. *Nature* **2009**, *461*, 74.

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

101. Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, *117*, 12584–12640.

102. Alivisatos, A.P.; Johnsson, K.P.; Peng, X.; Wilson, T.E.; Loweth, C.J.; Bruchez, M.P.; Schultz, P.G. Organization of 'Nanocrystal Molecules' Using DNA. *Nature* **1996**, *382*, 609–611.

103. Mirkin, C.A.; Letsinger, R.L.; Mucic, R.C.; Storhoff, J.J. A DNA-Based Method for Rationally Assembling Nanoparticles into Macroscopic Materials. *Nature* **1996**, *382*, 607–609.

104. Laramy, C.R.; O'Brien, M.N.; Mirkin, C.A. Crystal Engineering with DNA. *Nat. Rev. Mater.* **2019**, *4*, 201–224.

105. Seeman, N.C.; Sleiman, H.F. DNA Nanotechnology. *Nat. Rev. Mater.* **2017**, *3*, 17068.

106. McMillan, J.R.; Hayes, O.G.; Winegar, P.H.; Mirkin, C.A. Protein Materials Engineering with DNA. *Acc. Chem. Res.* **2019**, *52*, 1939–1948.

107. Zhang, F.; Nangreave, J.; Liu, Y.; Yan, H. Structural DNA Nanotechnology: State of the Art and Future Perspective. *J. Am. Chem. Soc.* **2014**, *136*, 11198–11211.

108. Jones, M.R.; Seeman, N.C.; Mirkin, C.A. Nanomaterials. Programmable Materials and the Nature of the DNA Bond. *Science* **2015**, *347*, 1260901.

109. Park, S.Y.; Lytton-Jean, A.K.R.; Lee, B.; Weigand, S.; Schatz, G.C.; Mirkin, C.A. DNA-Programmable Nanoparticle Crystallization. *Nature* **2008**, *451*, 553–556.

110. Macfarlane, R.J.; Lee, B.; Jones, M.R.; Harris, N.; Schatz, G.C.; Mirkin, C.A. Nanoparticle Superlattice Engineering with DNA. *Science* **2011**, *334*, 204–208.

111. Sunasee, R.; Narain, R., Covalent and Noncovalent Bioconjugation Strategies. In *Chemistry of Bioconjugates*, Narain, R., Ed. John Wiley & Sons, Inc.: 2014; pp 1–75.

112. 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.

113. Trads, J.B.; Tørring, T.; Gothelf, K.V. Site-Selective Conjugation of Native Proteins with DNA. *Acc. Chem. Res.* **2017**, *50*, 1367–1374.

114. Niemeyer, C.M.; Sano, T.; Smith, C.L.; Cantor, C.R. Oligonucleotide-Directed Self-Assembly of Proteins: Semisynthetic DNA—Streptavidin Hybrid Molecules as Connectors for the Generation of Macroscopic Arrays and the Construction of Supramolecular Bioconjugates. *Nucleic Acids Res.* **1994**, *22*, 5530–5539.

115. Corey, D.; Schultz, P. Generation of a Hybrid Sequence-Specific Single-Stranded Deoxyribonuclease. *Science* **1987**, *238*, 1401–1403.

116. Nakajima, N.; Ikada, Y. Mechanism of Amide Formation by Carbodiimide for Bioconjugation in Aqueous Media. *Bioconjugate Chem.* **1995**, *6*, 123–130.

117. Moura, A.; Savageau, M.A.; Alves, R. Relative Amino Acid Composition Signatures of Organisms and Environments. *PLoS One* **2013**, *8*, e77319.

118. Noren, C.J.; Anthony-Cahill, S.J.; Griffith, M.C.; Schultz, P.G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **1989**, *244*, 182–188.

119. McMillan, J.R.; Mirkin, C.A. DNA-Functionalized, Bivalent Proteins. *J. Am. Chem. Soc.* **2018**, *140*, 6776–6779.

120. Kashiwagi, D.; Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. Protein Nanotube Selectively Cleavable with DNA: Supramolecular Polymerization of "DNA-Appended Molecular Chaperones". *J. Am. Chem. Soc.* **2018**, *140*, 26–29.

121. McMillan, J.R.; Hayes, O.G.; Remis, J.P.; Mirkin, C.A. Programming Protein Polymerization with DNA. *J. Am. Chem. Soc.* **2018**, *140*, 15950–15956.

122. Figg, C.A.; Winegar, P.H.; Hayes, O.G.; Mirkin, C.A. Controlling the DNA Hybridization Chain Reaction. *J. Am. Chem. Soc.* **2020**, *142*, 8596–8601.

123. Coyle, M.P.; Xu, Q.; Chiang, S.; Francis, M.B.; Groves, J.T. DNA-Mediated Assembly of Protein Heterodimers on Membrane Surfaces. *J. Am. Chem. Soc.* **2013**, *135*, 5012–5016.

124. 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.

125. Mou, Y.; Yu, J.-Y.; Wannier, T.M.; Guo, C.-L.; Mayo, S.L. Computational Design of Co-Assembling Protein–DNA Nanowires. *Nature* **2015**, *525*, 230–233.

126. Lim, S.; Kim, J.; Kim, Y.; Xu, D.; Clark, D.S. Crispr/Cas-Directed Programmable Assembly of Multi-Enzyme Complexes. *Chem. Commun.* **2020**, *56*, 4950–4953.

127. Wilner, O.I.; Shimron, S.; Weizmann, Y.; Wang, Z.-G.; Willner, I. Self-Assembly of Enzymes on DNA Scaffolds: En Route to Biocatalytic Cascades and the Synthesis of Metallic Nanowires. *Nano Lett.* **2009**, *9*, 2040–2043.

128. Cheglakov, Z.; Weizmann, Y.; Braunschweig, A.B.; Wilner, O.I.; Willner, I. Increasing the Complexity of Periodic Protein Nanostructures by the Rolling-Circle-Amplified Synthesis of Aptamers. *Angew. Chem. Int. Ed.* **2008**, *47*, 126–130.

129. Erkelenz, M.; Kuo, C.-H.; Niemeyer, C.M. DNA-Mediated Assembly of Cytochrome P450 Bm3 Subdomains. *J. Am. Chem. Soc.* **2011**, *133*, 16111–16118.

130. Marczynke, M.; Gröger, K.; Seitz, O. Selective Binders of the Tandem Src Homology 2 Domains in Syk and Zap70 Protein Kinases by DNA-Programmed Spatial Screening. *Bioconj. Chem.* **2017**, *28*, 2384–2392.

131. Setyawati, M.I.; Kutty, R.V.; Leong, D.T. DNA Nanostructures Carrying Stoichiometrically Definable Antibodies. *Small* **2016**, *12*, 5601–5611.

132. Kazane, S.A.; Axup, J.Y.; Kim, C.H.; Ciobanu, M.; Wold, E.D.; Barluenga, S.; Hutchins, B.A.; Schultz, P.G.; Winssinger, N.; Smider, V.V. Self-Assembled Antibody Multimers through Peptide Nucleic Acid Conjugation. *J. Am. Chem. Soc.* **2013**, *135*, 340–346.

133. Meyer, R.; Niemeyer, C.M. Orthogonal Protein Decoration of DNA Nanostructures. *Small* **2011**, *7*, 3211–3218.

134. Pan, L.; Cao, C.; Run, C.; Zhou, L.; Chou, J.J. DNA-Mediated Assembly of Multispecific Antibodies for T Cell Engaging and Tumor Killing. *Adv. Sci.* **2020**, *7*, 1900973.

135. 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.

136. 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.

137. Liang, S.I.; McFarland, J.M.; Rabuka, D.; Gartner, Z.J. A Modular Approach for Assembling Aldehyde-Tagged Proteins on DNA Scaffolds. *J. Am. Chem. Soc.* **2014**, *136*, 10850–10853.

138. Nielsen, T.B.; Thomsen, R.P.; Mortensen, M.R.; Kjems, J.; Nielsen, P.F.; Nielsen, T.E.; Kodal, A.L.B.; Cló, E.; Gothelf, K.V. Peptide-Directed DNA-Templated Protein Labelling for the Assembly of a Pseudo-Igm. *Angew. Chem. Int. Ed.* **2019**, *58*, 9068–9072.

139. 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.

140. Aslan, H.; Krissanaprasit, A.; Besenbacher, F.; Gothelf, K.V.; Dong, M. Protein Patterning by a DNA Origami Framework. *Nanoscale* **2016**, *8*, 15233–15240.

141. Yan, H.; Park, S.H.; Finkelstein, G.; Reif, J.H.; LaBean, T.H. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* **2003**, *301*, 1882–1884.

142. Xu, Y.; Jiang, S.; Simmons, C.R.; Narayanan, R.P.; Zhang, F.; Aziz, A.-M.; Yan, H.; Stephanopoulos, N. Tunable Nanoscale Cages from Self-Assembling DNA and Protein Building Blocks. *ACS Nano* **2019**, *13*, 3545–3554.

143. Li, S.; Jiang, Q.; Liu, S.; Zhang, Y.; Tian, Y.; Song, C.; Wang, J.; Zou, Y.; Anderson, G.J.; Han, J.-Y.; Chang, Y.; Liu, Y.; Zhang, C.; Chen, L.; Zhou, G.; Nie, G.; Yan, H.; Ding, B.; Zhao, Y. A DNA Nanorobot Functions as a Cancer Therapeutic in Response to a Molecular Trigger in Vivo. *Nat. Biotechnol.* **2018**, *36*, 258–264.

144. Yang, Y.R.; Fu, J.; Wootten, S.; Qi, X.; Liu, M.; Yan, H.; Liu, Y. 2d Enzyme Cascade Network with Efficient Substrate Channeling by Swinging Arms. *ChemBioChem* **2018**, *19*, 212–216.

145. Praetorius, F.; Dietz, H. Self-Assembly of Genetically Encoded DNA-Protein Hybrid Nanoscale Shapes. *Science* **2017**, *355*, eaam5488.

146. Ke, G.; Liu, M.; Jiang, S.; Qi, X.; Yang, Y.R.; Wootten, S.; Zhang, F.; Zhu, Z.; Liu, Y.; Yang, C.J. Directional Regulation of Enzyme Pathways through the Control of Substrate Channeling on a DNA Origami Scaffold. *Angew. Chem.* **2016**, *128*, 7609–7612.

147. Wong, N.Y.; Xing, H.; Tan, L.H.; Lu, Y. Nano-Encrypted Morse Code: A Versatile Approach to Programmable and Reversible Nanoscale Assembly and Disassembly. *J. Am. Chem. Soc.* **2013**, *135*, 2931–2934.

148. 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.

149. 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. Chem. Int. Ed.* **2007**, *46*, 3051–3054.

150. 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.

151. 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.

152. 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.

153. Wacker, R.; Niemeyer, C.M. Ddi-Mfia—a Readily Configurable Microarray-Fluorescence Immunoassay Based on DNA-Directed Immobilization of Proteins. *ChemBioChem* **2004**, *5*, 453–459.

154. Douglas, S.M.; Bachelet, I.; Church, G.M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **2012**, *335*, 831–834.

155. Lacroix, A.; Edwardson, T.G.W.; Hancock, M.A.; Dore, M.D.; Sleiman, H.F. Development of DNA Nanostructures for High-Affinity Binding to Human Serum Albumin. *J. Am. Chem. Soc.* **2017**, *139*, 7355–7362.

156. Hayes, O.G.; Partridge, B.E.; Mirkin, C.A. Encoding Hierarchical Assembly Pathways of Proteins with DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2106808118.

157. Brodin, J.D.; Auyeung, E.; Mirkin, C.A. DNA-Mediated Engineering of Multicomponent Enzyme Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4564–4569.

158. McMillan, J.R.; Brodin, J.D.; Millan, J.A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C.A. Modulating Nanoparticle Superlattice Structure Using Proteins with Tunable Bond Distributions. *J. Am. Chem. Soc.* **2017**, *139*, 1754–1757.

159. Wang, S.-T.; Minevich, B.; Liu, J.; Zhang, H.; Nykypanchuk, D.; Byrnes, J.; Liu, W.; Bershadsky, L.; Liu, Q.; Wang, T.; Ren, G.; Gang, O. Designed and Biologically Active Protein Lattices. *Nat. Commun.* **2021**, *12*, 3702.

160. Hayes, O.G.; McMillan, J.R.; Lee, B.; Mirkin, C.A. DNA-Encoded Protein Janus Nanoparticles. *J. Am. Chem. Soc.* **2018**, *140*, 9269–9274.

161. Partridge, B.E.; Winegar, P.H.; Han, Z.; Mirkin, C.A. Redefining Protein Interfaces within Protein Single Crystals with DNA. *J. Am. Chem. Soc.* **2021**, *143*, 8925–8934.

162. Subramanian, R.H.; Smith, S.J.; Alberstein, R.G.; Bailey, J.B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T.S.; Tezcan, F.A. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions. *ACS Cent. Sci.* **2018**, *4*, 1578–1586.

163. Kashiwagi, D.; Shen, H.K.; Sim, S.; Sano, K.; Ishida, Y.; Kimura, A.; Niwa, T.; Taguchi, H.; Aida, T. Molecularly Engineered "Janus Groel": Application to Supramolecular Copolymerization with a Higher Level of Sequence Control. *J. Am. Chem. Soc.* **2020**, *142*, 13310–13315.

164. Strable, E.; Johnson, J.E.; Finn, M.G. Natural Nanochemical Building Blocks: Icosahedral Virus Particles Organized by Attached Oligonucleotides. *Nano Lett.* **2004**, *4*, 1385–1389.

165. Wang, M.X.; Brodin, J.D.; Millan, J.A.; Seo, S.E.; Girard, M.; Olvera de la Cruz, M.; Lee, B.; Mirkin, C.A. Altering DNA-Programmable Colloidal Crystallization Paths by Modulating Particle Repulsion. *Nano Lett.* **2017**, *17*, 5126–5132.

166. Zhang, C.; Tian, C.; Guo, F.; Liu, Z.; Jiang, W.; Mao, C. DNA-Directed Three-Dimensional Protein Organization. *Angew. Chem. Int. Ed.* **2012**, *51*, 3382–3385.

167. Engelen, W.; Sigl, C.; Kadletz, K.; Willner, E.M.; Dietz, H. Antigen-Triggered Logic-Gating of DNA Nanodevices. *J. Am. Chem. Soc.* **2021**, *143*, 21630–21636.

168. Saccà, B.; Niemeyer, C.M. Functionalization of DNA Nanostructures with Proteins. *Chem. Soc. Rev.* **2011**, *40*, 5910–5921.

169. Stephanopoulos, N. Peptide–Oligonucleotide Hybrid Molecules for Bioactive Nanomaterials. *Bioconj. Chem.* **2019**, *30*, 1915–1922.

170. Pinheiro, A.V.; Han, D.; Shih, W.M.; Yan, H. Challenges and Opportunities for Structural DNA Nanotechnology. *Nat. Nanotechnol.* **2011**, *6*, 763–772.

171. Bachand, G.D.; Spoerke, E.D.; Stevens, M.J. Microtubule-Based Nanomaterials: Exploiting Nature's Dynamic Biopolymers. *Biotechnol. Bioeng.* **2015**, *112*, 1065–1073.

172. McRee, D.E., *Practical Protein Crystallography*, 2nd ed.; Elsevier, **1999**.

173. Wlodawer, A.; Dauter, Z.; Jaskolski, M., *Protein Crystallography: Methods and Protocols*, Springer, **2017**.

174. Chothia, C.; Janin, J. Principles of Protein–Protein Recognition. *Nature* **1975**, *256*, 705–708.

175. Rohs, R.; Jin, X.; West, S.M.; Joshi, R.; Honig, B.; Mann, R.S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233–269.

176. Janin, J.; Bahadur, R.P.; Chakrabarti, P. Protein–Protein Interaction and Quaternary Structure. *Q. Rev. Biophys.* **2008**, *41*, 133–180.

177. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

178. Derewenda, Z.S.; Vekilov, P.G. Entropy and Surface Engineering in Protein Crystallization. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 116–124.

179. Wlodawer, A.; Dauter, Z.; Jaskolski, M. Protein Crystallography. *Methods Mol. Biol.* **2017**, *1607*, 595–610.

180. McPherson, A.; Gavira, J.A. Introduction to Protein Crystallization. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2013**, *70*, 2–20.

181. Durbin, S.D.; Feher, G. Protein Crystallization. *Annu. Rev. Phys. Chem.* **1996**, *47*, 171–204.

182. Russo Krauss, I.; Merlino, A.; Vergara, A.; Sica, F. An Overview of Biological Macromolecule Crystallization. *Int. J. Mol.* **2013**, *14*, 11643–11691.

183. Derewenda, Z. Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 604–615.

184. McPherson, A., Protein Crystallization. In *Protein Crystallography: Methods and Protocols*, Wlodawer, A.; Dauter, Z.; Jaskolski, M., Eds. Springer New York: 2017; pp 17–50.

185. Leibly, D.J.; Arbing, M.A.; Pashkov, I.; DeVore, N.; Waldo, G.S.; Terwilliger, T.C.; Yeates, T.O. A Suite of Engineered Gfp Molecules for Oligomeric Scaffolding. *Structure* **2015**, *23*, 1754–1768.

186. Arpino, J.A.J.; Rizkallah, P.J.; Jones, D.D. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. *PLoS One* **2012**, *7*, e47132.

187. Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R.E. Crystal Structure Analysis of a Complete Turn of B-DNA. *Nature* **1980**, *287*, 755–758.

188. Nygren, J.; Svanvik, N.; Kubista, M. The Interactions between the Fluorescent Dye Thiazole Orange and DNA. *Biopolymers* **1998**, *46*, 39–51.

189. Rye, H.S.; Yue, S.; Wemmer, D.E.; Quesada, M.A.; Haugland, R.P.; Mathies, R.A.; Glazer, A.N. Stable Fluorescent Complexes of Double-Stranded DNA with Bis-Intercalating Asymmetric Cyanine Dyes: Properties and Applications. *Nucleic Acids Res.* **1992**, *20*, 2803–2812.

190. Ohayon, Y.P.; Hernandez, C.; Chandrasekaran, A.R.; Wang, X.; Abdallah, H.O.; Jong, M.A.; Mohsen, M.G.; Sha, R.; Birktoft, J.J.; Lukeman, P.S.; Chaikin, P.M.; Ginell, S.L.; Mao, C.; Seeman, N.C. Designing Higher Resolution Self-Assembled 3d DNA Crystals Via Strand Terminus Modifications. *ACS Nano* **2019**, *13*, 7957–7965.

191. Krissinel, E.; Henrick, K. Protein Interfaces, Surfaces and Assemblies Service Pisa at European Bioinformatics Institute. *J. Mol. Biol.* **2007**, *372*, 774–797.

192. Doyle, D.A.; Cabral, J.M.; Pfuetzner, R.A.; Kuo, A.; Gulbis, J.M.; Cohen, S.L.; Chait, B.T.; MacKinnon, R. The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity. *Science* **1998**, *280*, 69–77.

193.   Nuñez-Prado, N.; Compte, M.; Harwood, S.; Álvarez-Méndez, A.; Lykkemark, S.; Sanz, L.; Álvarez-Vallina, L. The Coming of Age of Engineered Multivalent Antibodies. *Drug Discov.* **2015**, *20*, 588–594.

194.   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.

195.   Li, X.; Qiao, S.; Zhao, L.; Liu, S.; Li, F.; Yang, F.; Luo, Q.; Hou, C.; Xu, J.; Liu, J. Template-Free Construction of Highly Ordered Monolayered Fluorescent Protein Nanosheets: A Bioinspired Artificial Light-Harvesting System. *ACS Nano* **2019**, *13*, 1861–1869.

196.   Aydin, S. A Short History, Principles, and Types of Elisa, and Our Laboratory Experience with Peptide/Protein Analyses Using Elisa. *Peptides* **2015**, *72*, 4–15.

197.   Szijj, P.; Chudasama, V. The Renaissance of Chemically Generated Bispecific Antibodies. *Nat. Rev. Chem.* **2021**, *5*, 78–92.

198.   Yang, Y.R.; Liu, Y.; Yan, H. DNA Nanostructures as Programmable Biomolecular Scaffolds. *Bioconj. Chem.* **2015**, *26*, 1381–1395.

199.   Hamann, P.R.; Hinman, L.M.; Hollander, I.; Beyer, C.F.; Lindh, D.; Holcomb, R.; Hallett, W.; Tsou, H.-R.; Upeslacis, J.; Shochat, D.; Mountain, A.; Flowers, D.A.; Bernstein, I. Gemtuzumab Ozogamicin, a Potent and Selective Anti-Cd33 Antibody−Calicheamicin Conjugate for Treatment of Acute Myeloid Leukemia. *Bioconj. Chem.* **2002**, *13*, 47–58.

200.   Harris, L.J.; Larson, S.B.; Hasel, K.W.; McPherson, A. Refined Structure of an Intact Igg2a Monoclonal Antibody. *Biochemistry* **1997**, *36*, 1581–1597.

201.   Cui, H.; Chen, Z.; Zhong, S.; Wooley, K.L.; Pochan, D.J. Block Copolymer Assembly Via Kinetic Control. *Science* **2007**, *317*, 647–650.

202.   Figg, C.A.; Simula, A.; Gebre, K.A.; Tucker, B.S.; Haddleton, D.M.; Sumerlin, B.S. Polymerization-Induced Thermal Self-Assembly (Pitsa). *Chemical Science* **2015**, *6*, 1230–1236.

203. Malmsten, M.; Lindman, B. Self-Assembly in Aqueous Block Copolymer Solutions. *Macromolecules* **1992**, *25*, 5440–5445.

204. Nikos Hadjichristidis, S.P., George Floudas, Block Copolymer Applications. In *Block Copolymers: Synthetic Strategies, Physical Properties, and Applications*, 2002; pp 383–408.

205. Rizzuto, F.J.; Dore, M.D.; Rafique, M.G.; Luo, X.; Sleiman, H.F. DNA Sequence and Length Dictate the Assembly of Nucleic Acid Block Copolymers. *J. Am. Chem. Soc.* **2022**, *144*, 12272–12279.

206. Pan, G.; Jin, X.; Mou, Q.; Zhang, C. Recent Progress on DNA Block Copolymer. *Chin. Chem. Lett.* **2017**, *28*, 1822-1828.

