

NORTHWESTERN UNIVERSITY

Programming Assembly Pathways of Proteins Using DNA

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILMENTS OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemistry

By

Oliver George Hayes

EVANSTON, ILLINOIS

September 2021

© Copyright by Oliver George Hayes 2021
All Rights Reserved

ABSTRACT

Programming Assembly Pathways of Proteins Using DNA

Oliver George Hayes

The building blocks of life are proteins. These incredible nanostructures are responsible for forming the diverse infrastructure of living systems and for performing countless biological functions. In Nature, these materials and systems achieve structural complexity and function through highly regulated and controlled assembly of protein building blocks, driven by specific interactions encoded between protein surfaces (protein-protein interactions; PPIs). Despite their significance in Nature, the realization of synthetic protein materials displaying similar structural complexity and functions, remains challenging due to the difficulties in controlling the association and assembly pathways of proteins by design. While some methods have been established to reliably control the association of proteins, the knowledge for instructing and guiding the assembly of protein building blocks down a specific pathway has yet to be created. Here, we investigate how the programmability and high-information content of DNA interactions can be used to program the assembly pathways of proteins to realize novel protein-based materials. Chapter two describes how protein sequence and secondary structure can be manipulated to tune the energy barriers associated with protein association to direct polymerization along specific pathways. Chapter three explores how protein amino acid sequences can be designed to encode multiple, orthogonal DNA interactions onto protein surfaces to drive the assembly of multi-component systems into extended three-dimensional architectures. Chapter four combines both protein and DNA sequence design to encode directional interactions onto protein surfaces with designed interaction strengths that

enable proteins to traverse hierarchical assembly pathways, akin to those observed in Nature. Chapter 5 provides a summary of the key conclusions and lessons established through this research as well as an overview of important future directions. Collectively these chapters establish new methods to design and engineer the assembly pathways of proteins which will enable the synthesis and discovery of synthetic protein-based materials with the structural complexity and functions observed in Nature and beyond.

ACKNOWLEDGEMENTS

The work in this thesis was not conducted in a vacuum. Instead, it was made possible by the support, mentorship and collaboration from colleagues and faculty that I have had the pleasure of working alongside, as well as resources made available by Northwestern University. The past five years and the collective influences of so many brilliant role models have transformed me into a scientist and leader. Professor Chad Mirkin has created a rich scientific environment which provides students with genuine academic freedom to pursue interesting and important questions. His mentorship has been a tremendous source of personal growth, both as a scientist, mentor, and a leader, and I am incredibly grateful for his continued support and for constantly challenging me to be a better scientist. Professors Milan Mrksich and Michael Jewett both served as members of my committee, and I thank them for contributing their time and for being important sources of mentorship and scientific insight throughout my time as a graduate student.

I was fortunate to begin my PhD under the mentorship of a brilliant senior graduate student, Dr. Janet McMillan. She was an incredible mentor, collaborator, and friend throughout this journey and taught me to think deeply about projects, critically about my results and to craft good scientific questions. I know I will continue to carry many of her lessons with me far beyond my graduate studies. Dr. Benjamin Partridge, a post-doctoral scholar in the group, was also an incredible mentor, collaborator, and friend. It was an absolute joy and privilege to work alongside him on projects, have intense scientific discussions, and to learn from his boundless brilliance.

I learned a lot from collaborating with experts during various projects. Dr. Byeongdu Lee and Jonathan Remis provided valuable insight and technical skills that have been transformative

to several projects. I am incredibly grateful that we were able to work together to solve some of the challenges presented in this thesis.

Throughout my graduate studies, I had the honour of mentoring several students: Peter Winegar, Max Distler and Henry Han, all three of which are gifted graduate students and have likely taught me more than I have taught them. I want to thank them for making me excited to come to lab each day to work with such passionate, engaged colleagues and for asking the best questions! Experiments can often be frustrating and results confusing, however, working in the Mirkin group meant that I was constantly surrounded by people who are supportive, always willing to offer guidance or listen to my science/life problems. I give a special thanks to Dr. Andrea d'Aquino, Dr. Lam-Kiu Fong, Dr. Eileen Seo, Dr. Kacper Skakuj, Dr. Christine Laramy, Dr. Devleena Semanta, Dr. Michelle Teplensky, Dr. Zack Urbach and Dr. Edmund Cheng for their support, friendship and positivity throughout. I also acknowledge the immense support that I received from the Mirkin group office and research staff: Dr. Sara Rupich, Dr. Tanushri Sengupta, Dr. Sarah Petrosko, Pam Watson and Elizabeth Forest. This team of people provides advice, answered questions and ensured my success as a student throughout my entire graduate studies. Much of the work presented here utilizes synthetic oligonucleotides, which are synthesised in-house by the talented DNA technicians, Matt Capek, Robert Stawicki and Jennifer Delgado.

Finally, I give thanks to my parents and my sister, Scarlett, who support and inspire me to pursue passions, embrace challenges and work with integrity.

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 2016 and June 2021. 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.

McMillan, J. R.*; Hayes, O. G.*; Remis, J.; Mirkin, C.A. Programming Protein Polymerization with DNA. *J. Am. Chem. Soc.* 2018, *140*, 1595-15956.

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

Part of this work will be presented in:

Hayes, O. G.*; Partridge, B. P.*; Mirkin, C.A. Encoding Hierarchical Assembly Pathways of Proteins with DNA. *Under review*

*Equal author contribution

LIST OF ABBREVIATIONS AND NOMENCLATURE

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
A	Adenine
AB2	Lattice isostructural with aluminium diboride
Asp	Aspartic acid
AuNP	Gold nanoparticle
bp	Base pair
BCC	Body-centered cubic
C	Cytosine
Cryo-EM	Cryogenic electron microscopy
Cy3	Cyanine-3
Cy5	Cyanine-5
Cys	Cysteine
DBCO	Dibenzylcyclooctyne
DNA	Deoxyribonucleic acid
EM	Electron microscopy
FWHM	Full width at half maximum
G	Guanine
GFP	Green fluorescent protein
Glu	Glutamic acid
His	Histidine
Lys	Lysine
MALDI-MS	Matrix-assisted laser desorption-ionization mass spectroscopy
MBP	Maltose binding protein

M_n	Number-averaged molecular weight
M_w	Weight-averaged molecular weight
mGFP	Mutant green fluorescent protein
NHS	N-hydroxysuccinimide
NP	Nanoparticle
PCR	Polymerase chain reaction
PDI	Polydispersity index
PPIs	Protein-protein interactions
RNA	Ribonucleic acid
SAXS	Small-angle X-Ray scattering
SDS-PAGE	Sodium-dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Sp1	Stable protein one
SPDP	Succinimidyl 3 -(2 -pyridyldithio)propionate
T	Thymine
TEM	Transmission electron microscopy
TMV	Tobacco mosaic virus
UV-vis	Ultra -Violet visible

TABLE OF CONTENTS

ABSTRACT.....	3
ACKNOWLEDGEMENTS.....	5
PREFACE.....	7
LIST OF ABBREVIATIONS AND NOMENCLATURE.....	8
TABLE OF CONTENTS.....	10
LIST OF FIGURES.....	14
LIST OF TABLES.....	17
CHAPTER ONE.....	18

Protein Assembly in Nature and the Lab

1.1 The infrastructure of life assembled from proteins.....	19
1.1.1 Protein structure and Protein-Protein Interactions (PPIs).....	19
1.1.2 Materials in Nature from protein building blocks.....	21
1.1.3 Structural complexity and functions through regulated assembly.....	23
1.2 Strategies and challenges for controlling protein assembly.....	27
1.2.1 Exchanging PPIs for orthogonal supramolecular interactions.....	27
1.2.2 Engineering protein interfaces and <i>De Novo</i> design.....	30
1.3 Encoding assembly using the genetic code.....	32
1.3.1 The recognition properties of DNA.....	32
1.3.2 DNA-encoded assembly of materials and nanostructures.....	33
1.3.3 The anatomy of the Programmable Atom Equivalent (PAE).....	34
1.4 Protein-DNA conjugates.....	36
1.4.1 Bioconjugation of protein-DNA conjugates.....	36
1.4.2 Building materials and systems from protein-DNA conjugates.....	38
1.5 Thesis overview.....	40
CHAPTER TWO.....	45

Programming Protein Polymerization Pathways with DNA

2.1 Lessons from the polymerization of small molecules.....	46
---	----

2.2	Challenges and opportunities for supramolecular protein polymerization	47
2.3	Design and synthesis of protein-DNA conjugates for polymerization	50
2.4	Protein polymers via a step-growth pathway	54
2.4.1	Characterization of product structures and distribution	54
2.4.2	Commentary on cyclic structure formation	56
2.5	Protein polymers via a chain-growth pathway	57
2.5.1	Characterization of product structures and distribution	57
2.5.2	Chain extension of living polymer chain-ends	61
2.6	Conclusions	63
2.7	Materials and Methods	63
2.7.1	Oligonucleotide design, synthesis and purification	63
2.7.2	Synthesis and purification of mGFP-DNA monomers	64
2.7.3	Size exclusion characterization	66
2.7.4	Polymerization experiments	66
2.7.4	Cryo-TEM imaging and class averaging	67
2.7.5	Analysis of polymer length distributions	68
CHAPTER THREE	71
	<i>DNA-Encoded Protein Janus Nanoparticles</i>	
3.1	The diversity and multiplicity of protein interactions	72
3.2	Encoding multiple, orthogonal interactions at the nanoscale	72
3.3	DNA-encoded protein Janus nanoparticles	74
3.3.1	Designing DNA-encoded protein Janus nanoparticles	74
3.3.2	Synthesis and characterization of DNA-encoded protein Janus nanoparticles	76
3.4	Interrogating assembly behavior in the context of colloidal crystallization	79
3.4.1	Co-assembly of Janus particles with two sets of 10 nm AuNP	79
3.4.2	Determining the identity of particles within the unit cell	82
3.4.3	Confirming the orientation of Janus particles within the lattice	85
3.5	Conclusions	87
3.6	Materials and Methods	88
3.6.1	Protein mutation, expression, and purification	88
3.6.2	Oligonucleotide design and synthesis	90

	12
3.6.3 Synthesis and characterization of DNA-encoded Janus particle	92
3.6.4 DNA functionalization of Au & Ag nanoparticles	95
3.6.5 Crystallization and characterization of DNA-encoded protein Janus nanoparticle with Ag and Au NPs	96
3.6.6 Small angle X-ray scattering analysis	98
3.6.7 Calculation and comparison of lattice parameters derived from modulation of DNA “Bond”	99
3.6.8 Electron microscopy imaging of superlattices.....	101
CHAPTER FOUR.....	104
<i>Encoding Hierarchical Assembly Pathways of Proteins with DNA</i>	
4.1 Challenges and promise of encoding hierarchical assembly pathways.....	105
4.2 Design and Synthesis of Sp1m-DNA Building Blocks.....	108
4.3 Directional Assembly Encoded by Strong Axial or Equatorial DNA Interactions.....	111
4.4 Multi-Stage Assembly Encoded by Strong and Weak DNA Interactions	115
4.5 Programming Structural Outcomes via DNA Design	120
4.6 Conclusions	122
4.7 Materials and Methods	123
4.7.1 Protein Design and Expression.....	123
4.7.2 Oligonucleotide Design and Synthesis	125
4.7.3 Synthesis of TCO Phosphoramidite	128
4.7.4 Synthesis and Characterization of Sp1m-DNA Conjugates	129
4.7.5 Cross-Reactivity of TCO/Tetrazine and DBCO/Azide Conjugation Reactions.....	132
4.7.6 Proteomics Analysis	134
4.7.7 Donor-Quenching FRET Studies.....	137
4.7.8 Melting Temperature and Full-Width Half Maximum of Assembly Transitions	139
4.7.9 Transmission Electron Microscopy (TEM).....	140
4.7.10 Atomic Force Microscopy	145
4.7.11 Effect of Salt Concentration on Hierarchical Assembly of Sp1m-A _S E _w and Sp1m-A'sE _w	148
CHAPTER FIVE	150
<i>Conclusions and Outlook</i>	
5.1 Conclusions	151

5.2 Outlook.....	153
5.2.1 Summary and future steps	153
5.2.2 Leveraging symmetry and valency control to enhance protein crystallization	154
REFERENCES	161

LIST OF FIGURES

Figure 1.1 – Chemical complexity of protein-protein interactions.	21
Figure 1.2 – Protein materials build by Nature.	23
Figure 1.3 – Regulated and hierarchical assembly processes in Nature.	25
Figure 1.4 – Synthetic approaches for mediating and designing protein association.	28
Figure 1.5 – DNA interactions and the anatomy of the PAE.	34
Figure 1.6 – Programming protein polymerization with DNA conformation.	41
Figure 1.7 – DNA-encoded protein Janus nanoparticles.	42
Figure 1.8 – Programming hierarchical assembly pathways of proteins with DNA.	44
Figure 2.1 - Step-growth and chain-growth mGFP–DNA monomer sets and their reaction pathways.	51
Figure 2.2 – SDS PAGE and SEC analysis of monomers.	52
Figure 2.3 - UV-vis spectra of mGFP, free DNA, and mGFP DNA monomers.	54
Figure 2.4 - Step-growth polymerization of mGFP–DNA monomers, S_A and S_B .	55
Figure 2.5 – Characterization of mGFP-DNA hairpin monomers in the absence of initiator.	58
Figure 2.6 – Chain-growth polymerization of H_A and H_B monomers.	59
Figure 2.8 - Chain extension of polymers with active chain ends.	62
Figure 2.9 - Representative micrographs and analysis for samples analyzed by TEM.	70
Figure 3.1 - Schematic of a Protein-Based Janus Particle.	75
Figure 3.2 – DNA conjugation to Cys residue of mGFP.	77
Figure 3.3 – Domain synthesis and formation of Janus particles.	78
Figure 3.4 – Characterization of protein-DNA conjugates using size exclusion chromatography.	79

Figure 3.5 - Structural determination of a superlattice assembled from 10 nm AuNPs and protein Janus nanoparticles.	81
Figure 3.6 - Assembly of multicomponent superlattices revealed layered hexagonal unit cell.	84
Figure 3.7 – A comparison of lattice parameter length between lattices synthesized from Janus particles with different interprotein bonds.	86
Figure 3.8 – SDS PAGE analysis of DNA conjugation reactions with surface cysteine.	93
Figure 3.9 – MALDI mass characterization of mGFP and intermediates.	94
Figure 3.10 – Melting curve of aggregates and calculated T_m .	97
Figure 3.11 – Representative TEM images of silica embedded superlattices.	102
Figure 3.12 – Representative TEM images of a sectioned crystal system containing 5 and 10 nm Au NPs.	103
Figure 4.1 – Design of Sp1m chemical surface and proposed hierarchical assembly schemes	108
Figure 4.2 – Synthesis and characterization of Sp1m-DNA conjugates.	109
Figure 4.3 – Characterization of the assembly of Sp1m with strong axial ($A_s/A's$) interactions.	113
Figure 4.4 – Characterization of the assembly of Sp1m with strong equatorial ($E_s/E's$) interactions.	115
Figure 4.5 – FRET-based characterization of temperature-dependent hierarchical assembly processes.	117
Figure 4.6 – Characterization of assembly outcomes from axial-first, equatorial-second hierarchical assembly processes.	122
Figure 4.7 – Synthesis of TCO Phosphoramidite (7)	128
Figure 4.8 – Functionalization of Sp1m with Azide and Tetrazine Linkers	129
Figure 4.9 – Denaturing 4–15% PAGE of Sp1m-DNA conjugates containing A_s or $A's$ DNA.	132
Figure 4.10 – Denaturing 4–15% PAGE of Sp1m-DNA conjugates containing E_s or $E's$ DNA.	132

Figure 4.11 – Reactivity of TCO- and DBCO-modified DNA with Sp1m-N3 (2) and Sp1m-2L (3).	133
Figure 4.12 – Chromatogram from nanocapillary reverse-phase liquid chromatography of 3.	136
Figure 4.13 – Local chemical environment of K74	137
Figure 4.14 – Raw Cy3 intensity data used to determine fraction assembled vs temperature curves in Figures 4.3 and 4.4.	138
Figure 4.15 – Raw Cy3 intensity data used to determine fraction assembled vs temperature curves in Figure 4.5.	139
Figure 4.16 – Graphs showing first derivative of fraction assembled vs temperature	140
Figure 4.17 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-A _S E _{NC} and Sp1m-A' _S E _{NC} .	141
Figure 4.18 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-A _S E _{NC} .	141
Figure 4.19 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-E _S and Sp1m-E' _S .	142
Figure 4.20 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-E _S .	143
Figure 4.21 – TEM micrographs of structures tuned by equatorial DNA design.	144
Figure 4.22 – Representative cryo-TEM micrographs of slow cooled Sp1m-A _S E _{NC} and Sp1m-A' _S E _{NC} .	145
Figure 4.23 – AFM micrograph of slow-cooled Sp1m-E _S and Sp1m-E' _S .	146
Figure 4.24 – AFM characterization of hierarchical structures.	147
Figure 4.25 – Graphs showing the influence of salt concentration on the assembly of Sp1m-A _S E _W and Sp1m-A' _S E _W .	148
Figure 5.1 – Preparation of protein clusters with defined valency	156
Figure 5.2 – Preparation of protein clusters from monomeric proteins.	158

LIST OF TABLES

Table 2.1 – DNA sequences, molecular weights, and extinction coefficients.	64
Table 3.1 – Mutagenetic primer design of S148 to C148 mutation.	88
Table 3.2 – Sanger sequencing data of mutated EGFP plasmid compared to native plasmid.	89
Table 3.3 – Oligonucleotide sequences designed for the assembly of inorganic nanoparticles with DNA-encoded protein Janus nanoparticles.	92
Table 3.4 – Lattice parameters measured using SAXS for three different colloidal crystal systems, assembled using Janus particles with either an 18 or 24 base-pair DNA “bond”.	100
Table 3.5 – Calculated lattice parameter differences.	101
Table 4.1 – Protein and Gene Sequences for Sp1m.	124
Table 4.2 – Oligonucleotide Sequences, Extinction Coefficients, and Calculated and Observed Masses.	127
Table 4.3 – Observed and Theoretical Masses and Corresponding Linker Attachment Positions on Sp1m-2L.	136

CHAPTER ONE

Protein assembly in Nature and the lab

Material in this chapter is 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 The infrastructure of life assembled from proteins

1.1.1 Protein structure and Protein-Protein Interactions (PPIs)

Proteins are nanoscale, three-dimensional (3D) structures that arise from the precise synthesis and folding of linear, one-dimensional (1D) polypeptide molecules into defined shapes that allow them to perform intricate biological functions.¹ In Nature, proteins are manufactured by cellular machinery that interpret instructions written in the genetic code (genes) to stitch together a specific sequence of amino acids and synthesize a desired protein.² All the chemical information required to specify a protein's three-dimensional structure is contained within its amino-acid sequence.³ Evolutionary processes have given rise to countless numbers of unique proteins that span immense functional and structural diversity.⁴ The vast infrastructure of life is built from and sustained by these protein "building blocks", thus, understanding their structures in immense detail has been a central research theme at the nexus of chemistry, structural biology, and physics, for the past century.

However, it was not until 1934, when the first x-ray diffraction pattern of a protein crystal was reported by John Desmond Bernal and Dorothy Hodgkin, that methods to determine protein structure began to emerge.⁵ Throughout the following two decades, methods were refined and in 1958 the first structure of a protein, myoglobin, obtained from x-ray analysis was reported by John Kendrew.⁶ These Nobel winning, seminal contributions unlocked the ability to visualize protein structures with angstrom resolution, consequently providing molecular-level information regarding their properties and functions.⁷ Such information has uncovered, for example, how enzymes bind to their substrates,⁸ how electrons are transported through protein complexes,⁹ and how Nature builds materials and systems through Protein-Protein Interactions (PPIs).¹⁰

The study of protein structures reveals how Nature utilizes and designs protein building blocks to construct complex, functional materials and systems that compose the infrastructure of life. Over millennia, the surfaces of proteins have evolved to display well-defined ensembles of non-covalent contacts, capable of forming selective interfaces between proteins to drive the assembly of higher order materials and systems.¹¹ These contacts can broadly be grouped into distinct classes of chemical interaction based on the identities of the amino acids displayed at the protein's surface (Figure 1.1). This chemical diversity results in highly specific and sophisticated PPIs with a degree of complexity that continues to be the subject of intense experimental and theoretical investigation.^{12, 13} It is unsurprising that such sophistication has evolved because controlled association between proteins is critically important in Nature: many, if not all proteins interact with others to form complexes or higher-order assemblies that carry out all types of cellular processes.¹⁴ The evaluation of these interactions, made possible by structural elucidation of proteins, has advanced our understanding of fundamental biological mechanisms, and led to important discoveries, for example in the development of new therapeutics.¹⁵

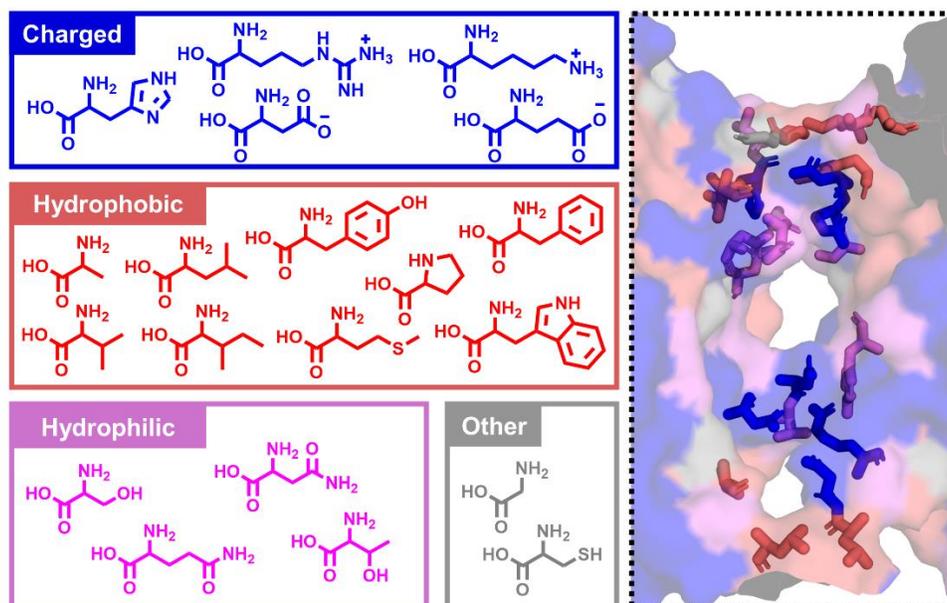


Figure 1.1 – Chemical complexity of protein-protein interactions. (Left) Amino acids grouped based on their side-chain chemical properties. (Right) A scheme showing the interface between two subunits of beta-glucosidase (PDB:1QOX) and the amino acids at the interface, color coded based on their chemical properties.

Analogous to how large, complex molecules can be built from simple building blocks, these developments in structural biology present the opportunity to mimic and exceed Nature by creating methods to program the assembly of materials and systems from protein building blocks. This chapter will highlight the types of materials that arise in Nature and how they form, as well as the synthetic efforts and challenges in designing and synthesizing unnatural protein-based materials.

1.1.2 Materials in Nature from protein building blocks

Materials and systems comprised of proteins in the natural world possess huge structural diversity, including polymeric filamentous architectures,¹⁶ two-dimensional (2D) sheets and 3D clusters,¹⁷ networks,¹⁸ cages,¹⁹ and crystals.²⁰ Arguably the most prevalent protein building block found in Nature is actin. Actin is the most abundant protein in nearly all eukaryotic cells and

participates in more PPIs than any known protein.²¹ The materials that form from the assembly of actin with itself and/or actin binding proteins play important roles in cell motility,²² maintaining cell structure,²³ and muscle contraction,²⁴ as well as many others. Such functions rely on the polymerization of actin monomers into filaments (Figure 1.2a).²⁵ For example, a high concentration of actin filaments can be found at the periphery of the cell where they form a 3D network beneath the plasma membrane to provide structural integrity and define the cell's shape.²⁶

Structural protein assemblies can also be found on the surfaces of many bacteria and archaea in the form of 2D crystalline arrays that coat the entirety of the cell surface, known as S-layers (Figure 1.2b). These materials form spontaneously *in vivo* and are typically composed of a single protein. The family of S-layer proteins is structurally diverse and examples of 2D S-layer crystals with oblique (p1, p2), square (p4) or hexagonal (p6) lattice symmetries have all been reported.^{27, 28} These structures are important to the growth and survival of cells, and their many functions include the preservation of cell integrity, presentation of enzymes, and in pathogens, interaction with the host and its immune system.²⁹ Additional functions and structural varieties of S-layers are still being discovered as our tools to interrogate these architectures improve, underscoring the variety and significance of this protein material in Nature.