207. Huang, A.; Paloni, J.M.; Wang, A.; Obermeyer, A.C.; Sureka, H.V.; Yao, H.; Olsen, B.D. Predicting Protein–Polymer Block Copolymer Self-Assembly from Protein Properties. *Biomacromolecules* **2019**, *20*, 3713–3723.

208. Olsen, B.D. Self-Assembly of Globular-Protein-Containing Block Copolymers. *Macromol. Chem. Phys.* **2013**, *214*, 1659–1668.

209. Lam, C.N.; Yao, H.; Olsen, B.D. The Effect of Protein Electrostatic Interactions on Globular Protein–Polymer Block Copolymer Self-Assembly. *Biomacromolecules* **2016**, *17*, 2820–2829.

210. Obermeyer, A.C.; Olsen, B.D. Synthesis and Application of Protein-Containing Block Copolymers. *ACS Macro Letters* **2015**, *4*, 101–110.

211. Romei, M.G.; Boxer, S.G. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annu. Rev. Biophys.* **2019**, *48*, 19–44.

212. Riggs, P. Expression and Purification of Maltose-Binding Protein Fusions. *Curr. Protoc. Mol. Biol.* **1994**, *28*, 16.16.11–16.16.14.

213. Kapust, R.B.; Tözsér, J.; Copeland, T.D.; Waugh, D.S. The P1′ Specificity of Tobacco Etch Virus Protease. *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949–955.

214. Kastritis, P.L.; Bonvin, A.M.J.J. On the Binding Affinity of Macromolecular Interactions: Daring to Ask Why Proteins Interact. *J. R. Soc. Interface* **2012**, *10*, 20120835–20120835.

215. Dale, R.; Kato, N. Truly Quantitative Analysis of the Firefly Luciferase Complementation Assay. *Curr. Plant Biol.* **2016**, *5*, 57–64.

216. Kato, N.; Jones, J., The Split Luciferase Complementation Assay. In *Plant Developmental Biology: Methods and Protocols*, Hennig, L.; Köhler, C., Eds. Humana Press: Totowa, NJ, 2010; pp 359–376.

217. Chumakov, S.P.; Kravchenko, Y.E.; Chumakov, P.M. Protein Complementation as Tool for Studying Protein-Protein Interactions in Living Cells. *Mol. Biol.* **2012**, *46*, 627–638.

218. Paulmurugan, R.; Gambhir, S.S. Monitoring Protein−Protein Interactions Using Split Synthetic Renilla Luciferase Protein-Fragment-Assisted Complementation. *Anal. Chem.* **2003**, *75*, 1584–1589.

219. Marsh, Joseph A.; Hernández, H.; Hall, Z.; Ahnert, Sebastian E.; Perica, T.; Robinson, Carol V.; Teichmann, Sarah A. Protein Complexes Are under Evolutionary Selection to Assemble Via Ordered Pathways. *Cell* **2013**, *153*, 461–470.

220. Zámocký, M.; Koller, F. Understanding the Structure and Function of Catalases: Clues from Molecular Evolution and in Vitro Mutagenesis. *Prog. Biophys. Mol. Biol.* **1999**, *72*, 19–66.

221. Sepasi Tehrani, H.; Moosavi-Movahedi, A.A. Catalase and Its Mysteries. *Prog. Biophys. Mol. Biol.* **2018**, *140*, 5–12.

222. Alfonso-Prieto, M.; Biarnés, X.; Vidossich, P.; Rovira, C. The Molecular Mechanism of the Catalase Reaction. *J. Am. Chem. Soc.* **2009**, *131*, 11751–11761.

223. Ahmed, E.M. Hydrogel: Preparation, Characterization, and Applications: A Review. *J. Adv. Res.* **2015**, *6*, 105–121.

224. Li, Y.; Xue, B.; Cao, Y. 100th Anniversary of Macromolecular Science Viewpoint: Synthetic Protein Hydrogels. *ACS Macro Lett.* **2020**, *9*, 512–524.

225. Bitterwolf, P.; Gallus, S.; Peschke, T.; Mittmann, E.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K.S.; Niemeyer, C.M. Valency Engineering of Monomeric Enzymes for Self-Assembling Biocatalytic Hydrogels. *Chem. Sci.* **2019**, *10*, 9752–9757.

226. Bitterwolf, P.; Ott, F.; Rabe, K.S.; Niemeyer, C.M. Imine Reductase Based All-Enzyme Hydrogel with Intrinsic Cofactor Regeneration for Flow Biocatalysis. *Micromachines* **2019**, *10*, 783.

227. Peschke, T.; Bitterwolf, P.; Gallus, S.; Hu, Y.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K.S.; Niemeyer, C.M. Self-Assembling All-Enzyme Hydrogels for Flow Biocatalysis. *Angew. Chem. Int. Ed.* **2018**, *130*, 17274–17278.

228. Wang, X.; Wang, Q. Enzyme-Laden Bioactive Hydrogel for Biocatalytic Monitoring and Regulation. *Acc. Chem. Res.* **2021**, *54*, 1274–1287.

229. Heo, J.; Crooks, R.M. Microfluidic Biosensor Based on an Array of Hydrogel-Entrapped Enzymes. *Anal. Chem.* **2005**, *77*, 6843–6851.

230. Russell, R.J.; Pishko, M.V.; Simonian, A.L.; Wild, J.R. Poly(Ethylene Glycol) Hydrogel-Encapsulated Fluorophore−Enzyme Conjugates for Direct Detection of Organophosphorus Neurotoxins. *Anal. Chem.* **1999**, *71*, 4909–4912.

231. Shao, Y.; Jia, H.; Cao, T.; Liu, D. Supramolecular Hydrogels Based on DNA Self-Assembly. *Acc. Chem. Res.* **2017**, *50*, 659–668.

232. Um, S.H.; Lee, J.B.; Park, N.; Kwon, S.Y.; Umbach, C.C.; Luo, D. Enzyme-Catalysed Assembly of DNA Hydrogel. *Nat. Mater.* **2006**, *5*, 797–801.

233. Kahn, J.S.; Hu, Y.; Willner, I. Stimuli-Responsive DNA-Based Hydrogels: From Basic Principles to Applications. *Acc. Chem. Res.* **2017**, *50*, 680–690.

234. Wang, C.; Fischer, A.; Ehrlich, A.; Nahmias, Y.; Willner, I. Biocatalytic Reversible Control of the Stiffness of DNA-Modified Responsive Hydrogels: Applications in Shape-Memory, Self-Healing and Autonomous Controlled Release of Insulin. *Chem. Sci.* **2020**, *11*, 4516–4524.

235. Wang, C.; Fadeev, M.; Vázquez-González, M.; Willner, I. Stimuli-Responsive Donor–Acceptor and DNA-Crosslinked Hydrogels: Application as Shape-Memory and Self-Healing Materials. *Adv. Funct. Mater.* **2018**, *28*, 1803111.

236. Li, C.; Zhou, X.; Shao, Y.; Chen, P.; Xing, Y.; Yang, Z.; Li, Z.; Liu, D. A Supramolecular Hydrogel with Identical Cross-Linking Point Density but Distinctive Rheological Properties. *Mater. Chem. Front.* **2017**, *1*, 654–659.

237. Hirano, Y.; Ikegami, M.; Kowata, K.; Komatsu, Y. Bienzyme Reactions on Cross-Linked DNA Scaffolds for Electrochemical Analysis. *Bioelectrochemistry* **2017**, *113*, 15–19.

238. Zhou, L.; Morel, M.; Rudiuk, S.; Baigl, D. Intramolecularly Protein-Crosslinked DNA Gels: New Biohybrid Nanomaterials with Controllable Size and Catalytic Activity. *Small* **2017**, *13*, 1700706.

239. Niemeyer, C.M.; Adler, M.; Gao, S.; Chi, L. Nanostructured DNA−Protein Aggregates Consisting of Covalent Oligonucleotide−Streptavidin Conjugates. *Bioconj. Chem.* **2001**, *12*, 364–371.

240. Chilkoti, A.; Stayton, P.S. Molecular Origins of the Slow Streptavidin-Biotin Dissociation Kinetics. *J. Am. Chem. Soc.* **1995**, *117*, 10622–10628.

241. Mann, D.A.; Kanai, M.; Maly, D.J.; Kiessling, L.L. Probing Low Affinity and Multivalent Interactions with Surface Plasmon Resonance: Ligands for Concanavalin A. *J. Am. Chem. Soc.* **1998**, *120*, 10575–10582.

242. De Kok, A.; Veeger, C. Studies on L-Amino-Acid Oxidase I. Effects of Ph and Competitive Inhibitors. *Biochim. Biophys. Acta* **1968**, *167*, 35–47.

243. Pawelek, P.D.; Cheah, J.; Coulombe, R.; Macheroux, P.; Ghisla, S.; Vrielink, A. The Structure of L-Amino Acid Oxidase Reveals the Substrate Trajectory into an Enantiomerically Conserved Active Site. *The EMBO Journal* **2000**, *19*, 4204–4215.

244. Fernandez-Castanon, J.; Bianchi, S.; Saglimbeni, F.; Di Leonardo, R.; Sciortino, F. Microrheology of DNA Hydrogel Gelling and Melting on Cooling. *Soft Matter* **2018**, *14*, 6431–6438.

245. McGlynn, J.A.; Wu, N.; Schultz, K.M. Multiple Particle Tracking Microrheological Characterization: Fundamentals, Emerging Techniques and Applications. *J. Appl. Phys.* **2020**, *127*, 201101.

246. Gibson, D.G.; Young, L.; Chuang, R.-Y.; Venter, J.C.; Hutchison Iii, C.A.; Smith, H.O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.

247. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463–5467.

248. Patterson, G.H.; Knobel, S.M.; Sharif, W.D.; Kain, S.R.; Piston, D.W. Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy. *Biophys. J.* **1997**, *73*, 2782–2790.

249. Gill, S.C.; von Hippel, P.H. Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* **1989**, *182*, 319–326.

250. Evans, P.R.; Murshudov, G.N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204–1214.

251. Winter, G. Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186–190.

252. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; McNicholas, S.J.; Murshudov, G.N.; Pannu, N.S.; Potterton, E.A.; Powell, H.R.; Read, R.J.; Vagin, A.; Wilson, K.S. Overview of the Ccp4 Suite and Current Developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235–242.

253. Battye, T.G.G.; Kontogiannis, L.; Johnson, O.; Powell, H.R.; Leslie, A.G.W. Imosflm: A New Graphical Interface for Diffraction-Image Processing with Mosflm. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 271–281.

254. Evans, P. Scaling and Assessment of Data Quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 72–82.

255. Evans, P. An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 282–292.

256. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

257. Kachalova, G.S., Popov, A.P., Simanovskaya, A.A., and Lipkin, A.V. Structure of Egfp(Enhanced Green Fluorescent Protein) Mutant - L232h at 0.153 Nm. *To Be Published* **2018**.

258. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501.

259. Murshudov, G.N.; Skubak, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. Refmac5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355–367.

260. Schrodinger, LLC The Pymol Molecular Graphics System, Version 1.8. **2015**.

261. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. Ucsf Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

262. Tarini, M.; Cignoni, P.; Montani, C. Ambient Occlusion and Edge Cueing for Enhancing Real Time Molecular Visualization. *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 1237–1244.

263. Ikebuchi, R.; Konnai, S.; Okagawa, T.; Yokoyama, K.; Nakajima, C.; Suzuki, Y.; Murata, S.; Ohashi, K. Blockade of Bovine Pd-1 Increases T Cell Function and Inhibits Bovine Leukemia Virus Expression in B Cells in Vitro. *Vet. Res.* **2013**, *44*, 59.

264. Agata, Y.; Kawasaki, A.; Nishimura, H.; Ishida, Y.; Tsubat, T.; Yagita, H.; Honjo, T. Expression of the Pd-1 Antigen on the Surface of Stimulated Mouse T and B Lymphocytes. *Int. Immunol.* **1996**, *8*, 765–772.

265. Cheng, Z.-J.J.; Karassina, N.; Grailer, J.; Hartnett, J.; Fan, F.; Cong, M. Novel Pd-1 Blockade Bioassay to Assess Therapeutic Antibodies in Pd-1 and Pd-L1 Immunotherapy Programs. *Cancer Res.* **2015**, *75*, abstr 5440.

266. Gilardi, G.; Mei, G.; Rosato, N.; Agrò, A.F.; Cass, A.E. Spectroscopic Properties of an Engineered Maltose Binding Protein. *Protein Eng. Des. Sel.* **1997**, *10*, 479–486.

APPENDICES

6.1. Appendix A: Materials and Methods for Chapter Two.

6.1.1. Protein Mutation, Recombinant Expression, and Purification.

A gene for C148 mGFP (Table 2.1) was cloned and transformed into One Shot® BL21(DE3) Chemically Competent E. coli (Thermo Fisher) in previous work.[160] Genes for C176 mGFP and C191 mGFP (Table 2.1, Integrated DNA Technologies) were cloned into the pET28 vector backbone using Gibson Assembly.[246] The assembled plasmids for C176 mGFP and C191 mGFP were transformed into BL21(DE3) electrically competent cells (ThermoFisher) with electroporation. After recovery in S.O.C. Medium (ThermoFisher) for 1 hour at 37 °C with 300 rpm shaking, cells were grown overnight on LB Agar plates with antibiotic (50 µg/mL kanamycin). Single colonies were selected and cultured in 8 mL of LB broth with antibiotic (50 µg/mL kanamycin) overnight at 37 °C with 200 rpm shaking. After cell growth, glycerol stocks of the cells were prepared and stored at -80 °C. Plasmids were extracted from cells using the QIAprep Spin Miniprep Kit (Qiagen) and the correct plasmid sequences were confirmed using Sanger Sequencing (ACGT).[247]

Cultures in 8 mL of LB broth with antibiotic (100 µg/mL ampicillin for C148 mGFP and 50 µg/mL kanamycin for C176 mGFP, and C191 mGFP) were inoculated using glycerol stocks and grown overnight at 37 °C with 200 rpm shaking. Next, these cultures were added to 1 L of 2x YTP broth with antibiotic (100 µg/mL ampicillin for C148 mGFP and 50 µg/mL kanamycin for C176 mGFP, and C191 mGFP) and grown at 37 °C with 200 rpm shaking until a cell OD at 600 nm of 0.6 (~4 h). Cultures were induced (0.2 % [w/w] L-arabinose for C148 mGFP and 1 mM

isopropyl β-D-1-thiogalactopyranoside (IPTG) for C176 mGFP, and C191 mGFP) and grown overnight at 17°C with 200 rpm shaking. Cells were pelleted (6000 g, 20 min, 4 °C), resuspended in 1× PBS, and lysed with a high-pressure homogenizer. The insoluble fraction was removed with centrifugation (15000 g, 20 min, 4 °C).

The mGFP mutants contain a poly-histidine tag, which was used to isolate the mutants from cell lysate using nickel affinity chromatography. Proteins were loaded onto a column packed with Profinity™ IMAC Resin (Bio-Rad). The column was washed with 100 mL of 1× PBS with 12.5 mM imidazole and proteins were eluted with 15 mL of 1× PBS with 250 mM imidazole. The mGFP mutants were separated from the imidazole using anion exchange chromatography. The proteins were then loaded onto a column packed with Macro-Prep® DEAE resin (Bio-Rad). The column was washed with 40 mL of 1× PBS and proteins were eluted with 15 mL of 1× PBS with an additional 250 mM NaCl. Protein purity was confirmed with SDS-PAGE, showing mGFP primarily as monomers with small impurities of dimers that are formed from the oxidation of cysteine to form a disulfide bond. The concentrations of mGFP were identified using UV-VIS absorption spectroscopy using an established extinction coefficients ($\varepsilon_{GFP, 488 nm}$ = 55,000 $M^{-1}cm^{-1}$).[248]

6.1.2. Oligonucleotide Design, Synthesis, and Purification.

Nine DNA sequences or pairs of complementary DNA sequences were designed to study how DNA interactions can influence protein crystallization and packing into single crystals (Table 2.2). DNA designs varied between self-complementary (scDNA), complementary (cDNA), and non-complementary (ncDNA). DNA length varied between 6 and 18 bases. The sites with the DNA for attachment to mGFP was either at an internal or external position on the DNA strand.

Oligonucleotides utilized herein were synthesized on solid supports using reagents obtained from Glen Research and standard protocols. Products were cleaved from the solid support using 15% (w/v) ammonium hydroxide (aq) and 20% (w/v) methyl amine for 20 min at 55 °C and purified using reverse-phase high-pressure liquid chromatography (HPLC) with a gradient of 0 to 75 percent acetonitrile in triethylammonium acetate buffer over 45 min. Dimethoxytrityl or monomethoxytrityl groups were cleaved with 20% (v/v) acetic acid for 2 h and extracted with ethyl acetate. The masses of the oligonucleotides were MALDI-TOF MS using 3-hydroxypicolinic acid, 2′,5′-dihydroxyacetophenone, or 2′,4′,6′-trihydroxyacetophenone monohydrate as a matrix. All observed masses of synthesized DNA were within 30 Da of the expected mass.

Scheme 6.1. Structures of amino-modifier phosphoramidites.
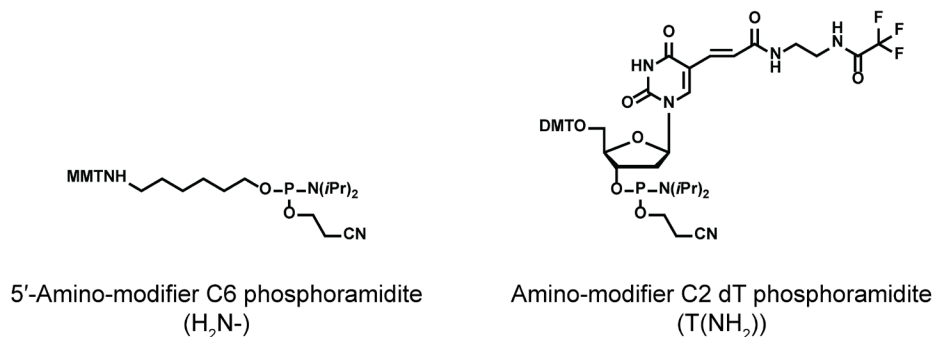
5′-Amino-modifier C6 phosphoramidite
(H₂N-)

Amino-modifier C2 dT phosphoramidite
(T(NH₂))

These phosphoramidites contain 4,4'-dimethoxytrityl (DMT), *N*,*N*-diisopropyl (N(*i*Pr)₂), ß-cyanoethyl, and trifluoroacetyl, and monomethoxytrityl (MMT) protecting groups that are removed during DNA workup and purification.

6.1.3. Synthesis, Purification, and Characterization of mGFP–DNA Conjugates.

Conjugation of mGFP and DNA was performed according to a previously published procedure.[160] Linkage structures for mGFP–DNA are depicted in Scheme 6.2. Amine-modified DNA (3,000 nmol) was reacted with 30–50 equiv. of succinimidyl 3-(2-pyridyldithio)propionate (SPDP, ThermoFisher) in 50:50 DMF:1× PBS, pH 7.4 for 1 h at RT. DNA was purified from

excess SPDP with two consecutive illustra NAP Columns (GE Healthcare Life Sciences). The purified DNA was reacted with mGFP (300 nmol) overnight at RT with 300 rpm shaking. The reaction mixture was loaded onto a column packed with Profinity™ IMAC Resin (Bio-Rad). To remove unreacted DNA, the column was washed with 40 mL of 1× PBS. Protein and protein–DNA conjugates were eluted with 15 mL of 1× PBS with 250 mM imidazole. The eluent was then loaded onto a column packed with Macro-Prep® DEAE resin (Bio-Rad). The column was washed with 40 mL of 1× PBS and 30 mL of 1× PBS with an additional 200 mM NaCl to remove thiol and disulfide forms of mGFP. Conjugates of mGFP–DNA were eluted with 15 mL of 1× PBS with an additional 500 mM NaCl.

Synthesis and purity of mGFP–DNA conjugates were confirmed with UV-vis absorption spectroscopy, SDS-PAGE, and MALDI-TOF MS. The C148 mGFP, C176 mGFP, and C191 mGFP mutants show absorption maxima at 488 nm ($\varepsilon_{488\ nm} = 55{,}000\ M^{-1}cm^{-1}$) due to the mGFP chromophore[248] and at 280 nm due to aromatic amino acid side chains.[249] DNA shows an absorption maxima around 260 nm and extinction coefficients at 260 nm were calculated with the IDT OligoAnalyzer Tool (Integrated DNA Technologies). After purification of mGFP–DNA conjugates, the number of DNA per mGFP in solution was quantified by comparing the relative absorption at 488 nm and 260 nm for mGFP and mGFP–DNA. The increase in mass of mGFP–DNA conjugates after DNA functionalization and sample purity was confirmed with SDS-PAGE using 4-15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) and a Precision Plus Protein™ All Blue Prestained Protein Standard (Bio-Rad). The increase in mass of mGFP–DNA conjugates after DNA conjugation and sample purity was also confirmed with MALDI-TOF MS. Before MALDI-TOF MS, conjugates of mGFP–DNA were transferred to water six times using 30

kDa cutoff Amicon® Ultra-0.5 mL Centrifugal Filters (Millepore Sigma) and mixed with MALDI matrix 2′,5′-dihydroxyacetophenone, 2′,4′,6′-trihydroxyacetophenone monohydrate, or sinapinic acid.
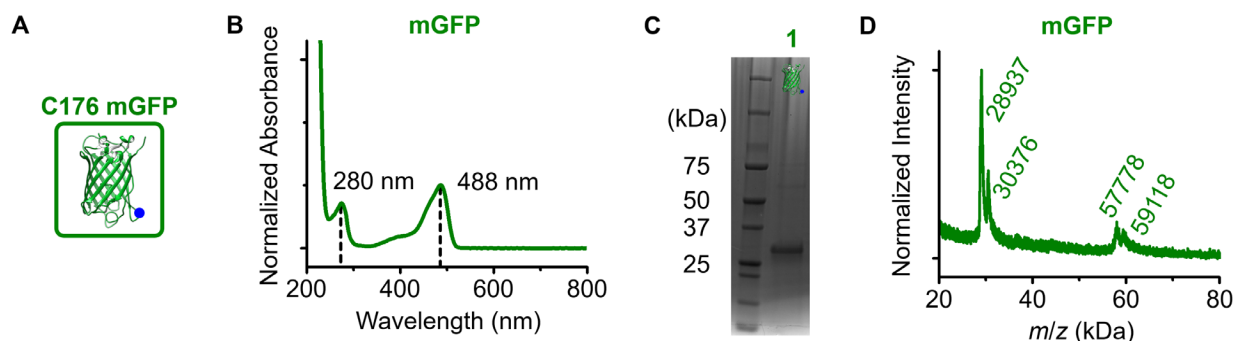
Figure 6.1. Characterization of C176 mGFP. (A) Schematic of C176 mGFP (green). (B) A UV-vis absorption spectrum that is normalized to the absorbance of the mGFP chromophore at 488 nm. The spectrum also contains a characteristic protein absorbance peak at 280 nm. (C) SDS-PAGE analysis shows C176 mGFP (lane 1) primarily in the thiol form (~30 kDa) with a small amount in the disulfide form (~60 kDa). (D) MALDI-TOF MS characterization shows an experimental C176 mGFP (green) mass of ~29.0 and 30.5 kDa.

Figure 6.2. Characterization of C191 mGFP. (A) Schematic of C191 mGFP (green). (B) A UV-vis absorption spectrum that is normalized to the absorbance of the mGFP chromophore at 488 nm. The spectrum also contains a characteristic protein absorbance peak at 280 nm. (C) SDS-PAGE analysis shows C191 mGFP (lane 1) primarily in the thiol form (~30 kDa) with a small amount in the disulfide form (~60 kDa). (D) MALDI-TOF MS characterization shows an experimental C191 mGFP (green) mass of ~30.5 kDa.

Scheme 6.2. Linkage chemical structures in mGFP–DNA conjugates.

Linkage chemical structures in mGFP–DNA conjugates with (A) external and (B) internal DNA attachment positions. Atoms from mGFP and DNA are colored in blue and atoms from SPDP are colored in black.
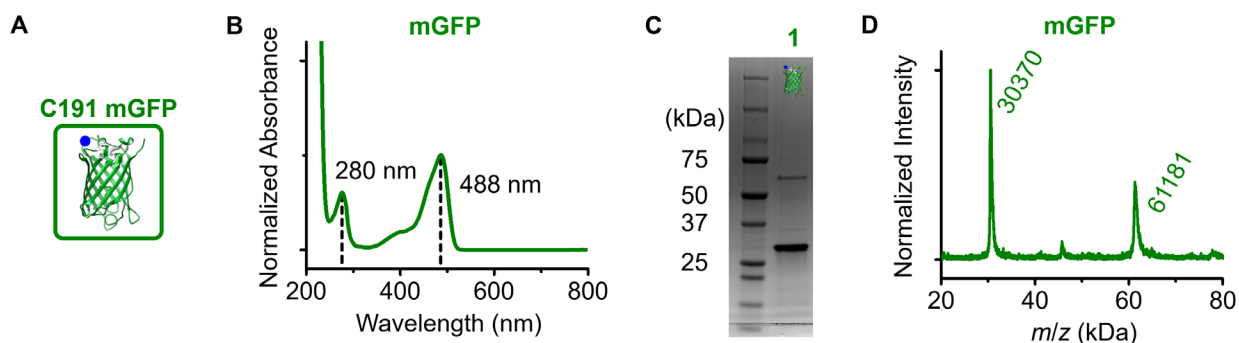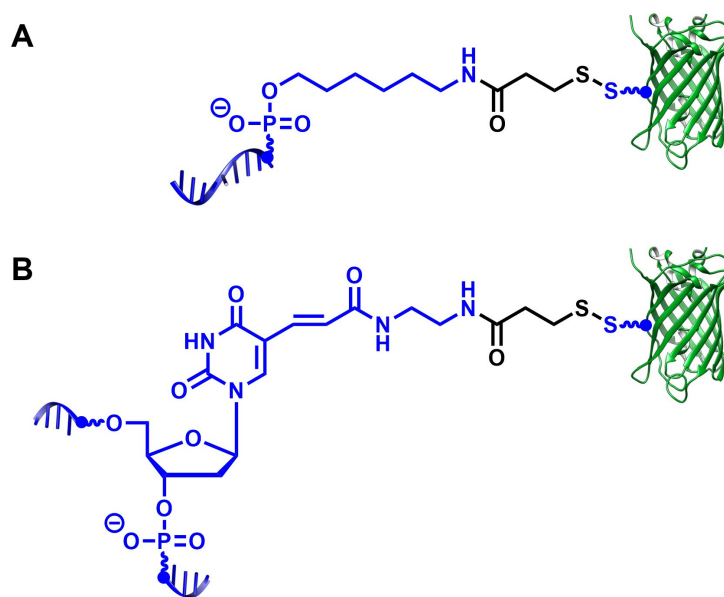
Figure 6.3. Characterization of C148 mGFP–cDNA-1 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–cDNA-1 (red and purple). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-1 (red and purple) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–cDNA-1 compared to C148 mGFP corresponds to 1.0 and 1.1 cDNA-1 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–cDNA-1 (lanes 2 and 3) show the conjugation of one cDNA-1 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–cDNA-1 (red and purple) shows an experimental mass increases of 2002 and 1942 Da that are consistent with the theoretical mass increase of mono-functionalization (2098 Da = 2012 Da (cDNA-1) + 86 Da (linker molecule) and 2018 Da = 1932 Da (cDNA-1) + 86 Da (linker molecule), respectively).

Figure 6.4. Characterization of C148 mGFP–cDNA-2 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–cDNA-2 (red and purple). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-2 (red and purple) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–cDNA-2 compared to C148 mGFP corresponds to 0.9 and 1.3 cDNA-2 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–cDNA-2 (lanes 2 and 3) show the conjugation of one cDNA-1 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–cDNA-2 (red and purple) shows an experimental mass increases of 2002 and 1942 Da that are consistent with the theoretical mass increase of mono-functionalization (2130 Da = 2044 Da (cDNA-2) + 86 Da (linker molecule) and 1983 Da = 1897 Da (cDNA-2) + 86 Da (linker molecule), respectively).

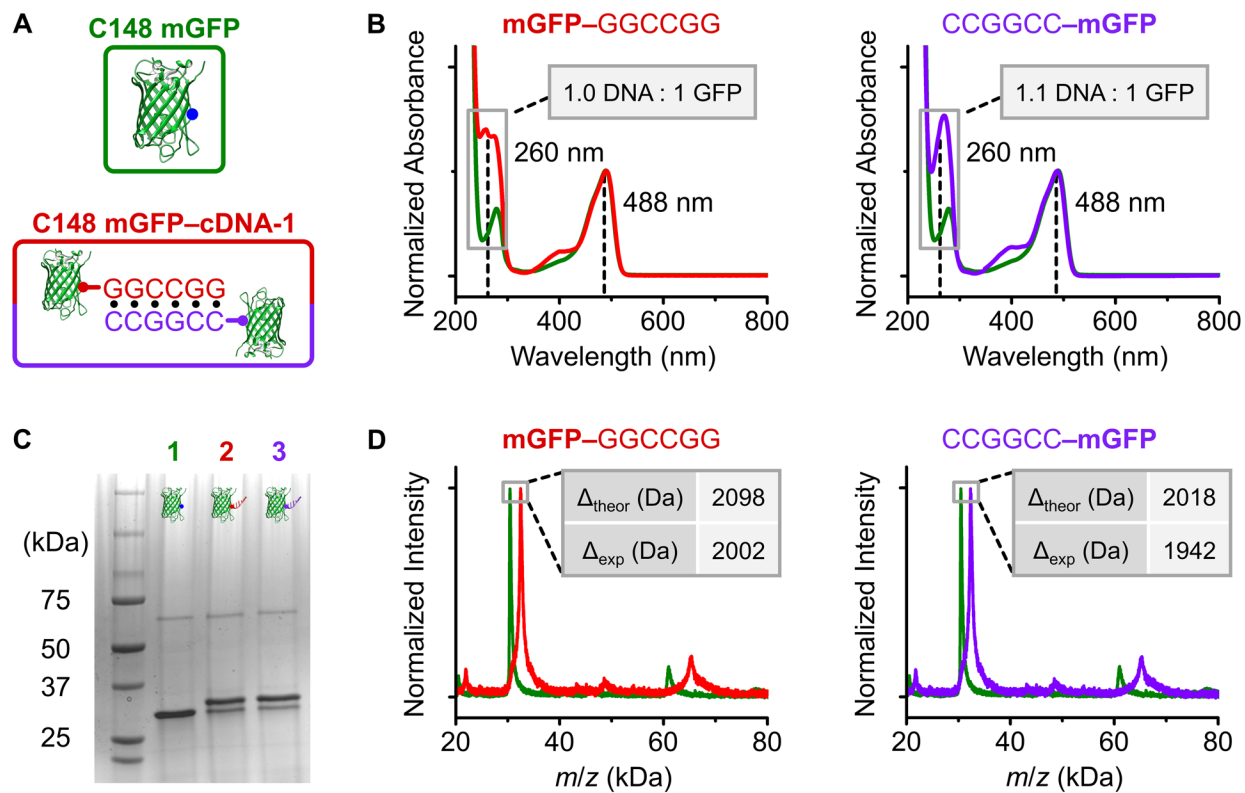**Figure 6.5. Characterization of C148 mGFP–ncDNA-1 conjugates.** (A) Schematics of C148 mGFP (green) and C148 mGFP–ncDNA-1 (orange). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–ncDNA-1 (orange) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–ncDNA-1 compared to C148 mGFP corresponds to 1.1 ncDNA-1 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–ncDNA-1 (lane 2) show the conjugation of one ncDNA-1 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–ncDNA-1 (orange) shows an experimental mass increase of 1967 Da that is consistent with the theoretical mass increase of mono-functionalization (2028 Da = 1942 Da (ncDNA-1) + 86 Da (linker molecule)).
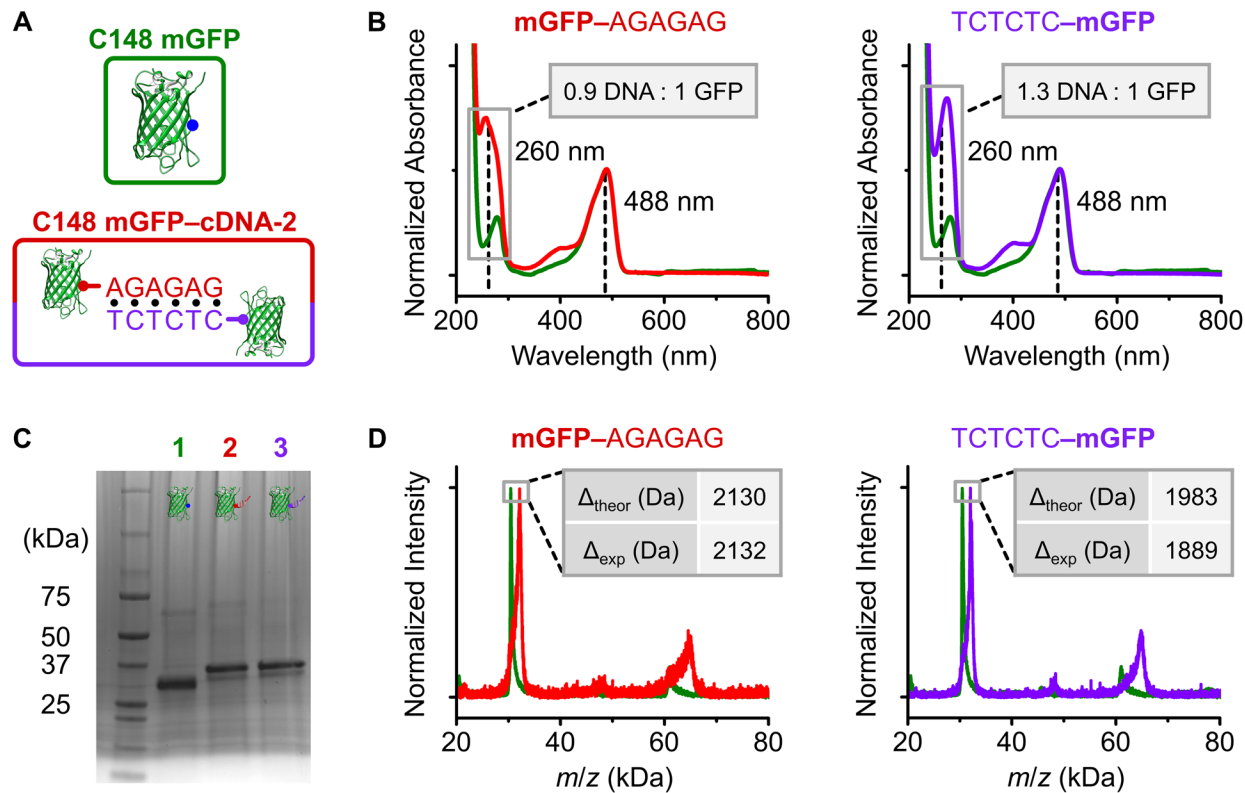
Figure 6.6. Characterization of C148 mGFP–cDNA-3 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–cDNA-3 (red and purple). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-3 (red and purple) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–cDNA-3 compared to C148 mGFP corresponds to 1.0 and 1.1 cDNA-3 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–cDNA-3 (lanes 2 and 3) show the conjugation of one cDNA-3 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–cDNA-3 (red and purple) shows an experimental mass increases of 3141 and 2789 Da that are consistent with the theoretical mass increase of mono-functionalization (3086 Da = 3000 Da (cDNA-3) + 86 Da (linker molecule) and 2881 Da = 2795 Da (cDNA-3) + 86 Da (linker molecule), respectively).
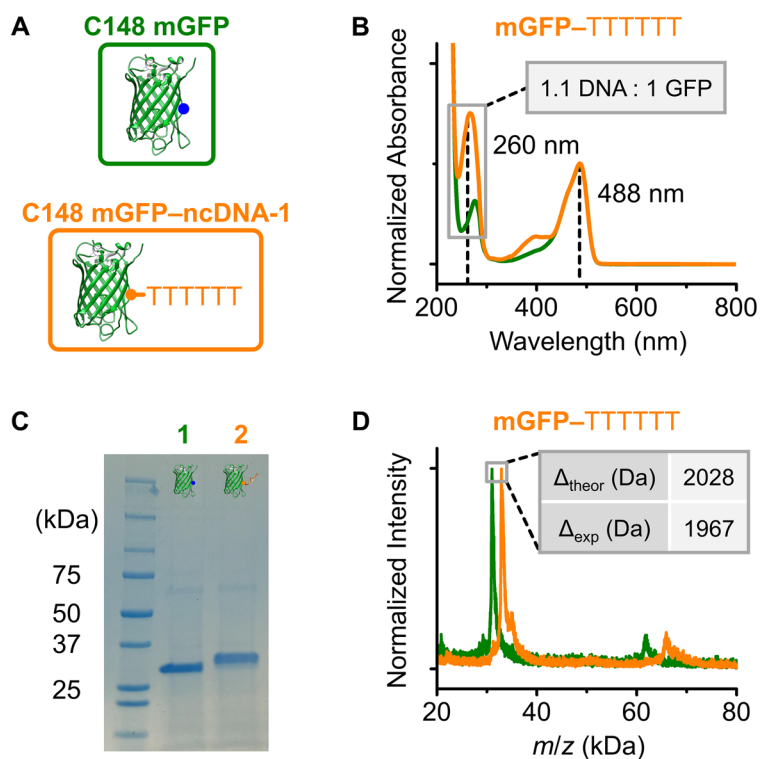
Figure 6.7. Characterization of C148 mGFP–cDNA-4 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–cDNA-4 (red and purple). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-4 (red and purple) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–cDNA-4 compared to C148 mGFP corresponds to 1.0 and 1.1 cDNA-4 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–cDNA-4 (lanes 2 and 3) show the conjugation of one cDNA-4 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–cDNA-4 (red and purple) shows an experimental mass increases of 3887 and 3603 Da that are consistent with the theoretical mass increase of mono-functionalization (4058 Da = 3972 Da (cDNA-4) + 86 Da (linker molecule) and 3763 Da = 3677 Da (cDNA-4) + 86 Da (linker molecule), respectively).
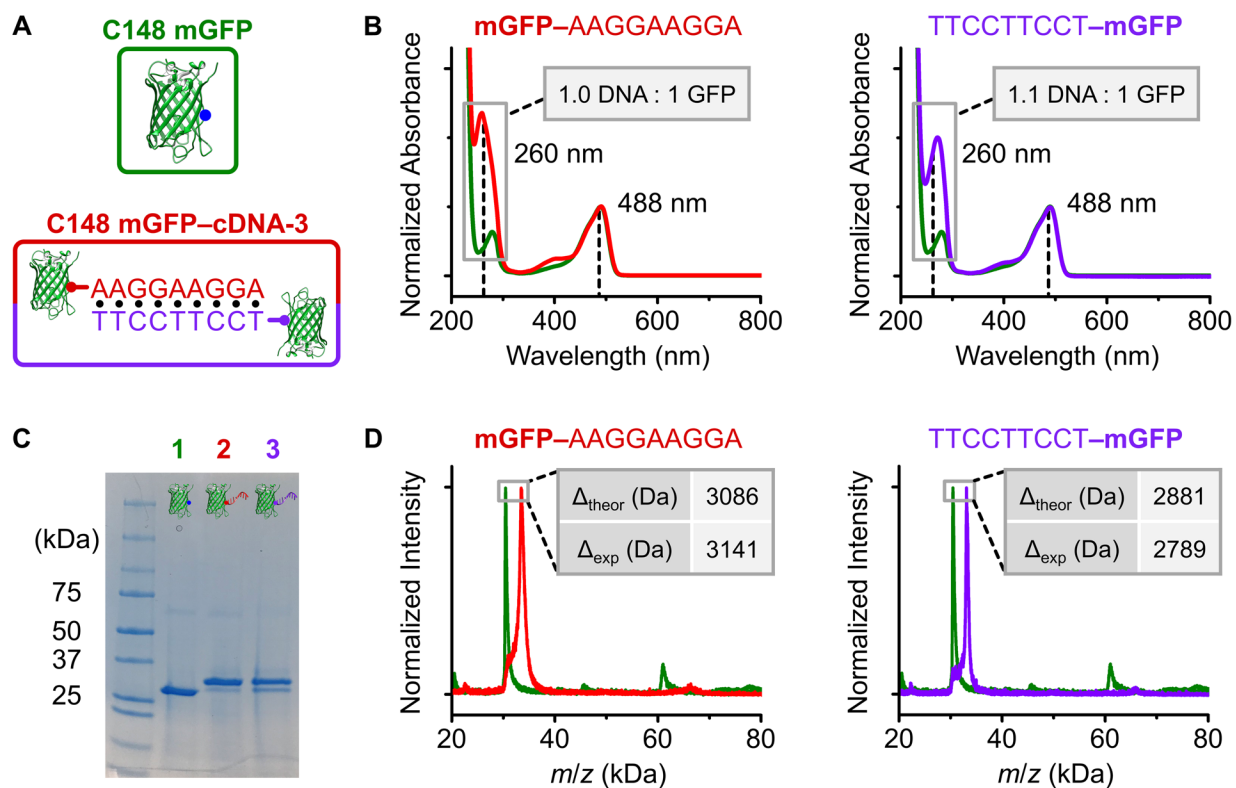
Figure 6.8. Characterization of C148 mGFP–cDNA-5 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–cDNA-5 (red and purple). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-5 (red and purple) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–cDNA-5 compared to C148 mGFP corresponds to 1.1 and 1.0 cDNA-5 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–cDNA-5 (lanes 2 and 3) show the conjugation of one cDNA-5 to C148 mGFP.
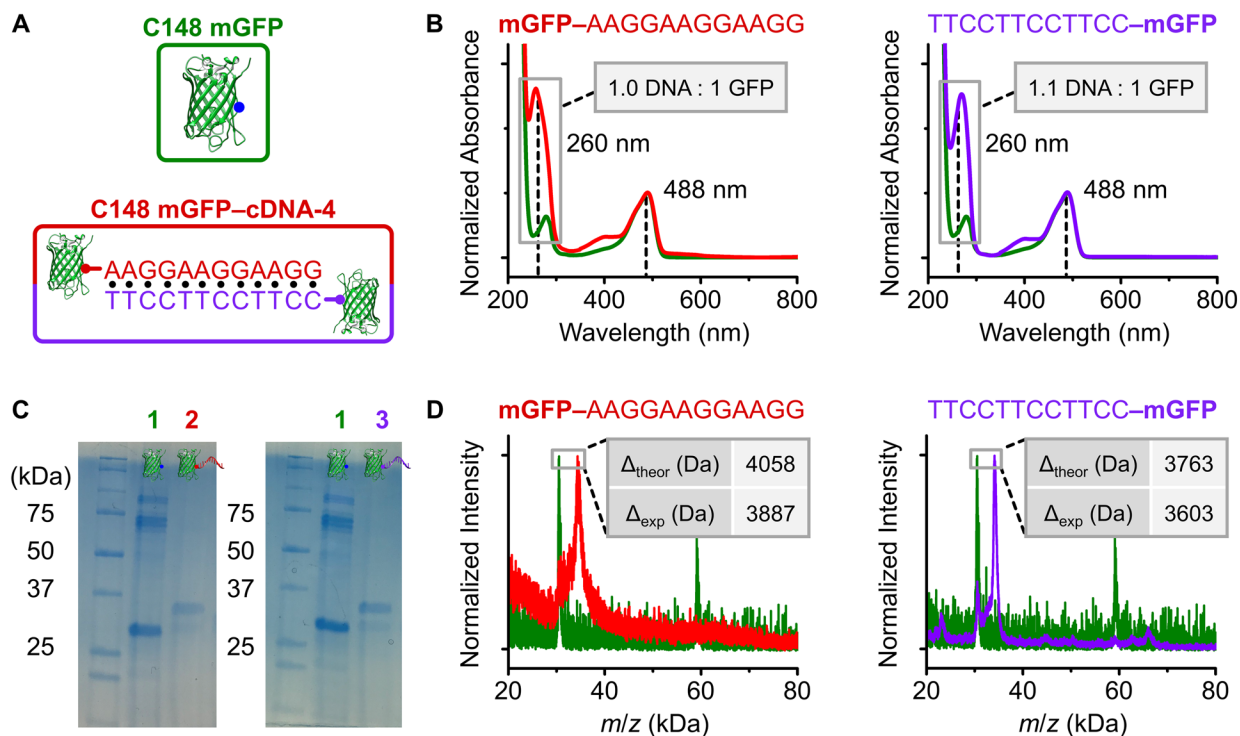
Figure 6.9. Characterization of C148 mGFP–ncDNA-2 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–ncDNA-2 (orange). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–ncDNA-2 (orange) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–ncDNA-2 compared to C148 mGFP corresponds to 1.5 ncDNA-2 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–ncDNA-2 (lane 2) show the conjugation of one ncDNA-2 to C148 mGFP. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–ncDNA-2 (orange) shows an experimental mass increase of 4510 Da that is consistent with the theoretical mass increase of mono-functionalization (2941 Da = 2855 Da (ncDNA-2) + 86 Da (linker molecule)).
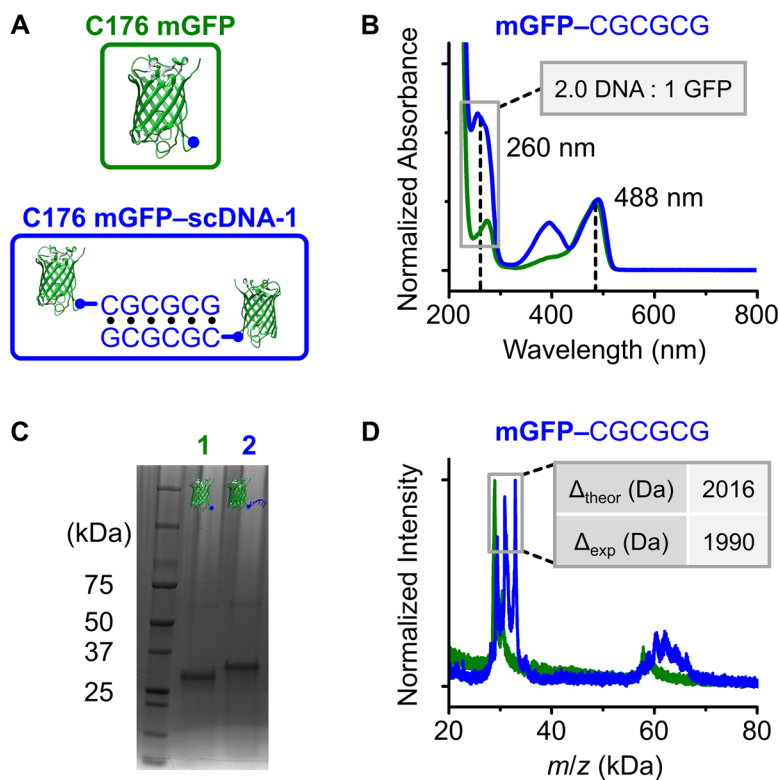
Figure 6.10. Characterization of C176 mGFP–scDNA-1 conjugates. (A) Schematics of C176 mGFP (green) and C176 mGFP–scDNA-1 (blue). (B) UV-vis absorption spectra of C176 mGFP (green) and C176 mGFP–scDNA-1 (blue) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C176 mGFP–scDNA-1 compared to C176 mGFP corresponds to 2.0 scDNA-1 per C176 mGFP. The ratio of scDNA-1 to C176 mGFP is high because some of the chromophores in C176 mGFP–scDNA-1 were protonated as indicated by the absorbance peak at 395 nm. (C) SDS-PAGE analysis of C176 mGFP (lane 1) to C176 mGFP–scDNA-1 (lane 2) show the conjugation of one scDNA-1 to C176 mGFP. (D) MALDI-TOF MS characterization of C176 mGFP (green) C176 mGFP–scDNA-1 (blue) shows an experimental mass increase of 1990 Da that is consistent with the theoretical mass increase of mono-functionalization (2016 Da = 1930 Da (scDNA-1) + 86 Da (linker molecule)).
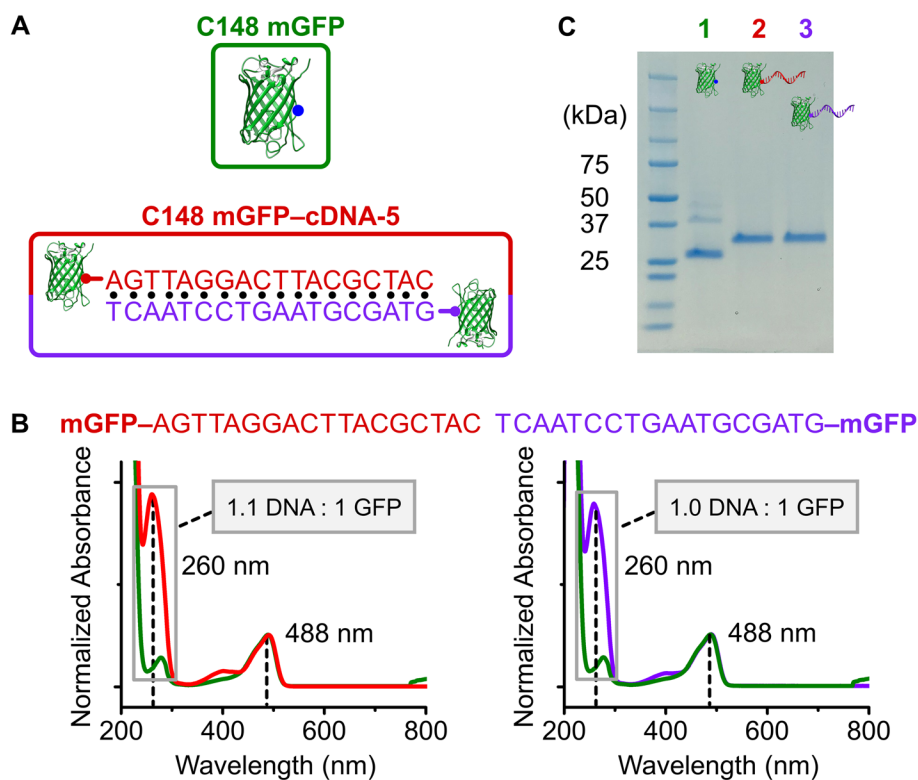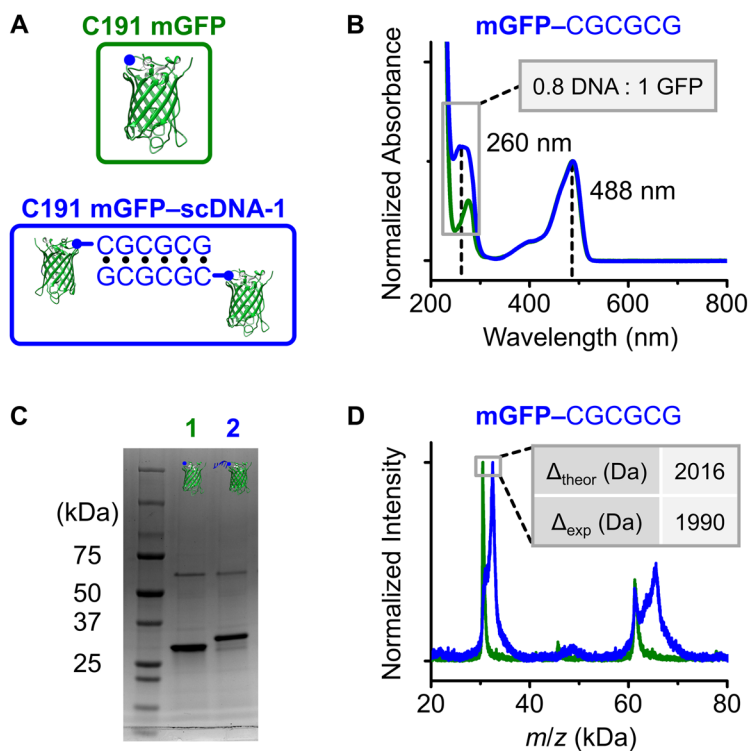
Figure 6.11. Characterization of C191 mGFP–scDNA-1 conjugates. (A) Schematics of C191 mGFP (green) and C191 mGFP–scDNA-1 (blue). (B) UV-vis absorption spectra of C191 mGFP (green) and C191 mGFP–scDNA-1 (blue) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C191 mGFP–scDNA-1 compared to C191 mGFP corresponds to 0.8 scDNA-1 per C191 mGFP. (C) SDS-PAGE analysis of C191 mGFP (lane 1) to C191 mGFP–scDNA-1 (lane 2) show the conjugation of one scDNA-1 to C191 mGFP. (D) MALDI-TOF MS characterization of C191 mGFP (green) C191 mGFP–scDNA-1 (blue) shows an experimental mass increase of 1990 Da that is consistent with the theoretical mass increase of mono-functionalization (2016 Da = 1930 Da (scDNA-1) + 86 Da (linker molecule)).
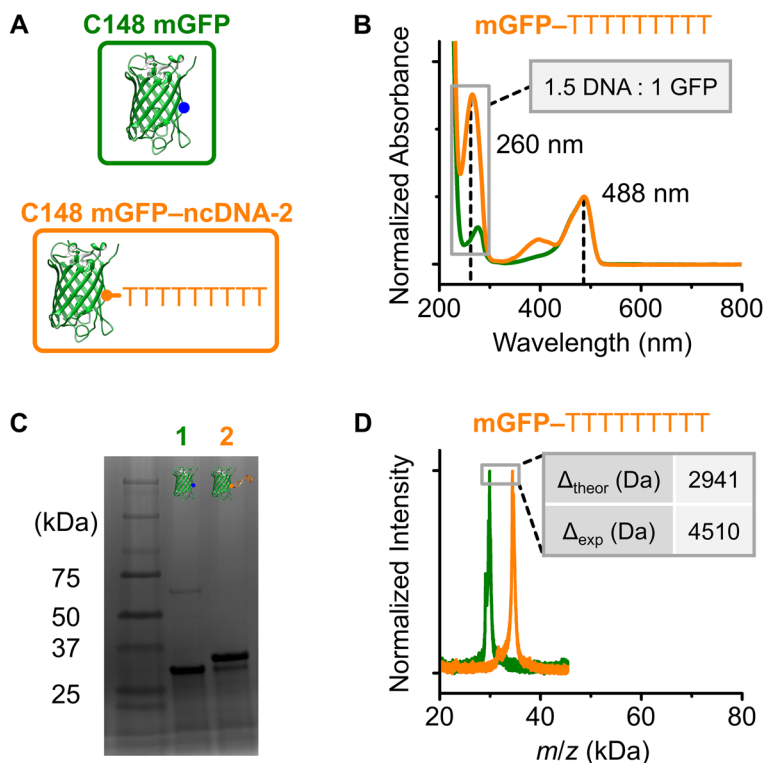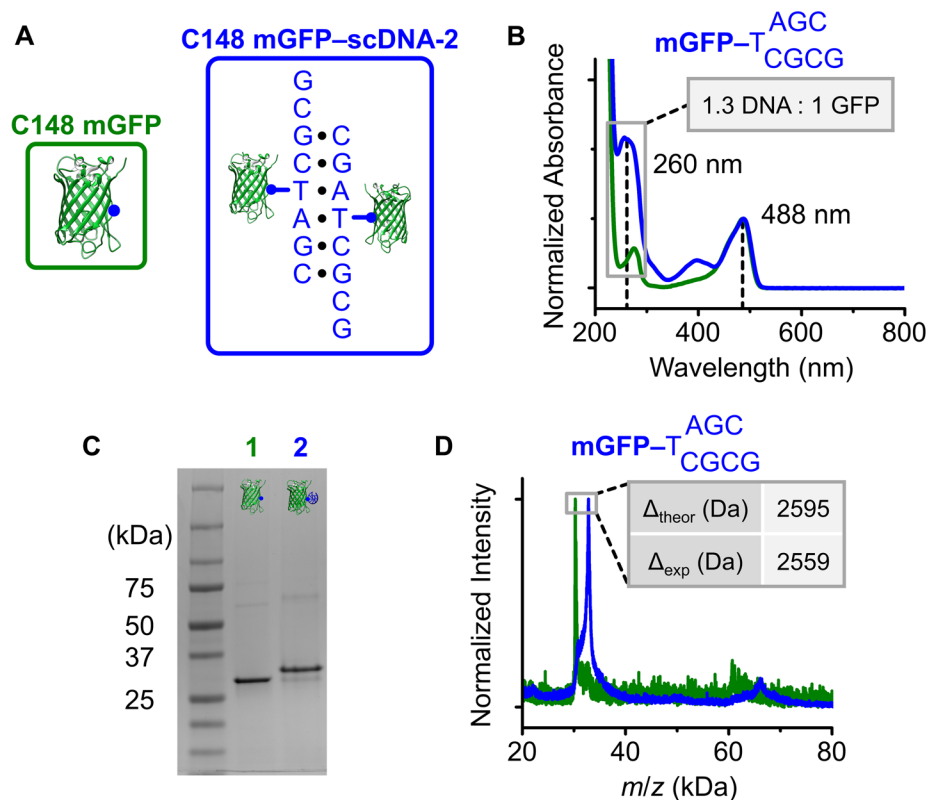
Figure 6.12. Characterization of C148 mGFP–scDNA-2 conjugates. (A) Schematics of C148 mGFP (green) and C148 mGFP–scDNA-1 (blue). (B) UV-vis absorption spectra of C148 mGFP (green) and C148 mGFP–scDNA-1 (blue) that are normalized to the absorbance of the mGFP chromophore at 488 nm. The increase in absorbance at 260 nm of C148 mGFP–scDNA-1 compared to C148 mGFP corresponds to 1.3 scDNA-1 per C148 mGFP. (C) SDS-PAGE analysis of C148 mGFP (lane 1) to C148 mGFP–scDNA-1 (lane 2) show the conjugation of one scDNA-1 to C148 mGFP. The ratio of scDNA-1 to C148 mGFP is high because some of the chromophores in C148 mGFP–scDNA-1 were protonated as indicated by the absorbance peak at 395 nm. (D) MALDI-TOF MS characterization of C148 mGFP (green) C148 mGFP–scDNA-1 (blue) shows an experimental mass increase of 2559 Da that is consistent with the theoretical mass increase of mono-functionalization (2595 Da = 2509 Da (scDNA-1) + 86 Da (linker molecule)).