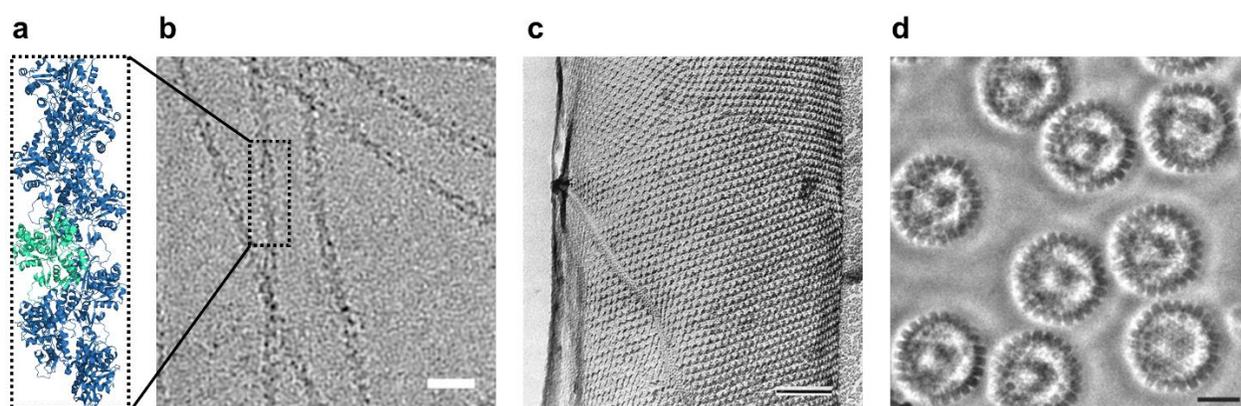


Figure 1.2 – Protein materials build by Nature. (a) Crystal structure of actin filament showing monomer in green (PDB:6BNO). (b) Cryo-EM micrograph of actin filaments. Scale bar = 20 nm Adapted from ref. [25]. (c) SEM Micrograph of bacterial s-layer proteins. Scale bar = 100 nm. Adapted from ref. [27]. Cryo-EM micrograph of herpes simplex virus. Scale bar = 50 nm. Adapted from ref. [31].

Spherical protein cage-architectures are a common structural motif in Nature and fulfil a number of important functions by taking advantage of the cage's internal cavity. Typically, multiple copies of only a single, or just a few, protein building blocks are required to assemble into these monodispersed, intricate 3D cage structures.³⁰ Virus capsids, which act as protective containers and delivery vehicles for genetic material, represent a diverse and large contingent of the cage-forming protein family (Figure 1.2c).³¹ Viruses are by far the most abundant, by number, biological entity on the planet, with an estimated 4×10^{30} virus particles contained within the world's oceans alone.³² Another important class of protein cage is the ferritins. Ferritins are iron storage proteins that play critical role in iron metabolism across many living species (Figure 1.2c). The importance of this iron sequestering material is emphasized by its universal presence in cells across all kingdoms of life as well as its highly conserved 24-subunit cage structure, with the capacity to store 4500 Fe(III) atoms in its cavity.³³

These examples represent a handful of key materials that arise from the assembly of protein building blocks in Nature. The structural and functional diversity of materials and systems in the natural world is immense and they provide insight and inspiration for tackling challenges in the synthesis of new materials, mimicking those found in Nature, through well-controlled supramolecular assembly of macromolecules.

1.1.3 Structural complexity and functions through regulated assembly

Proteins have been crafted, through epochs of evolutionary processes, into highly designed building blocks that undergo spontaneous organization to form ordered materials and systems without the need of external intervention. However, in addition to utilizing well-designed building blocks with specific shapes and PPIs, structural diversity and complexity in Nature is achieved through highly regulated assembly mechanisms that enable materials to have multiple levels of hierarchy and properties that depend on regulated and dynamic assembly and disassembly processes.

The importance of regulated assembly processes is captured well by the dynamic nature of microtubule structures. Microtubules are cytoskeletal filaments with hollow cylindrical structures formed from the polymerization of α -tubulin and β -tubulin protein building blocks.³⁴ They serve broadly as scaffolding for transport and mediate morphological changes within the cell but are also involved in many other cellular functions. Assembly of α -tubulin and β -tubulin protein building blocks is highly regulated in these materials through a mechanism called “dynamic instability”.³⁵ This mechanism allows microtubules to change their structures in response to cellular cues by switching between periods of elongation and truncation at the two ends of the cylindrical structure (Figure 1.3a).^{36, 37} The relative rates of microtubule growth at each end of the tube have also been shown to be tightly regulated which allows for directional, localized structural changes of these materials.^{38, 39} Such assembly mechanisms are intrinsic to materials in living systems that have to self-regulate in response to changes in environment and equilibriums.

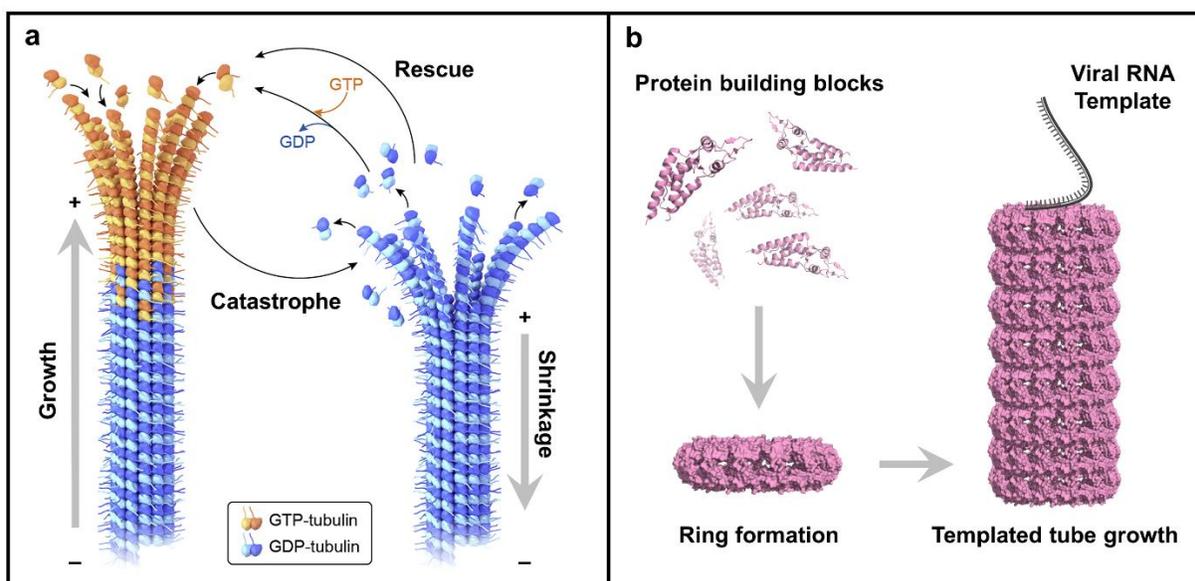


Figure 1.3 – Regulated and hierarchical assembly processes in Nature. (a) Scheme representing the dynamic instability mechanism of microtubule growth. Adapted from ref. [37]. (b) Scheme showing stepwise assembly process of TMV virus proteins to afford a hierarchical structure.

In concert with dynamic assembly/disassembly supramolecular polymerization processes, structural complexity and function in Nature is achieved by the formation of materials through sequential, hierarchical assembly processes.^{40, 41} By creating pathways whereby building blocks undergo multiple stages of assembly, novel material architectures can be accessed with specific functions. This concept can be simply demonstrated by considering the formation of the tobacco mosaic virus (TMV), comprising ~2130 protein monomers arranged helically, enclosing the single-stranded viral RNA sequence, to form a rigid rod-like structure (Figure 1.3b).⁴² The assembly of this structure begins with the formation of discs, composed of 34 protein monomers, that associate with a specific recognition sequence at the terminus of the viral RNA.⁴³ Binding of the RNA molecule to the protein disc leads to the growth and elongation of the TMV rod architecture, templated by the RNA.⁴⁴

At the opposite end of the scale to the TMV, in terms of structural complexity, is the hierarchical structure of skeletal muscle tissue.⁴⁵ Skeletal muscle has multiple levels of hierarchy with the top level being the muscle fiber. These fibers are built from numerous tubular organelles called myofibrils that are composed of repeating units of sarcomeres. Sarcomeres contain long, highly ordered networks of protein filaments and are the basic contractile unit of muscle. The two main protein building blocks that form these filaments are actin and myosin and their highly ordered cross-linked structure is fundamental for facilitating the longitudinal transfer of mechanical energy through the muscle fiber.⁴⁶ Muscle tissue is an intricate biological material and a clear example of how structural hierarchy and complexity facilitates its primary function: converting chemical energy into mechanical energy.

1.2 Strategies and challenges for controlling protein assembly

Using the term “building block” to describe proteins seems incredibly appropriate given the highly sophisticated and vast number of different materials and systems that arise in Nature from these fundamental components. However, this term implies that one should be able to readily assemble proteins into arbitrary structures. While there are some rare examples of a protein having multiple assembly outcomes, the majority follow a very specific assembly behaviour to form one specified structure.⁴⁷ The specificity of assembly, encoded by highly designed PPIs, is essential, in a biological context, for avoiding unwanted and potentially deleterious formation of off-target structures.⁴⁸ Unfortunately, this makes it difficult to modulate the assembly behaviour of protein building blocks to synthetically create materials, by design.

At the centre of this challenge is that the surfaces of proteins are highly chemically heterogeneous, comprising an ensemble of amino acids with different chemical properties. Due to this chemical complexity, the PPIs that arise from these intricate surfaces are challenging to characterize, predict and alter. To address these challenges, significant efforts have been made over the past two decades to program the association of proteins with designed interactions by introducing controlled molecular interactions on protein surfaces, and through computational design of protein interfaces.

1.2.1 Exchanging PPIs for orthogonal supramolecular interactions

An early strategy for controlling the association of proteins focussed on exploiting the innate receptor-ligand interactions between protein binding pockets and their small molecule, high affinity, binding partners. Interactions between proteins are induced by either modifying a protein’s surface with a specific ligand or by crosslinking two proteins binding pockets using bi-

functional ligands. For example, Hayashi et al. have utilized heme-binding characteristics of hemeproteins to program protein interactions.⁴⁹ By appending an external heme moiety at the opposite site to the native heme binding pocket, end-to-end polymerization of heme proteins can be programmed, resulting in the formation of one-dimensional nanowires (Figure 1.4a). Other material architectures have also been accessed using similar approaches, for example, the synthesis of 2D protein arrays mediated by biotin-streptavidin interactions and 3D protein crystals from the crosslinking of tetrameric lectins with bivalent sugar molecules.^{50, 51} While this is clearly a powerful strategy, the scope of proteins and type of interactions amenable to this method is severely limited.

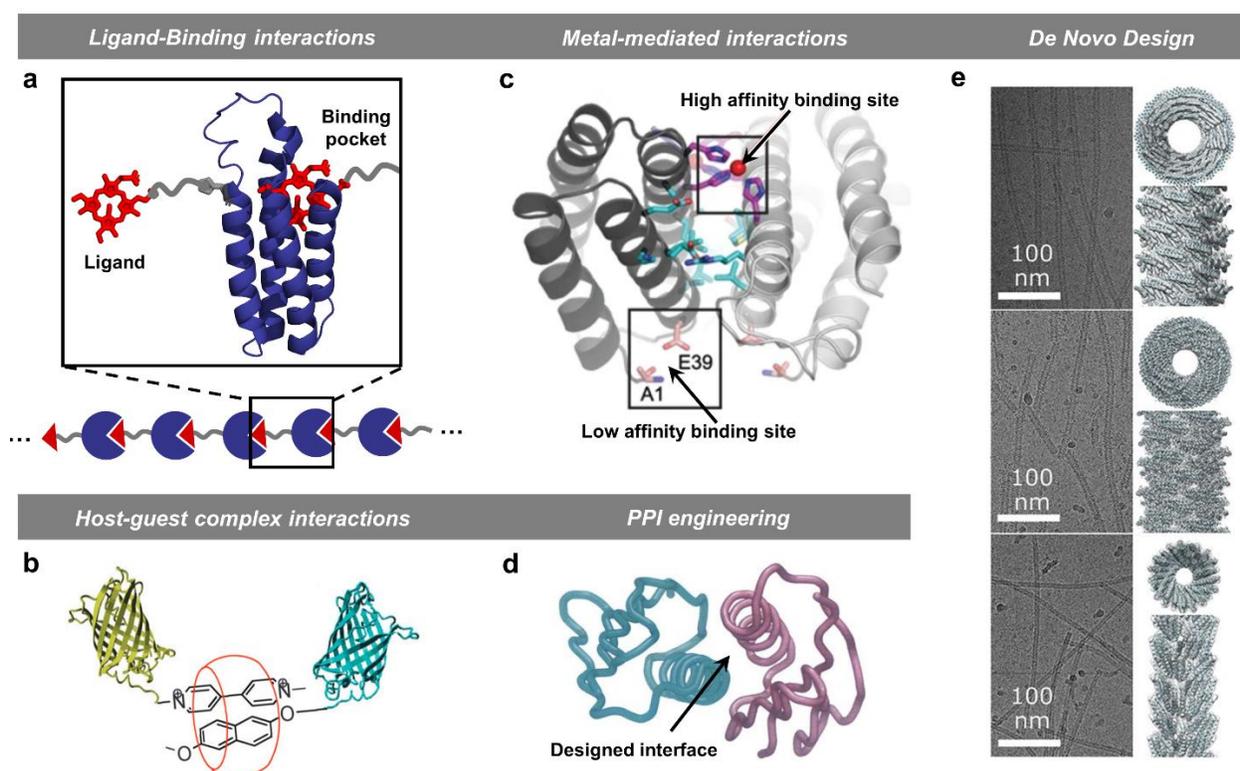


Figure 1.4 – Synthetic approaches for mediating and designing protein association. (a) Ligand binding of heme proteins to direct formation of nanowires. Adapted from ref. [49]. (b) Protein dimerization through cucurbit[8]uril host-guest interactions. Adapted from ref. [55]. (c) Metal-mediated protein association by engineering high-affinity binding sites. Adapted from ref. [63]. (d) Surface-engineered interface of monomeric protein to form

dimer. Adapted from ref. [65]. (e) Micrographs and models of filamentous de novo designed protein structures. Adapted from ref. [74].

Advances in supramolecular chemistry have contributed widely to the field of protein assembly through the discovery and creation of simple, predictable recognition motifs that can manipulate the association of large biological macromolecular structures.⁵² Brunsveld and co-workers carried out foundational work in studying protein-dimerization, mediated by a series of different host–guest interactions.^{53,54,55} The strong host–guest pair, β -cyclodextrin and lithocholic acids, with a binding constant in the sub-micromolar range in aqueous media, was initially used to successfully drive protein association,⁵⁶ but this molecular toolbox has since been expanded. More recently, higher order, 1D protein assemblies have also been realized using host-guest interactions (Figure 1.4b).⁵⁷ Host-guest interactions provide a high degree of specificity, reminiscent of native PPIs, however, the relatively fixed binding constant for given host-guest complex limits the ability to finely tune and control these binding events.

Metal-coordination supramolecular chemistry has also emerged as a powerful method in driving and controlling the assembly of proteins. In Nature, metal ions play critical roles in stabilizing quaternary protein superstructures and in the regulation of transient protein–protein interactions.⁵⁸ As a consequence, the metal-coordination properties of protein surfaces are well understood and derive from the chelating abilities of amino acids such as histidine (His), cysteine (Cys), aspartic acid (Asp), and glutamic acid (Glu). Tezcan and co-workers pioneered the strategy of directing protein assembly by metal-coordination, initially through the formation of discrete oligomeric structures,⁵⁹ and then later, by expanding the approach to access 1D,⁶⁰ 2D,⁶¹ and 3D protein arrays and crystals.⁶² Typically, sites containing multivalent His residues are introduced to the protein's surface, by mutagenesis, to create high-affinity sites for divalent transition metal ions

(Zn(II), Cu(II) and Ni(II)) that bind and template the association of proteins (Figure 1.4c).⁶³ Metal-coordination supramolecular chemistry has contributed significantly to the synthesis of novel protein-based materials in-part due to the advantageous tailorable properties of metal-ligand interactions (geometry, strength, directionality, and stereo-chemistry) which allow fine structural control. Still, metal coordination events are not always specific, and this strategy requires carefully controlled conditions (e.g ion concentration and pH) to prevent off-target amino acid ligands from coordinating to the metal ions leading to undesired protein interactions.

1.2.2 Engineering protein interfaces and *De Novo* design

The notion of redesigning protein surfaces to reprogram PPIs for controlled assembly or simply designing protein building blocks from scratch, is an attractive one. However, the intrinsic chemical complexity of protein-protein interfaces has made this approach challenging since understanding, modulating, and predicting the impact of extensive amino acid changes is nontrivial. Despite this, and aided by major advances in computational biology, impressive progress has been made in engineering protein interfaces to direct protein assembly, and in the *de novo* design of protein materials.

Early work by Bolan and co-workers demonstrated the potential of interface design by successfully engineering a variant of a protein that forms an asymmetric heterodimer instead of its native symmetric homodimer structure.⁶⁴ Computational design of protein interfaces has since grown in sophistication,⁶⁵ enabling the design of diverse protein materials, including 1D protein nano wires and 3D cage structures, by simply redesigning existing interfaces (Figure 1.4d).^{66, 67} Native proteins, however, are often marginally stable, and as a result sequence changes can cause

unfolding or aggregation. Therefore, the de novo design of proteins with novel functions, assembly properties and thermodynamic stabilities that far exceed those seen in nature has become an important tool in the creating of protein materials. Baker and co-workers have led the field through the development of the Rosetta software for de novo design,^{68, 69} enabling multicomponent two-dimensional lattices,⁷⁰ discrete oligomers,⁷¹ polyhedral cages,⁷² transmembrane protein pores,⁷³ filamentous protein polymers (Figure 1.4e),⁷⁴ and protein switches to be realized.⁷⁵

The design of proteins from scratch represents an exciting and virtually limitless opportunity to discover novel protein materials and functions. However, the approach inherently circumvents and diverges from naturally occurring proteins, which are critical building blocks in the development of protein-based materials.

1.3 Encoding assembly using the genetic code

1.3.1 The recognition properties of DNA

Deoxyribonucleic acid (DNA) is an important biopolymer in Nature that is responsible for storing and transmitting information in the form of the genetic code, written in the sequence of four nucleobases (adenine (A), guanine (G), cytosine (C) and thymine (T)). The DNA code interacts through Watson-Crick base pairing, a hydrogen-bond molecular recognition system which ensures that nucleobases interact preferentially with their corresponding binding partner, that is A with T and G with C (Figure 1.5a).⁷⁶ The consistency of information transfer in cellular replication processes and genetic inheritance is a testament to the high-fidelity of supramolecular DNA interactions.

Compared to the chemical complexity and unintuitive nature of PPIs, the interactions between complementary DNA strands are relatively simplistic and highly predictable. In addition, advances in synthetic organic chemistry have given rise to robust methods for the facile synthesis of DNA molecules in the lab, which can now be designed with an arbitrary sequence and built base by base.⁷⁷ The modular chemistry of oligonucleotides results in the straightforward design of sequences with predetermined properties such as melting temperature, complementarity, and length. Synthetic access to oligonucleotides, in tandem with their simplistic, yet highly programmable recognition properties have led to the rapid manifestation of DNA nanotechnologies and, pertinently, the prolific use of DNA as a ligand to direct the assembly of nanomaterials.⁷⁸

1.3.2 DNA-encoded assembly of materials and nanostructures

Over the past four decades, oligonucleotides have become synonymous with controlling structure, architecture, and function of materials at the nanoscale. The field of DNA nanotechnology finds its origins in the creative work of Ned Seeman, who used DNA to construct structural motifs, such as immobile DNA junctions, to build DNA frameworks and extended lattices.^{79, 80} These structural principals were further developed with the invention of DNA origami by Paul Rothman in 2006, a method by which DNA structures with arbitrary architecture and complexity can be formed through the folding of a single, long oligonucleotide sequence by short staple strands.⁸¹ With these innovations, virtually any material composed solely of DNA can be conceived and synthesized.

In parallel, Alivisatos,⁸² Mirkin,⁸³ and coworkers were developing methods to control the assembly of non-DNA nanomaterials, such as gold nanoparticles, into discrete and extended structures by using DNA surface ligands to program particle interactions. The versatile chemistry of synthetic oligonucleotides allows simple chemical strategies to densely functionalize the surface of gold nanoparticles with a layer of radially oriented oligonucleotides that facilitate the reversible organization of particles into ordered arrays and crystals. The independent modularity of DNA sequence, length and interaction strength enables angstrom level architectural control over the resulting macroscopic structures. Since these initial discoveries, the field has rapidly matured and key concepts have been refined to now include textbook methods for engineering colloidal crystals from a wide variety of DNA-modified constructs.⁸⁴ These constructs have been termed “programmable atom equivalents” (PAEs) to describe the bond-like characteristics of the

collective DNA interactions connecting nanoparticle “cores” in colloidal crystals, analogous to atomic lattices.

1.3.3 The anatomy of the Programmable Atom Equivalent (PAE)

A key requirement of the PAE core is that it possesses a surface that is amenable to the functionalization of multi-valent DNA ligands. There are a wide variety of nanostructures that satisfy this requirement and have been utilized as PAE cores, ranging from inorganic nanoparticles and metal-organic frameworks to proteins and polymer micelles. This simple constraint allows for essentially any type of nanomaterial to constitute a PAE core, allowing one to select from a wide scope of material properties.

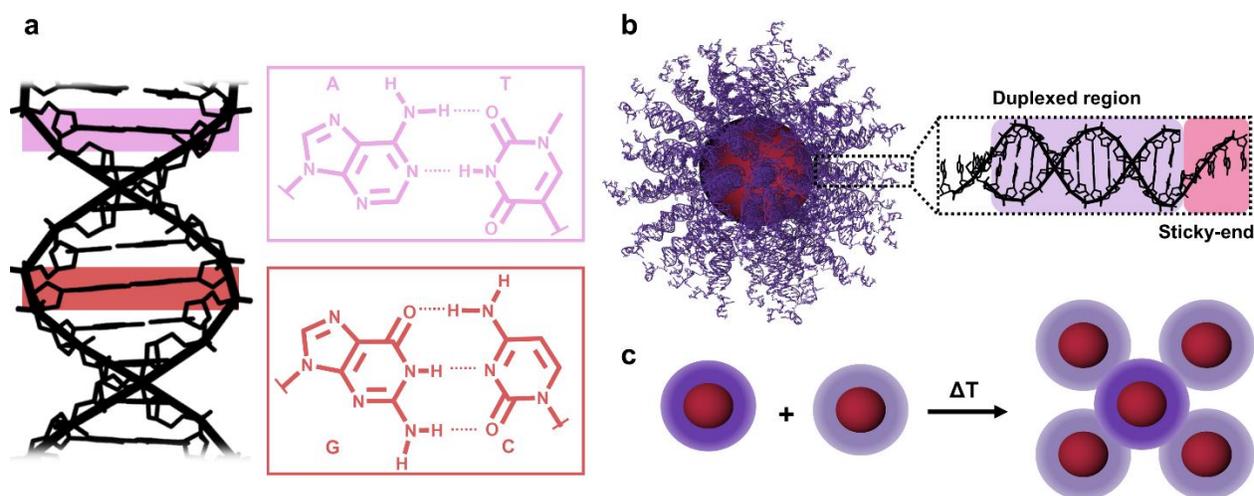


Figure 1.5 – DNA interactions and the anatomy of the PAE. (a) Hydrogen bond mediated base-pairing between the four bases of DNA. (b) Representation of PAE structure highlighting DNA design. (c) Assembly scheme showing formation of ordered PAE lattice mediated by temperature control.

To minimize the formation of undesired assembly pathways, and to encourage error-correcting reorganization of PAEs, careful consideration has been given to the DNA design of these constructs. Typically, two types of DNA strands are involved: the anchor strands, which are

chemically tethered to the surface of the particle core, and the linker strand, which form a fully complementary duplex with the anchor strands and possess a region of unpaired bases at the DNA terminus furthest away from the surface of the particle. This short region of unpaired bases is referred to as the “sticky-end” and multiple sticky-ends act collectively to drive the formation of bonds between PAEs (Figure 1.5b). The design of the sticky-end sequence determines the specific interactions between PAEs, for example, a palindromic sequence is inherently self-complementary and particles displaying a palindromic sticky-end will associate with themselves. The double stranded region, formed through hybridization of the anchor and linker strands, is an intentional and essential design feature of the PAE. Double stranded (ds) DNA is more rigid than single stranded (ss) DNA, owing to their differences in persistence length, and therefore pre-orient sticky-ends radially at the particles surface for hybridization and facilitates collective interactions that drive PAE bond formation. The slow cooling of PAEs through a temperature gradient, from above to below the melting temperature of the system, allows particles to dynamically sample their environment and encourages the formation of the most thermodynamically stable product (Figure 1.5c). Owing to the enthalpic driving forces of DNA hybridization, the most stable structures that form are highly ordered, crystalline materials where particles can maximize the number of hybridization events with neighbouring PAEs. Guided by this fundamental principle, predictive models, such as the Complementary Contact Model (CCM) and others, have been established to project structure outcomes based on the features of PAE inputs.⁸⁵

The predictability and programmability of DNA interactions, coupled with the ability to thermally control assembly, has revolutionized the bottom-up fabrication of nanomaterials, particularly in the field of colloidal crystal engineering. Comparing the structural control and

powerful ability to design materials that is provided by DNA to the current methods employed to program the assembly of proteins, reveals a vast difference. However, it also presents an opportunity to explore the development of DNA-based methods to direct the assembly of protein building blocks.