6.1.4. Protein Crystallization and X-ray Crystallography.

Using 30 kDa cutoff Amicon® Ultra-15 Centrifugal Filter Units (Millepore Sigma), all mGFP–DNA conjugates were buffer exchanged 4 times to 10 mM Tris with 137 mM NaCl and concentrated to 5 mg/mL (protein concentration). High-throughput sitting drop vapor diffusion experiments were set up with Crystal Gryphon (Art Robbins Instruments) or mosquito® crystal (TTP Labtech) liquid handlers in 96-3 well INTELLI-PLATE® trays (Art Robbins Instruments). The reservoirs consisted of 70 µL of crystallization condition and the sitting drops consisted of 1 µL of sample and 1 µL of crystallization condition. Crystallization conditions from the PACT, JCSG+, Classics II, and PEGs II Suites (Qiagen) were screened. These condition suites vary salt identity and concentration, buffer identity and concentration, pH, and precipitant identity and concentration. Crystallization experiments at both 4 and 22 °C proceeded for 2 weeks undisturbed. Obtained crystals were transferred to nylon loops and frozen in liquid nitrogen. X-ray diffraction experiments were performed at the Life Sciences Collaborative Access Team beamlines 21-ID-D, 21-ID-F, and 21-ID-G at the Advanced Photon Source, Argonne National Laboratory.

Diffraction data were processed with programs run through Xia2[250-251] or programs from the CCP4 software suite.[252] Data were indexed and integrated with iMosflm,[253] and space group and unit cell parameters were confirmed with Pointless.[254] Data was scaled and merged with SCALA using resolution cut-off criteria for the highest-resolution bin (a mean $I/\sigma(I)$ of approximately 1.5 and a R_{sym} value of approximately 0.7).[255] Structures were determined by molecular replacement with PhaserMR,[256] using GFP (5N9O or 4EUL) as the starting model.[186, 257] After successive rounds of manual model building and addition of water molecules with Coot[258] and refinement with Refmac5,[259] structures were deemed finalized when R_{work}/R_{free} values

plateaued. For the solved crystal structures, there are 32–38 unmodeled N-terminal residues and 7–9 unmodeled C-terminal residues due to poor electron density for these regions. Analogous residues in other GFP crystal structures are disordered and unmodeled.[186] Protein and water B-factor analyses were performed using the bavarage module in the CCP4 software suite. Graphics for protein crystal structures were generated using PyMOL,[260] UCSF Chimera,[261] and QuteMol.[262]

Crystallization conditions for all samples are disclosed in Tables 6.1 and 6.2. Samples were screened over the same set of crystallization conditions. Each sample crystallizes in different conditions, indicating contributions from the specific condition. However, the changes in protein packing within single crystals as a function of the DNA design indicates that DNA influences protein packing within the crystals irrespective of specific crystallization conditions. Summaries of crystallographic data can be found in Tables 6.3–6.5 and refinement summaries can be found in Tables 6.6–6.8.

Table 6.1. Crystallization conditions for mGFP and mGFP–DNA crystals.

Crystal Number	Sample	Condition Name	Chemical Components of Screen Condition	Incubation Temp. (°C)
1	*C148 mGFP*	*PEGs II Suite F8*	*0.1 M sodium acetate, 25 % (w/v) PEG 4000, 8 % (w/v) isopropanol*	*22*
2	*C176 mGFP*	*JCSG+ Suite F8*	*2.1 M DL-malic acid pH 7.0*	*22*
3	*C148 mGFP–scDNA-1*	*PACT Suite A5*	*0.1 M SPG buffer pH 8, 25 % (w/v) PEG 1500*	*22*
4	C148 mGFP–scDNA-1	PACT Suite C9	0.2 M lithium sulfate, 0.1 M HEPES buffer pH 7, 20 % (w/v) PEG 6000	22
5	C148 mGFP–scDNA-1	JCSG+ Suite H7	0.1 M ammonium sulfate, 0.1 M bis-Tris buffer pH 5.5, 25 % (w/v) PEG 3350	22
6	C148 mGFP–scDNA-1	JCSG+ Suite H9	0.2 M lithium sulfate, 0.1 M bis-Tris buffer pH 5.5, 25 % (w/v) PEG 3350	22
7	C148 mGFP–scDNA-1	JCSG+ Suite H10	0.2 M ammonium sulfate 0.1 M bis-Tris buffer pH 5.5, 25 % (w/v) PEG 3350	22
8	C148 mGFP–scDNA-1	PEGs II Suite G2	0.2 M ammonium sulfate, 0.1 M MES buffer pH 6.5, 30 % (w/v) PEG 5000 MME	4
9	*C148 mGFP and scDNA-1*	*PEGs II Suite A5*	*0.1 M sodium acetate, 25 % (w/v) PEG 4000, 8 % (w/v) isopropanol*	*22*
10	*C148 mGFP–cDNA-1*	*JCSG+ Suite H9*	*0.2 M lithium sulfate, 0.1 M bis-Tris buffer pH 5.5, 25 % (w/v) PEG 3350*	*22*
11	C148 mGFP–cDNA-1	JCSG+ Suite A5	0.2 M magnesium formate, 20% (w/v) PEG 3350	22
12	C148 mGFP–cDNA-1	PACT Suite B4	0.1 M MIB buffer pH 7, 25 % (w/v) PEG 1500	22
13	C148 mGFP–cDNA-1	PACT Suite D3	0.1 M MMT buffer pH 6, 25 % (w/v) 1500	22
14	C148 mGFP–scDNA-1	PACT Suite E1	0.2 M sodium fluoride, 20 % (w/v) PEG 3350	22
15	*C148 mGFP–cDNA-2*	*JCSG+ Suite G10*	*0.15 M potassium bromide, 30 % (w/v) PEG MME 2000*	*22*
16	*C148 mGFP–ncDNA-1*	*Classics II Suite F10*	*0.2 M sodium chloride, 0.1 M bis-Tris buffer pH 5-5, 25 % (w/v) PEG 3350*	*4*
17	*C148 mGFP–cDNA-3*	*PEGs II Suite D7*	*0.1 M MES buffer pH 6.5, 30 % (w/v) PEG 4000*	*22*
18	*C148 mGFP–scDNA-2*	*Classics II Suite H12*	*0.15 M potassium bromide, 30 % (w/v) PEG 2000 MME*	*22*

For each sample, conditions led to crystals with the same space group and unit cell parameters. Conditions with the highest resolution (italicized) were processed to give the crystal structures **6UHJ**–**6UHR**. Unit cell parameters for other conditions (not italicized) are based on preliminary analysis. The crystallization temperature and unit cell parameters for each crystal are listed in Table 6.2.

Table 6.2. Unit cell parameters for mGFP and mGFP–DNA crystals.

Crystal Number	Space Group	Cell Parameters (Å)	Cell Parameters (°)
1	*P2₁2₁2₁*	*52, 63, 69*	*90, 90, 90*
2	*I222*	*89, 92, 152*	*90, 90, 90*
3	*P12₁1*	*65, 52, 87*	*90, 94, 90*
4	*P12₁1*	64, 52, 91	90, 97, 90
5	*P12₁1*	64, 52, 87	90, 94, 90
6	*P12₁1*	65, 52, 87	90, 94, 90
7	*P12₁1*	66, 52, 87	90, 95, 90
8	*P12₁1*	65, 52, 91	90, 97, 90
9	*P2₁2₁2₁*	*58, 62, 135*	*90, 90, 90*
10	*P12₁1*	*65, 52, 87*	*90, 94, 90*
11	*P12₁1*	65, 52, 88	90, 92, 90
12	*P12₁1*	65, 52, 87	90, 94, 90
13	*P12₁1*	64, 53, 86	90, 94, 90
14	*P12₁1*	65, 51, 87	90, 94, 90
15	*P12₁1*	*65, 52, 86*	*90, 94, 90*
16	*P12₁1*	59, 52, 100	90, 107, 90
17	*C121*	*107, 51, 57*	*90, 110, 90*
18	*P2₁2₁2₁*	*51, 51, 209*	*90, 90, 90*

For each sample, conditions led to crystals with the same space group and unit cell parameters. Conditions with the highest resolution (italicized) were processed to give the crystal structures **6UHJ–6UHR**. Unit cell parameters for other conditions (not italicized) are based on preliminary analysis. All unit cell parameters were rounded for consistency.