1.4 Protein-DNA conjugates

1.4.1 Bioconjugation of protein-DNA conjugates

Methods to synthesize protein-DNA conjugates are becoming increasingly available, motivated by the desire to create constructs that combine the recognition properties of DNA and the various functions of proteins. These constructs have been successfully synthesized using both covalent conjugation and non-covalent conjugation strategies. The most common approach to covalently join proteins with oligonucleotides is the use of heterobifunctional cross-linkers. These small molecules mediate the covalent conjugation of proteins and oligonucleotides by targeting a specific functional group on the surface of the protein and a different reactive species on the oligonucleotide ligand. An early example, reported by Niemeyer et al, demonstrated this method using sulfo-SMPB (sulfosuccinimidyl 4-(N-maleimidophenyl)butyrate) to crosslink thiolated-DNA and lysine residues on streptavidin proteins.⁸⁶ The high abundance of lysine residues (i.e primary amines) on the surfaces of proteins makes them a favourable choice for conjugation, however, this typically results in non-specific and multiple conjugation events. Cysteine, conversely, is relatively low in abundance on the surface of most proteins and its side chain, thiol, can react orthogonally to primary amines, making in an attractive candidate for highly specific conjugation events. New chemistries are being explored to target other low-frequency amino acids,

such as tryptophan and tyrosine,^{87, 88} in order to expand the toolbox of cross-linking conjugation methods. Additionally, non-canonical amino acids with biorthogonal reactivities can also be engineered at specific locations on a protein's surface to provide unique chemical handles, such as azides and tetrazines, for site-specific conjugation of oligonucleotides.⁸⁹ Finally, covalent coupling of proteins and oligonucleotides has also been achieved using enzymes that crosslink proteins tagged with nucleotide triphosphates and the 3' hydroxyl group of DNA ligands.⁹⁰

Protein-DNA conjugates can also be synthesized via non-covalent interactions, such as: native ligand binding; electrostatic; metal-mediated coordination; hydrophobic; and hydrogen bonding. A prevalent example is the use of the high affinity, native ligand interaction between biotin and streptavidin, the strongest known biological, non-covalent interaction, with a $K_a = 10^{15}$ / M between protein and ligand.⁹¹ This stability and the chemical capability to simply label oligonucleotides with biotin enables wide application of this technique in the preparation of protein-DNA conjugates, for example, by Kim et al, to synthesize and study the assembly of multi-valent streptavidin-DNA conjugates.⁹² Recombinantly expressed proteins are frequently purified through the non-covalent interactions between poly histidine-tags (His-tags), located at the N- or C-terminus, and metal ions. Therefore, methods to prepare protein-DNA conjugates by harnessing the strong coordination between His-tags and nickel (Ni^{2+}) ions have also been developed. These typically involve modifying oligonucleotides with chelating moieties such as nitrilotriacetic acid (NTA) that form complexes with Ni^{2+} and His-residues.⁹³

1.4.2 Building materials and systems from protein-DNA conjugates

Protein-DNA conjugates are highly modular structures, representing a union of two functional moieties, and have found applications in detection, sensing, as therapeutics and in the supramolecular assembly of protein-based nanomaterials. For example, antibody-DNA conjugates provide the benefits of extremely high-affinity and specificity of antibody-substrate binding in concert with the signal amplification properties of DNA, by means of the polymerase chain reaction (PCR), to create effective detection platforms. Such constructs have been investigated as immunoassays that demonstrated significantly enhanced performance (1000-10,000 fold-enhanced detection limit) over the conventional enzyme-linked immunosorbent assay (ELISA).⁹⁴

The programmability of DNA interactions enables the organization, design and application of functional protein-based materials and systems. Several approaches have been taken to impart proteins with the programmable association properties of DNA to control their assembly into higher-order architectures and materials. These efforts can be categorized into two groups: template-directed approaches, where DNA nanostructures define positions onto which proteins are immobilized, and surface-directed approaches, where the collective structural properties of protein surfaces, modified with DNA, dictate the assembly outcome. Template-directed methods are well established and typically utilize sophisticated, DNA tiles, cages, or origami-scaffolds with predetermined architectures to program the spatial organization of proteins.⁹⁵ Protein identity, number and relative positioning can all be defined by the template, which are important structural features in the development of functional materials, for example, multi-step enzyme cascade nanoreactors.⁹⁶

Surface-directed approaches are relatively less well developed compared to template-directed methods. This is likely due to the challenges associated with controlling the structural outcomes of surface-directed assembly events, compared to the definitive architectural control that a DNA template provides. However, the large body of research on controlling the assembly of inorganic nanoparticles using surface DNA ligands provides critical blueprints for establishing surface-directed assembly methods of protein-based materials. Foundational work by Mirkin and coworkers revealed that proteins, whose surfaces are modified with a dense layer of radially oriented DNA ligands (protein-PAEs), akin to inorganic nanoparticle PAEs, can indeed form 3D protein crystalline lattices.⁹⁷ Proteins, however, offer design handles that inorganic nanoparticles do not: their chemically diverse, mutable, and well-defined surface chemistries allow the spatial distribution, number, and type of DNA strands to be tuned with precision not currently possible with inorganic particles.

The additional design parameters afforded by proteins offer a rich opportunity to develop methods to control the assembly of protein building blocks with judicious design and placement of DNA surface ligands. This potential is alluded to by two recent examples from Aida, Mirkin and coworkers, that report the successful design of directional DNA interactions between proteins to program their assembly into 1D linear fibers.^{98,99} These structures are reminiscent of the naturally occurring filamentous materials formed by the polymerization of proteins such as actin and myosin. Moreover, the DNA interactions introduced to the axial surfaces of these proteins, GroEl and β -galactosidase, define clear interfaces through which assembly occurs, akin to interfaces defined by PPIs in naturally occurring protein assemblies.

While there is clear potential to program the assembly of proteins using DNA ligands to define interfaces between proteins, we know that Nature builds complex and functional materials by controlling both the interactions and the assembly pathway of proteins. The next frontiers in surface-direct protein assembly, using DNA, is to discover how to encode multiple orthogonal surface interactions and to determine how assembly pathways of protein building blocks can be programmed. This knowledge would dramatically expand the architectural control, diversity and potential applications of these materials and systems.

1.5 Thesis overview

The body of work presented herein focusses on addressing the fundamental question: how can the assembly pathway of proteins be programmed? To this end, we explore protein-DNA conjugates as modular building blocks and determine key design parameters that encode specific protein interactions, assembly pathways and final architecture. This thesis teaches how the amino acid sequence of proteins, in concert with the sequence and chemical conjugation strategies of the oligonucleotide ligands can be designed to direct and regulate the association of proteins building blocks, define specific assembly pathways and control novel material outcomes. These lessons advance the field of protein-materials engineering with DNA by addressing challenges associated with defining multiple, orthogonal interactions on protein building blocks, controlling the pathway of protein polymerization, and constructing hierarchically structured protein materials.

Chapter two explores the concept of programming the polymerization pathway of proteins through DNA sequence design. While it has previously been established that even a single oligonucleotide surface modification can mediate the association of nanoparticles and proteins in the formation of dimeric or extended assemblies, methods to control the mechanism of protein

assembly, to influence structural outcome, had yet to be realized. Inspired by decades of research dedicated to controlling the covalent and supramolecular polymerization pathways of small molecules, we investigated how the energy barriers to protein association can be tuned by designing the secondary structure of the DNA strand conjugated to proteins. We examined and compared systems where DNA adopts either a “closed” hairpin or “open” single-stranded conformation, guided by the hypothesis that these secondary structures would encode chain-growth and step-growth polymerization pathways, respectively (Figure 1.6). With these findings, a method to program protein polymerization pathway, molecular weight, and product distribution by designing DNA secondary structure is reported.

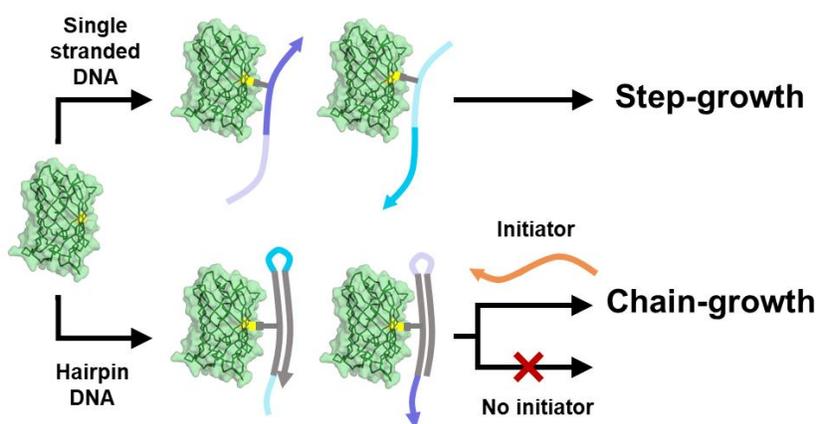


Figure 1.6 – Programming protein polymerization with DNA conformation. This scheme shows two sets of protein monomers modified with either single-stranded (top) or hairpin (bottom) DNA ligands. The distinct conformation, single-stranded and hairpin, dictates the polymerization pathway: spontaneous step-growth or initiated chain growth, respectively.

The work presented in the third chapter transitions to more intricately designed protein-DNA conjugates to investigate the assembly behaviour of proteins encoded with multiple, orthogonal DNA interactions that mimic the complexity of PPIs in Nature. Directing the association and assembly of systems containing many unique building blocks at the nanoscale

presents significant challenges, synthetically, but is achieved frequently and successfully in Nature to create hierarchical, functional architectures. We begin to address these challenges by designing and synthesizing a protein-based Janus nanoparticle comprised of two proteins encoded with sequence-specific nucleic acid domains, tethered together by an inter-protein “DNA bond”. The assembly behavior of these novel structures is interrogated in the context of colloidal crystallization with inorganic PAEs. Owing to the Janus particle’s asymmetric, orthogonal DNA interactions, two unique groups of inorganic PAE can associate and co-crystallize to realize a new class of three-dimensional superlattice, only possible when two sides of the Janus particle are modified with orthogonal oligonucleotide sequence (Figure 1.7). The synthesis and assembly behavior reported here represents a significant milestone in developing DNA-based tools to program the assembly of complex, multicomponent hierarchically structured protein-based materials.

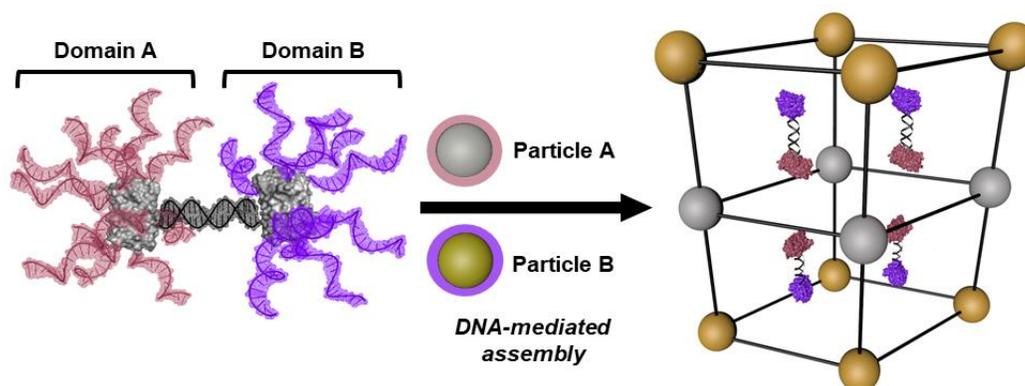


Figure 1.7 – DNA-encoded protein Janus nanoparticles. Scheme representing the asymmetric structure of DNA functionalized protein-based Janus particles and their assembly behavior in multi-component colloidal crystals.

Building on concepts studied in chapter three, chapter four focusses on the question: how can we encode hierarchical assembly pathways of proteins with DNA? In this work, we

investigated if stepwise assembly profiles of proteins could be encoded by modifying their surfaces with multiple, orthogonal DNA interactions that differ in relative binding “strength” (Figure 1.8a). The underlying hypothesis of this work is that building blocks with both sufficiently strong and weak surface interactions will be able to traverse a hierarchical assembly pathway where assembly steps are separated and controlled by the difference in interaction strengths (Figure 1.8b). We selected a model protein possessing a ring-like geometry and used mutagenesis to map the chemical anisotropy of the protein’s surface to the shape anisotropy, creating chemically well-defined axial and equatorial faces. The axial and equatorial faces were functionalized with either a strong or weak DNA sequence and the assembly behavior of these constructs was characterized through fluorescence-based techniques and microscopy. Directional (axial or equatorial) interactions arise in the first assembly step, mediated by the location of strong interactions, which result in the formation of an initial structure. Upon the formation of the initial structure, the weak interactions are displayed multivalently, resulting in a highly cooperative emergent interaction that initiates a second assembly step and subsequent architectures. By bringing together DNA design and protein engineering, we have created protein-DNA building blocks encoded with information that instructs their assembly along a specific multi-step pathway. This work is a powerful display of how protein and DNA sequence design can be harnessed to engineer pre-determined hierarchical assembly pathways for proteins that give rise to previously inaccessible structures.

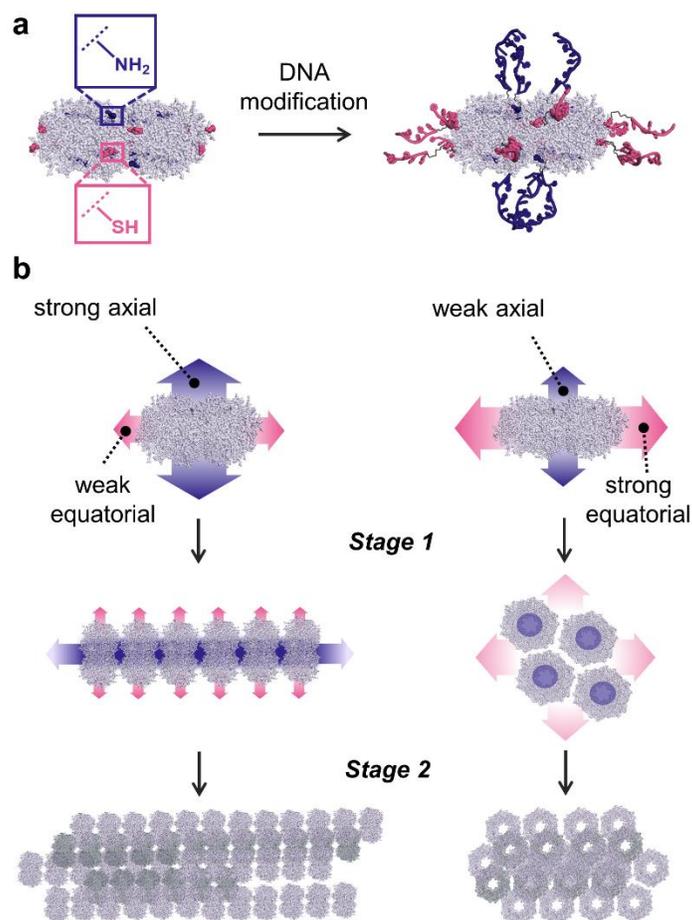


Figure 1.8 – Programming hierarchical assembly pathways of proteins with DNA. (a) DNA modification of chemically anisotropic protein with two different oligonucleotide sequences leads to protein-DNA conjugates with regiospecific functionalization pattern. (b) Proteins encoded with strong and weak axial and equatorial interactions, or vice versa, undergo distinct multi-step assembly pathways.

Chapter five summarizes the key conclusions from the body of work presented in this thesis and provides a reflection on how this work has impacted the fields of protein-materials engineering with DNA and, more broadly, DNA nanotechnology. The remaining challenges in directing protein assembly with DNA are discussed and an important future research initiative is proposed.

CHAPTER TWO

Programming Protein Polymerization Pathways with DNA

Material in this chapter is based upon published work:

McMillan, J. R.*; Hayes, O. G.*; Remis, J.; Mirkin, C.A. Programming Protein Polymerization with DNA. *J. Am. Chem. Soc.*, 2018, *140*, 1595-15956.

*Equal author contribution

2.1 Lessons from the polymerization of small molecules

The mechanisms by which chemists describe the polymerization of molecular monomers can be broadly classified into step-growth, chain-growth or ring-opening polymerizations pathways.^{100, 101} The molecular weight, dispersity and structure of polymers produced can, in general, be explained by the growth mechanism associated with a specific system.¹⁰² In step-growth polymerization, all species in solution are considered “active”, in the sense that monomer addition can occur with any free monomer or polymer chain-end, resulting in uncontrolled growth and a broad distribution of polymer chain lengths. The situation is different in chain-growth polymerization, since only the growing end of a polymer chain is “active”, so monomer addition is limited to a specific chain-end.¹⁰³ Polymerization therefore proceeds in a more controlled manner and dispersity is often very low. Controlling the mechanism is consequently key to determining the nature of the polymer material produced, with respect to dispersity of chain lengths and rate of polymerization. Living polymerization, a type of chain-growth polymerization, has also enabled chemists to realize polymeric structures of high structural complexity, such as block-copolymers, brush polymers and star-polymers.¹⁰⁴ Controlling the polymerization of small molecules, namely via living polymerization processes, has revolutionized polymer science by providing synthetic access to complex macro-molecules with precisely defined compositions and architectures, and therefore structures with uniform properties and specific functionalities.¹⁰⁵

Understanding and controlling polymer synthesis via covalent bond formation have been focusses of the chemistry community for the past century and extensive research efforts have led to the invention of many different techniques to covalently polymerize small molecules and control their architectural outcome.^{106, 107} Comparatively, methods to control supramolecular (non-

covalent) polymerization of small molecules are in their infancy and it was only recently discovered how the size and dispersity of these structures can be mediated.¹⁰⁸ A key distinction between covalent bonds and supramolecular bonds are their bond energies: covalent bonds are typically on the order of 100s of kJ / mol whereas supramolecular bonds are usually much lower in energy, for example hydrogen bonds range between 6 - 30 kJ / mol.¹⁰⁹ The low energy, weakly interacting, nature of supramolecular bonds gives rise to dynamic, adaptive, and responsive polymers and materials with distinct properties and applications from their covalent counterparts.¹¹⁰ However, developing strategies to control innately dynamic and weak bonds is inherently difficult and accessing diverse supramolecular polymer architectures still represents a significant challenge in the field of small molecule polymer synthesis.^{111, 112}

2.2 Challenges and opportunities for supramolecular protein polymerization

Supramolecular interactions are ubiquitous in biological systems and, unsurprisingly, some of the best instances of supramolecular polymers can be found in Nature. Controlled polymerization of proteins, such as actin and tubulin, is critical to the formation of specific structures in response to the ever-changing needs to the cell and for performing functions such as cell division, organization of intracellular structure, intracellular transport, and as ciliary and flagellar motility.^{113, 114} It has been shown that actin filaments and microtubules polymerize via highly regulated nucleation-growth mechanisms, similar to chain-growth mechanisms of small molecules, which are central to their functions.¹¹⁵ Polymeric protein structures range from simple repeating units of helical motifs, such as collagen, to intricate ensembles of multiple different proteins, such as the polyketide synthase, an oligomeric enzyme assembly-line responsible for the

synthesis of chemically diverse polyketide natural products that form the basis for nearly one-third of pharmaceuticals.¹¹⁶

Inspired by these examples of functional materials in Nature, supramolecular protein polymers have become important synthetic targets with a wide variety of potential applications in biology, medicine, and catalysis.¹¹⁷ However, directly translating *in vivo* polymerization pathways to *in vitro*, synthetic systems is not feasible, since the assembly, disassembly and reorganization of protein polymers in biological systems are carefully orchestrated by a host of complex binding events.^{25, 118, 119} To date, synthetic strategies to polymerize proteins have largely focused on introducing relatively simple supramolecular interactions to the surfaces of proteins, such as host-guest partners and metal coordination complexes, to mediate the formation of 1D linear or cyclic protein polymers.^{120, 121, 122} These approaches are all demonstrations of step-growth polymerization and collectively lack the capability to deliberately alter the pathways by which proteins polymerize. Synthetically programming the polymerization pathways of proteins, beyond step-growth, would provide access to novel protein material with architectures and functions, akin to those observed in well-controlled molecular and biological systems.

Despite the large body of literature devoted to honing pathway control over the polymerization of small-molecule monomers, the extension of these concepts to building blocks at larger length scales, such as proteins, has not been explored. Indeed, while examples of protein and nanoparticle polymerization by a spontaneous step-growth process have been reported,¹²³ the ability to deliberately control the polymerization process of nanoscale building blocks presents a significant challenge due to the inherent difficulties of finely controlling interactions on this length scale. In the field of supramolecular polymerization, recent examples have demonstrated that the

conformation or aggregation state of monomers in solution can dictate whether polymerization occurs spontaneously via a step-growth process, or a chain-growth pathway. By designing monomers that exist in kinetically trapped states that only undergo polymerization in the presence of an initiating molecule that eliminate the kinetic barrier to polymerization, chain-growth mechanisms have been realized.^{110, 124, 125} Thus, in general, the kinetic barrier toward polymerization, or lack thereof, dictates whether a system follows a spontaneous step-growth pathway, or whether the possibility for chain-growth exists.

DNA has emerged as a highly tailorable bonding motif for controlling the assembly of nanoscale building blocks, including proteins, into both crystalline and polymeric architectures.^{97-99, 126} In these systems, sequence specificity and carefully designed sticky ends, along with ligand placement, are employed as design handles to control particle association and therefore the final thermodynamic structure of an assembly. However, oligonucleotide conformation also represents an additional and promising design handle for controlling assembly events of proteins. Pertinently, conformational switching of oligonucleotides is frequently exploited in biological systems in the form of riboswitches,¹²⁷ and G-quadruplexes for regulating cell processes.¹²⁸ This chapter explores the hypothesis that DNA conformation can be designed to modulate the energetic barriers of protein assembly in a manner reminiscent of the supramolecular strategies that manipulate polymerization pathways of small molecules by designing kinetic barriers to polymerization. A general strategy is established by which the assembly pathway of proteins, or in principle any nanoscale building block, can be finely controlled using DNA interactions to synthesize protein polymers with controlled product distributions.

2.3 Design and synthesis of protein-DNA conjugates for polymerization

To direct the pathway of DNA-mediated protein polymerization, two distinct sets of DNA sequences were designed that, although identical in their overall complementarity, differ in the energy barrier that exists for polymerization. The DNA design for protein monomers expected to engage in a step-growth process (Figure 2.1a), consists of two 48 base pair (bp) strands that possess minimal secondary structure, and therefore a minimal energetic barrier for monomer association. Polymerization of the step-growth monomers is driven by the staggered complementary overlap between two halves of each of the 48 bp DNA sequences (Figure 2.1c). Therefore, the indefinite association of alternating A and B strands in one dimension is theoretically possible. To realize a chain-growth polymerization pathway (Figure 2.1b), DNA sequences where monomers would remain kinetically trapped until the addition of an initiator sequence were utilized. To this end, we employed the hybridization chain reaction, a DNA reaction scheme where a set of two hairpins can be induced to polymerize upon the addition of an initiator sequence.¹²⁹ Here, two 48 bp hairpins were used, with a 18 bp stem and orthogonal 6 bp toeholds such that the loop of hairpin A was complementary to the toehold of hairpin B. Polymerization will only occur when an initiator strand opens hairpin A, thereby exposing its loop sequence that is complementary to the toehold of hairpin B, thus inducing a cascade of hairpin opening (Figure 2.1c). Overall, each set of DNA sequences employed possesses an identical length and duplexation pattern, with 65% of A- and B-type sequences being identical between step- and chain-growth DNA (Materials and Methods Table 2.1). They differ, however, in the designed conformation and conditions required to initiate polymerization.

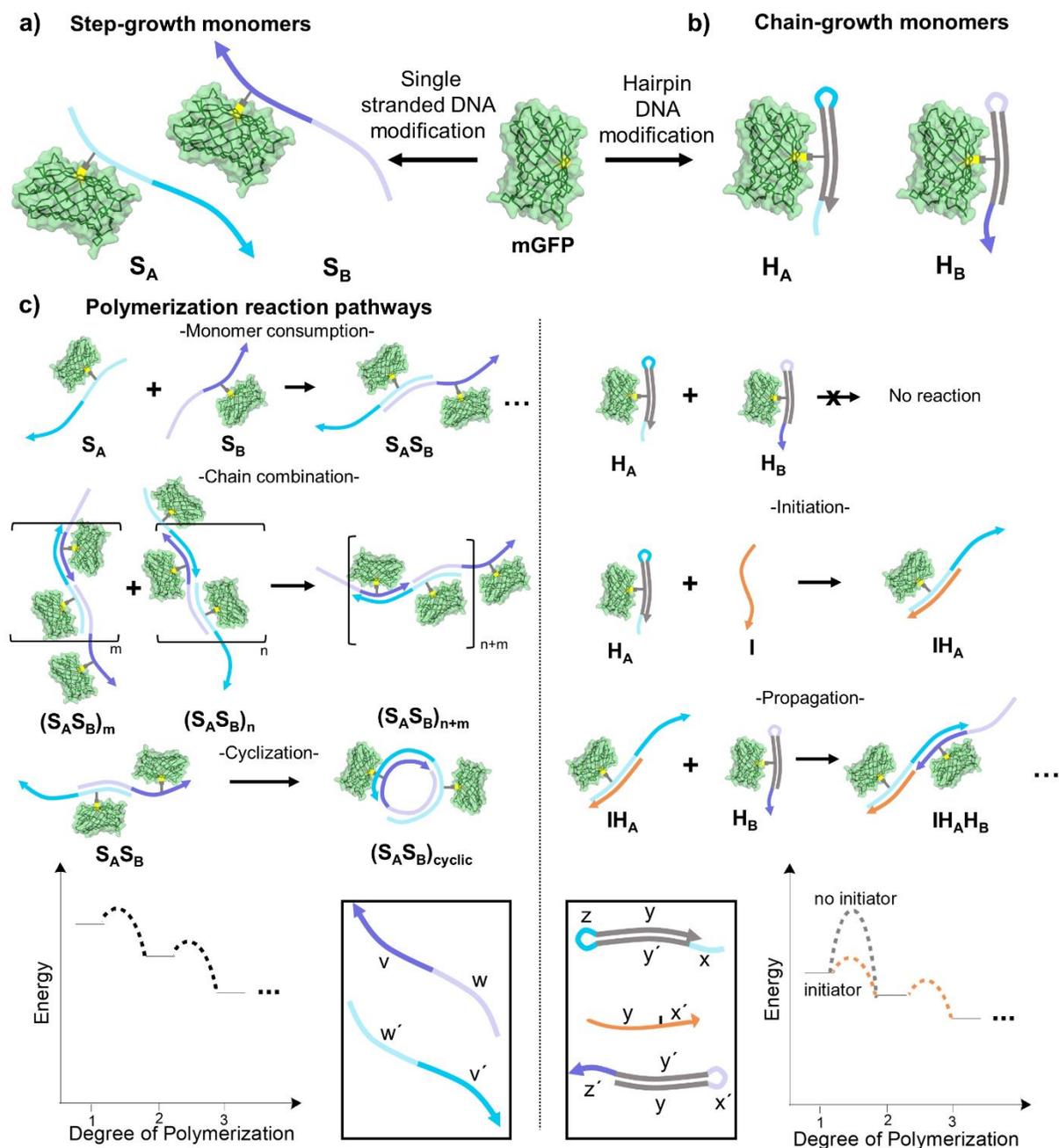


Figure 2.1 - Step-growth and chain-growth mGFP–DNA monomer sets and their reaction pathways. (a) Step-growth monomers S_A and S_B with a single-stranded DNA modification and therefore no kinetic barrier to polymerization. (b) Chain-growth monomers H_A and H_B possessing a hairpin DNA modification, and therefore an insurmountable kinetic barrier to polymerization in the absence of an initiator strand. (c) Proposed association pathways for step- (left) and chain-growth (right) monomer systems based on the DNA sequence design (bottom, boxes). Proposed system free energy diagrams for polymerization events are also shown.

A mutant, green fluorescent protein (mGFP) was chosen as a model system to explore how DNA sequence and conformation can be used to program the polymerization pathway of protein monomers. Its monomeric oligomerization state and solvent accessible cysteine residue (C148) enable the preparation of protein–DNA conjugates with a single modification of the designed oligonucleotides. For all the systems studied, mGFP–DNA monomers were prepared using an identical procedure (see Materials and Methods section 2.7.2). Briefly, an excess of pyridyl disulfide-functionalized oligonucleotide was incubated with mGFP overnight, followed by purification by anion-exchange to remove any unreacted protein, and nickel affinity to remove excess DNA.

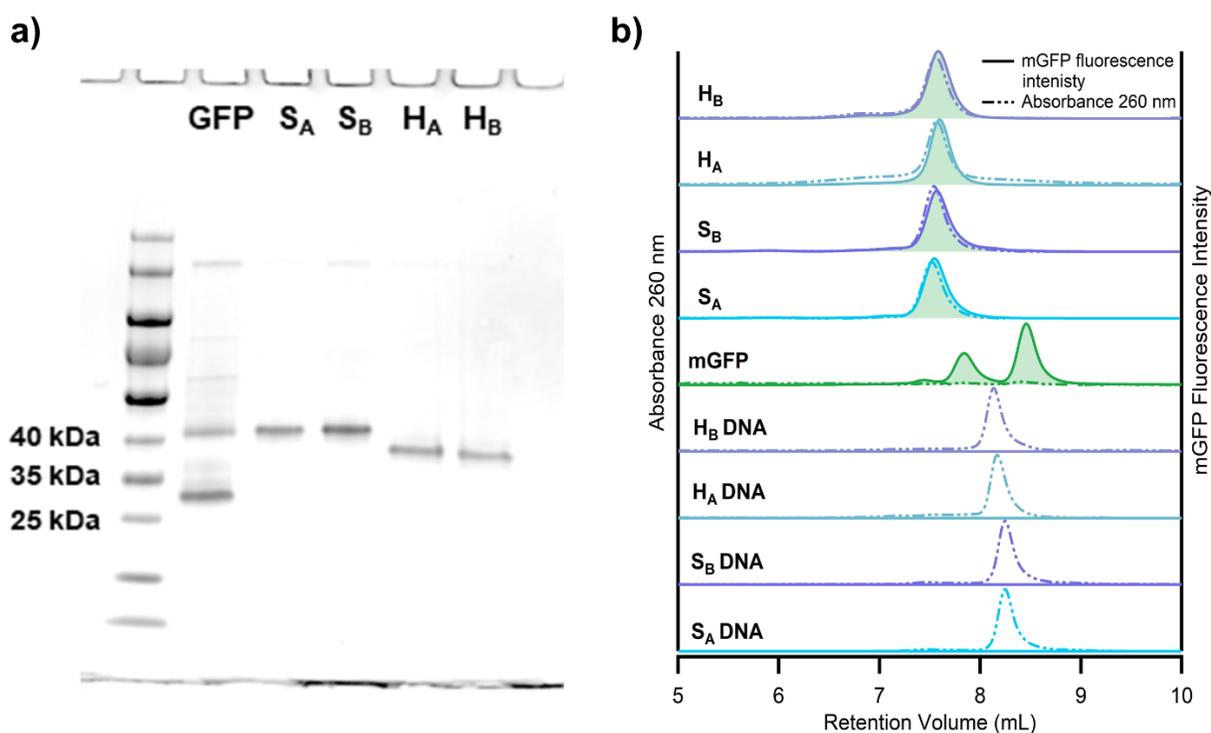


Figure 2.2 – SDS PAGE and SEC analysis of monomers. (a) SDS PAGE displaying bands with electrophoretic mobilities that corresponds well to the addition of a single oligonucleotide to the surface of the protein. (b) SEC chromatograms of native mGFP, free DNA, and mGFP–DNA monomers. Data confirms the absence of free DNA and unconjugated mGFP from purified monomer samples. The chromatogram for mGFP shows a higher molecular weight peak that corresponds to the oxidized dimer of the protein that is removed upon anion exchange purification of the DNA conjugates.

SDS–PAGE analysis of both the single-stranded protein–DNA conjugates, S_A and S_B , and the hairpin protein–DNA conjugates, H_A and H_B , revealed single protein bands with a decrease in electrophoretic mobility, consistent with the incorporation of a single 48 bp DNA modification (Figure 2.2a). Importantly, both H_A and H_B displayed slightly higher mobilities than S_A and S_B , consistent with the more compact DNA conformation resulting from the hairpin sequences employed. Furthermore, analytical size-exclusion chromatography (SEC) of all monomers showed discrete peaks that confirmed the expected mass increase, as well as the absence of any free DNA or aggregated protein (Figure 2.2b). Finally, UV–vis spectra of the conjugates revealed ratios of mGFP chromophore absorbance (488 nm) to DNA absorbance (260 nm) that were consistent with the conjugation of a single strand of DNA to each protein (Figure 2.3).

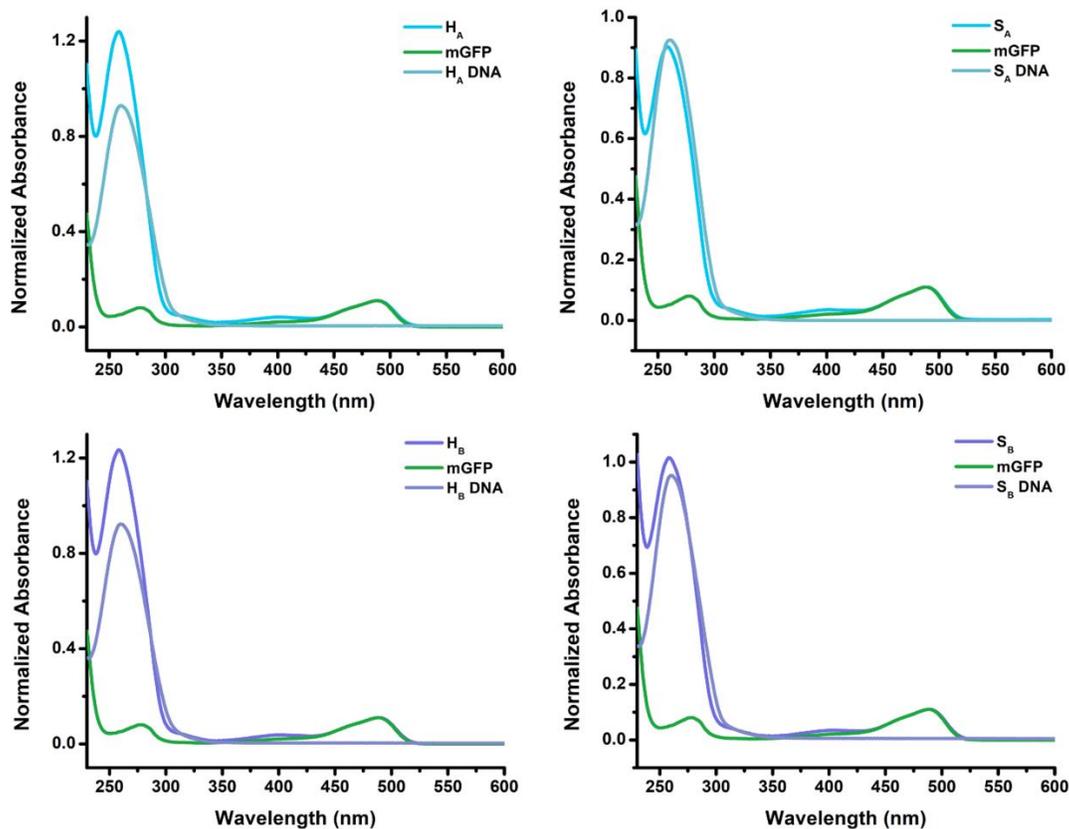


Figure 2.3 - UV-vis spectra of mGFP, free DNA, and mGFP DNA monomers. Each plot shows the spectra for unmodified mGFP (green), free DNA and purified mGFP-DNA conjugates for each monomer. All spectra on each plot are normalized to a concentration of 2 μM and give an approximate ratio of 1 DNA:1 mGFP for mGFP-DNA conjugates.

Taken together, these data unambiguously confirm the synthesis and purification of the desired protein–DNA conjugates. Significantly, each set of monomers synthesized are nearly identical in their overall mass, and the appended DNA strands possess identical staggered complementarity between A and B monomers, differing only in the conformation of the DNA modification. The central hypothesis of our work, therefore, is that this small difference in sequence, and thereby conformation, of the protein-appended DNA will alter the underlying pathway of polymerization of the monomers between a spontaneous, step-growth process and an initiated, chain-growth one.

2.4 Protein polymers via a step-growth pathway

2.4.1 Characterization of product structures and distribution

The polymerization of single-stranded mGFP–DNA monomers (Figure 2.4a) was initially examined using analytical SEC as an effective method of characterizing the aggregation state of mGFP. The combination and overnight incubation of equimolar amounts of S_A and S_B monomers at room temperature resulted in size exclusion profiles indicative of near complete monomer consumption, and the presence of higher-order aggregates (Figure 2.4c). While the majority of species in solution were above the exclusion limit of the column employed, low molecular weight species were also present. The lower molecular weight species that persisted in the sample, even after several days, suggested the presence of cyclic products. To better characterize the product distribution, we analyzed the samples by cryo-EM to enable the direct characterization and

quantification of product distribution and possible cyclic products, in a close-to-solution-state environment that avoids drying artifacts. Obtaining images with sufficient contrast to enable the conclusive identification of species composed of mGFP monomers, a protein much smaller than those routinely visualized via cryo-EM, connected through a double-stranded DNA backbone is nontrivial. Indeed, even when employing large defocus with a direct-electron detector camera, the synthesized structures could barely be discerned. To improve the contrast in these images, we employed a Volta phase plate, a thin continuous carbon film which phase shifts the scattered electron beam, increasing in-focus phase contrast,¹³⁰ and thereby greatly enhancing the signal-to-noise ratio in our images.^{131, 132}

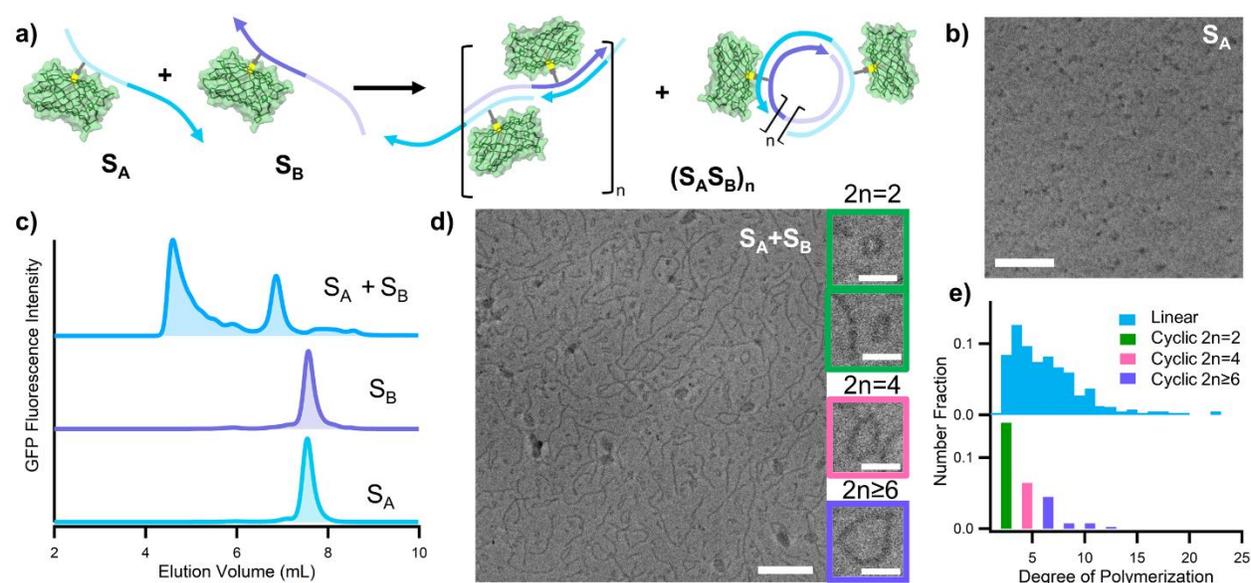


Figure 2.4 - Step-growth polymerization of mGFP–DNA monomers, S_A and S_B . (a) Scheme showing the spontaneous polymerization of single-stranded monomers into linear and cyclic products. (b) Cryo-EM micrograph of S_A monomer. Scale bar = 50 nm. (c) SEC profiles of S_A and S_B monomers, and polymerization products after incubation for 24 h. (d) Cryo-EM micrograph of polymers grown from S_A and S_B monomers within sets showing dominant cyclic products. Scale bar = 50 nm (10 nm in cyclic insets). (e) Histogram of number fraction degree of polymerization of linear (top) and cyclic species (bottom).

The phase plate enabled the double-stranded DNA backbone to be clearly visualized, and in certain images, small spots of electron density corresponding to mGFP could also be visualized

(Figure 2.4b, d). The micrographs clearly revealed a mixture of linear and cyclic products, which were quantified using a fiber analysis software (see Materials and Methods section 2.7.5).¹³³ This analysis revealed that cyclic products, formed through intra-chain hybridization of terminal complementary overhangs, accounted for 28 number percent of the overall product distribution. Quantification of cycle circumference enabled us to determine that the dominant cyclic product formed (15 number percent) is through the dimerization of S_A and S_B (Figure 2.4e).

2.4.2 Commentary on cyclic structure formation

Cyclic oligomers are a commonly observed side product of both covalent and supramolecular polymerizations that undergo a step-growth mechanism, where both ends of a growing polymer chain are reactive, and therefore the possibility of cyclization exists. Indeed, the presence of cyclic products has been posited in DNA-only polymerization systems with similar staggered DNA designs but have never been observed directly.¹³⁴ Our observed distribution of cyclic products, dominated by a 48 bp cyclic dimer having a 15 nm diameter may appear surprising at first given the widely reported persistence length of DNA of ~ 50 nm.^{135, 136, 137} However, the bending of double-stranded DNA well below its persistence length has been reported: DNA as short as 63 bp in length has been shown to form cyclic structures spontaneously for double strands containing a 10 bp single-stranded overhang region that hybridizes upon cyclization (compared to 24 bp in this system),^{138, 139} and template-directed ligation approaches have been reported to result in un-nicked cycles as small as 42 bp.^{140, 141} Furthermore, sharply bent DNA can be explained by the presence of kinks,¹⁴² which format DNA nick sites.¹⁴³ Interestingly, cyclic dimers can be observed with both circular conformations, and more oblate conformations, where it appears that sharp DNA bending maybe occurring at nick sites (Figure 2.4d). The cryo-EM techniques

employed have enabled the thorough characterization of products resulting from our mGFP monomers with single-stranded DNA modifications, demonstrating a distribution consistent with the designed step-growth formation process. This EM study also suggests that cryo-EM coupled with phase plate technology is a powerful platform to readily observe the conformations of sharply bent DNA and give insight into the topology of small DNA minicircles.¹⁴⁴

2.5 Protein polymers via a chain-growth pathway

2.5.1 Characterization of product structures and distribution

Having shown that DNA can mediate the spontaneous polymerization of proteins resulting in product distributions consistent with a step-growth process, we next tested the overarching hypothesis of this work: that the underlying pathway of protein-monomer polymerization can be controlled by the secondary structure of the appended DNA sequence, which in turn controls the energy barrier to polymerization. First, we combined H_A and H_B monomers under identical conditions to those studied in the step-growth system, to test whether the hairpin DNA design impeded the spontaneous polymerization of monomers as desired. Indeed, we observed SEC profiles that were indistinguishable from the individual monomers, even after 1 week of incubation at room temperature (Figure 2.5a, b). Furthermore, the absence of any polymerized species was evident from cryo-EM images (Figure 2.5c). While the structure of the mGFP–hairpin monomers is not immediately obvious upon inspection, 2D class averages of ~250 particles clearly show electron density corresponding to both mGFP and the hairpin appendage (Figure 2.5d). Importantly, previously reported attempts to apply the hybridization chain reaction to control the association of proteins were unsuccessful due to the challenge of annealing hairpins conjugated to

thermally unstable proteins.¹⁴⁵ Here, however, we circumvent this problem by snap-cooling the hairpin DNA prior to the protein conjugation reaction described above.

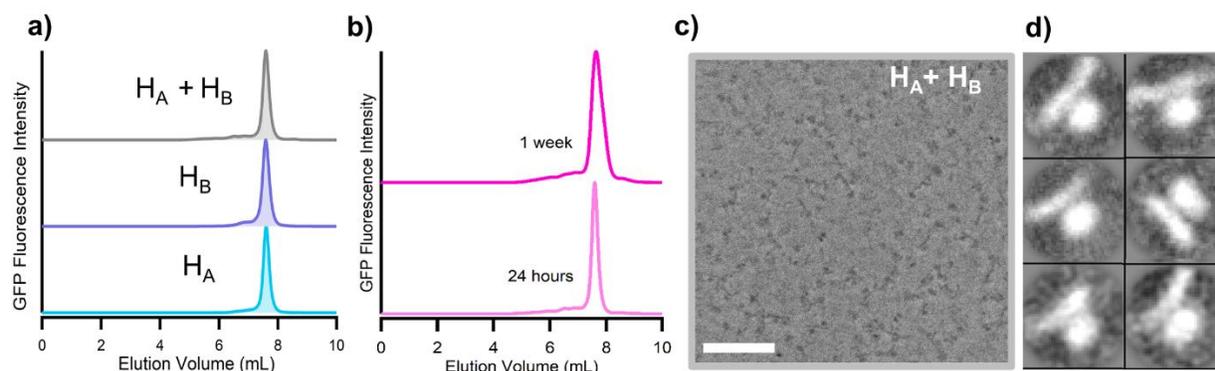


Figure 2.5 – Characterization of mGFP-DNA hairpin monomers in the absence of initiator. (a) SEC profiles of H_A and H_B monomers separately and together after incubation for 24 h without initiator. (b) SEC chromatogram of H_A and H_B monomers after incubation for 24 hours, and after 1 week of incubation at room temperature. (c) Cryo-EM micrograph of H_A and H_B monomers. Scale bar = 50 nm. (d) 6 classes that were generated from data processing showing multiple orientations of the protein-hairpin DNA conjugate.

The addition of the initiator strand induces the polymerization of mGFP–DNA monomers (Figure 2.6d), as evidenced by SEC (Figure 2.6a). Varying the equivalents of initiator strand with respect to monomer dramatically changes the molecular weight distribution of aggregates observed by SEC (Figure 2.6a). Qualitatively, these chromatograms show that the molecular weight distribution decreases with increasing equivalents of initiator, with species below the exclusion limit of the column becoming more prominent at higher initiator concentrations, consistent with a chain-growth polymerization process. Cryo-EM analysis of these samples allows this change to be quantified: a steady increase in both number and weight-average degree of polymerization from 3.7 and 4.9, to 6.9 and 10 units was observed from 1 to 0.4 equiv. of initiator, respectively (Figure 2.6b, c and e). Importantly, these images also reveal the presence of only linear products for all initiator concentrations tested, in stark contrast with the large population of cyclic products observed for the step-growth system. Since polymers growing via a chain-growth process contain

only one single-stranded “active end”, with the other end remaining fully duplexed with initiator, cyclization events are not kinetically accessible. This change in product distribution from a mixture of both cyclic and linear species, to exclusively linear, therefore reflects the change in polymer formation pathway.

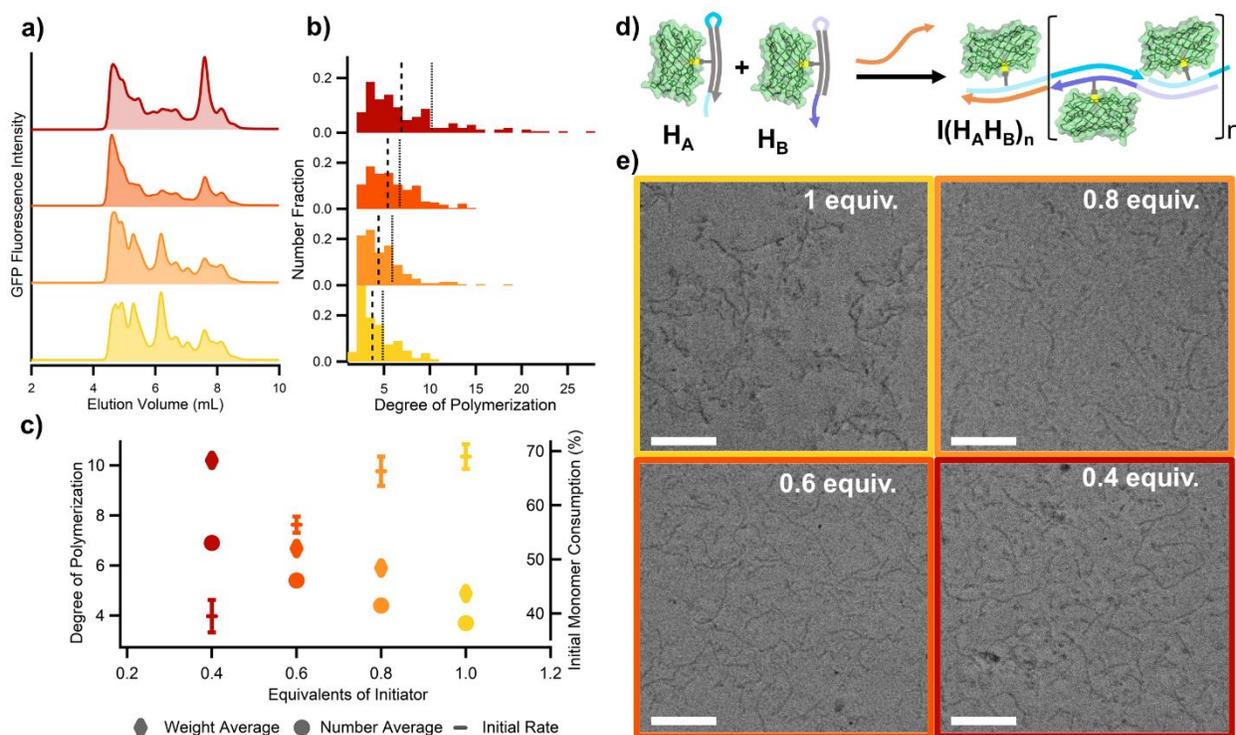


Figure 2.6 – Chain-growth polymerization of H_A and H_B monomers. (a) SEC profiles of chain-growth polymerization products with 0.4, 0.6, 0.8, and 1.0 equiv. of initiator. (b) Quantitative analysis of degree of polymerization for monomers with 0.4, 0.6, 0.8, and 1.0 equiv. of initiator (top to bottom). Long dashed lines indicate number-average degree of polymerization, and short dashed lines indicate weight-average degree of polymerization. (c) Weight and number-average degree of polymerization (left axis) and % initial monomer consumption (right axis) as a function of equivalents of initiator added. (d) Scheme showing the initiated polymerization of chain-growth monomers. (e) Cryo-EM micrographs of samples prepared with different concentrations of initiator

We also estimated the initial rate of monomer consumption via SEC, which increased with increasing initiator concentration, another key characteristic of chain-growth pathways at the molecular scale (Figure 2.6c, and Materials and Methods section 2.7.4). Furthermore, the product distribution of the system could also be shifted by changes in the timing of initiator addition,

similar to molecular polymerization techniques.¹⁴⁶ When 1 equiv. of initiator was added in five aliquots over 25 or 75 min, an SEC profile with a significantly larger fraction of high-molecular-weight products was observed, with the percentage of species eluting with a retention volume below 5 mL increasing from 27%, to 31% and 43% of the overall integrated area of the mGFP fluorescence signal, respectively (Figure 2.7). This suggests that directing protein polymerization via the hybridization chain reaction enables control over both molecular weight and polydispersity of the resulting protein polymers.