Table 6.3. Crystallographic data summary for 6UHJ, 6UHK, 6UHL, and 6UHM.

Sample	C148 mGFP (Thiol Form)	C176 mGFP (Disulfide Form)	C148 mGFP–scDNA-1 (6 bp)	C148 mGFP and scDNA-1 (6 bp)
PDB Code	**6UHJ**	**6UHK**	**6UHL**	**6UHM**
Cell parameters (Å)	51.51, 62.90, 69.40	88.93, 91.76, 151.71	64.87, 52.29, 86.80	58.28, 61.76, 135.32
Cell parameters (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 94.13, 90.00	90.00, 90.00, 90.00
Space group	$P2_12_12_1$	$I222$	$P12_11$	$P2_12_12_1$
Crystallization condition	PEGs II Suite F8 at 22°C [0.1 M sodium acetate, 25 % (w/v) PEG 4000, 8 % (w/v) isopropanol]	JCSG+ Suite F8 at 22°C [2.1 M DL-malic acid pH 7.0]	PACT Suite A5 at 22°C [0.1 M SPG buffer pH 8, 25 % (w/v) PEG 1500]	PEGs II Suite F8 at 22°C [0.1 M sodium acetate, 25 % (w/v) PEG 4000, 8 % (w/v) isopropanol]
Resolution range (Å)	51.51-1.50 (1.58-1.50)	78.51-1.90 (2.00-1.90)	43.29-1.91 (1.96-1.91)	61.76-2.10 (2.21-2.10)
Wavelength (Å)	0.97857	0.97872	0.97872	0.97872
Observed reflections	163324	727590	435564	250177
Unique reflections	36338	96006	43945	29215
Redundancy	4.5 (3.9)	7.6 (7.5)	9.9 (5.8)	8.6 (8.7)
Completeness (%)	98.8 (98.0)	100.0 (99.9)	96.9 (67.2)	99.8 (99.4)
Mean (I/σ(I))	10.7 (3.6)	14.7 (3.8)	12.1 (1.3)	14.5 (3.7)
R_{sym} (%)	0.083 (0.347)	0.082 (0.515)	0.117 (1.189)	0.074 (0.503)

Numbers in parentheses refer to the highest-resolution shell. $R_{sym} = \Sigma_h \Sigma_i |I_1(h) - <I(h)|/\Sigma_h \Sigma_i I_1(h)$.

Table 6.4. Crystallographic data summary for 6UHN, 6UHO, 6UHP, and 6UHQ.

Sample	C148 mGFP–cDNA-1 (6 bp)	C148 mGFP–cDNA-2 (6 bp)	C148 mGFP–ncDNA-1 (6 bp)	C148 mGFP–cDNA-3 (9 bp)
PDB Code	**6UHN**	**6UHO**	**6UHP**	**6UHQ**
Cell parameters (Å)	64.92, 52.18, 86.47	64.71, 52.22, 86.44	59.05, 51.60, 100.41	106.63, 50.58, 56.69
Cell parameters (°)	90.00, 94.24, 90.00	90.00, 94.23, 90.00	90.00, 106.97, 90.00	90.00, 110.33, 90.00
Space group	*P12$_1$1*	*P12$_1$1*	*P12$_1$1*	*C121*
Crystallization condition	JCSG+ Suite H9 at 22°C [0.2 M lithium sulfate, 0.1 M bis-Tris buffer pH 5.5, 25 % (w/v) PEG 3350]	JCSG+ Suite G10 at 22°C [0.15 M potassium bromide, 30 % (w/v) PEG MME 2000]	Classics II Suite F10 at 4°C [0.2 M sodium chloride, 0.1 M bis-Tris buffer pH 5-5, 25 % (w/v) PEG 3350]	PEGs II Suite D7 at 22°C [0.1 M MES buffer pH 6.5, 30 % (w/v) PEG 4000]
Resolution range (Å)	64.07-1.92 (2.02-1.92)	64.53-1.95 (2.06-1.95)	56.48-2.90 (3.06-2.90)	53.16-2.85 (3.00-2.85)
Wavelength (Å)	0.97872	0.97872	0.97857	0.97872
Observed reflections	192150	179992	66203	21886
Unique reflections	44394	41971	13070	6723
Redundancy	4.3 (4.3)	4.3 (4.3)	5.1 (5.2)	3.3 (3.2)
Completeness (%)	100.0 (99.9)	99.2 (100.0)	99.8 (100.0)	99.3 (99.2)
Mean (I/σ(I))	7.8 (3.1)	9.4 (3.3)	7.9 (2.8)	5.8 (2.1)
R$_{sym}$ (%)	0.112 (0.442)	0.094 (0.406)	0.147 (0.657)	0.178 (0.677)

Numbers in parentheses refer to the highest-resolution shell. $R_{sym} = \Sigma_h \Sigma_i |I_1(h) - <I(h)|/\Sigma_h \Sigma_i I_1(h)$.

Table 6.5. Crystallographic data summary for 6UHR.

Sample	C148 mGFP–scDNA-2 (8 bp, int. DNA attach.)
PDB Code	**6UHR**
Cell parameters (Å)	50.58, 50.89, 209.19
Cell parameters (°)	90.00, 90.00, 90.00
Space group	$P2_12_12_1$
Crystallization condition	Classics II Suite H12 at 22°C [0.15 M potassium bromide, 30 % (w/v) PEG 2000 MME]
Resolution range (Å)	69.73-3.00 (3.16-3.00)
Wavelength (Å)	1.12710
Observed reflections	79211
Unique reflections	11425
Redundancy	6.9 (7.1)
Completeness (%)	99.7 (100.0)
Mean (I/σ(I))	8.1 (2.7)
R_{sym} (%)	0.194 (0.716)

Numbers in parentheses refer to the highest-resolution shell. $R_{sym} = \Sigma_h \Sigma_i |I_1(h) - <I(h)|/\Sigma_h \Sigma_i I_1(h)$.

Table 6.6. Model refinement summary for 6UHJ, 6UHK, 6UHL, and 6UHM.

Sample	C148 mGFP (Thiol Form)	C176 mGFP (Disulfide Form)	C148 mGFP–scDNA-1 (6 bp)	C148 mGFP and scDNA-1 (6 bp)
PDB Code	**6UHJ**	**6UHK**	**6UHL**	**6UHM**
Resolution range (Å)	46.61-1.50 (1.54-1.50)	75.97-1.90 (1.95-1.90)	43.29-1.91 (1.96-1.91)	56.25-2.10 (2.16-2.10)
No. of reflections	36338	96006	43939	29157
R factor	15.3	18.6	22.8	21.1
R_{free}	18.5	22.9	27.3	27.0
RMSD bond lengths (Å)	0.013	0.012	0.0094	0.008
RMSD bond angles (°)	1.88	1.86	1.78	1.67
Average B value protein ($Å^2$)	10.5	32.4	21.1	52.2
Average B value water ($Å^2$)	25.8	39.9	30.1	53.9
Ramachandran plot:				
Favored and allowed regions (%)	96.7	95.4	96.8	93.9
Generously allowed regions (%)	3.3	4.4	3.2	5.4
Disallowed regions (%)	0.0	0.2	0.0	0.7
Solvent content (%)	33	52	49	38

R factor = $\Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$. R free is calculated using the same equation as that for R factor, but 5.0 % of reflections were chosen randomly and omitted from the refinement.

Table 6.7. Model refinement summary for 6UHN, 6UHO, 6UHP, and 6UHQ.

Sample	C148 mGFP–cDNA-1 (6 bp)	C148 mGFP–cDNA-2 (6 bp)	C148 mGFP–ncDNA-1 (6 bp)	C148 mGFP–cDNA-3 (9 bp)
PDB Code	**6UHN**	**6UHO**	**6UHP**	**6UHQ**
Resolution range (Å)	64.60-1.92 (1.97-1.92)	64.53-1.95 (2.00-1.95)	56.48-2.90 (2.98-2.90)	50.04-2.85 (2.92-2.85)
No. of reflections	44378	41801	12900	6722
R factor	21.6	21.4	34.1	18.1
R_{free}	24.7	25.1	37.3	27.4
RMSD bond lengths (Å)	0.011	0.010	0.007	0.007
RMSD bond angles (°)	1.83	1.78	1.63	1.75
Average B value protein ($Å^2$)	24.5	27.1	55.4	33.3
Average B value water ($Å^2$)	28.6	36.0	*n/a*	24.0
Ramachandran plot:				
Favored and allowed regions (%)	96.5	95.9	82.3	92.8
Generously allowed regions (%)	3.5	3.6	12.3	7.2
Disallowed regions (%)	0.0	0.5	5.5	0.0
Solvent content (%)	49	48	49	48

R factor = $\Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$. R free is calculated using the same equation as that for R factor, but 5.0 % of reflections were chosen randomly and omitted from the refinement.

Table 6.8. Model refinement summary for 6UHR.

Sample	C148 mGFP–scDNA-2 (8 bp, int. DNA attach.)
PDB Code	**6UHR**
Resolution range (Å)	52.30-3.00 (3.08-3.00)
No. of reflections	11431
R factor	22.6
R_{free}	30.9
RMSD bond lengths (Å)	0.008
RMSD bond angles (°)	1.82
Average B value protein ($Å^2$)	47.4
Average B value water ($Å^2$)	40.7
Ramachandran plot:	
Favored and allowed regions (%)	84.4
Generously allowed regions (%)	11.8
Disallowed regions (%)	3.9
Solvent content (%)	44

R factor $= \Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$. R free is calculated using the same equation as that for R factor, but 5.0 % of reflections were chosen randomly and omitted from the refinement.

Figure 6.13. Extended packing arrangement in C148 mGFP crystals. The extended packing arrangement of mGFP (protein in teal, surface-exposed cysteine in red) is depicted in the C148 mGFP crystal structure (PDB code: **6UHJ**).

Figure 6.14. Extended packing arrangement in C176 mGFP crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C176 mGFP crystal structure (PDB code: **6UHK**).
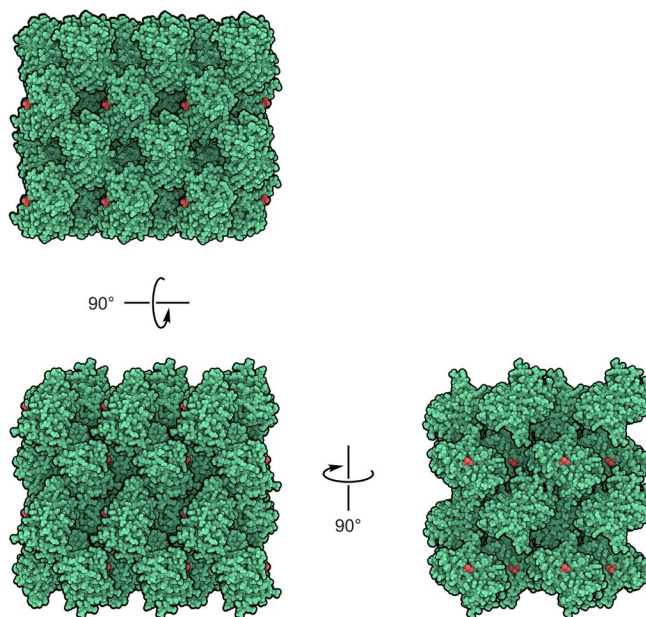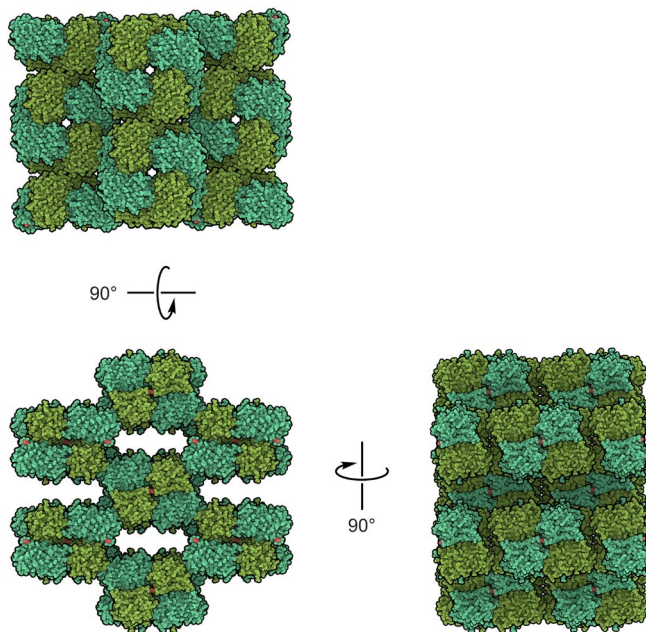
Figure 6.15. Extended packing arrangement in C148 mGFP–scDNA-1 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C148 mGFP–scDNA-1 crystal structure (PDB code: **6UHL**).

Figure 6.16. Extended packing arrangement in C148 mGFP and scDNA-1 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the unconjugated C148 mGFP and scDNA-1 crystal structure (PDB code: **6UHM**).

163

Figure 6.17. Extended packing arrangement in C148 mGFP–cDNA-1 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C148 mGFP–cDNA-1 crystal structure (PDB code: **6UHN**).

Figure 6.18. Extended packing arrangement in C148 mGFP–cDNA-2 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C148 mGFP–cDNA-2 crystal structure (PDB code: **6UHO**).
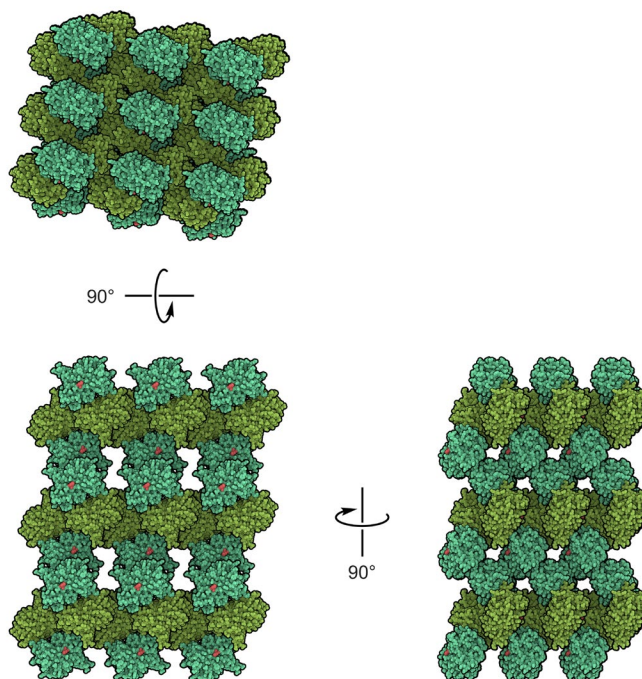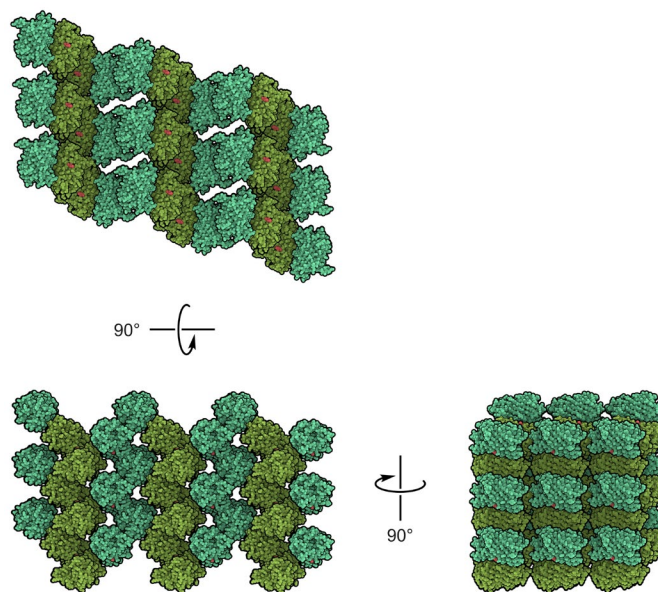
Figure 6.19. Extended packing arrangement in C148 mGFP–ncDNA-1 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C148 mGFP–ncDNA-1 crystal structure (PDB code: **6UHP**).
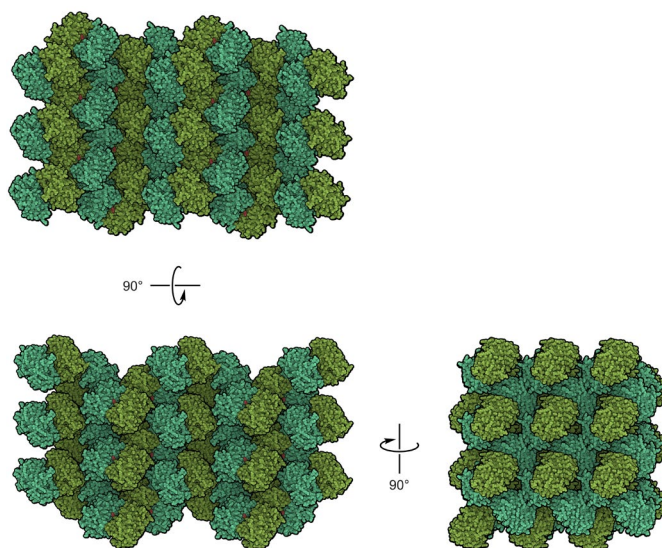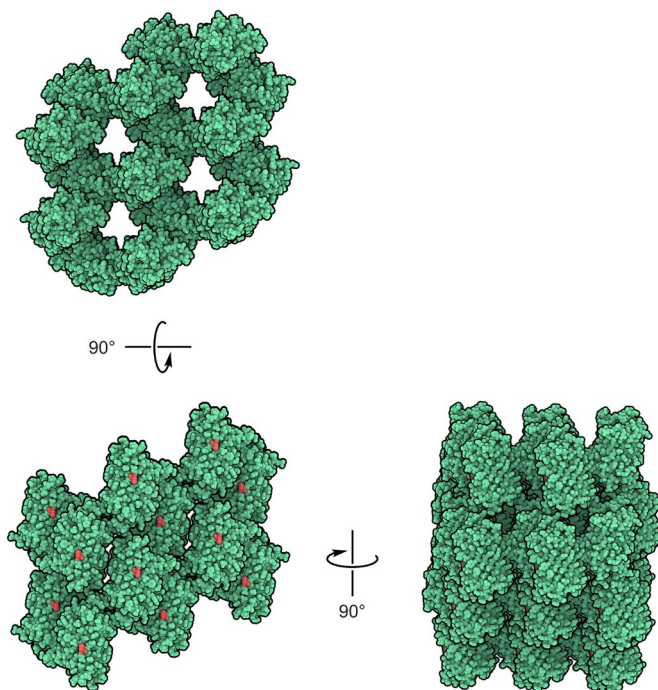

Figure 6.20. Extended packing arrangement in C148 mGFP–cDNA-3 crystals. The extended packing arrangement of mGFP (proteins in teal, surface-exposed cysteine in red) is depicted in the C148 mGFP–cDNA-3 crystal structure (PDB code: **6UHQ**).
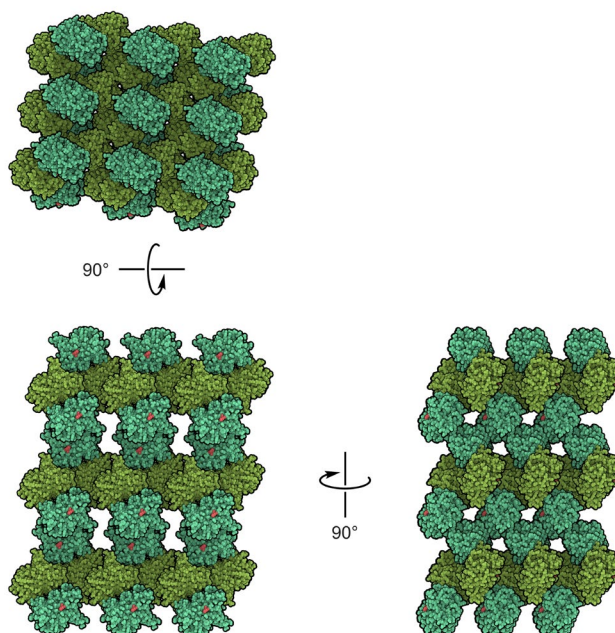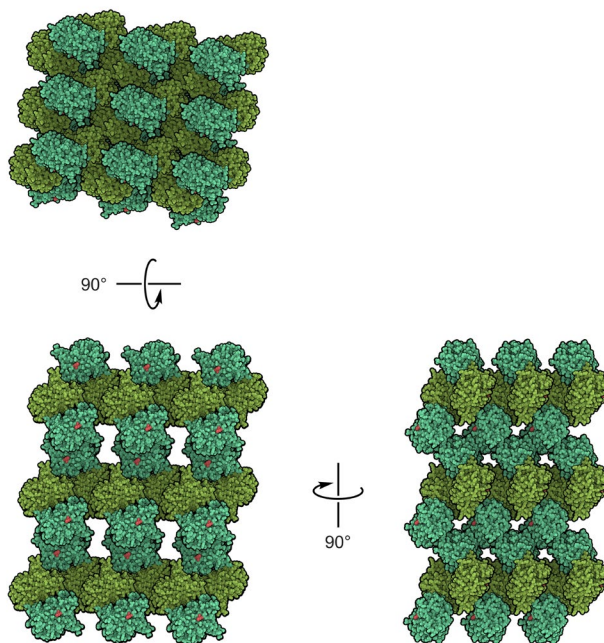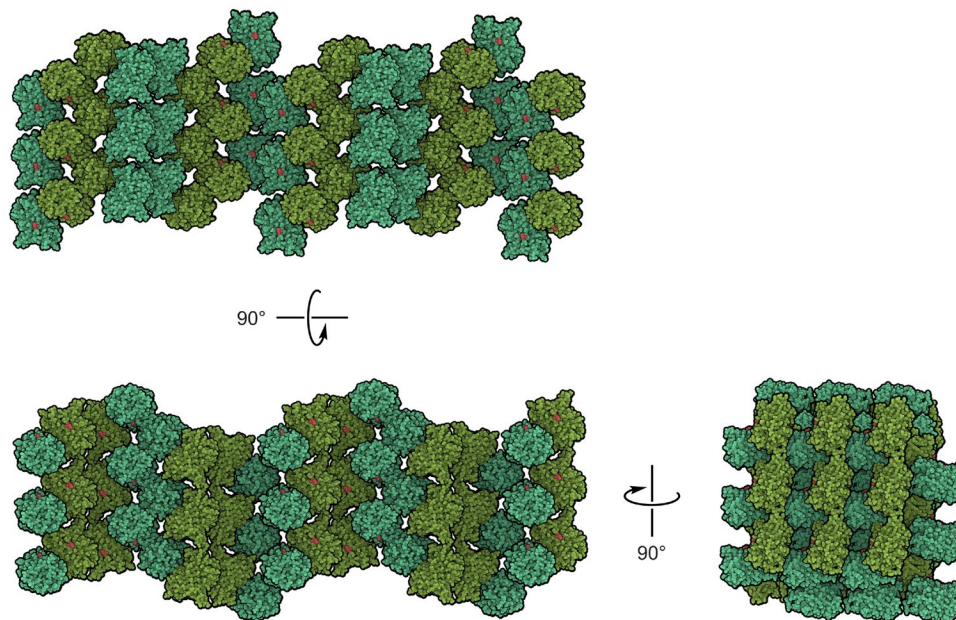
Figure 6.21. Extended packing arrangement in C148 mGFP–scDNA-2 crystals. The extended packing arrangement of mGFP (proteins in teal and green, surface-exposed cysteine in red) is depicted in the C148 mGFP–scDNA-2 crystal structure (PDB code: **6UHR**).