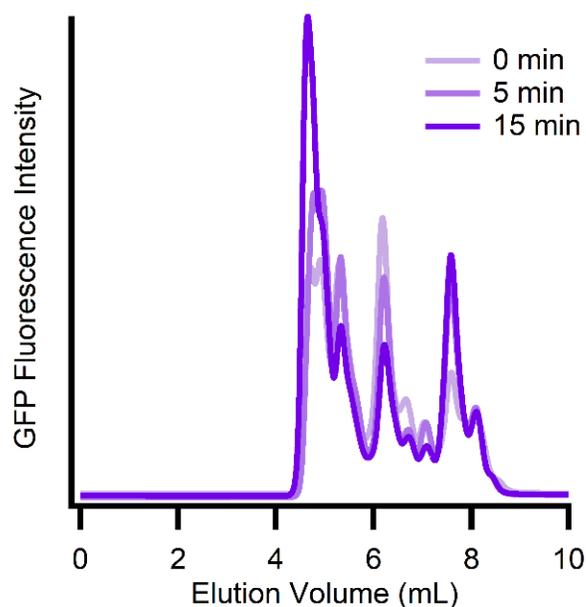


Figure 2.7 - Effect of initiator addition timing on polymer distribution. SEC of H_A and H_B with 1 equivalent of initiator added over 5 additions at different time intervals. The legend refers to the time interval between each addition: the experiment was conducted by adding 1 equivalent of initiator all at once (0 min), or 0.2 equivalents every 5 min or 15 min until 1 equivalent total had been added to the sample.

Ultimately this system displays some important differences from an idealized chain-growth polymerization. In an ideal chain-growth reaction, the rate of initiation is fast relative to propagation and $M_n = [M]_0/[I]$. In this system, however, M_n is much greater than predicted from the $[M]_0/[I]$, suggesting that the initiation reaction does not reach completion before monomer is

depleted. In contrast with typical chain-growth processes, for example atom-transfer radical polymerization (ATRP),¹⁴⁷ where the rate of initiation is much faster than the rate of propagation, the rate of initiation in this system is likely similar to the rate of propagation, owing to the identical chemical nature of these two reactions from a DNA perspective. In addition, with initiator concentrations below 0.6 equiv. we observe a decrease in conversion from ~90 to 74% that persists even after several weeks. We compared these results to the free DNA system polymerized under identical conditions and observed almost complete consumption of monomer (90%) with 0.4 equiv. of initiator, which suggests the incomplete conversion observed for low initiator concentrations is not a thermodynamic consequence. Rather, this may be a result of a mass-transfer or chain-end accessibility problem, which will be the subject of future investigations to enable access to well-controlled complex macromolecular architectures.

2.5.2 Chain extension of living polymer chain-ends

Certain classes of covalent and supra-molecular chain-growth polymerizations display a living character, where chain termination events are absent. In these systems, because active chain ends persist indefinitely, the addition of fresh monomer to a sample of polymer results in the consumption of the monomer, and a subsequent increase in molecular weight distribution of the polymer sample (Figure 2.8a). The hybridization chain reaction employed herein has been proposed to possess a living polymerization character,¹²⁹ and based on the DNA sequences, no chain termination or combination events should be possible. Therefore, to test the living character of the chain-growth system, we added a polymerized solution of H_A and H_B with 0.6 equiv. of initiator to an equal volume of metastable monomer solution containing no initiator. Monitoring the monomer fraction in solution after the addition of the polymer, we observe the consumption

of the monomer over time via SEC (Figure S7), demonstrating that polymerization continues and suggesting chain extension. To characterize the change in molecular weight distribution after the addition of fresh monomer, we conducted cryo-EM analysis on this sample, which revealed a substantial increase in the number- and weight-average degree of polymerization from 5.4 to 7.3, and 6.7 to 14, respectively (Figure 3B,C). This excludes the possibility that the monomer consumption observed via SEC is solely a result of excess initiator strands reacting with fresh monomer, and conclusively demonstrates that the DNA-mediated chain-growth polymerization of proteins reported herein possesses a living character.

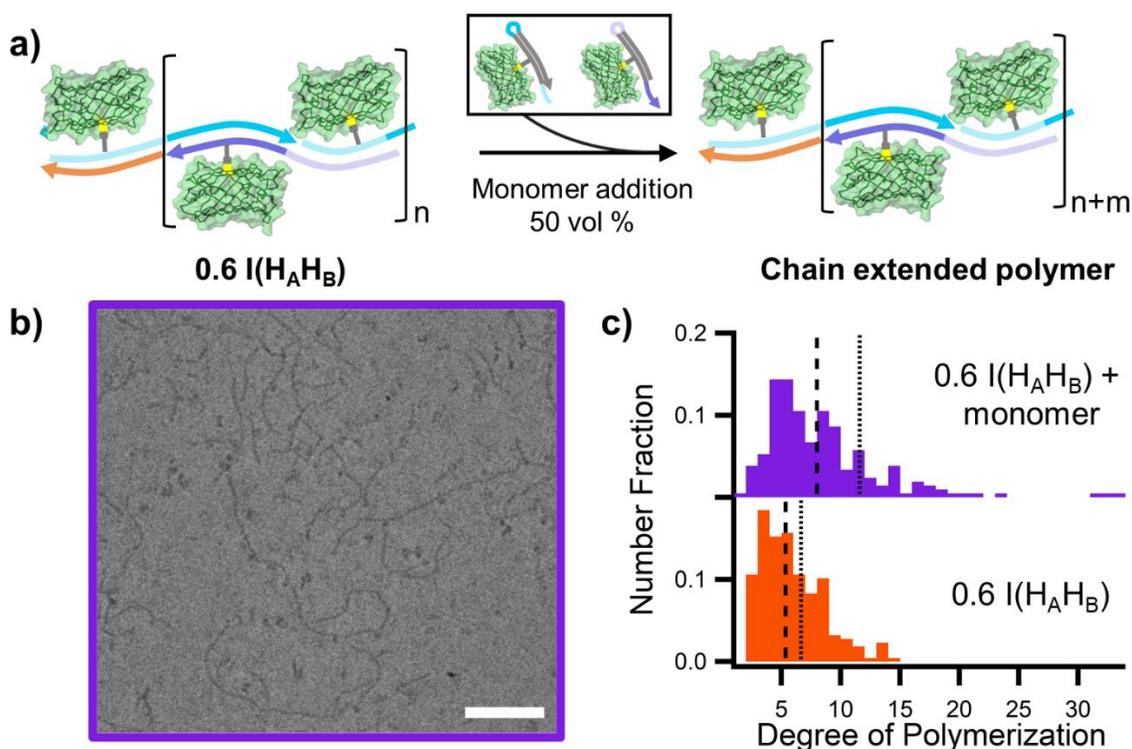


Figure 2.8 - Chain extension of polymers with active chain ends. (a) Scheme showing addition of fresh monomer to sample with active chain ends. (b) Cryo-EM micrograph of resulting chain extension products. Scale bar = 50 nm. (c) Histograms showing an increase in average degree of polymerization before (orange) and after (purple) chain extension. Long dashed lines indicate number-average, and short dashed lines indicate weight-average degree of polymerization.

2.6 Conclusions

The complexity observed in the assembly processes of proteins into highly intricate and functional polymeric architectures in Nature has been unparalleled in the synthetic space. Herein, we have reported an initial step in this direction by providing the first demonstration of designed protein polymerization path-way control. This work will yield access to new classes of protein polymer architectures, defined, and differentiated by sequence, branching type, and function. These controlled and complex architectures, as well as the reported cyclic structures, could represent a new biomaterial space for investigating different applications, including catalysis, sensing, tissue engineering, and pharmaceutical development.

In addition, this study stands as a powerful demonstration of how DNA can be used to precisely tune the energy landscape, and thereby assembly pathways, of nanoscale building blocks, and will enable the synthesis of entirely new classes of protein-based materials. The work presented in this chapter also highlights additional avenues of research into understanding how these polymerization pathways can be better controlled to afford materials with highly specified architectures and distributions. For example, we have recently reported on how a small change to the hairpin structure, through the incorporation of a mismatch in the stem region, can greatly enhance the control over the hybridization chain reaction.¹⁴⁸ This methodology facilitated the synthesis of discrete protein oligomers.

2.7 Materials and Methods

2.7.1 Oligonucleotide design, synthesis and purification

Oligonucleotides were synthesized on solid supports using reagents obtained from Glen Research and standard protocols. Products were cleaved from the solid support using 30% NH₃ (aq) for 16 h at room temperature, and purified using reverse-phase HPLC with a gradient of 0 to 75 % acetonitrile in triethylammonium acetate buffer over 45 min. After HPLC purification, the dimethoxytrityl group was removed in 20% acetic acid for 2 h, followed by an extraction in ethylacetate. The masses of the oligonucleotides were confirmed using matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) using 3-hydroxypicolinic acid as a matrix. For the chain-growth system, previously reported hairpin sequences were employed.¹²⁹ In the case of the step-growth system, sequences were designed using the IDT oligo analyzer tool, where the sequence of a single domain was iterated until the sequence afforded no secondary structure elements that displayed a predicted melting temperature above 25 °C.

Name	Sequence (5' → 3')	MW expected (Da)	MW observed (Da)	ε ₂₆₀ (M ⁻¹ cm ⁻¹)
H_A	TTAACCCACGCCGAATCCTAGACTCA AAGTAGTCTAGGAT NH ₂ TCGGCGTG	14890	14811	463800
H_B	AGTCTAGGATT NH ₂ CGGCGTGGGTT AACACGCCGAACCAGACTACTTTG	14953	14982	461500
I	AGTCTAGGATTCGGCGTGGGTTAA	7464	7444	239600
S_A	TTAGTCGTCTCTCATCATGTGTTACAA AGTAGTCTAGGAT NH ₂ TCGGCGTG	14949	14960	461700
S_B	TAACACATGAT NH ₂ GAGAGACGACT AA CACGCCGAATCCTAGACTACTTTG	14892	14845	476300

T NH₂ = C6 Amino dT modifier from Glen Research

Table 2.1 - DNA sequences, molecular weights, and extinction coefficients.

2.7.2 Synthesis and purification of mGFP-DNA monomers

The mutated plasmid containing the gene for the mutated EGFP (mGFP) was transformed into One Shot®BL21(DE3) Chemically Competent *E. coli* (Thermo Fisher) by heat shock, and cells were grown overnight on LB Agar plates with 100 µg/mL ampicillin. Single colonies were picked, and 7 mL cultures were grown overnight at 37 °C in LB broth with 100 µg/mL Ampicillin. These cultures were added to 1 L of Terrific Broth (Thermo Fisher) with 1 % glycerol and 100 µg/mL ampicillin, and cells were grown at 37 °C to an optical density of 0.6, then induced with 0.02 wt% arabinose overnight at 17 °C. Cells were spun down (6000 g, 30 min) and resuspended in 100 mL of 1x PBS, then lysed using a high-pressure homogenizer. The cell lysate was clarified by centrifugation at 30,000 g for 30 min and loaded onto a Bio-Scale™ Mini Profinity™ IMAC Cartridge (Bio-Rad). The column was washed with 100 mL of resuspension buffer, then eluted in the same buffer with 250 mM imidazole. The eluted fraction was further purified by loading on to Macro-Prep®DEAE Resin (Bio-Rad) and washing with 20 mL of 1xPBS. mGFP was eluted with a solution of 1xPBS + 0.25 M NaCl.

DNA conjugation was carried out immediately after purification. Amine terminated DNA (300 nmoles) was reacted with 50 equivalents of SPDP (Thermo Fischer Scientific) crosslinker in 50 % DMF, 1x PBS + 1 mM EDTA for 1 h at room temperature. Excess SPDP was removed from the DNA by two rounds of size exclusion using NAP10 and NAP25 columns (GE Healthcare) equilibrated with PBS (pH 7.4), consecutively. Ten equivalents of the resulting pyridyl disulfide terminated DNA was added to 1.5 mL of 20 µM protein solution, and the reaction allowed to proceed for 16 h at room temperature. For hairpin DNA –mGFP conjugation reactions, hairpin DNA was snap cooled after SPDP conjugation, but before being added to mGFP. This consisted

of heating the DNA solutions to 95°C for 4 mins, then 3 mins at 4 °C. The DNA solutions were then equilibrated at room temperature for 5 mins before adding to the protein solution.

mGFP-DNA monomers were purified using a two-step protocol to ensure removal of both unreacted DNA and protein. First, samples were loaded on Ni-NTA column, and washed with 30 mL of 1x PBS to ensure removal of excess DNA. The protein sample was then eluted with a solution of 1x PBS + 250 mM imidazole. This eluent was then loaded on Macro-Prep® DEAE Resin, and washed with 20 mLs of 1xPBS, and 1xPBS + 0.25 M NaCl. Subsequently, mGFP-DNA conjugates were eluted with a solution of 1xPBS + 0.5 M NaCl, and analyzed via SDS-PAGE to ensure successful DNA conjugation and purification.

2.7.3 Size exclusion characterization

Size-exclusion chromatograms were collected using an Agilent 1260 Infinity HPLC equipped with an Advanced Bio SEC 300Å column (Agilent). All chromatograms reported in this work were monitored at 260 nm, and using a fluorescence detector with an excitation at 488 nm and an emission of 520 nm. Samples were measured with an injection volume of 5 µL at a flow rate of 1 mL/min. For monomer characterization, samples were injected at concentrations between 2 and 5 µM. For polymer characterization, samples were injected at the concentration of assembly.

2.7.4 Polymerization experiments

All mGFP-DNA polymers studied were assembled at 1 µM of each building block (2 µM total protein concentration) in 1x PBS + 0.5 M NaCl at room temperature. For all characterization data presented, samples were incubated for a minimum of 12 hours at room temperature prior to analysis. For the chain-growth system, both monomers were combined and mixed in solution prior to the addition of the initiator strand. In this system, equivalents of initiator reported refer to

equivalents with respect to a single building block (i.e.: for 0.4 equiv. initiator, sample contains 0.4 μM initiator, 1 μM H_A and 1 μM H_B).

Kinetic measurements were conducted by adding initiator to a sample immediately (~15 seconds) prior to SEC injection and calculating the integrated area percent of the monomer peak after this first injection as an estimate of the initial rate of polymerization. The error bars reported in the main text report the standard deviation from triplicate measurements.

2.7.4 Cryo-TEM imaging and class averaging

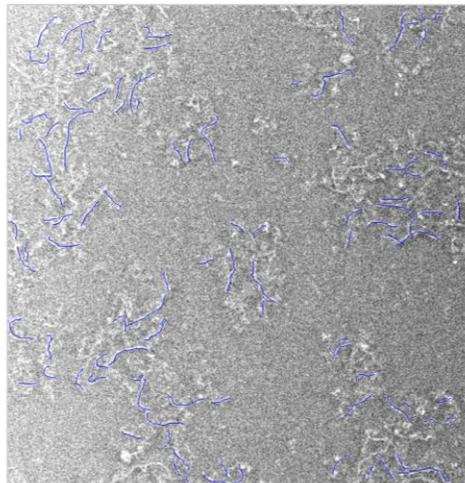
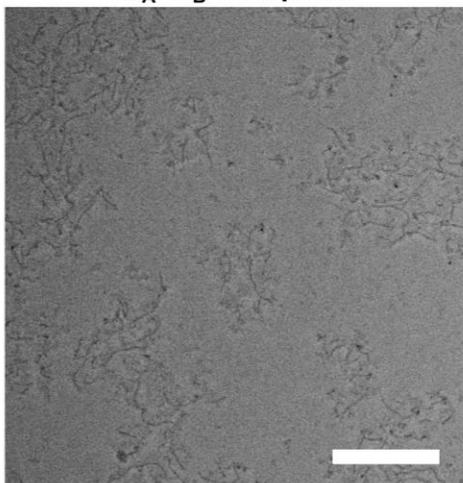
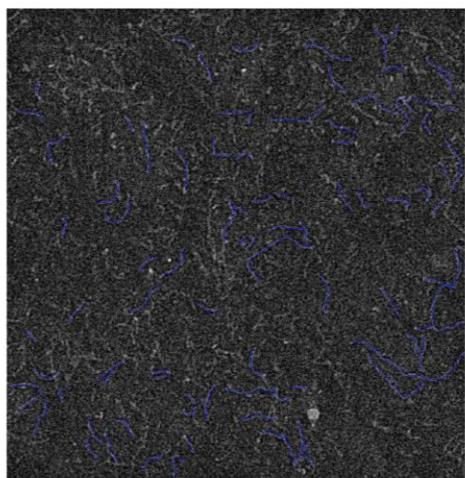
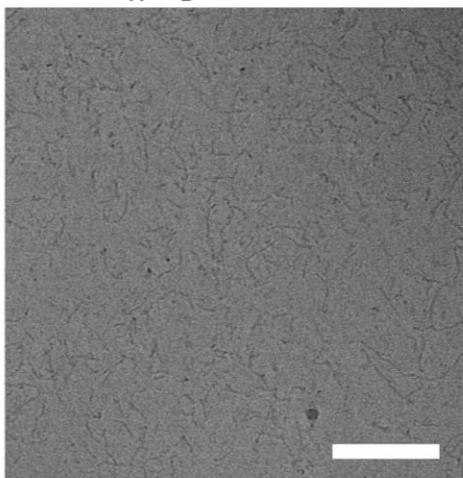
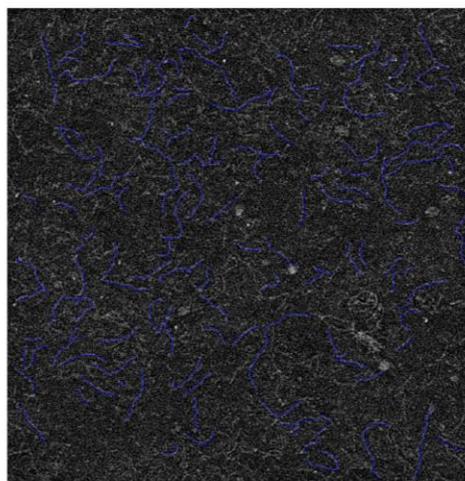
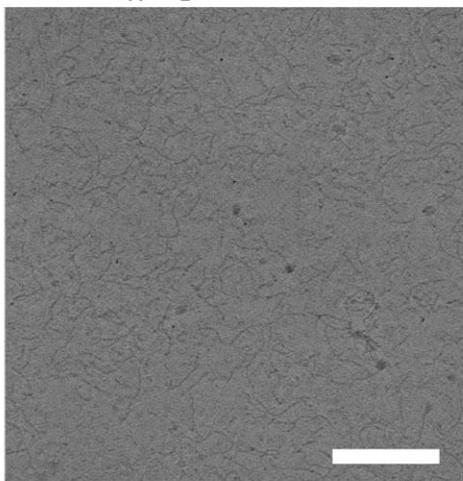
Sample solutions were deposited onto 400 mesh 1.2/1.3 C-Flat grids (Protochips) and were plunge frozen into liquid ethane using a VitrobotTM Mark IV. The grids were imaged using a JEOL 3200FS microscope operating at 300 kV equipped with a Volta phase plate and Omega energy filter. The microscope was aligned and adjusted to give 90° phase shift in acquired images. Movies were acquired on a K2 summit camera (Gatan) with a defocus range between 0.1-1.0 μm using counting mode with a pixel size of 1.1 Angstrom. The dose rate that was used was approximately $10\text{e}^-/\text{pix}/\text{s}$ (equivalent to $8.3\text{e}^-/\text{\AA}^2/\text{s}$ on the plane of the sample) for a total exposure of 6 seconds.

Images in Figure 2.5d were generated using the following class average processing procedure: 12 recorded movies were subjected to motion correction with MotionCor2.¹⁴⁹ Following CTF estimation with CTFFIND4,¹⁵⁰ 8 micrographs with the best quality were then selected for further processing. ~1500 particles were picked with a box size of 96 Angstroms, extracted, and 2D classification was all done within RELION-2 software package.¹⁵¹

2.7.5 Analysis of polymer length distributions

Polymer lengths were analyzed using FiberApp.¹³³ The relatively large noise level in the images necessitated that the polymers be identified visually. Only fibers where clear beginning and end points could be identified were counted, and every identifiable fiber was counted in each image analyzed. Images were binned and inverted prior to analysis in FiberApp to make fibers easier to visualize. For all samples 2-3 images were analyzed to give polymer number counts greater than 200. The calculated length generated by FiberApp was then converted to degree of polymerization (DP) using the following conversion based on the rise-per-base pair of double stranded DNA and then rounded to the nearest whole number:

$$(1) \quad DP = \frac{\text{length (nm)}}{24 \text{ bp} \times 0.332 \text{ nm/bp}}$$

$H_A+H_B+ 1 \text{ equiv. I}$  $H_A+H_B+ 0.8 \text{ equiv. I}$  $H_A+H_B+ 0.6 \text{ equiv. I}$ 

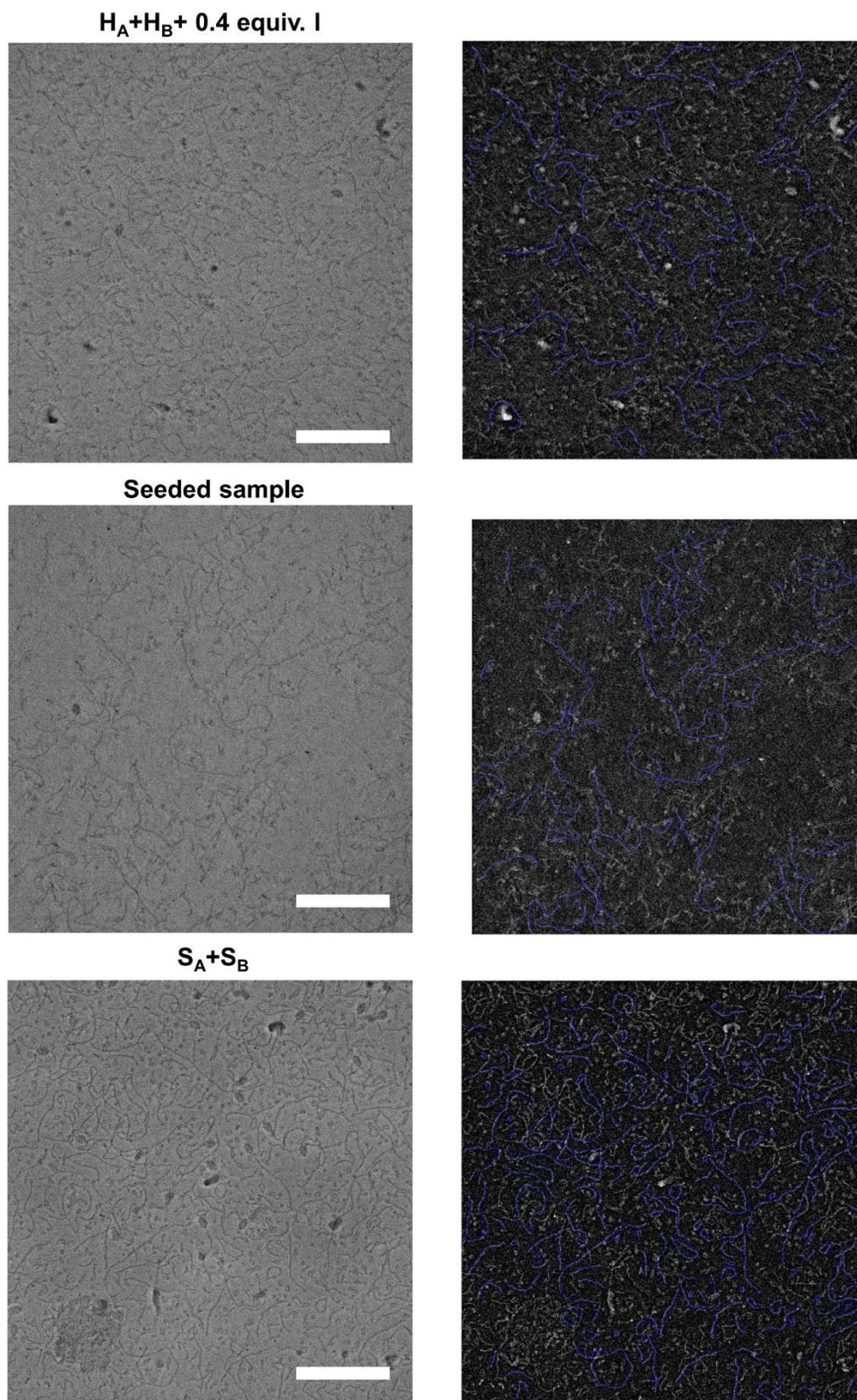


Figure 2.9 - Representative micrographs and analysis for samples analyzed by TEM. Left: original image (scale bars = 100 nm), Right: analyzed image with fibers traced in blue.