6.1.5. Confocal Microscopy Analysis of mGFP and mGFP–DNA Crystals.

Crystals were transferred from sitting drops to a 7 µL drop of crystallization condition on a confocal microscopy dish. Crystals were imaged with a Nikon A1R confocal microscope using a 20x objective with bright field and two laser channels. The first channel for the mGFP chromophore (488 nm excitation maximum, 509 emission maximum) was excited with a 485 nm laser and had an emission filter of 500–550 nm. The second channel for the DNA intercalating dye, TOTO-3, (642 nm excitation maximum, 662 emission maximum)[188-189] was excited with a 640 nm laser and had an emission filter of 663–738 nm. After imaging the crystals, TOTO-3 (1 mM in DMSO, Biotium) was diluted to 0.1 mM in 10 mM Tris with 137 mM NaCl and 0.5 µL of the diluted dye was added to the drop containing the crystals. After waiting 30 minutes for the dye to diffuse through the crystals, the crystals were imaged again with the same bright field and laser channels.

6.2. Appendix B: Materials and Methods for Chapter Three.

6.2.1. Design, Synthesis, and Characterization of DNA.

A set of six DNA strands (Table 3.1) was designed to constitute a modular DNA scaffold (Table 3.1). Three DNA strands were designed for conjugation to proteins (Scheme 3.1.A, *i.e.*, S2–DBCO–Cy3–S3, S4–DBCO–Cy5–S5, and S6–DBCO–FITC–S1) and three DNA strands were designed to template the assembly of protein–DNA conjugates (Scheme 3.1.B, *i.e.*, S1'–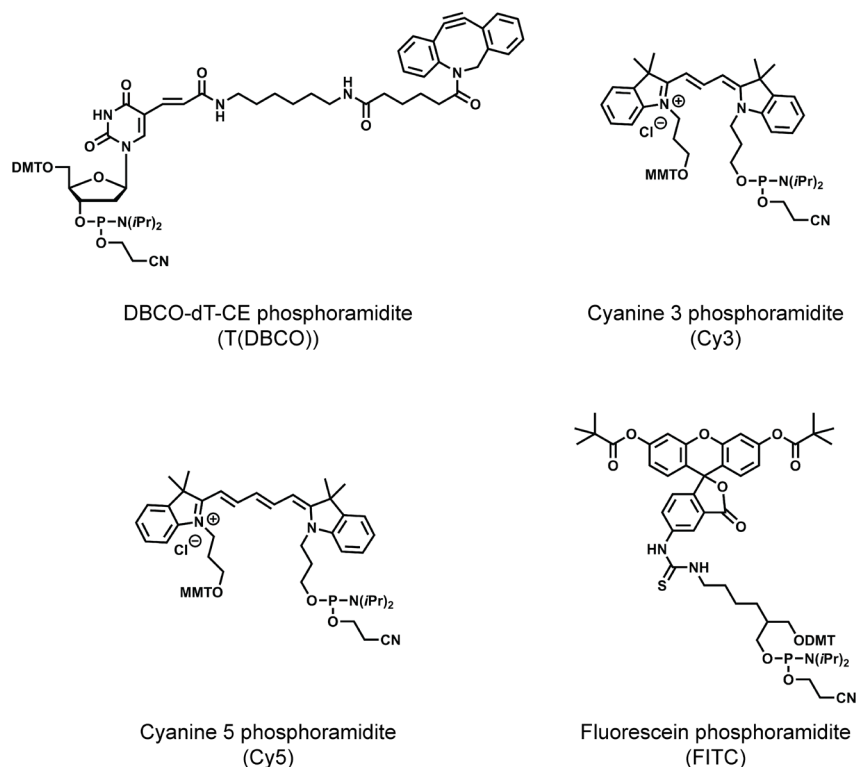S2', S3'–S4', and S5'–S6'). Stars represent dye modifications to DNA strands in the scheme (*i.e.*, red stars represent Cyanine 3 (Cy3), blue stars represent Cyanine 5 (Cy5), and green stars represent fluorescein (FITC)). Template DNA strands were designed to include 8-base toeholds to facilitate the removal of template DNA strands from oligomers via toehold-mediated strand displacement reactions, however, these reactions were not utilized in this work. DNA–DNA interactions between protein conjugation strands and template strands are used to organize proteins that are functionalized with DNA into sequence-encoded oligomers (Scheme 3.1.C, Table 3.3). Labels on the DNA strands represent distinct nucleic acid sequences and pairs of sequences that have and lack the prime symbol are complementary and will interact specifically (*e.g.*, S3 and S3' are complementary DNA sequences and S4 and S4' are complementary DNA sequences). Complementary DNA sequences were designed with melting temperatures between 60 and 70 °C in 1× PBS to facilitate assembly at room temperature at micromolar concentrations.

All materials for oligonucleotide synthesis were obtained from Glen Research and were used as received. Oligonucleotides were synthesized on a 5 μmol scale on a Bio Automation MerMade 12 oligonucleotide synthesizer or an ABI 394 DNA synthesizer using standard solid phase synthetic protocols on controlled pore glass (CPG) beads. Successfully synthesized

oligonucleotides were isolated using reverse-phase HPLC on an Agilent Technologies 1260 Infinity II HPLC using an Agilent Dynamax Microsorb 300-10 C4 or Agilent Dynamax Microsorb 300-5 C18 column. Successful synthesis of desired oligonucleotides was confirmed by MALDI-TOF MS. Samples were mixed with a 2′,6′-dihydroxyacetophenone matrix containing diammonium hydrogen citrate and characterized using a Bruker MALDI Rapiflex Tissue Typer instrument in linear mode with negative ion detection. Oligonucleotide concentrations were quantified by measuring the absorbances of oligonucleotide samples at 260 nm within the linear range on an Agilent Technologies Cary 60 UV-vis spectrophotometer.

Scheme 6.3. Structures of T(DBCO), Cy3, Cy5, and FITC phosphoramidites.

DBCO-dT-CE phosphoramidite
(T(DBCO))

Cyanine 3 phosphoramidite
(Cy3)

Cyanine 5 phosphoramidite
(Cy5)

Fluorescein phosphoramidite
(FITC)

These phosphoramidites contain 4,4'-dimethoxytrityl (DMT), monomethoxytrityl (MMT), *N,N*-diisopropyl (N(*i*Pr)$_2$), ß-cyanoethyl, and acetyl protecting groups that are removed during DNA workup and purification.

6.2.2. Synthesis and Characterization of Protein–DNA Conjugates.

InVivoMAb **anti-mouse-PD-1** (CD279, clone RMP1-14), *InVivoMAb* **anti-mouse-TIGIT** (clone 1G9), *InVivoMAb* **anti-mouse-CTLA-4** (CD152, clone 9H10), and *InVivoSIM* **anti-human-PD-1** (Nivolumab Biosimilar) were obtained from Bio X Cell and used as received. To install a single DNA strand onto antibodies (*i.e.*, **anti-mouse-PD-1 (A)**, **anti-mouse-TIGIT (B)**, **anti-mouse-CTLA-4 (C)**, and **anti-human-PD-1 (D)**), 1 equiv. of antibody was added to 2 equiv. of a molecule containing an *N*-hydroxysuccinimide activated ester and an azide (NHS–PEG$_{12}$–N$_3$, Quanta Biodesign) at an antibody concentration of 35–50 µM in 10 mM TRIS:HCl (pH 7.4) buffer containing 500 mM NaCl and 10% (v/v) glycerol.[199] The reaction was allowed to proceed for 45 min at room temperature. Next, excess unreacted NHS–PEG$_{12}$–N$_3$ ester was removed from the reaction mixture using a Cytiva NAP DNA purification column that was equilibrated in 10 mM TRIS:HCl (pH 7.4) buffer containing 500 mM NaCl and 10% (v/v) glycerol. To functionalize azide-modified antibodies with DNA, 1 equiv. of azide-modified antibodies was added to 5 equiv. of DBCO-modified DNA (DNA strands for protein conjugation, *i.e.*, S2–DBCO–Cy3–S3, S4–DBCO–Cy5–S5, and S6–DBCO–FITC–S1) and the resulting strain promoted azide-alkyne cycloaddition was allowed to proceed for 16 h at room temperature.

Excess unreacted DNA was removed from the reaction mixture using two successive rounds of preparative SEC. The volumes of antibody–DNA conjugation reaction mixtures were reduced to ~300 µL using Amicon Ultra centrifugal filters with a 30 kDa cutoff. Next, samples were purified on a Bio-Rad ENrich^{TM} SEC 650 column on a Bio-Rad NGC Quest 10 Plus Chromatography System. SEC purification was performed in 10 mM TRIS:HCl (pH 7.4) buffer containing 500 mM NaCl at a flow rate of 0.33 mL/min.

The complete removal of DNA was confirmed using analytical SDS-PAGE. Samples were diluted with Bio-Rad Laemmi buffer and heated to 85 °C for 5 minutes. Next, samples were loaded into Bio-Rad Mini-Protean® TGX™ precast SDS-PAGE gels and gels were run in 1× Tris/Glycine/SDS buffer from Bio-Rad for 35 minutes at 200 V. Gels were washed three times with heated water and imaged for Cy3, Cy5, and FITC fluorescence on a Bio-Rad ChemiDoc MP Imaging System. Gels were then stained with Thermo Fisher Scientific SimplyBlue™ SafeStain and imaged for SimplyBlue™ SafeStain absorbance or fluorescence on a Bio-Rad ChemiDoc MP Imaging System.

Excess unreacted antibodies or antibodies functionalized with multiple DNA strands were removed using preparative anion exchange chromatography. Antibody–DNA conjugation reaction mixtures were diluted to ~50 mL in 10 mM TRIS:HCl (pH 7.4) buffer containing 10 mM NaCl. Next, samples were purified by anion exchange on a Bio-Rad NGC Quest 10 Plus Chromatography System. Samples were loaded at a flow rate of 4.0 mL/min onto a column packed with Macro-Prep DEAE resin that was obtained from Bio-Rad. Antibodies were eluted from the column with 10 mM TRIS:HCl (pH 7.4) buffer containing 200 mM NaCl at a flow rate of 5.0 mL/min. Antibody–DNA conjugates were eluted from the column using a 20 minute gradient from 10 mM TRIS:HCl (pH 7.4) buffer containing 400 mM NaCl to 10 mM TRIS:HCl (pH 7.4) buffer containing 800 mM NaCl at a flow rate of 5.0 mL/min.

Antibody and antibody–DNA conjugate samples were characterized by analytical SEC on an Agilent Technologies 1260 Infinity HPLC. Samples were run through an AdvanceBio SEC 300 Å 2.7 μm column in 1× PBS (pH 7.4) at a flow rate of 0.5 mL/min. Sample elution from the SEC column was tracked by measuring absorbance at 280 nm and fluorescence of dye modifications

(*i.e.*, Cy3 at 550/568 nm excitation (ex)/emission (em), Cy5 at 650/666 nm ex/em, and FITC at 495/525 nm ex/em). Analytical SDS-PAGE was also used to confirm the isolation of antibodies that were functionalized with a single DNA strand. Antibody–DNA conjugate samples were exchanged four times into 1× PBS (pH 7.4) buffer containing 10% (v/v) glycerol using Amicon Ultra centrifugal filters with a 30 kDa cutoff. Antibody–DNA conjugate concentrations were quantified using UV-vis absorbance spectroscopy by measuring the absorbances of dye modifications (*i.e.*, Cy3, Cy5, and FITC) on an Agilent Technologies Cary 60 UV-vis spectrophotometer.
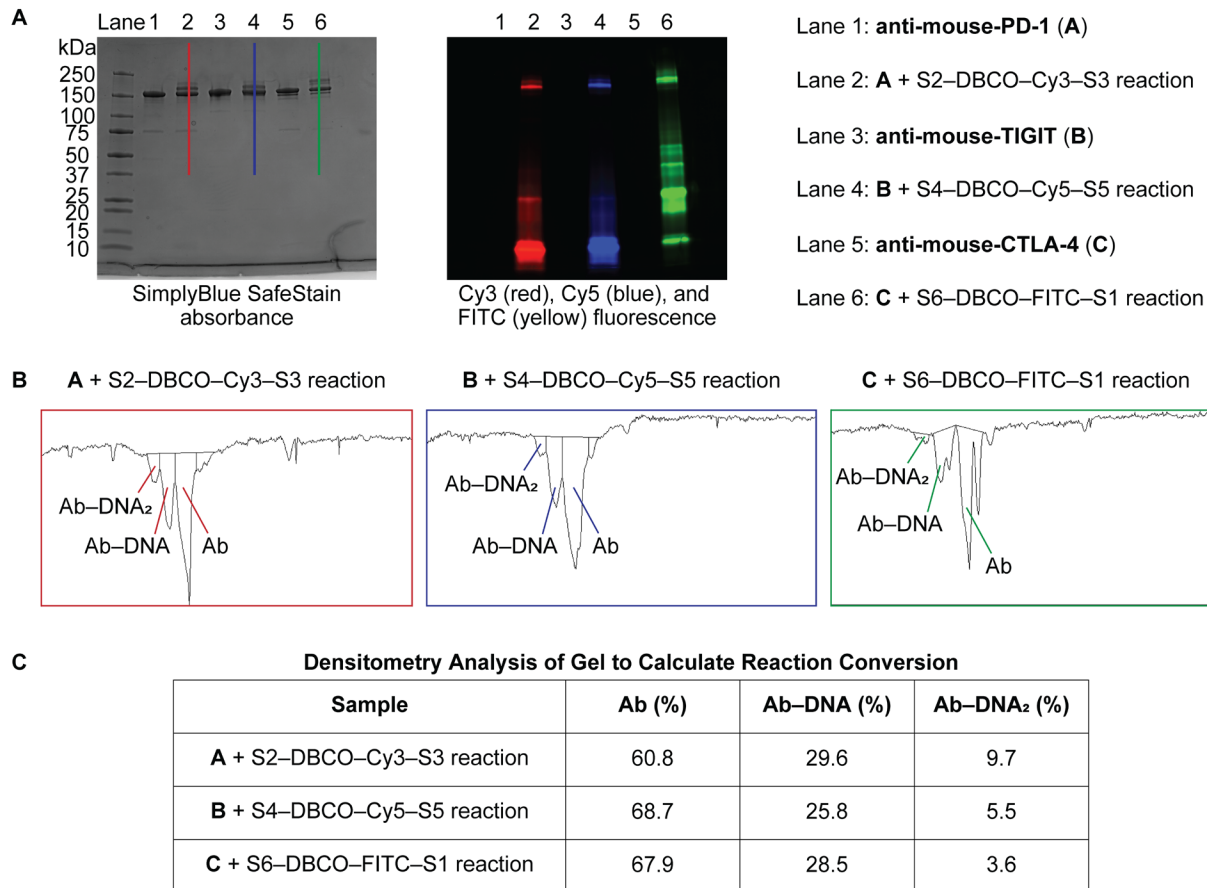
Figure 6.22. Quantifying reaction conversion of antibody–DNA conjugation. Mouse antibodies (*i.e.*, sample 1: **anti-mouse-PD-1** (**A**), sample 3: **anti-mouse-TIGIT** (**B**), and sample 5: **anti-mouse-CTLA-4** (**C**)) and the reaction mixtures of antibody–DNA conjugation (*i.e.*, sample 2: **A** + S2–DBCO–Cy3–S3, sample 4: **B** + S4–DBCO–Cy5–S5, and sample 6: **C** + S6–DBCO–FITC–S1) were characterized by SDS-PAGE. A single gel was imaged for SimplyBlue SafeStain absorbance as well as Cy3, Cy5, and FITC fluorescence. The SimplyBlue SafeStain absorbance image of the gel was studied via (B) densitometry analysis to (C) calculate reaction conversions.

Figure 6.23. Characterization of a synthesized human antibody–DNA conjugates. (A) Human antibodies (*i.e.*, lane 1: **anti-human-PD-1 (D)**) and antibody–DNA conjugates (*i.e.*, lane 2: S2–**D**–Cy3–S3, lane 3: S4–**D**–Cy5–S5, and lane 4: S6–**D**–FITC–S1) were characterized by SDS-PAGE. A single gel was imaged for SimplyBlue SafeStain, Cy3, Cy5, and FITC fluorescence. (B) The same samples were characterized by analytical SEC using absorbance at 280 nm (left) and fluorescence of DNA-conjugated fluorophores (right). In both graphs, the elution times fro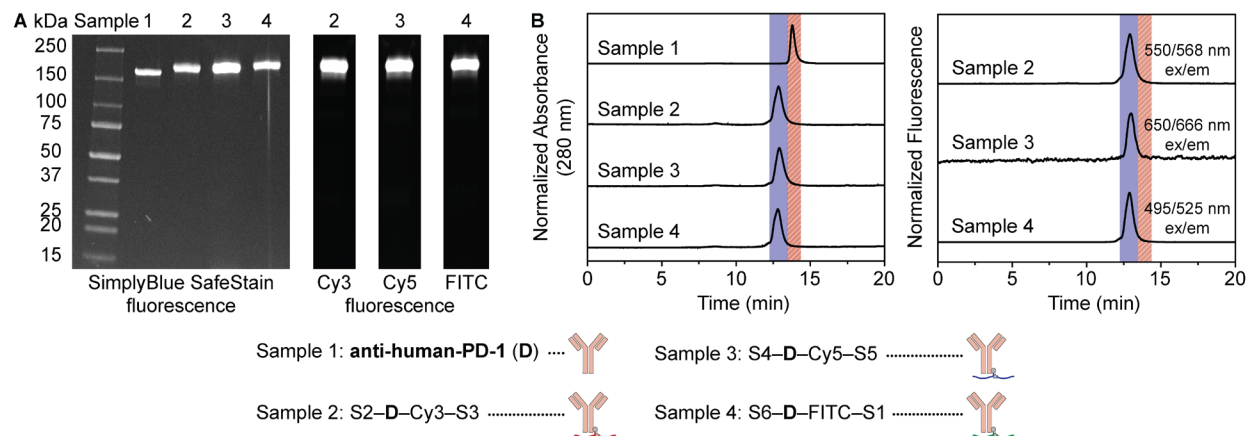m 12.25–13.50 min are highlighted in blue (shaded area) and the elution times from 13.50–14.30 min are highlighted in red (shaded area with diagonal lines).

6.2.3. Synthesis and Characterization of Protein Oligomers.

To assemble antibody–DNA conjugates into antibody dimers and sequence-encoded antibody trimers, equimolar antibody–DNA conjugates and template DNA strands with complementary DNA sequences were mixed at reagent concentrations of 1.5–2 μM in 1× PBS (pH 7.4) buffer containing 10% (v/v) glycerol. The assembly was allowed to proceed for 1 h at room temperature. The volumes of antibody dimer and sequence-encoded antibody trimer assembly mixtures were reduced to ~300 μL using Amicon Ultra centrifugal filters with a 30 kDa cutoff. Next, samples were purified on a GE Healthcare Superose™ 6 Increase 10/300 GL column on a Bio-Rad NGC Quest 10 Plus Chromatography System. SEC purification was performed in 1× PBS (pH 7.4) buffer at a flow rate of 0.20 mL/min. Samples exchanged four times into 1× PBS (pH 7.4) buffer containing 10% (v/v) glycerol using Amicon Ultra centrifugal filters with a 30 kDa cutoff. Analytical agarose gel electrophoresis and analytical SEC were used to confirm the

isolation of monodisperse, sequence-encoded antibody dimers and trimers. Gels were cast containing 2% (w/w) agarose in 1× Tris-acetate buffer containing SDS from ThermoFisher Scientific. Next, samples were loaded into gels and the gels were run in 1× Tris-acetate buffer containing SDS for 90 minutes at 120 V. Gels were imaged for Cy3, Cy5, and FITC fluorescence on a Bio-Rad ChemiDoc MP Imaging System. Antibody–DNA conjugate and antibody oligomer samples were also characterized by analytical SEC on an Agilent Technologies 1260 Infinity HPLC. Samples were run through an AdvanceBio SEC 300 Å 2.7 μm column in 1× PBS (pH 7.4) at a flow rate of 0.5 mL/min. Sample elution from the SEC column was tracked by measuring absorbance at 280 nm and fluorescence of dye modifications (*i.e.*, Cy3 at 550/568 nm ex/em, Cy5 at 650/666 nm ex/em, and FITC at 495/525 nm ex/em).

Oligomer concentrations were quantified using UV-vis absorbance spectroscopy by measuring the absorbances of Cy5 dye modifications within the linear range on an Agilent Technologies Cary 60 UV-vis spectrophotometer. Antibody dimers and trimers with complementary single stranded DNA sequences were mixed at reagent concentrations of 125 nM in 1× PBS (pH 7.4) buffer containing 10% (v/v) glycerol to target the assembly of sequence-encoded antibody pentamers. The assembly was allowed to proceed for 18 h at room temperature. Next, sequence-encoded antibody pentamers were isolated using preparative SEC and exchanged four times into 1× PBS (pH 7.4) buffer containing 10% (v/v) glycerol using Amicon Ultra centrifugal filters with a 30 kDa cutoff. Analytical agarose gel electrophoresis were used to confirm the isolation of monodisperse, sequence-encoded antibody pentamers.

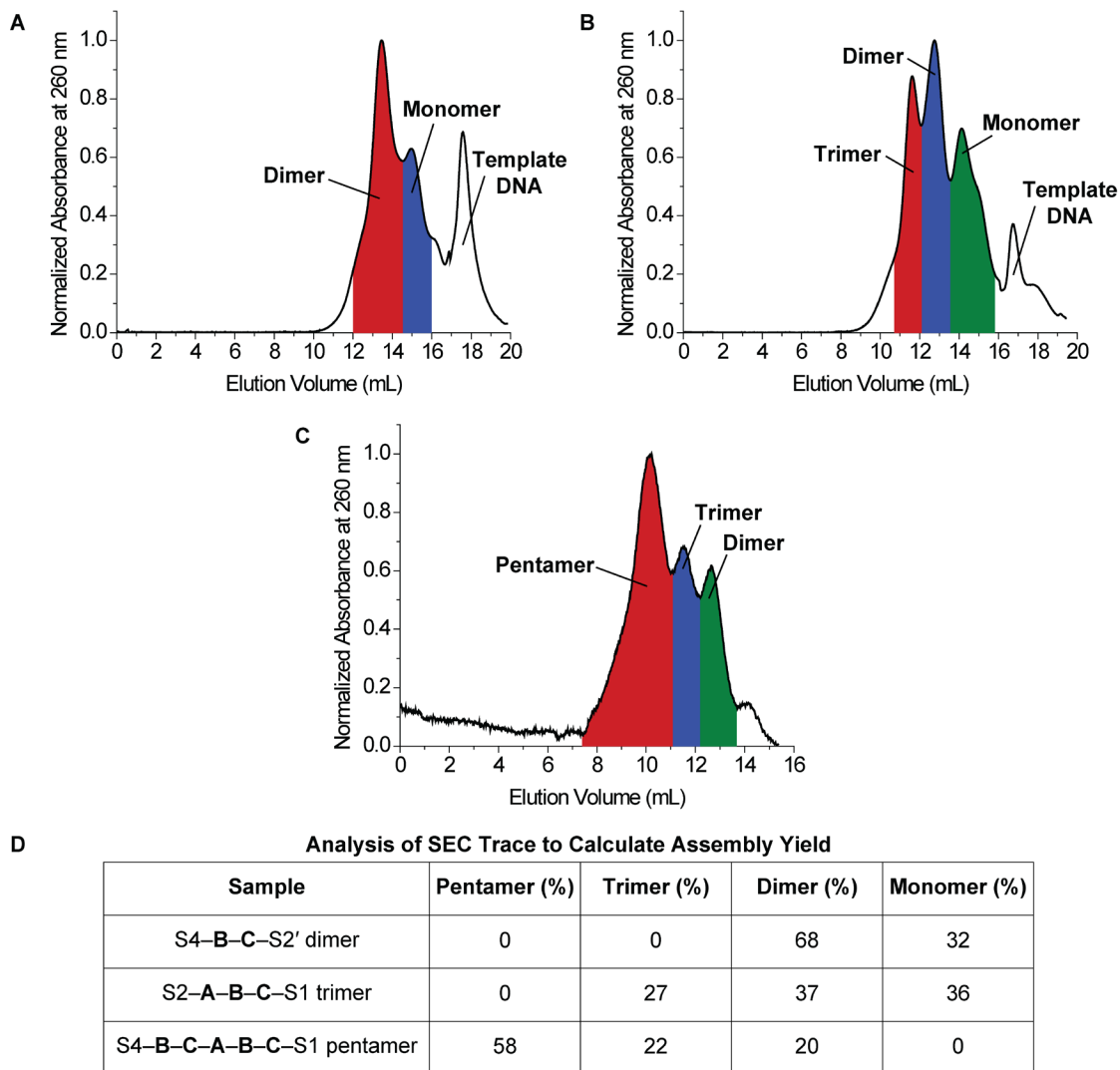**Figure 6.24. Quantifying antibody–DNA assembly yields.** Antibody–DNA assembly mixtures of (A) S4–**B**–**C**–S2' protein dimers, (B) S2–**A**–**B**–**C**–S1 protein trimers, and (C) S4–**B**–**C**–**A**–**B**–**C**–S1 protein pentamers were purified via preparative SEC using absorbance at 260 nm to track the elution of sample. (D) SEC traces were then integrated to calculate assembly yields of oligomers.

Figure 6.25. Characterization of a sequence-encoded mouse antibody trimer. (A) Mouse antibody–DNA conjugates (*i.e.*, sample 1: S2–**anti-mouse-PD-1** (**A**)–Cy3–S3, sample 2: S4–**anti-mouse-TIGIT** (**B**)–Cy5–S5, and sample 3: S6–**B**–FITC–S1) and template DNA strands (*i.e.*, S3'–S4' and S5'–S6') were assembled into a mouse antibody trimer (*i.e.*, sample 4: S2–**A**–**B**–**B**–S1) using DNA–DNA interactions. (B) Agarose gel characterization of mouse antibody–DNA conjugates and a mouse antibody trimer. A single gel was imaged for Cy3 (left), Cy5 (center), and FITC fluorescence (right). (C) Analytical SEC characterization using absorbance at 280 nm (left) and fluorescence of DNA-conjugated fluorophores (right). In both graphs, the elution times from 9.75–10.60 min are highlighted in green (shaded area with squares) and the elution times from 11.80–13.60 min are highlighted in blue (shaded area).
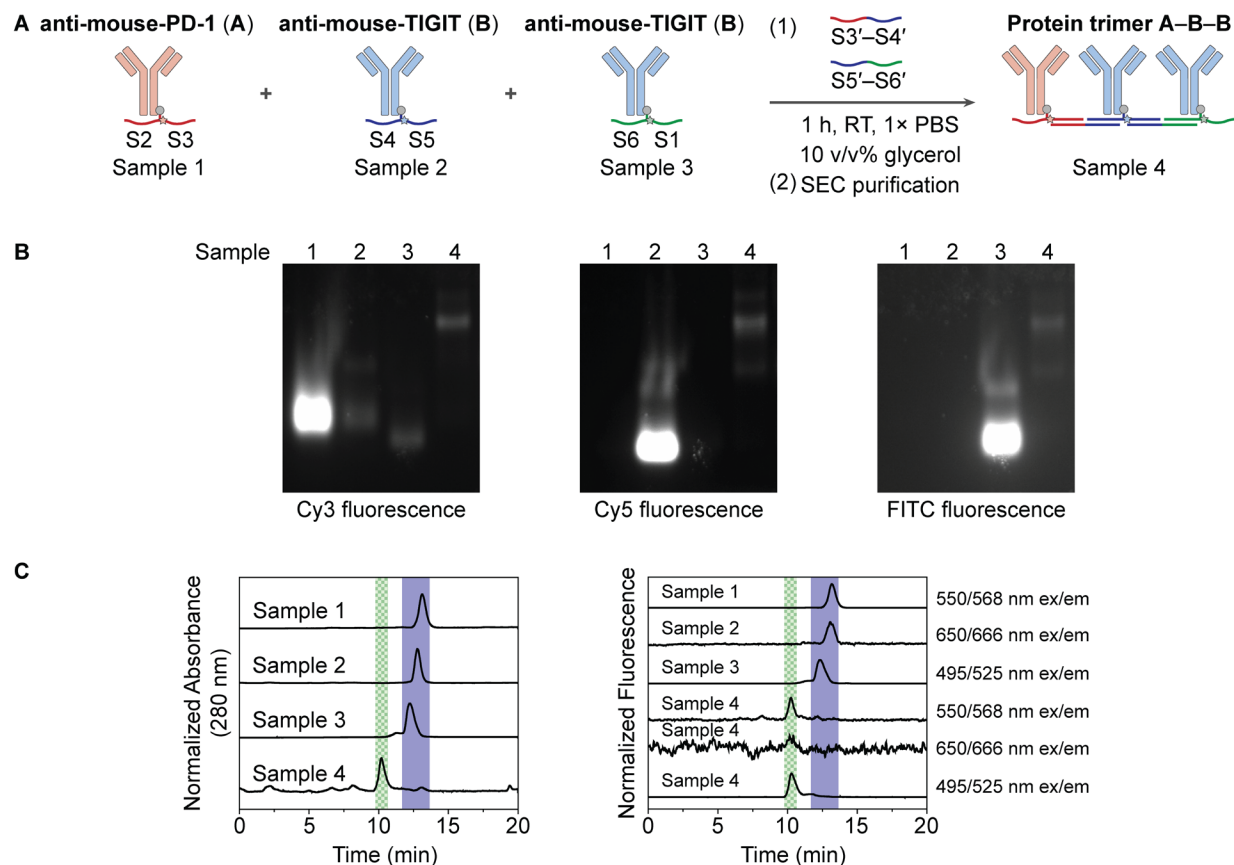
Figure 6.26. Characterization of a sequence-encoded human antibody dimer. (A) Human antibody–DNA conjugates (*i.e.*, sample 1: S2–**anti-human-PD-1** (**D**)–Cy3–S3 and sample 2: S4–**D**–Cy5–S5) and template DNA strands (*i.e.*, S3'–S4') were assembled into a human antibody dimer (*i.e.*, sample 3: S2–**D**–**D**–S5) using DNA–DNA interactions. (B) Agarose gel characterization of human antibody–DNA conjugates and a human antibody dimer. A single gel was imaged for Cy3 (left) and Cy5 fluorescence (right). (C) Analytical SEC characterization using absorbance at 280 nm and fluorescence of DNA-conjugated fluorophores. In both graphs, the elution times from 10.30–11.40 min are highlighted in green (shaded area with squares) and the elution times from 12.20–13.80 min are highlighted in blue (shaded area).
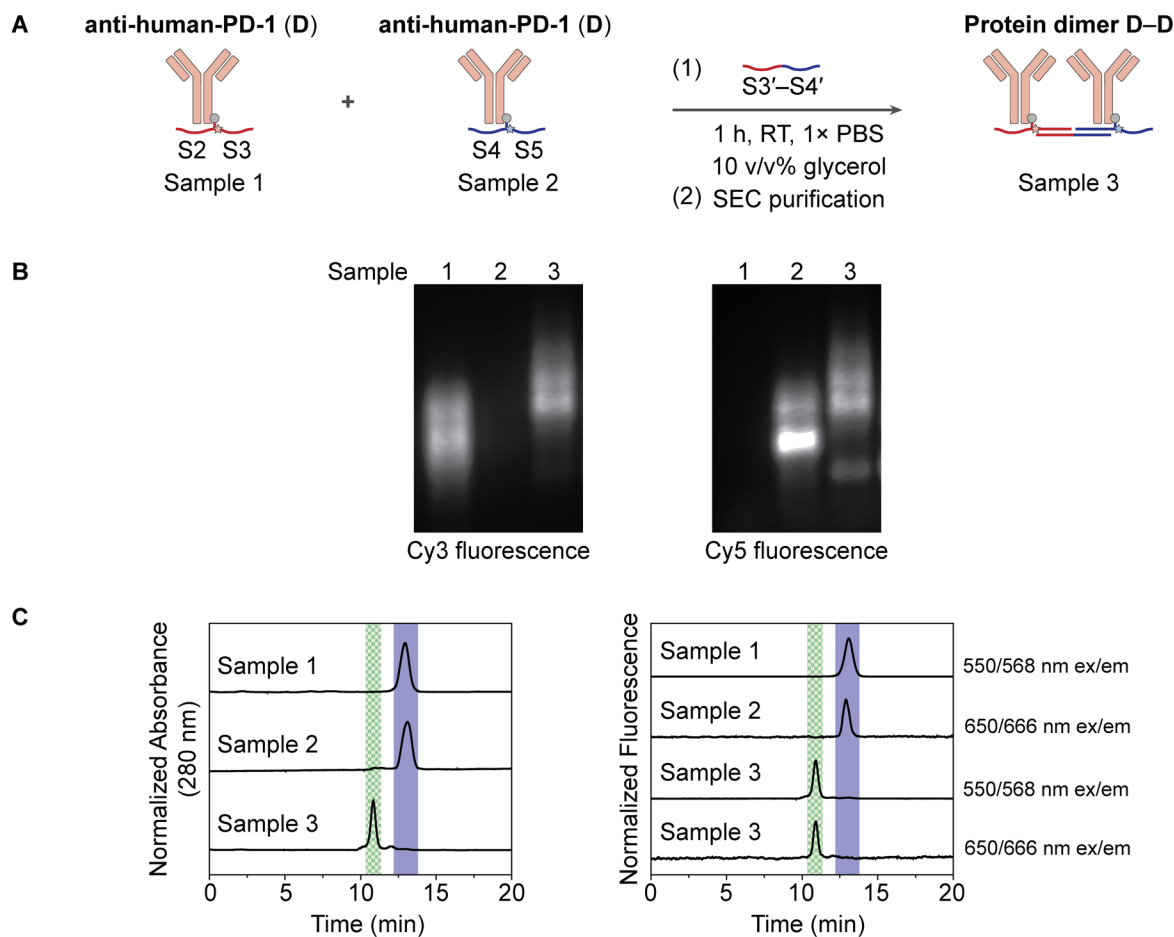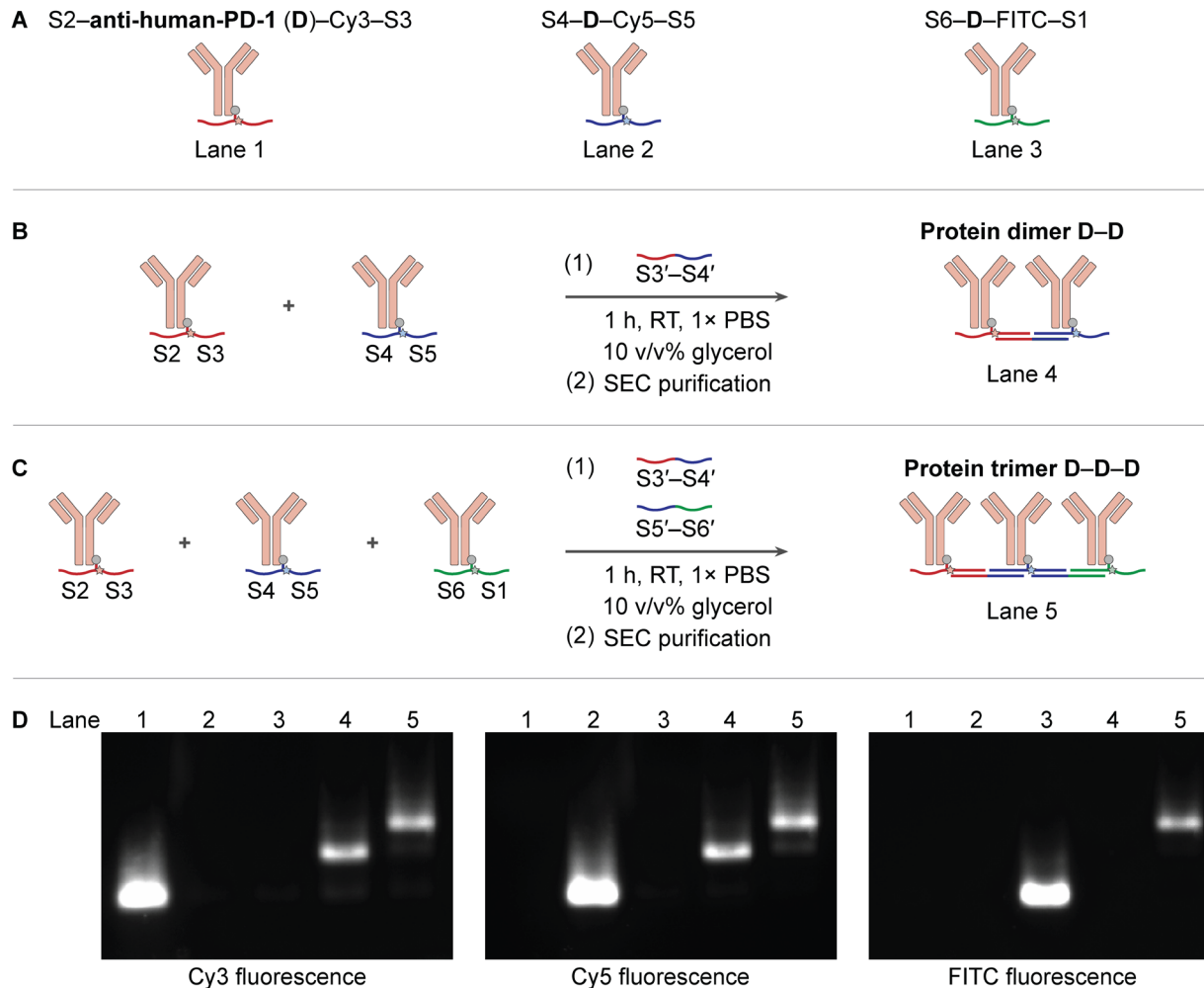
Figure 6.27. Characterization of sequence-encoded human antibody dimers and trimers. (A) Human antibody–DNA conjugates and template DNA strands were assembled into sequence-encoded antibody (B) dimers and (C) trimers using DNA–DNA interactions. (D) Agarose gel characterization of human antibody–DNA conjugates (*i.e.*, lane 1: S2–**anti-human-PD-1** (**D**)–Cy3–S3, lane 2: S4–**D**–Cy5–S5, and lane 3: S6–**D**–FITC–S1), a human antibody dimer (*i.e.*, lane 4: S2–**D**–**D**–S5), and a human antibody trimer (*i.e.*, lane 5: S2–**D**–**D**–**D**–S1). A single gel was imaged for Cy3, Cy5, and FITC fluorescence and these images are merged into one composite image in Figure 6.35.

Merged Cy3, Cy5, and FITC fluorescence

Figure 6.28. Merged agarose gel image of sequence-encoded mouse antibody oligomers.
Agarose gel characterization of antibody–DNA conjugates (*i.e.*, lane 1: S2–**anti-mouse-PD-1** (**A**)–Cy3–S3, lane 2: S4–**anti-mouse-TIGIT** (**B**)–Cy5–S5, and lane 3: S6–**anti-mouse-CTLA-4** (**C**)–FITC–S1) along with sequence-encoded antibody dimers (*i.e.*, lane 4: S4–**B**–**C**–S2'), antibody trimers (*i.e.*, lane 5: S2–**A**–**B**–**C**–S1), and antibody pentamers (*i.e.*, lane 6: S4–**B**–**C**–**A**–**B**–**C**–S1). A single gel was imaged for Cy3, Cy5, and FITC fluorescence and the Cy3 (red), Cy5 (blue), and FITC (green) images (shown in Figure 3.6.E) were merged here into one composite image.

Merged Cy3, Cy5, and FITC fluorescence

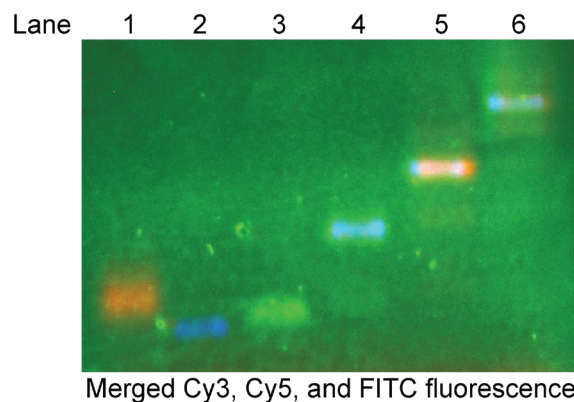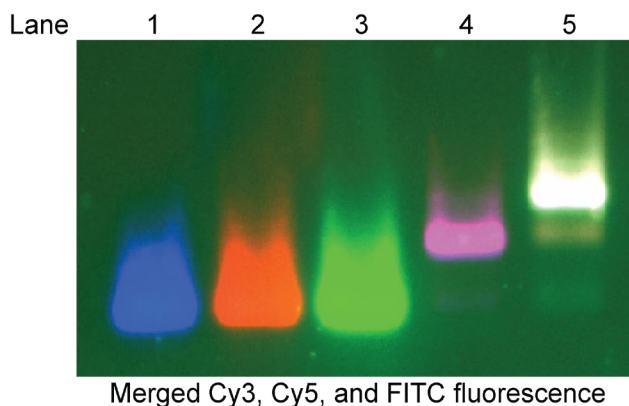**Figure 6.29. Merged agarose gel image of sequence-encoded human antibody oligomers.**
Agarose gel characterization of antibody–DNA conjugates (*i.e.*, lane 1: S2–**anti-human-PD-1** (**D**)–Cy3–S3, lane 2: S4–**D**–Cy5–S5, and lane 3: S6–**D**–FITC–S1) along with antibody dimers (*i.e.*, lane 4: S2–**D**–**D**–S5), and antibody trimers (*i.e.*, lane 5: S2–**D**–**D**–**D**–S1). A single gel was imaged for Cy3, Cy5, and FITC fluorescence and the Cy3 (blue), Cy5 (red), and FITC (green) images (shown in Figure 6.27) were merged here into one composite image.