CHAPTER THREE

DNA-Encoded Protein Janus Nanoparticles

Material in this chapter is based upon published work:

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

3.1 The diversity and multiplicity of protein interactions

Ribosomes are the macromolecular machines that synthesize proteins within cells. Within eukaryotes, these complex structures are assembled from 78 ribosomal proteins and 4 ribosomal RNAs and represent one of the most intricate examples of a naturally occurring protein-based architecture.¹⁵² Elucidating how each component fits together and how the assembly process of ribosomes is regulated remains an active area of research in the field of structural biology.^{153, 154,}¹⁵⁵ However, several key advances, such as the structural elucidation of ribosomal subunits using cryo-EM, clearly reveal that ribosomal proteins and RNA associate through a web of interactions.¹⁵⁶ Further studies, that interrogate the interactome of specific ribosomal proteins, reveal that these proteins typically interact, simultaneously, with multiple binding partners, including other structural ribosomal proteins, ribosomal RNA, and transiently associating protein species.¹⁵⁷

Structural insight into complex protein architectures, such as the ribosome, indicates that many proteins have evolved surfaces that display multiple orthogonal, specific binding sites. These building blocks are therefore encoded with many regions of orthogonal and directional interactions that drive the assembly of sophisticated, hierarchical structures containing multiple components. This is a powerful design principle as it allows distinct functionalities and properties from different building blocks to be brought together in an ensemble to perform functions collectively.^{158, 159} Developing methods to achieve this control, synthetically, at the nanoscale would drastically enhance the bottom-up fabrication of multi-component, multi-functional materials, and systems.

3.2 Encoding multiple, orthogonal interactions at the nanoscale

Synthetic structures fabricated from nanoscale components, such as proteins and inorganic nanoparticles still face obstacles to realizing well-defined, multicomponent, hierarchical structures that exhibit the structural and functional complexity of materials and systems orchestrated in Nature.^{160, 161, 162} In particle-based assembly, the challenge of programming symmetries, directionality, periodicities, and composition of superstructures is a synthetic one that requires monodispersed building blocks with designed shape anisotropy or anisotropic interaction patterns (“patches”), akin to proteins.¹⁶³ While there has been a major effort to design and synthesize monodispersed patchy particles with defined shape and interaction patterns, even for particles measuring tens of nanometers, using techniques like masking and heterogeneous nucleation, methods are still very limited and not generalizable.^{164, 165, 166} Inorganic particles usually have chemically homogenous surfaces, which makes it difficult to chemically differentiate and encode multiple, anisotropic interaction patterns on their surfaces. For example, the association of PAEs is typically mediated by an isotropic layer of radially oriented DNA ligands with identical sequences that reflects the chemical homogeneity of the inorganic particle PAE core’s surface.

Although an impressive design space has been uncovered with building blocks isotropically functionalized with single DNA sequences, it is clear that anisotropically functionalized building blocks with multiple, orthogonal DNA ligand domains will dramatically increase the scope of possible structures outcomes.^{167, 168} Indeed, it has recently been reported that one can chemically anchor one type of DNA ligand at specific sites on a protein’s surface,¹⁶⁹ including opposing ends,^{97, 98} enabling a novel set of building blocks for both one- and three-dimensional assembly. A major advance would be to develop methods to place different oligonucleotides at specific locations on a nanoparticle’s surface. Realizing these types of building

blocks requires the ability to spatially control the positions of multiple different oligonucleotides, a formidable challenge for inorganic nanoparticles that present chemically isotropic surfaces.^{170, 171, 172, 173} In the case of proteins, however, orthogonal chemistries of surface amino acid residues provide spatial control over surface modifications that is difficult to achieve with other classes of nanoscale building blocks. Therefore, in principle, we can harness the mutability of a protein's surface chemistry to solve a significant unmet challenge: the realization of a DNA-encoded Janus nanoparticle. The assembly properties of these particles can then be investigated to synthesize complex, multicomponent materials that are dictated by directional, asymmetric interactions.

3.3 DNA-encoded protein Janus nanoparticles

3.3.1 Designing DNA-encoded protein Janus nanoparticles

One approach to synthesizing this new class of building blocks could involve the functionalization of proteins that natively possess a high density of asymmetrically distributed orthogonal surface amino acids, such that chemically distinct “faces” could be defined. Albeit common for some classes of proteins, such as transmembrane proteins, to exhibit stark asymmetry with respect to surface residue distribution,^{174, 175} to our knowledge, a candidate protein with a Janus-type amino acid distribution does not exist. Alternatively, a protein's surface chemistry can be manipulated through site-directed mutagenesis, and this technique may be employed to introduce a single chemically orthogonal amino acid at the protein's surface. Herein, we report a method for synthesizing a protein-based Janus nanoparticle functionalized with nucleic acid domains by tethering two, separately functionalized, proteins together with a DNA bond (Figure 3.1).

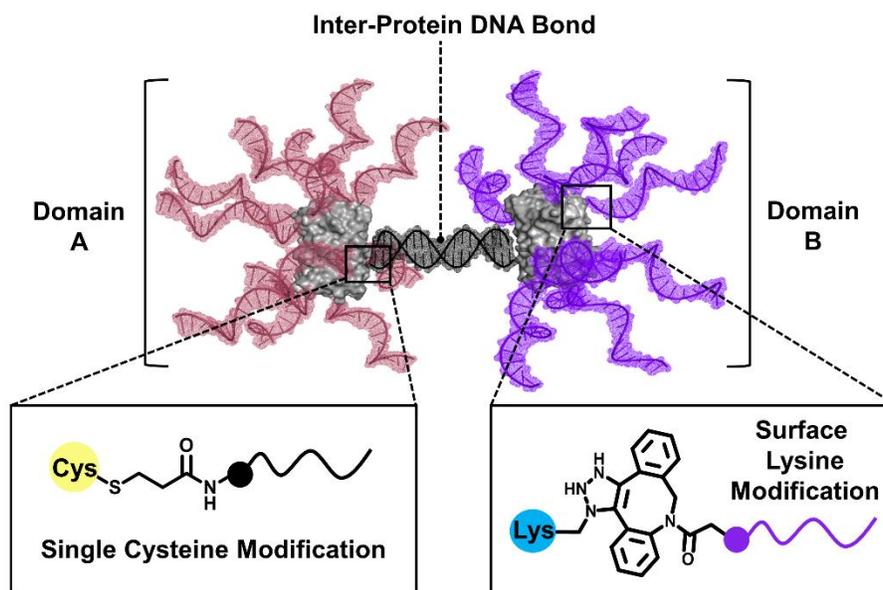


Figure 3.1 - Schematic of a Protein-Based Janus Particle. Two mutated green fluorescent proteins, functionalized with dense shells of nucleic acids of different sequences, tethered together through an interprotein DNA bond.

The design of our DNA-encoded protein Janus nanoparticle circumvents the synthetic challenges posed by inorganic counterparts while minimizing protein mutagenesis and introduces an additional design parameter to the Janus particle: an interprotein DNA bond. A general hypothesis guiding this work is that the co-assembly of a protein Janus particle with two sets of spherical nanoparticles, each functionalized with face-complementary orthogonal nucleic acid sequences, would form architectures that arise from selective interactions between each face of the Janus particle and each set of spherical NPs. We expected that this novel assembly behavior would enable the synthesis of multicomponent crystalline lattices where the unit cells consist of various types (sizes, compositions) of nanoparticles at defined positions dictated by the Janus particle. Through the elucidation of colloidal crystal architectures using small-angle X-ray scattering (SAXS) and transmission electron microscopy (TEM) techniques, we aimed to reveal a new class

of materials with uniquely defined particle positions reliant upon the interactions dictated by the protein Janus particle.

As proof-of-concept, a protein-based DNA-functionalized Janus particle was designed by exploiting the orthogonal reactivity of surface lysine (amine) and cysteine (thiol) residues on an enhanced green fluorescent protein mutant (mGFP) (Figure 3.1). GFP was chosen as a model construct because it is a monomeric protein that contains an even distribution of surface lysine (Lys) residues, which is optimal for dense functionalization of its surface with DNA. In addition, it is stable with respect to a host of surface mutations,^{176, 177} allowing one to place a single cysteine (Cys) residue at a discrete location on the protein's surface (Materials and Methods section 3.6.1). This protein building block was then modified in two different ways to create the “domains” of the Janus construct. One domain of the Janus particle was generated by functionalization with a single strand of DNA at the Cys site of the mutated protein, and subsequent functionalization of a second, orthogonal DNA sequence at the Lys residues. The second face of protein was modified in a similar manner but with different DNA strands, such that the DNA placed at the Cys site was complementary to the Cys-functionalized DNA of the first face, and a fourth, orthogonal sequence was functionalized on the Lys sites. When combined below the melting temperature (T_m) of the Cys DNA duplex, these two protein building blocks form a Janus particle through DNA hybridization, consisting of two orthogonally functionalized protein “domains”.

3.3.2 Synthesis and characterization of DNA-encoded protein Janus nanoparticles

In a typical experiment, complementary pyridyl disulfide-terminated oligonucleotides were synthesized and then reacted with the surface cysteine of mGFP to form monofunctionalized mGFP-DNA conjugates (Figure 3.2a, Materials and Methods section 3.6.3). The mGFP-DNA

conjugates were purified from unreacted protein by anion exchange chromatography (Materials and Methods Figure 3.8), and the purity and mass of monofunctionalized mGFP-DNA conjugates were confirmed using SDS-PAGE analysis (Figure 3.2b).

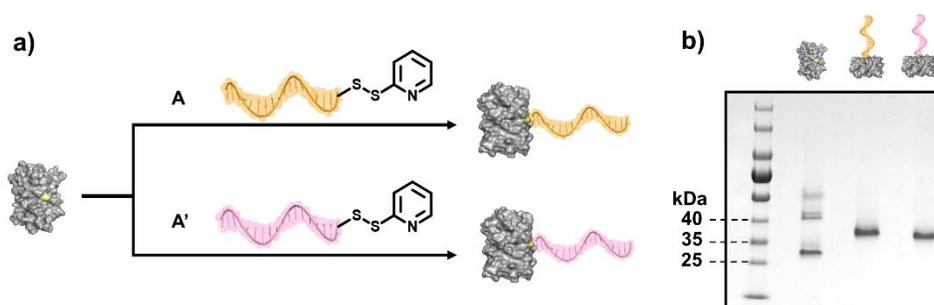


Figure 3.2 – DNA conjugation to Cys residue of mGFP. (a) Complementary DNA strands were conjugated independently to an EGFP mutant, mGFP, through pyridyl disulfide exchange with a single surface cysteine residue. (b) SDS PAGE gel of native mGFP and purified mGFP-DNA conjugates demonstrates successful mono functionalization.

Surface-accessible lysine residues were then converted to azide functionality in high yield (74% with respect to 19 surface-accessible lysine residues, which was confirmed by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Materials and Methods Figure 3.9)). The azide-modified mGFP-DNA conjugates were then separately functionalized with two different oligonucleotides via a strain-promoted cycloaddition (Cu-free “click-chemistry”) between the surface-bound azides and dibenzocyclooctyne (DBCO) moieties at the 5’termini of synthetic oligonucleotides (Figure 3.3a).¹⁷⁸

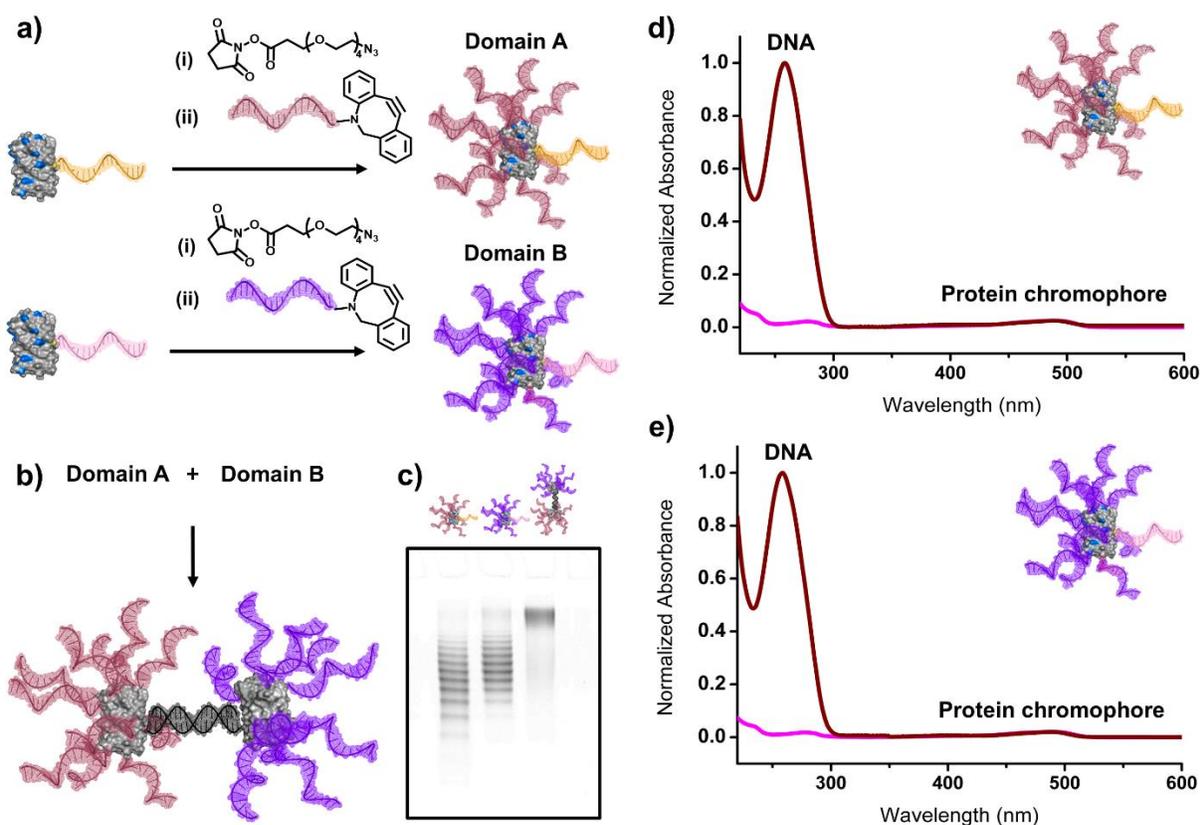


Figure 3.3 – Domain synthesis and formation of Janus particles. (a) Surface-accessible lysines were modified with azides using the NHS ester functionality of the linker group (i), after which the covalently attached azides were conjugated to two distinct 5'-DBCO-modified DNA strands via a copper-free "click chemistry" reaction (ii). (b) Hybridization of the two complementary sequences enabled formation of a protein Janus particle. (c) Native PAGE gel of the Janus particle and each separately functionalized domain shows formation of a protein dimer. The gel also reveals a distribution in number of DNA strands on each domain. (d and e) Determination of the number of DNA strands (DBCO-modified strand C (red) and D (blue)) per protein from concentration ratios measured using UV-vis spectroscopy. Protein and DNA concentrations were calculated from the absorbance at 488 and 260 nm, respectively. All traces of a single plot are normalized to the same protein concentration.

Densely functionalized products were separated from the reaction mixture using size exclusion chromatography (Materials and Methods section 3.6.3). UV-vis quantification of the DNA to protein ratio, by monitoring at 260 and 488 nm, respectively, showed an average of 14 DNA strands per protein (Figure 3.3 d and e). This synthetic route yielded two distinctly functionalized protein nucleic acid "domains", each possessing a dense shell of oligonucleotides in addition to a single orthogonal nucleic acid strand. When both domains were combined, Janus-

type particles formed through the hybridization of the single complementary strands on each domain (Figure 3.3b). Native PAGE gels confirmed the hybridization of the two protein “domains”, as evidenced by a large decrease in electrophoretic mobility as compared to a single protein “domain” (Figure 3.3c). Products from each synthetic step were further characterized using size exclusion chromatography: the elution profiles correlate with increasing hydrodynamic radii (Figure 3.4). Taken together, these data demonstrate the successful synthesis of our target protein-based Janus building block.

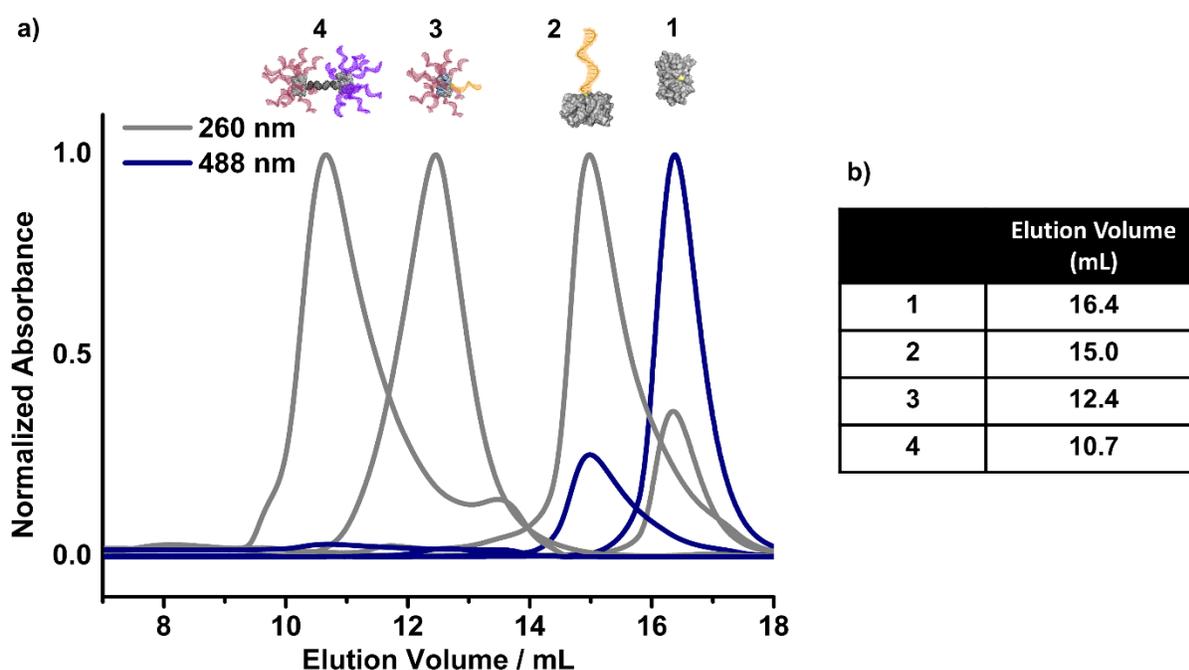


Figure 3.4 – Characterization of protein-DNA conjugates using size exclusion chromatography. (a) SEC traces of the (1) native protein, (2 & 3) protein-DNA conjugates, and (4) protein Janus particle. (b) Table listing specific elution volumes of each species.

3.4 Interrogating assembly behavior in the context of colloidal crystallization

3.4.1 Co-assembly of Janus particles with two sets of 10 nm AuNP

A major advantage of DNA-programmed crystallization is the specificity with which interactions between components can be engineered. Indeed, we have explored this advantage by utilizing the two orthogonal DNA “domains” of the protein Janus particle to direct the crystallization of nanoparticles designed to interact with the protein Janus particle through face-complementarity. DNA sequences were designed and synthesized such that each domain of the protein Janus particle could be duplexed with distinct complementary linker strands containing orthogonal 6-base-pair single strand overhangs (“sticky-ends”) at the 5' termini (Materials and Methods Table 3.3). Two sets of inorganic, spherical nanoparticles were independently functionalized with DNA containing different sequences such that, upon hybridization with their corresponding linker strand, the sticky-ends on each spherical nanoparticle would be complementary to only a single domain of the Janus particle. This design ensured that the building blocks would specifically interact with the Janus particle based on the asymmetry of the DNA functionalization. The assembly of two sets of 10 nm AuNPs, identical except for the DNA sequence with which they were functionalized, with the protein Janus particles was investigated (Figure 3.5). The combination of the two sets of AuNPs with the protein Janus particles, preduplexed with their respective linker strands, induced the precipitation of aggregates, which were then heated beyond their dissociation temperature (Materials and Methods Figure 3.10) and slowly cooled (0.01°C/min) to allow the reorganization of the system, presumably into its thermodynamically favored configuration.

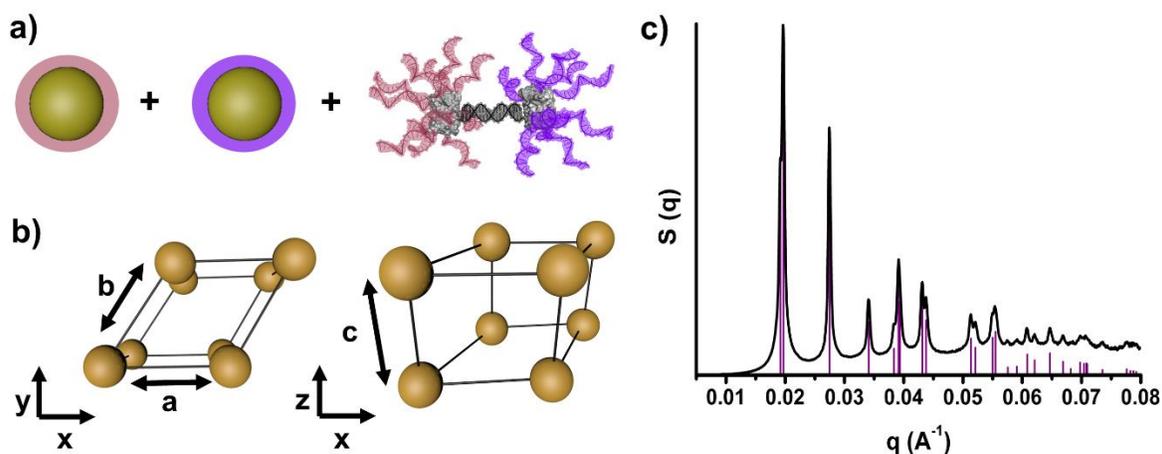


Figure 3.5 - Structural determination of a superlattice assembled from 10 nm AuNPs and protein Janus nanoparticles. (a) Scheme showing two sets of 10 nm AuNPs functionalized independently with different DNA sequences, each complementary to one side of the protein Janus nanoparticles. (b) The arrangement of AuNPs into a hexagonal unit cell is shown along two sets of axes. (b) SAXS data revealed that these components formed ordered three-dimensional structures with hexagonal symmetry, with respect to the AuNPs, upon addition of their respective linker strands and slow cooling.

Synchrotron-based SAXS was utilized to determine the positions of the AuNPs in the resulting superlattice (Figure 3.5c). Because of the vast difference in electron density between the 10 nm AuNPs and the protein Janus particles, scattering from the AuNPs dominates, thereby only providing information regarding the AuNP position. Remarkably, the SAXS patterns of the resulting aggregates revealed a highly crystalline structure, despite 10 distinct DNA sequences being involved in the assembly. From the position and relative intensities of the scattering peaks, we determined that the AuNPs organize into a primitive hexagonal lattice (Figure 3.5b). Given the design rules established for the crystallization of nanoparticles of varying sizes and DNA linker number,⁸⁵ and geometric arguments based on the size of the Janus particles relative to the AuNPs, we assign this structure as an AB₂ lattice with respect to both the AuNPs and the protein Janus particles. Upon consideration of the symmetry of the Janus particle's DNA functionalization, a perpendicular alignment of the Janus particle with respect to hexagonally ordered planes of

orthogonally functionalized AuNPs would maximize the number of interactions between a given domain and its respective complementary set of AuNPs. Based on the symmetry of the Janus particle and the thermodynamic driving force to maximize hybridization events, the possible location and orientation of the protein Janus particle within the unit cell is restricted. The Janus particle must adopt a vertical orientation thus supporting the hypothesis of forming an AB₂ lattice since two Janus particles are required to maximize interactions between nanoparticles occupying (0,0,0) and (0,0,1/2) positions. This structure has been previously observed with comparable AuNPs and proteins with similar relative sizes and DNA linker numbers. However, while this superlattice contains two sets of orthogonally functionalized AuNPs with identical inorganic cores, the SAXS data do not enable us to determine the specific location of each set of particles or their connectivity with the Janus particles.

3.4.2 Determining the identity of particles within the unit cell

A powerful aspect of using Janus particles in DNA-mediated crystallization is the ability to co-assemble with two sets of differently functionalized nanoparticles. The versatile chemistry of DNA also enables us to vary both the composition and the size of the nanoparticles, independent of the DNA ligand identity responsible for dictating particle association. Therefore, we next explored the assembly of systems formed from two sets of inorganic particles of different composition or size with the protein Janus particles. We expected that the assembly of multicomponent systems would lead to an isostructural superlattice, and thus similar peak positions in the SAXS data, but present different peak intensities (Figure 3.6) due to the significantly different electron densities between two particles of different composition or size. Therefore, we prepared lattices where one set of the 10 nm AuNPs was replaced with 10 nm silver

nanoparticles (AgNPs), while maintaining the same DNA design (Figure 3.6a). Remarkably, the SAXS pattern revealed a layered primitive hexagonal lattice structure with a layer of 10 nm AuNPs at the (0,0,0) sites and a layer of 10 nm AgNPs occupying the (0,0,1/2) sites. Simulated SAXS data are in good agreement with the experimentally obtained pattern for a multicomponent layered structure with this composition. Furthermore, isostructural lattices were obtained from the assembly of 5 nm AuNPs and 10 nm AuNPs (Figure 3.6b) with Janus particles, demonstrating the versatility of this system. The multicomponent 5 nm AuNP and 10 nm AuNP superlattice was embedded in silica following literature methods (Materials and Methods, section 3.6.8),¹⁷⁹ and the layered structure was visualized using TEM both in a silica embedded sample (Figure 3.cc) and after sectioning the lattices (Figure 3.6d). These data reveal layered architectures between the two particle sizes, where nanoparticles functionalized with the same DNA occupy lattice sites in the same layer, and they conclusively support the interpretation of the SAXS data. By designing structures containing particles with different electron densities, we were able to precisely determine the position and identity of nanoparticles within the primitive hexagonal crystals formed. Significantly, these data validate the realization of a new class of nanoparticle superlattices, where multiple types of inorganic nanoparticles can be programmed to occupy distinct positions within the same unit cell through co-assembly with protein Janus particles into architectures, which cannot yet be realized through interactions between the two inorganic nanoparticles alone.