6.2.4. Antigen Binding and Checkpoint Inhibitor Activity Cellular Assays.

Human peripheral blood mononuclear cells (hPBMCs, Zenbio, SER-PBMC-P-F) were taken from a liquid nitrogen dewar and thawed in a water bath. Next, cells were added to 9 mL of RPMI media containing 10% (v/v) heat-inactivated fetal bovine serum and 1% (w/v) penicillin-streptomycin (herein termed RPMI +/+ media). Cells were pelleted by centrifugation at 300 × g

for 10 min. The supernatant was removed by aspiration and cells were resuspended in 3 mL of RPMI +/+ media. Cells were counted using a Vi-CELL BLU Cell Viability Analyzer. Cells were adjusted to a concentration of 1×10^6 cells/mL through dilution with RPMI +/+ media containing 20 ng/mL phorbol myristate acetate (PMA) and 1000 ng/mL ionomycin to stimulate overexpression of PD-1.[263-264] A volume of 200 µL of the diluted cells was added to each well of a 96-well plate with round bottom wells. Plates containing cells were incubated at 37 °C with 5% CO_2 for 48 h. After incubation, cells and solution were transferred from each well into flow inserts and washed with RPMI +/+ media. Flow inserts were centrifuged at 1200 rpm for 5 min and samples were aspirated and resuspended in 200 µL of treatment. Treatment involved the addition of an antibody (**anti-human-PD-1 (D)** reacted with Alexa Fluor™ 546 (AF546) NHS Ester (Thermo Fisher Scientific) and characterized by UV-vis spectroscopy to have ~2.6 AF546 modifications per antibody), antibody–DNA conjugate (*i.e.*, S2–**D**–Cy3–S3), antibody dimer (*i.e.*, S2–**D**–**D**–S5), or antibody trimer (*i.e.*, S2–**D**–**D**–**D**–S1) in RPMI +/+ media at an antibody concentration of 50.0 nM. Samples were then incubated at 37 °C with 5% CO_2 for 6 h. After incubation, samples were washed once with 1× PBS (pH 7.4) buffer and resuspended in 100 µL 1× PBS (pH 7.4) buffer containing 0.5 µL of a BV421 CD8 antibody stain (Biolegend clone RPA-T8 #301036). Cells were incubated for 15 min at 4 °C. Next, samples were washed once with 1× PBS (pH 7.4) buffer and resuspended in 100 µL fixation buffer (Biolegend #420801). Samples were stored at 4 °C prior to analysis.

After treated hPBMC samples were stained and fixed, the samples were centrifuged at 300 × g for 5 min and cell concentration was adjusted to 2×10^6 cells/mL. A volume of 50 µL of each sample was mounted onto microscopy slides using ProLong Glass Anti-Fade Mountant

(Invitrogen, #P36984) and allowed to cure overnight. Fluorescent confocal microscopy was performed using a Zeiss LSM800 microscope (40× objective, GaAsP PMT detectors) to visualize antibody binding to CD8 cells in the hPBMC samples. Cells were imaged in the DAPI channel (λ = 405 nm), FITC channel (λ = 488 nm), Cy3 channel (λ = 561 nm) and Cy5 channel (λ = 640 nm) using the same image acquisition parameters for each sample (*e.g.*, laser power, master gain, pinhole size, scan speed, offset). All images were analyzed identically using ImageJ.

A PD-1/PD-L1 Blockade Bioassay (J1250) was purchased from Promega and was performed according to the commercial protocol.[265] In brief, CHO-K1 artificial antigen presenting cells (aAPC) expressing human PD-L1 and Jurkat T Cells expressing human PD-1 and NFAT-induced luciferase were treated with an antibody (**D** reacted with Alexa Fluor™ 546 (AF546) NHS Ester (Thermo Fisher Scientific) and characterized by UV-vis spectroscopy to have ~2.6 AF546 modifications per antibody), antibody–DNA conjugate (*i.e.*, S2–**D**–Cy3–S3), antibody dimer (*i.e.*, S2–**D**–**D**–S5), or antibody trimer (*i.e.*, S2–**D**–**D**–**D**–S1). Cells were treated with each sample at antibody concentrations of 167, 66.7, 26.7, 10.7, 4.27, 1.71, 0.683, and 0.273 μM and were incubated at 37 °C for 6 h. The addition of antibodies that block the checkpoint interaction between PD-L1 and PD-1 results in T cell receptor signaling and NFAT-mediated luciferase expression. A luciferase substrate was added to the cell mixture and luciferase activity was quantified by measuring luminescence on a BioTek Cytation 5 plate reader.

An antibody dimer (*i.e.*, S2–**A**–**B**–S5) was diluted to antibody concentrations of 125, 50.0, 20.0, 8.00, 3.20, and 1.28 nM in 1× PBS (pH 7.4) buffer containing 10% fetal bovine serum. Samples were incubated at 37 °C for 6 h. After incubation, the degradation of antibody dimers was quantified using analytical agarose gel electrophoresis, where images of agarose gels were

analyzed by densitometry using ImageJ to quantify the proportion of dimer and monomer in each sample after incubation.
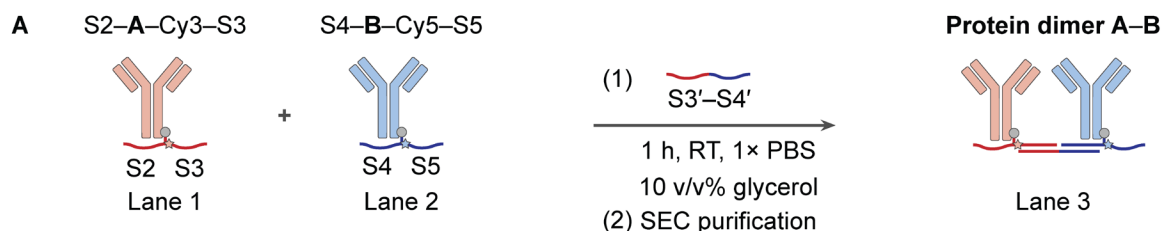
B Protein dimer **A**–**B** was incubated in 10% fetal bovine serum at 37 °C for 6 hours at antibody concentrations of 125 (lane 4), 50.0 (lane 5), 20.0 (lane 6), 8.00 (lane 7), 3.20 (lane 8), and 1.28 (lane 9) nM.
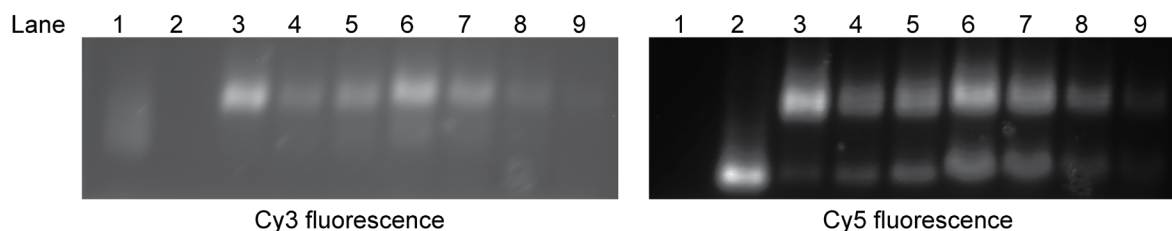
C Densitometry Analysis of Gel to Characterize Oligomer Stability

Sample	Cy3 fluorescence		Cy5 fluorescence	
	Monomer (%)	Dimer (%)	Monomer (%)	Dimer (%)
S2–**A**–Cy3–S3 (lane 1)	100	0	N/A	N/A
S4–**B**–Cy5–S5 (lane 2)	N/A	N/A	100	0
Protein dimer **A**–**B** (lane 3)	3	97	3	97
A–**B** incubated at 125 nM (lane 4)	16	84	17	85
A–**B** incubated at 50.0 nM (lane 5)	13	87	17	83
A–**B** incubated at 20.0 nM (lane 6)	14	86	24	76
A–**B** incubated at 8.00 nM (lane 7)	14	86	26	74
A–**B** incubated at 3.20 nM (lane 8)	12	88	27	73
A–**B** incubated at 1.28 nM (lane 9)	15	85	23	77

Figure 6.30. Stability study of antibody oligomers in cellular media. Agarose gel characterization of mouse antibody–DNA conjugates (*i.e.*, lane 1: S2–**anti-mouse-PD-1** (**A**)–Cy3–S3 and lane 2: S4–**anti-mouse-TIGIT** (**B**)–Cy5–S5), a mouse antibody dimer (*i.e.*, lane 3: S2–**A**–**B**–S5), and S2–**A**–**B**–S5 incubated in 10% FBS at 37 °C for 6 hours at antibody concentrations of 125 (lane 4), 50.0 (lane 5), 20.0 (lane 6), 8.00 (lane 7), 3.20 (lane 8), and 1.28 (lane 9) nM. Samples were then characterized on an agarose gel that was imaged for Cy3 and Cy5 fluorescence and analyzed by densitometry. The proportion of dimer in the sample decreased from 97% before incubation to 70–85% after incubation for each sample concentration.

6.3. Appendix C: Materials and Methods for Chapter Four.

6.3.1. Protein Mutation, Recombinant Expression, and Purification.

Genes for C191 GFP and C75 MBP that contain a tobacco etch virus (TEV) protease-cleavable poly-histidine tag (Table 4.1, Integrated DNA Technologies) were cloned into the pET28 vector backbone using Gibson Assembly.[246] The assembled plasmids for C191 GFP and C75 MBP were transformed into BL21(DE3) electrically competent cells (ThermoFisher) with electroporation. After recovery in S.O.C. Medium (ThermoFisher) for 1 hour at 37 °C with 300 rpm shaking, cells were grown overnight on LB Agar plates with antibiotic (50 μg/mL kanamycin). Single colonies were selected and cultured in 8 mL of LB broth with antibiotic (50 μg/mL kanamycin) overnight at 37 °C with 200 rpm shaking. After cell growth, glycerol stocks of the cells were prepared and stored at -80 °C. Plasmids were extracted from cells using the QIAprep Spin Miniprep Kit (Qiagen) and the correct plasmid sequences were confirmed using Sanger Sequencing (ACGT).[247]

Cultures in 8 mL of LB broth with antibiotic (50 μg/mL kanamycin) were inoculated using glycerol stocks and grown overnight at 37 °C with 200 rpm shaking. Next, these cultures were added to 1 L of 2x YTP broth with antibiotic (50 μg/mL kanamycin) and grown at 37 °C with 200 rpm shaking until a cell OD at 600 nm of 0.6 (~4 h). Cultures were induced (1 mM IPTG) and grown overnight at 17°C with 200 rpm shaking. Cells were pelleted (6000 g, 20 min, 4 °C), resuspended in 1× PBS, and lysed with a high-pressure homogenizer. The insoluble fraction was removed with centrifugation (15000 g, 20 min, 4 °C).

The GFP and MBP mutants contain a poly-histidine tag, which was used to isolate the mutants from cell lysate using nickel affinity chromatography. Proteins were loaded onto a column

packed with Profinity™ IMAC Resin (Bio-Rad). The column was washed with 100 mL of $1\times$ PBS with 12.5 mM imidazole and proteins were eluted with 15 mL of $1\times$ PBS with 250 mM imidazole. The buffer of the GFP and MBP mutants were exchanged to $1\times$ PBS six times using 30 kDa cutoff Amicon® Ultra-0.5 mL Centrifugal Filters (Millepore Sigma). The concentrations of GFP and MBP were identified using UV-VIS absorption spectroscopy using established extinction coefficients ($\varepsilon_{GFP, 488\ nm} = 55,000\ M^{-1}cm^{-1}$, $\varepsilon_{MBP, 280\ nm} = 77,600\ M^{-1}cm^{-1}$).[248, 266] BSA was obtained from Millepore Sigma as a lyophilized powder and was used without purification after reconstitution in $1\times$ PBS.

6.3.2. Oligonucleotide Design, Synthesis, and Purification.

Three DNA sequences (Table 4.2, *i.e.*, I2$_4$, H1$_4$–NH$_2$, and H2$_4$–NH$_2$) were synthesized on solid supports using reagents obtained from Glen Research and standard protocols. Products were cleaved from the solid support using 15% (w/v) ammonium hydroxide (aq) and 20% (w/v) methyl amine for 20 min at 55 °C and purified using reverse-phase HPLC with a gradient of 0 to 75 percent acetonitrile in triethylammonium acetate buffer over 45 min. Dimethoxytrityl or monomethoxytrityl groups were cleaved with 20% (v/v) acetic acid for 2 h and extracted with ethyl acetate. The masses of the oligonucleotides were MALDI-TOF MS using 2′,6′-dihydroxyacetophenone matrix containing diammonium hydrogen citrate as a matrix. All observed masses of synthesized DNA were within 30 Da of the expected mass.

Scheme 6.4. Structures of amino-modifier phosphoramidites.
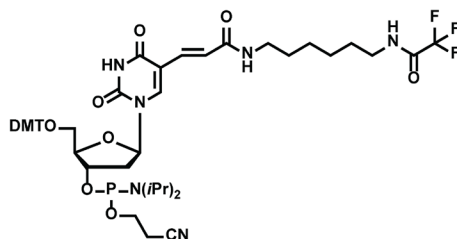
Amino-modifier C6 dT phosphoramidite
(T(NH$_2$))

These phosphoramidites contain 4,4'-dimethoxytrityl (DMT), *N,N*-diisopropyl (N(*i*Pr)$_2$), ß-cyanoethyl, and trifluoroacetyl protecting groups that are removed during DNA workup and purification.

6.3.3. Synthesis, Purification, and Characterization of Protein–DNA Conjugates.

Amine-modified DNA (*i.e.*, H1$_4$–NH$_2$ or H2$_4$–NH$_2$) was then snap-cooled in 1× PBS by heating to 95 °C for 4 min, cooled to 4 °C over 1 minute, held at 4 °C for 4 min, then brought to room temperature in an Applied Biosystems SimpliAmp Thermal Cycler. The concentration of DNA after snap-cooling was then confirmed using UV-vis absorption spectroscopy and the DNA extinction coefficient at 260 nm. Next, the DNA (3,000 nmol) was reacted with 50 equiv. of succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC, ThermoFisher) in 50:50 DMF:1× PBS, pH 7.4 for 1 h at RT. DNA was purified from excess SMCC with two consecutive illustra NAP Columns (GE Healthcare Life Sciences). The purified DNA was reacted with GFP, MBP, or BSA (300 nmol) overnight at RT. GFP–DNA, MBP–DNA, or BSA–DNA conjugates were isolated from unreacted protein and DNA using anion exchange chromatography on a DEAE column (Bio-Rad).

Synthesis and purity of GFP–DNA, MBP–DNA, and BSA–DNA conjugates were confirmed with UV-vis absorption spectroscopy and SDS-PAGE. The GFP–DNA conjugates show absorption maxima at 488 nm ($\varepsilon_{GFP, 488\ nm}$ = 55,000 M^{-1}cm^{-1}) due to the GFP chromophore and at 280 nm due to aromatic amino acid side chains.[249] DNA shows an absorption maxima

around 260 nm and extinction coefficients at 260 nm were calculated with the IDT OligoAnalyzer Tool (Integrated DNA Technologies). The number of DNA per GFP in solution was quantified by comparing the relative absorption at 488 nm and 260 nm for GFP and GFP–DNA. In contrast, MBP–DNA and BSA–DNA conjugates only show absorption maxima at 280 nm. The number of DNA per MBP or BSA in solution was quantified by comparing the increase in absorption at 260 nm for the protein–DNA conjugate compared to the protein. The increase in mass of protein–DNA conjugates after DNA functionalization and sample purity was confirmed with SDS-PAGE using 4-15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) and a Precision Plus Protein™ All Blue Prestained Protein Standard (Bio-Rad). The increase in masses of the GFP–DNA, MBP–DNA, and BSA–DNA conjugates after the conjugation reaction corresponds to their modification with one DNA strand.

VITA

PETER HENRY WINEGAR

Department of Chemistry, Northwestern University
2145 Sheridan Road, Evanston, IL, 60208
Email: peterwinegar@gmail.com

EDUCATION

Northwestern University, Evanston, IL   2017–2022
    Ph.D. Candidate in Chemistry (Advisor: Prof. Chad A. Mirkin)
    Graduate Course GPA: 3.93/4.00

Michigan Technological University, Houghton, MI   2013–2017
    B.S. in Chemistry, Minors in Math and Computer Science
    Undergraduate Course GPA: 3.97/4.00

PROFESSIONAL EXPERIENCE

University of California, Berkeley, Berkeley, CA   09/2022
    Postdoctoral Researcher (Advisor: Prof. Jay D. Keasling)
    Research Topic: Designing polyketide synthases in single cell hosts for *in vitro* synthesis of therapeutic molecules.

Northwestern University, Evanston, IL   2017–2022
    Graduate Research Assistant (Advisor: Prof. Chad A. Mirkin)
    Dissertation Title: Programming Protein Organization into Sequence-Encoded Architectures Using DNA
    Research Topic: Programming the organization of proteins into single crystals, sequence-encoded oligomers, and block co-oligomers using DNA for applications in protein structure determination, therapeutic discovery, and protein hierarchical assembly, respectively.

OTHER RESEARCH EXPERIENCE

Michigan Technological University, Houghton, MI   2014–2016
    Undergraduate Research Assistant (Advisor: Prof. Loredana Valenzano)
    Research Topic: Computationally studied nanoscale molecular machines to inform the design of experimental targets.

Paul-Drude-Institut, Berlin, Germany Summer 2016
 DAAD RISE Award Research Assistant (Advisor: Dr. João Marcelo Lopes)
 Research Topic: Developed new methodologies to transfer epitaxially-grown two-dimensional materials from growth substrate to inert surfaces for applications in device assembly.

University of Minnesota, Twin Cities, Minneapolis, MN Summer 2015
 LANDO/NSF REU in Chemistry Research Assistant (Advisor: Prof. Laura Gagliardi)
 Research Topic: Computationally studied C–H bond activation on copper-deposited metal-organic frameworks (MOFs) in a search for experimental catalyst targets for conversion of methane into methanol.

Case Western Reserve University, Cleveland, OH Summer 2014
 NSF REU in Polymers Research Assistant (Advisor: Prof. Rigoberto Advincula)
 Research Topic: Synthesized devices composed of layers of charged polymers for applications in the detection of viruses.

AWARDS AND HONORS

Ludo Frevel Crystallography Scholarship Award (Internationally Competitive) 2021

Ryan Graduate Fellowship, Northwestern University 2020

Honorable Mention, NDSEG and NSF GRFP 2018

Dean's List, Michigan Technological University 2013–2017

Chemistry Department Outstanding Senior, Michigan Technological University 2017

Barry Goldwater Scholarship (Nationally Competitive) 2016

ACS Undergraduate Award in Inorganic Chemistry, Michigan Technological University 2016

Chemistry Departmental Scholar, Michigan Technological University 2015

OUTREACH, LEADERSHIP, AND TEACHING EXPERIENCE

Mentor, Mirkin Research Group 2020–Present
 Mentored four graduate students in the Mirkin Research Group: Yinglun Ma, Catherine M. Davidson, Rachel Chan, Jennifer Delgado.

Teaching assistant, Northwestern University and Michigan Technological University 2015–2020
  General Chemistry Laboratory (2 quarters), Advanced Analytical Chemistry Laboratory (2 quarters), Physical Chemistry Lecture (1 semester).

Elected Officer, Phi Lambda Upsilon, Northwestern University   2019–2020
  Served as president of this chemistry honor society and actively engaged in outreach to a local elementary school and philanthropy related to COVID-19 pandemic relief.

Member, American Chemical Society   2017–Present

PUBLICATIONS (*Equal Author Contribution)

9. **Winegar, P. H.**; Figg, C. A.; Ma, Y.; Chan, R. R.; Landy, K. M.; Mirkin, C. A. Protein Block Co-Polymer Assembly Using DNA and Metal Coordination. *Manuscript in Preparation*.

8. **Winegar, P. H.**;* Figg, C. A.;* Teplensky, M. H.; Ramani, N.; Mirkin, C. A. Modular Nucleic Acid Scaffolds for Synthesizing Monodisperse and Sequence-Encoded Antibody Oligomers. *Chem* **2022**, https://doi.org/10.1016/j.chempr.2022.07.003.

7. Partridge, B. E.; **Winegar, P. H.**; Han, Z.; Mirkin, C. A. Redefining Protein Interfaces within Protein Single Crystals with DNA. *J. Am. Chem. Soc.* **2021**, *143*, 8925–8934.

6. Figg, C. A.; **Winegar, P. H.**; Hayes, O. G.; Mirkin, C. A. Controlling the DNA Hybridization Chain Reaction. *J. Am. Chem. Soc.* **2020**, *142*, 8596–8601.

5. **Winegar, P. H.**; Hayes, O. G.; McMillan, J. R.; Focia, P. J.; Figg, C. A.; Mirkin, C. A. DNA-Directed Protein Packing within Single Crystals. *Chem* **2020**, *6*, 1007–1017.

4. **Winegar, P. H.** How to Become a Protein Crystallographer in a Nanoscience Lab. *Chem* **2020**, *6*, 803–805.

3. McMillan, J. R.; Hayes, O. G.; **Winegar, P. H.**; Mirkin, C. A. Protein Materials Engineering With DNA. *Acc. Chem. Res.* **2019**, *52*, 1939–1948.

2. Liu, L.; Li, L.; DeGayner, J. A.; **Winegar, P. H.**; Fang, Y.; Harris, T. D. Harnessing Structural Dynamics in a 2D Manganese–Benzoquinoid Framework to Dramatically Accelerate Metal Transport in Diffusion-Limited Metal Exchange Reactions. *J. Am. Chem. Soc.* **2018**, *140*, 11444–11453.

1. Pahls, D. R.; Ortuño, M. A.; **Winegar, P. H.**; Cramer, C. J.; Gagliardi, L. Computational Screening of Bimetal-Functionalized Zr6O8 MOF Nodes for Methane C–H Bond Activation. *Inorg. Chem.* **2017**, *56*, 8739–8743.

PATENTS

2. Mirkin, C. A.; **<u>Winegar, P. H.</u>**; Figg, C. A. Synthetic Strategy to Polymerize Protein into Molecularly Defined Polymers. International Patent Application PCT/US2022/021359, March 22, 2021.

1. Mirkin, C. A.; McMillan, J. R.; Hayes, O. G.; **<u>Winegar, P. H.</u>** Protein Crystal Engineering Through DNA Hybridization Interactions. International Patent Application PCT/US2019/065078, December 6, 2019.

PRESENTATIONS

6. Vannevar Bush Faculty Fellowship Annual Meeting, Washington, D.C.   2019
   Poster presentation: Protein Materials Engineering with DNA

5. DAAD RISE Meeting, Heidelberg, Germany   2016
   Oral presentation: Transfer and Characterization of Hexagonal Boron Nitride

4. Summer Undergraduate Research Expo, University of Minnesota, Minneapolis, MN 2015
   Poster presentation: Methane to Methanol Catalysis with Cu_3O_3

3. Joint Great Lakes/Central Regional ACS Meeting, Grand Rapids, MI   2015
   Poster presentation: Driving Nanocars on Graphene Sheets: A Quantum Chemical Investigation

2. Undergraduate Research Expo, Michigan Technological University, Houghton, MI 2015
   Poster presentation: Driving Nanocars on Graphene Sheets: A Quantum Chemical Investigation

1. Undergraduate Research Symposium, Kent State University, Kent, OH   2014
   Poster presentation: Investigation of the Anticorrosion Properties of Hyperbranched Polymers