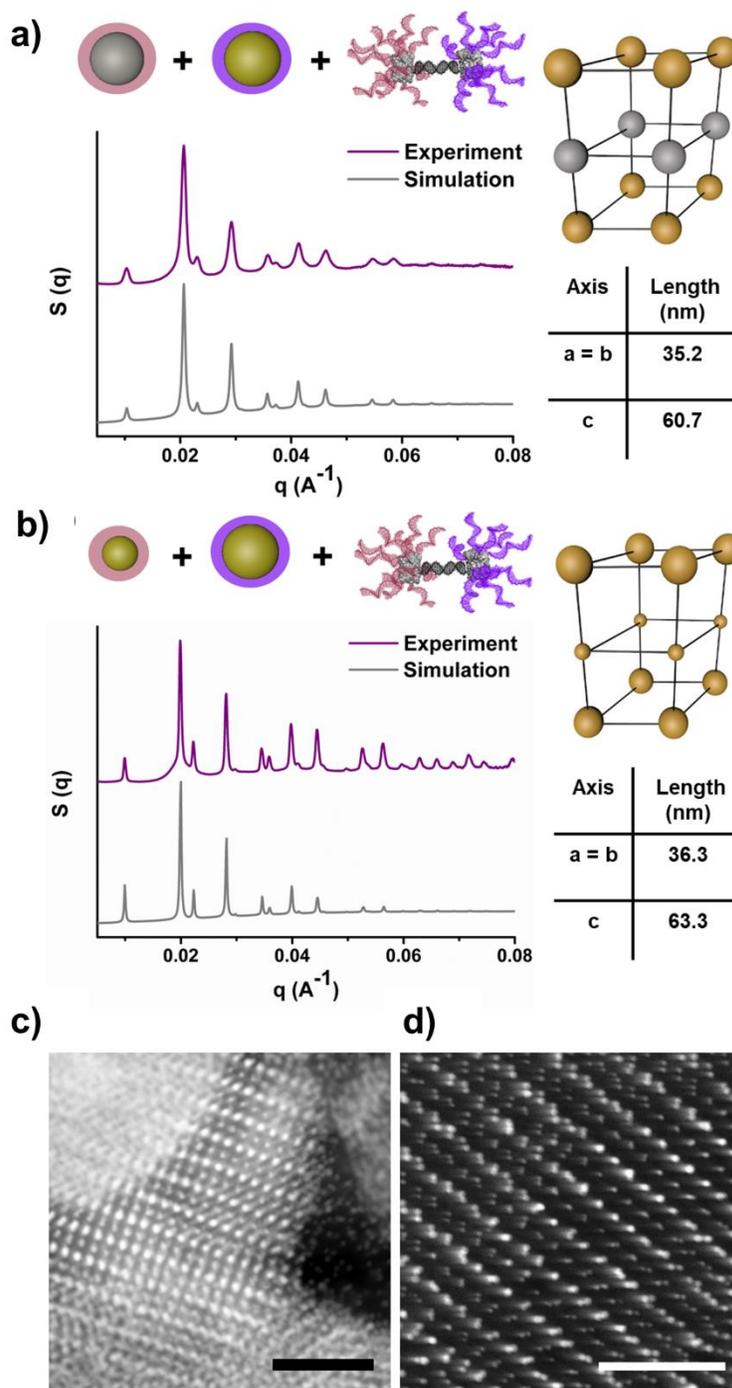


Figure 3.6 - Assembly of multicomponent superlattices revealed layered hexagonal unit cell.

(a, b) Schemes showing the assembly of (a) 10 nm Ag and Au nanoparticles with protein Janus particles and (b) 5 nm and 10 nm Au nanoparticles with protein Janus particles into a layered, hexagonal unit cell where each layer consists of one type of metal nanoparticle. Experimental and simulated SAXS patterns from the assembly of the particles shown above. (c, d) TEM images of silica embedded crystals of (b) reveal layers of 5 and 10 nm AuNPs. (d) Samples of (b) were thinly sectioned (60 nm) to capture a few layers of the crystal lattice. Scale bars are 150 nm.

3.4.3 Confirming the orientation of Janus particles within the lattice

We can infer from the layered structures that the Janus particle is most likely orientated vertically in the unit cell (Figure 3.7b). This orientation maximizes the number of DNA interactions made by each DNA “domain” of the Janus particle with the corresponding complementary particles. The interprotein DNA bond, unique to the Janus particle design, is highly programmable in the sense that the distance between each domain of the Janus particle can be programmed through the number of base-pairs incorporated into the sequence. We postulated, on the basis of the proposed orientation of the Janus particle within the unit cell, that the elongation of the interprotein DNA “bond” would lead to a Janus particle with a greater aspect ratio and that this would be reflected in the selective elongation of the c-axis obtained from SAXS data. To test this hypothesis, a second protein Janus particle was synthesized (Materials and Methods, section 3.6.3) with a 24 base-pair interprotein DNA “bond” (long Janus) instead of the 18 base-pairs (short Janus) used in the original design (Figure 3.7a). Isostructural lattices were obtained with both short and long Janus particles, irrespective of the types (sizes, compositions) of NPs incorporated. Triplicate measurements of the lattice parameters of superlattices assembled from either short or long Janus particles obtained from the SAXS patterns of each structure revealed that lattices that contained long Janus particles, indeed, had unit cells with a larger c parameter (Figure 3.7c and Materials and Methods Table 3.4). Moreover, when we compared the difference in length of the a and c lattice parameters between structures formed from short and long Janus particles, we observed that expansion along the c-axis was significantly greater (up to 2.31 nm) than the expansion along the a-axis (0.25 nm) (Figure 3.7c).

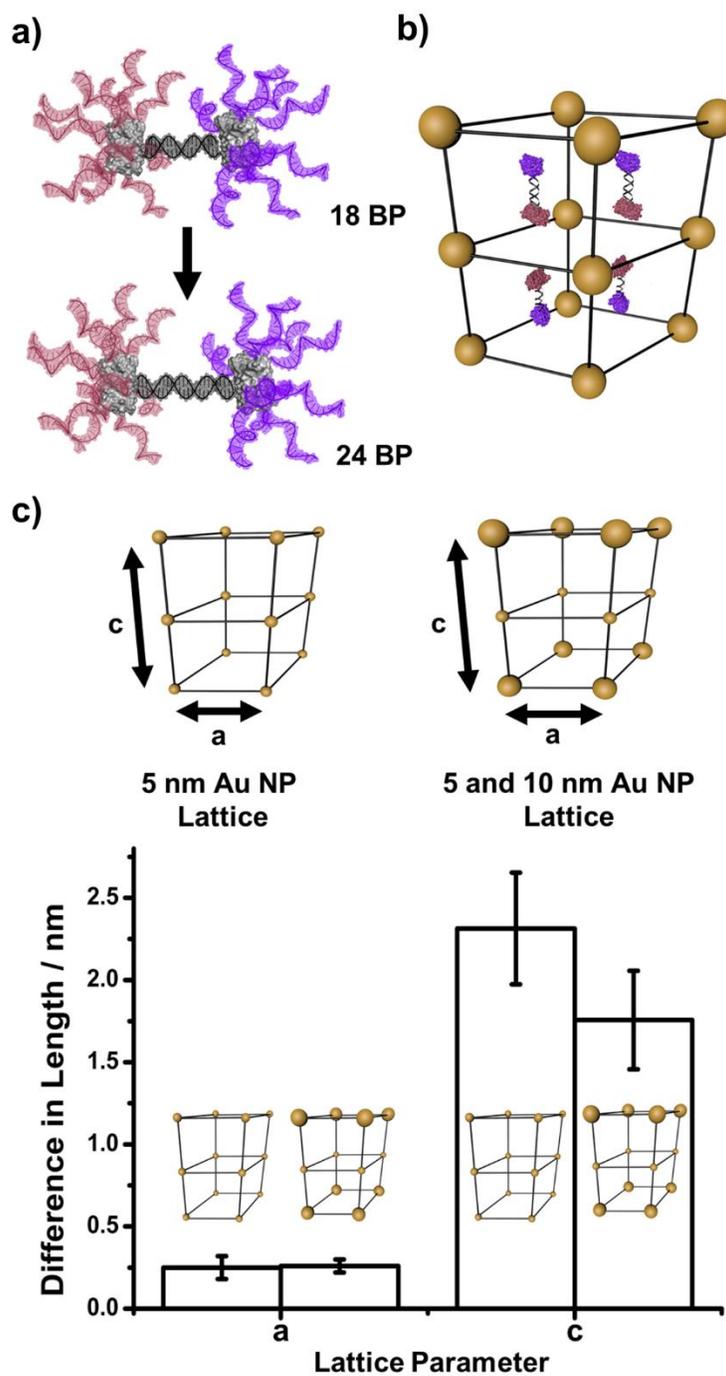


Figure 3.7 – A comparison of lattice parameter length between lattices synthesized from Janus particles with different interprotein bonds. (a) Two Janus particles prepared with interprotein bonds of 18 and 24 base-pairs (BP). (b) Hypothesized unit cell containing 4 Janus particles analogous to an AB₂ structure. (c) Depiction of the unit cells of superlattices assembled from both long and short protein Janus particles, as well as a graphical representation of the difference in lattice parameter length between lattices formed with 18 and 24 BP interprotein bonds. Lattice expansion data were measured using SAXS.

These results support the hypothesis that the Janus particle is orientated with the interprotein DNA bond along the c-axis of the unit cell. These experiments also highlight an unprecedented degree of structural control, enabled by the unique design of these protein Janus nanoparticles, which allows anisotropic manipulation of lattice dimensions. On the basis of the structural evidence reported thus far, and established DNA mediated colloidal crystal design rules,¹⁸⁰ we propose that four protein Janus particles occupy the unit cell (Figure 3.7b) and that the resulting structure is analogous to an AB₂ lattice.

3.5 Conclusions

Undeniably, the encoding of a particle's surface with discrete regions of orthogonal DNA is critical in determining the particle's behavior and properties. These findings are a convincing demonstration of how protein-based Janus constructs, which result from functionalization chemistry, can be used to engineer previously inaccessible crystalline nano-particle architectures, and, more widely, how the rational design of DNA-encoded nanoparticles can be used to program the assembly of novel colloidal crystal architectures. In addition, our findings represent the first example of structural tunability of colloidal crystal architectures through elongation of the unit cell selectively in one direction: architectural control that may be translated to the fine-tuning of the colloidal crystal's material properties, or the realization of NP crystals that can be anisotropically and reversibly actuated. This new class of DNA-encoded protein nanoparticles, introduced herein, will form the basis of a broad scope of future investigations including probing the potentially powerful biological properties of these constructs,^{181, 182, 183} and pioneering the next generation of DNA-mediated colloidal crystal engineering by realizing architectures with previously inaccessible complexity that may possess unique chemical and physical properties.

3.6 Materials and Methods

3.6.1 Protein mutation, expression, and purification

Site-specific mutagenesis of the EGFP-pBAD vector (EGFP-pBAD was a gift from Michael Davidson, Addgene plasmid #54762) was achieved using a QuickChange II Site-Directed Mutagenesis Kit (Agilent). This technique employs PCR-thermocycling reaction using mutagenic primers, followed by removal of the original plasmid template by *DpnI* digestion. Mutagenic primers were designed to introduce a single codon mutation into the gene. To facilitate annealing, primers contained at least 40% GC, had melting temperatures (T_m) above 60 °C, and exhibited one or more Gs and Cs at the 5' or 3' ends (Table 3.1). Two consecutive mutations were performed to convert the surface serine (S148) to a cysteine (C148) and also mutate an unwanted, partially surface accessible cysteine (C49) to a leucine (L49). Successful mutations were confirmed using traditional Sanger Sequencing (Table 3.2).

Primer	Sequence (5' to 3')	GC content	T_m / °C
S148 → C148 (A)	TGA TAT AGA CGT TGT GGC AGT TGT AGT TGT ACT CCA G	43.2 %	62.5
S148 → C148 (B)	CTG GAG TAC AAG TAC AAC TGC CAC AAC GTC TAT ATC A	43.2 %	62.5

Table 3.1 – Mutagenetic primer design of S148 to C148 mutation.

	Protein Sequence
Native	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGE GEGDATYGKLTLLKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNYNSHNVYIMADKQKNGIK VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSTQSALS KDPNEKRDHMLLEFVTAA GITLGMDELYK
C49 to L49	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGE GEGDATYGKLTLLKFI <u>L</u> TTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNYNSHNVYIMADKQKNGIK VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSTQSALS KDPNEKRDHMLLEFVTAA GITLGMDELYK
S148 to C148	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGE GEGDATYGKLTLLKFI <u>L</u> TTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF KEDGNILGHKLEYNYN <u>C</u> HNVYIMADKQKNGIK VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSTQSALS KDPNEKRDHMLLEFVTAA GITLGMDELYK

Table 3.2 – Sanger sequencing data of mutated EGFP plasmid compared to native plasmid.

The mutated plasmid containing the gene for the mutated EGFP was transformed into One Shot® BL21(DE3) Chemically Competent E. coli (Thermo Fisher) by heat shock, and cells were

grown overnight on LB Agar plates with 100 $\mu\text{g}/\text{mL}$ ampicillin. Single colonies were picked, and 7 mL cultures were grown for 6 h at 37 °C in LB broth with 100 $\mu\text{g}/\text{mL}$ Ampicillin. These cultures were added to 1 L of Terrific Broth (Thermo Fisher) with 1 % glycerol and 100 $\mu\text{g}/\text{mL}$ ampicillin, and cells were grown at 37 °C to an optical density of 0.6, then induced with 0.02 wt% arabinose overnight at 17 °C. Cells were spun down (6000 g, 30 min) and resuspended in 100 mL of 1x PBS, then lysed using a high-pressure homogenizer. The cell lysate was clarified by centrifugation at 30,000 g for 30 min and loaded onto a Bio-Scale™ Mini Profinity™ IMAC Cartridge (Bio-Rad). The column was washed with 100 mL of resuspension buffer, then eluted in the same buffer with 250 mM imidazole. The eluted fraction was immediately exchanged 5 times into 1x PBS using 30 kDa molecular weight cut-off filters (Millipore). After concentrating to 3 mL, the protein was further purified using a size exclusion column (SEC650, Bio-Rad), and fractions with molecular weights corresponding to mutant EGFP (mGFP) were stored at 4 °C for up to two months.

3.6.2 Oligonucleotide design and synthesis

Six orthogonal oligonucleotide sequences were designed specifically for the protein Janus particles. Amine terminated 18 and 24 base-pair sequences, as well as their respective complementary strands, were designed to form interprotein DNA bonds of different lengths. Two dibenzocyclooctyne (DBCO) functionalized strands, containing orthogonal sequences, were used to decorate each domain of the Janus particle. A further two orthogonal sequences were designed for the functionalization of the inorganic nanoparticles: of which, both thiol and dithiol-modified versions were prepared. Four linker strands were required: one complementary strand for each

domain of the protein Janus particle, as well as the corresponding complimentary strands for orthogonally functionalized inorganic nanoparticles.

All oligonucleotides were synthesized on solid supports using reagents obtained from Glen Research and standard protocols (Table 3.3). Products were cleaved from the solid support using 30% NH₃ (aq) for 16 hours at room temperature and purified using reverse-phase HPLC with a gradient of 0 to 75 % acetonitrile in triethylammonium acetate buffer over 45 min. The masses of the oligonucleotides were confirmed using matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) using 3-hydroxypicolinic acid as a matrix. All masses were within 20 Da of the expected calculated mass. Oligonucleotide concentration was calculated based on the extinction at 260 nm, using the extinction coefficients given in Table S3 that were calculated using the IDT Oligo Analyzer tool.

Name	Sequence (5' to 3')	ϵ (M ⁻¹ cm ⁻¹)	MW (Da)
DNA-Encoded Janus Particles			
Interprotein Bond 18 BP	Amino C6 -AGT TAG GAC TTA CGC TAC	176900	5680
Interprotein Bond 18 BP	Amino C6 -GTA GCG TAA GTC CTA ACT	177100	5682
Interprotein Bond 24 BP	Amino C6 -AGT TAG GAC TTA CGC TAC TTT TTT	225200	7503
Interprotein Bond 24 BP	Amino C6 -AAA AAA GTA GCG TAA GTC CTA ACT	250600	7557
DBCO-modified C	DBCO-Sp₂ -GAA TAT TGA CGT AAA TCT	185100	6780
DBCO-modified D	DBCO-Sp₂ -CAA TTA AAG TCA TAC AGA	191800	6766
Inorganic Nanoparticles			

Thiol-modified A	C6 SS-Sp₂ -AAC GAC TCA TAT TAA CAA	188300	6477
Thiol-modified B	C6 SS-Sp₂ -AAG ACG AAT ATT TAA GAA	200500	6578
Cyclic disulfide modified A	DS-Sp₂ -AAC GAC TCA TAT TAA CAA	188300	6561
Cyclic disulfide modified B	DS-Sp₂ -AAG ACG AAT ATT TAA GAA	200500	6664
Linker Strands			
Linker A	AAGGAA- Sp -TTG TTA ATA TGA GTC GTT	248200	7785
Linker B	GAAAGA- Sp -TTC TTA AAT ATT CGT CTT	237700	7682
Linker C	TTCCTT- Sp -AGA TTT ACG TCA ATA TTC	227000	7616
Linker D	TCTTTC- Sp -TCT GTA TGA CTT TAA TTG	218000	7619

Sp = hexaethylene glycol spacer unit

C6 SS = 5' C6 thiol modifier

DS = 5' Dithiol serinol modifier

Table 3.3 – Oligonucleotide sequences designed for the assembly of inorganic nanoparticles with DNA-encoded protein Janus nanoparticles. Extinction coefficients calculated using IDT OligoAnalyser Tool. Molecular weight (MW) measured using MALDI spectroscopy.

3.6.3 Synthesis and characterization of DNA-encoded Janus particle

DNA-Encoded Janus particles were prepared in a stepwise fashion. Each domain (A&B) of the Janus particle was prepared separately, in parallel, using orthogonal oligonucleotide sequences. Janus particles containing either an 18 or 24 base-pair DNA “bond” were synthesized using identical procedures.

Amine terminated DNA (300 nmoles) was reacted with 50 equivalents of SPDP (Thermo Fischer Scientific) crosslinker in 50 % DMF, 1x PBS + 1 mM EDTA for 1 h at room temperature.

Excess SPDP was removed from the DNA by two rounds of size exclusion using NAP10 and NAP25 columns (GE Healthcare) equilibrated with PBS (pH 7.4), consecutively. Ten equivalents of the resulting pyridyl disulfide terminated DNA was added to 1.5 mL of 20 μ M protein solution, and the reaction allowed to proceed for 16 h at room temperature. Excess DNA was removed from the reaction by 5 rounds of centrifugation in 30 kDa molecular weight cut-off filters (Millipore). Anion exchange chromatography was performed on the mGFP-DNA conjugate, using Macro-Prep® DEAE Resin, to remove any unreacted protein. SDS-PAGE analysis (Figure 3.8) confirmed elution of unreacted protein and pure mGFP-DNA conjugate in buffers containing 0.25 M and 0.50 M NaCl, respectively.

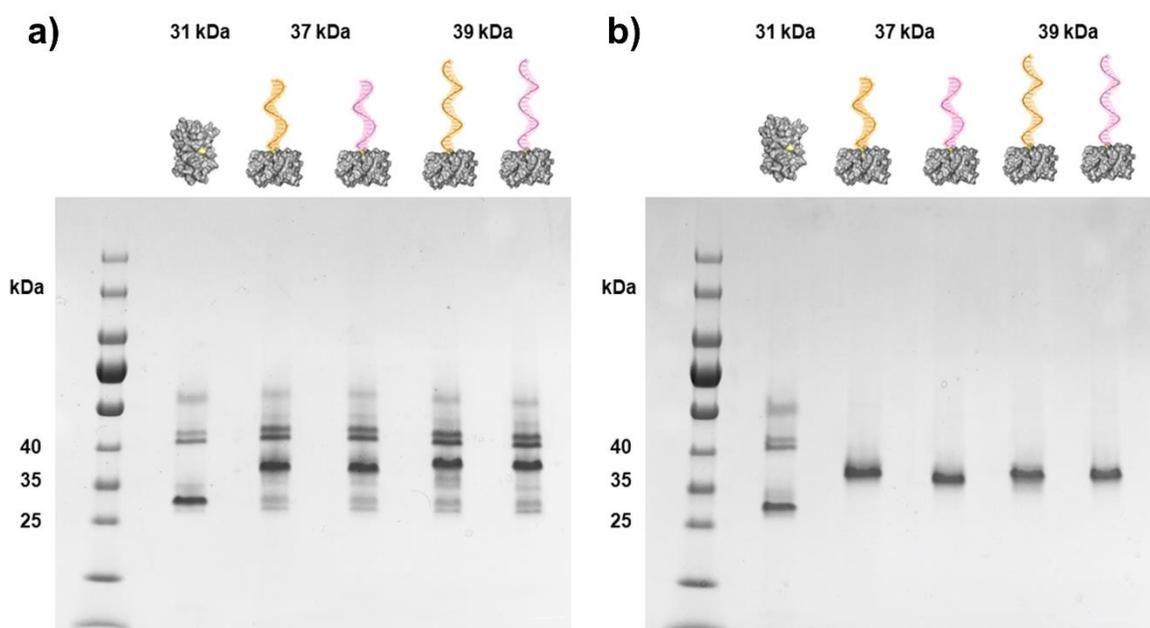


Figure 3.8 – SDS PAGE analysis of DNA conjugation reactions with surface cysteine. (a) Unpurified and (b) purified reactions.

A 1.5 mL sample of 10 μ M mGFP-DNA was reacted with 1000 equivalents of NHS-PEG₄-N₃ (Thermo Scientific). The reaction between surface accessible lysine residues and NHS-PEG₄-N₃ was allowed to proceed at 25 °C for 3 h while shaking at 800 rpm on a Benchmark Multi-therm

shaker. Azide labelled mGFP-DNA was purified by size exclusion chromatography using NAP25 columns (GE Healthcare), equilibrated with PBS (pH 7.4), and the extent of reaction was quantified using MALDI-TOF (Figure 3.9a). Subsequently, the pendent azides were then then reacted with 150 equivalents of 5' DBCO functionalized DNA, for 48 h at room temperature, at a protein concentration of 4.5 μ M. Unreacted DNA was then removed via 10 rounds of centrifugation in 50 kDa molecular weight cut-off filters (Millipore). Janus domains A and B were further purified using size exclusion chromatography (SEC650, Bio-Rad) (main text, Figure 3.4), and the fractions corresponding to the highest MW sample was collected. The DNA:Protein ratio of Janus domains was calculated, using UV-vis absorbance spectroscopy (Varian Cary 5000) (main text, Figure 3.3e and d), to be 14 strands on average per protein. Native PAGE analysis of each “domain” (main text, Figure 3.3c) revealed several discreet bands, representative of a distribution in number of DNA strands per protein.

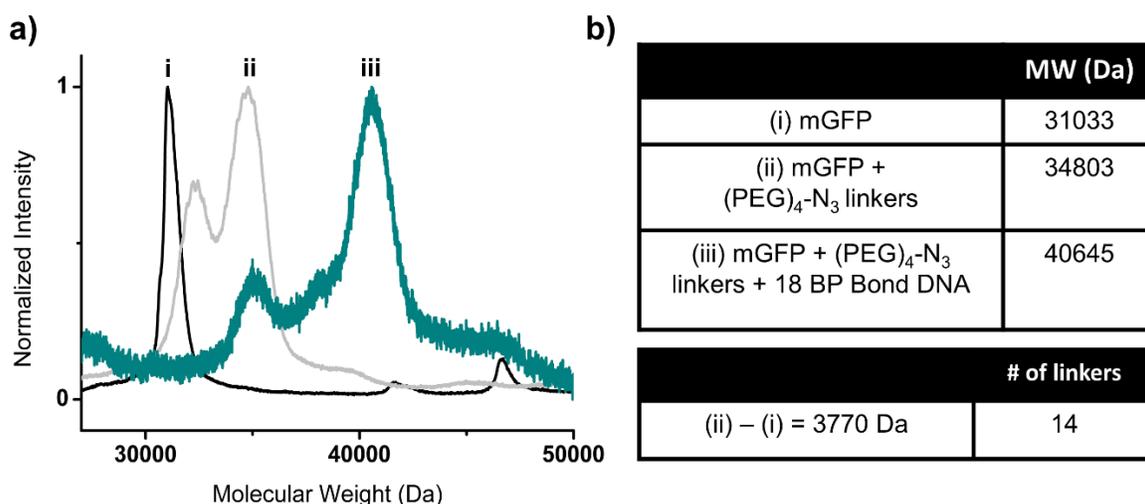


Figure 3.9 – MALDI mass characterization of mGFP and intermediates. (a) Native mGFP in (i) sinapinic acid (SA) matrix, (ii) mGFP after functionalization of surface lysine residues with (PEG)₄-N₃ linkers in SA matrix, and (iii) mGFP + linkers and surface cysteine functionalization with DNA in 3-Hydroxypicolinic acid matrix. (b) Table shows masses of (i-iii) obtained from MALDI. Number of (PEG)₄-N₃ linkers calculated using molecular weight of linker: 274 Da.

3.6.4 DNA functionalization of Au & Ag nanoparticles

Citrate-capped AuNPs with 10 and 5 nm diameters were obtained from Ted Pella and functionalized with 5' thiol terminated DNA (Table 3.3). Firstly, disulfide (C6 SS) DNA was reduced in 100 mM dithiothreitol (DTT) (Invotrogen) for 1 h at room temperature, after which, reduced DNA was purified using a NAP25 size exclusion column (GE Healthcare). To two batches of each NP size, ~5 nmols of the two different 5'-thiolated oligonucleotide were added per mL of AuNPs and the resulting solution was incubated for 4 h at room temperature. Aliquots of 5 M NaCl were added to the solution in 0.1 M increments over the course of 3 h to reach a final concentration of 0.5 M NaCl. This solution was then allowed to incubate for 48 h, at room temperature, to maximize DNA loading on the surface of the AuNPs. The DNA-functionalized particles were purified by five rounds of centrifugation at $21,130 \times g$, in 100 kDa molecular weight cut-off filters (Millipore), followed by resuspension of the resulting pellet in 1 mL of 1x PBS. Particle concentrations were determined based on UV-vis absorbance spectra (Varian Cary 5000) using molar extinction coefficients provided by Ted Pella; www.tedpella.com/gold_html/gold-tec.htm.

Citrate-capped AgNPs with 10 nm diameters were obtained from nanoComposix and functionalized with 5' cyclic disulfide terminated DNA (Table 3.3). To two batches of each NP size, ~5 nmols of the two different 5' cyclic disulfide modified oligonucleotide were added per mL of AgNPs and the resulting solution was incubated for 24 h at 4 °C in the dark. Aliquots of 5 M NaCl were added to the solution in 0.1 M increments over the course of 3 h to reach a final concentration of 0.5 M NaCl. This solution was sonicated for 10 min after each aliquot of 5M NaCl. The final solution was incubated for 48 h, at 4 °C in the dark, to maximize DNA loading on the surface of the AgNPs. The DNA-functionalized particles were purified by five rounds of

centrifugation at $21,130 \times g$, in 100 kDa molecular weight cut-off filters (Millipore), followed by resuspension of the resulting pellet in 1 mL of 1x PBS. Particle concentrations were determined based on UV-vis absorbance spectra (Varian Cary 5000) using reported molar extinction coefficients.¹⁸⁴

3.6.5 Crystallization and characterization of DNA-encoded protein Janus nanoparticle with Ag and Au NPs

Samples for SAXS and electron microscopy (EM) analysis were prepared by the co-assembly of two orthogonally functionalized inorganic nanoparticles with protein Janus nanoparticles. 50 equivalents of each Janus particle linker strand (Table 3.3) were added to a 400 nM solution of Janus particles in 1xPBS and 0.5M NaCl. 200 equivalents of both inorganic NP linkers (Table 3.3) were added to a solution containing 100 nM concentration of each orthogonally functionalized inorganic NP (A and B) in 1xPBS and 0.5M NaCl. The formation of aggregates was observed upon the combination of 50 μ L of each solution. Samples were heated to a few degrees above their melting temperature (Figure 3.10) and cooled at a rate of 0.01 $^{\circ}$ C/min to 20 $^{\circ}$ C using a ProFlexTM PCR system (Applied Biosystems). Resultant crystals were characterized by synchrotron SAXS experiments conducted at the Advanced Photon Source at Argonne National Laboratory.

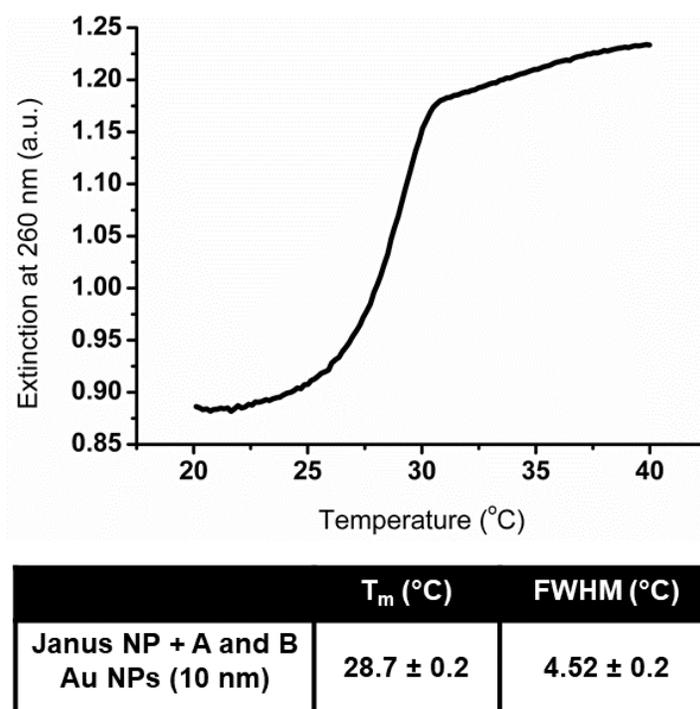


Figure 3.10 – Melting curve of aggregates and calculated T_m . Aggregates formed from orthogonally functionalized 10 nm Au NPs and DNA encoded Janus particle (18 BP bond).

Melt experiments were conducted using a temperature-controlled UV-vis spectrometer (Cary5000). Solutions containing 5 nM particle and 10 nM Janus particle concentration, with 50 equivalents of both protein linkers (linkers C and D) and 200 equivalents of both AuNP linkers (linkers A and B), in 1xPBS + 0.5 M NaCl, were prepared. In order to ensure that aggregates did not form in a kinetically trapped state, samples were first heated to 40 °C and cooled to room temperature at a rate of 0.1 °C/min. Subsequently, the temperature was increased, at a rate of 0.1 °C/min, and an extinction measurement recorded at 0.1 °C intervals at 260 nm to yield the above melting curves. The melting temperature was determined by taking the maximum of the first derivative of the curve, and the full width half max (FWHM) of the first derivative is reported. Triplicate measurements were recorded to ensure reproducibility of melting character.

3.6.6 Small angle X-ray scattering analysis

SAXS characterization was carried out at the DuPont-Northwestern-Dow Collaborative Access Team (DND-CAT) beamline of Argonne National Laboratory's Advanced Photon Source (APS). X-rays of wavelength 1.24 Å (10 keV) were used, and the system was calibrated using silver behenate as a standard. Two sets of slits were used to define and collimate the X-ray beam; parasitic scattering was removed via a pinhole. Slow-cooled samples were transferred into 1.5 mm quartz capillary tubes (Charles Supper) and the scattering was collected with typical exposure times varying between 0.1 and 0.5 s, depending upon the sample. The two-dimensional scattering patterns (collected on a CCD area detector) were azimuthally averaged to yield a 1-dimensional plot of intensity ($I(q)$) as a function of the scattering vector, q , where q is given as:

$$(1) \quad q = 4\pi \sin(\theta)/\lambda$$

Here, λ is the wavelength of incident radiation and θ is half the scattering angle, 2θ . The structure factor, $S(q)$, was obtained by dividing $I(q)$ by the form factor of free 10 nm or 5 nm DNA functionalized inorganic NPs in 1xPBS + 0.5 M NaCl (collected at 100 nM with a 0.5 second exposure time).

SAXS data consist of two basic types of scattering: form factor ($P(q)$), which is scattering inherent to the types of nanoparticles that are in solution, and structure factor ($S(q)$), which arises as a function of the arrangement of these particles relative to one another. The overall scattering of X-rays (denoted as $I(q)$) is a combination of these two types of scattering, and it is these values of $I(q)$ that were measured experimentally:

$$(2) \quad I(q) = kP(q)S(q)$$

The constant, k , is a factor proportional to the square of the number of particles composing a single crystal domain, and the number of crystalline domains in the path of the X-ray beam. Since the electron density of the metal nanoparticles, particularly 10 nm AuNPs, is significantly higher with respect to other organic components in the system, the SAXS data can be used to determine, solely, the position of the nanoparticles. Colloidal crystals assembled from combinations of orthogonally functionalized (Thiol-modified oligos A and B) inorganic nanoparticles and DNA encoded protein Janus nanoparticles all adopted hexagonal lattice symmetry of $P6/mmm$ space group. To account for differences in scattering intensity arising from inorganic nanoparticles of different composition and size, SAXS data was modeled to confirm structural assignment, as presented in Figures 3.6a and b of the main text. Simulated peak positions and intensities were generated using relative electron densities and sizes of Au and Ag NPs.

3.6.7 Calculation and comparison of lattice parameters derived from modulation of DNA “Bond”

Lattice expansion data was obtained from the crystallization of protein Janus particles with both an 18 and 24 base-pair DNA “bond”. Samples were prepared as previously stated and SAXS patterns were recorded in triplicate, from which lattice parameters were extracted (Table 3.4). TEM data were also collected as secondary evidence, but were not used for determination of assembly parameters, since the embedding procedure used to obtain TEM images can result in a reduction of crystal lattice parameters and decrease in crystal ordering.

Lattice Parameters Measured from SAXS (nm)				
1) 10 nm Au NPs (A and B) and Janus particle				
18 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	37.29	37.29	37.33	37.30
c	65.04	65.18	65.18	65.13
24 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	37.93	37.92	37.93	37.93
c	67.30	67.30	67.14	67.25
2) 5 nm Au NPs (A and B) and Janus particle				
18 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	35.85	35.85	35.77	35.82
c	61.02	61.14	61.14	61.10
24 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	36.06	36.06	36.09	36.07
c	62.86	63.24	64.14	63.41
3) 10 nm Au NP (A), 5 nm Au NPs (B) and Janus particle				
18 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	36.39	36.39	36.46	36.41
c	63.51	63.38	63.51	63.47
24 BP “Bond”	Sample 1	Sample 2	Sample 3	Average
a=b	36.67	36.67	36.68	36.67
c	65.05	65.05	65.57	65.22

Table 3.4 – Lattice parameters measured using SAXS for three different colloidal crystal systems, assembled using Janus particles with either an 18 or 24 base-pair DNA “bond”. Slow cooled samples, prepared as stated in section 3.6.5, were measured in triplicate. Each of the three crystal systems assembled using both 18 and 24 BP Janus particles. From these data, lattice parameters were extracted and an average lattice parameter for each system was calculated.

	1	2	3
a=b	0.62 nm \pm 0.02 nm	0.25 \pm 0.07 nm	0.26 \pm 0.04 nm
c	2.11 nm \pm 0.15 nm	2.31 \pm 0.68 nm	1.76 \pm 0.30 nm

Table 3.5 – Calculated lattice parameter differences. Difference measured between the average length of the lattice parameters measured for crystals containing Janus particles with either an 18 or 24 base-pair bond, across each of the three crystal systems.

3.6.8 Electron microscopy imaging of superlattices

Lattices were first embedded in silica, using a previously reported procedure,¹⁷⁹ to preserve their solution structure. To a suspension of slow cooled superlattice samples (100 μ L) in 0.5 mL 1xPBS + 0.5 M NaCl, 2 μ L of the ammonium salt, N-trimethoxysilylpropyl-N,N,N-trimethylammonium chloride (TMSPA) was added. After 15 min of gentle rotation of the Eppendorf tubes, 1.5 μ L of triethoxysilane was added. This mixture was shaken for 24 h at 600 rpm, after which it was centrifuged at 10,000 rpm (10 s) to encourage aggregates to pellet at the bottom of the tube. The supernatant was removed and replaced with water and this process was repeated 5 times to wash the silica embedded samples. Lattices were then deposited onto a carbon coated Cu-mesh TEM grid (Ted Pella). Images were obtained using a 200 keV acceleration voltage on a Hitachi HD-2300 STEM.

Samples of slow cooled, silica embedded superlattices were enrobed in 2% low melt Agar in water. The agar, containing superlattice, was extruded into small tubes using a glass pipette tip. These tubes were then transferred to 2 mL of 30% ethanol in water, followed by slow replacement of the solvent to obtain samples in pure ethanol. Next, the solvent was slowly replaced by pure propylene oxide, and once again by resin (Embed 812, Electron Microscopy Sciences); the slow transfer process ensured that the resin fully infiltrated the sample prior to solidification. Samples

were heated at 65 °C for 48 h to solidify the resin. The resin encased superlattices were sectioned into ~60 nm thick slices that could be imaged using a 200 keV acceleration voltage on a Hitachi HD-2300 STEM.

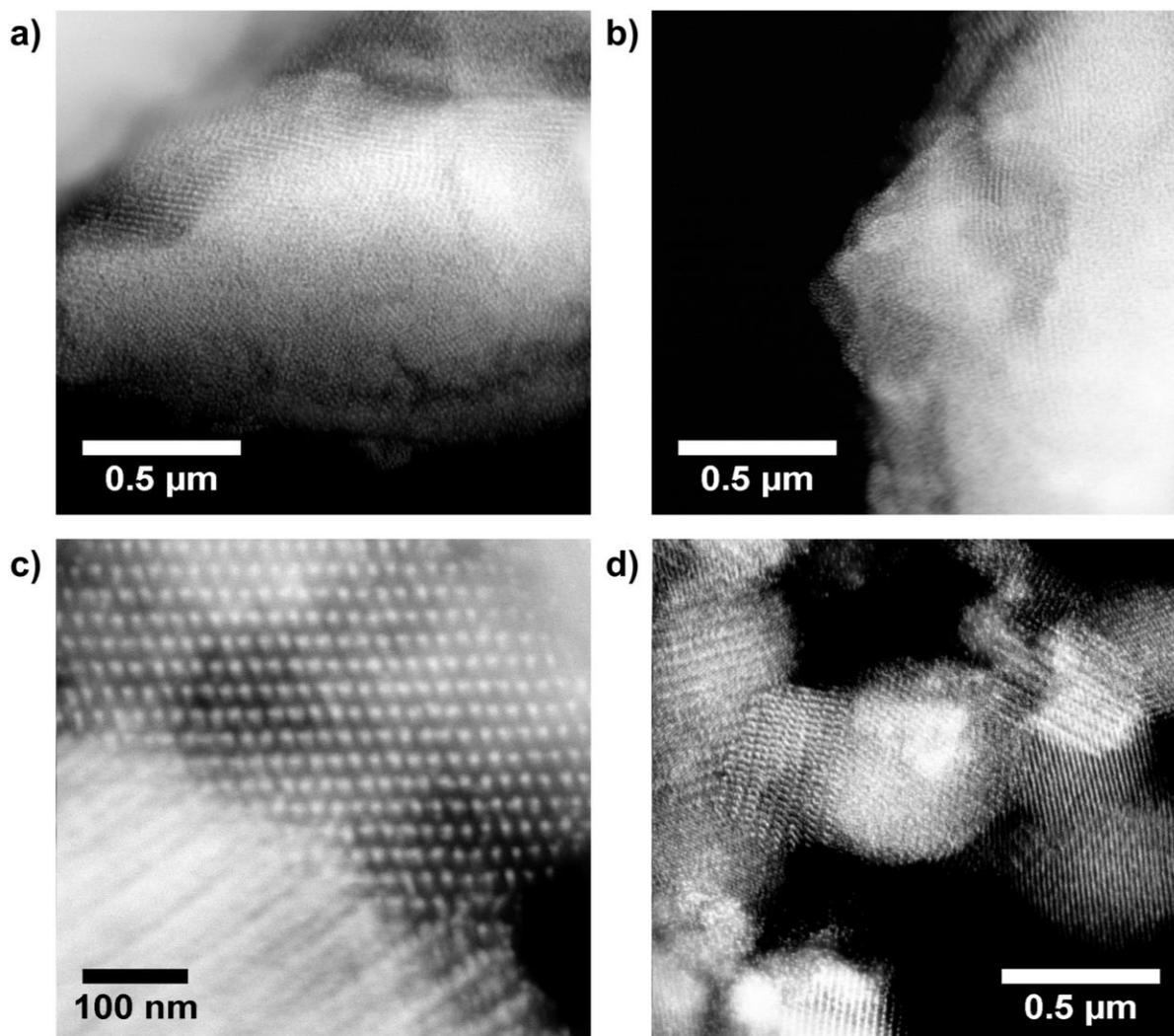


Figure 3.11 – Representative TEM images of silica embedded superlattices. (a) and (b) depict long range ordering of two orthogonally functionalized 10 nm Au NPs, arising from the DNA mediated interactions with protein Janus particles. In (c) and (d) a crystal assembled from orthogonally functionalized 5 and 10 nm AuNPs is visualized using STEM, from which we observe both the hexagonally packed and layered planes, characteristic of the $A1B_2$ structure type.

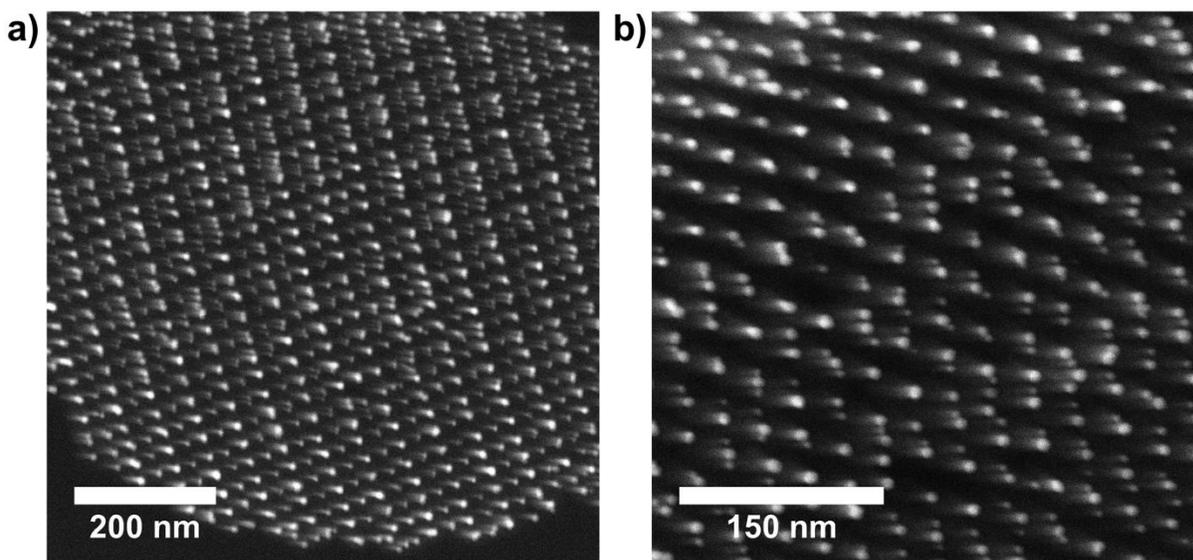


Figure 3.12 – Representative TEM images of a sectioned crystal system containing 5 and 10 nm Au NPs. (a) reveals the layers of 5 and 10 nm Au NPs, (b) shows the hexagonally packed planes.

CHAPTER FOUR

Encoding Hierarchical Assembly Pathways of Proteins with DNA

Material in this chapter is based upon:

Hayes, O. G.*; Partridge, B. P.*; Mirkin, C.A. Encoding Hierarchical Assembly Pathways of Proteins with DNA. *Under review*

*Equal author contribution

4.1 Challenges and promise of encoding hierarchical assembly pathways

Hierarchical assembly is integral to the structural complexity and function of materials and systems that occur in Nature. Muscle tissue,¹⁸⁵ amyloid fibrils,¹⁸⁶ and collagen networks¹⁸⁷ are all examples of highly organized supramolecular architectures that arise from bottom-up, multi-step, regulated assembly processes. The well-controlled sequence of assembly steps along a given pathway and the specificity of interactions between components are critical to the observed structural complexity and diversity.^{41, 188} While nanoscale hierarchical assembly is prevalent and important in Nature, and our ability to control the bottom-up assembly of synthetic nanoscale building blocks has been transformed over the past two decades,^{83, 84, 85} we are still limited in what can be programmed through hierarchical mechanisms.^{189, 190} This is due to difficulties in defining the number, type, and location of multiple interactions on synthetic building blocks, as well as limitations in controlling the interplay between orthogonal interactions to achieve a desired assembly pathway.¹⁹¹ The development of tools and strategies to program multi-step assembly pathways of nanoscale building blocks would redefine how we control the bottom-up synthesis of materials and accelerate the discovery of novel structures with desirable properties and functions.^{192, 193} In this work, we address this gap by spatially encoding programmable interacting ligands (DNA) onto the surface of chemically addressable building blocks (proteins).

Proteins are an important class of nanoscale building block because of their structural and functional roles in biology. As such, developing methods to synthetically engineer new materials from proteins is a common goal in the fields of synthetic biology, chemistry, and materials science. The chemical complexity of protein surfaces defines specific recognition between protein interfaces and is key to the hierarchical assembly processes observed in Nature. However, their

complex surfaces make it challenging to design protein building blocks that will transform into targeted materials by traversing an intended assembly pathway. While powerful *de novo* design strategies have been utilized to create proteins with predetermined interfaces and assembly outcomes, this approach inherently deviates from the pool of naturally occurring protein building blocks that could be utilized for materials engineering.^{74, 194} Other strategies have relied on introducing controlled molecular interactions to the surfaces of proteins ranging from metal coordination chemistries^{62, 195, 196} to hydrophobic¹⁹⁷ and host-guest interactions.^{122, 198} However, achieving specificity, orthogonality, and reconfigurability through these means can be challenging. Despite significant innovation in manipulating surface interactions through chemical modifications, less attention has been paid to designing protein building blocks that can undergo multi-step assembly pathways mimicking those in Nature.^{199, 200, 201} Although a growing body of literature has examined assembly pathways in the context of protein crystal polymorphism,^{202, 203,}²⁰⁴ the ability to design directional, multi-step assembly processes remains elusive. Methods to define interaction location and type on the surface of a building block, in conjunction with an understanding of how to control and regulate each interaction independently, are needed to successfully program hierarchical assembly pathways.

Our group and others have shown that DNA ligands can be chemically tethered to the surfaces of proteins, at specific locations, to drive the assembly of proteins into one- and three-dimensional assemblies and crystals.^{97-99, 169, 205, 206, 207, 208, 209} Protein mutagenesis has been used to site-specifically encode multiple, orthogonal DNA interactions onto protein surfaces to program directional assembly.^{210, 211} Furthermore, the programmable recognition properties of DNA surface ligands have been utilized to control the polymerization pathway of proteins.²¹² We hypothesized

that, if we could define the specificity, strength, and spatial distribution of multiple specific DNA interactions on the surface of a protein, we would be able to synthesize protein building blocks that undergo programmed, multi-step assembly processes. Here, by defining the chemical anisotropy of a protein's surface via mutagenesis, we define DNA interactions spatially, that is, axially or equatorially with respect to the geometry of an anisotropic protein (Figure 4.1a). Through careful DNA design, we modulate the relative interaction strengths of the axial and equatorial faces to confine each assembly step to a single direction, thereby directing proteins to assemble hierarchically along specific, multi-step pathways (Figure 4.1b).

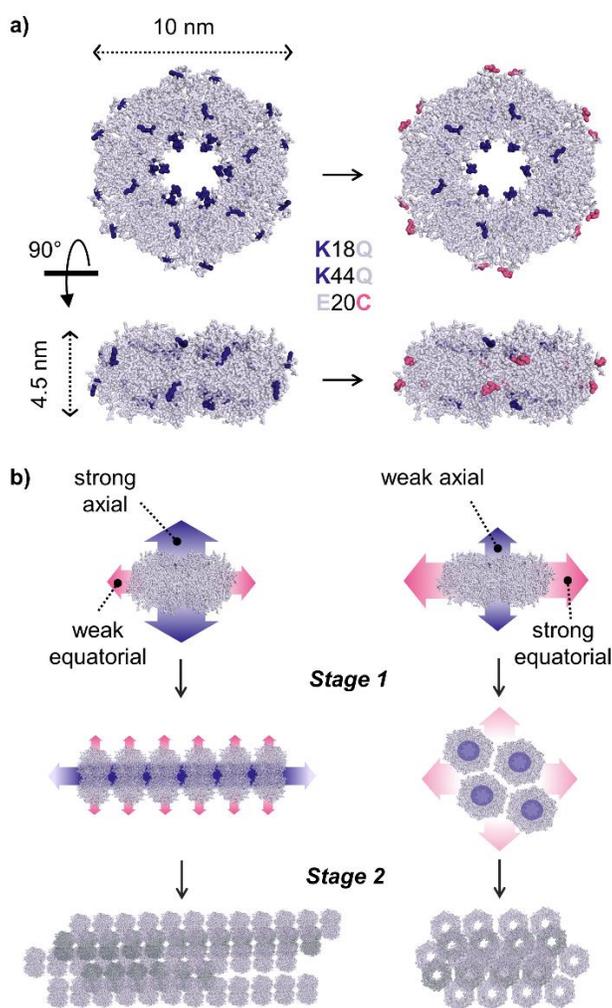


Figure 4.1 –Design of Sp1m chemical surface and proposed hierarchical assembly schemes.

(a) Native Sp1 (left) presents multiple primary amines (lysines and *N*-termini, blue) and no cysteines (red) on its surface. Three mutations were designed to remove two native lysines and introduce one cysteine. Due to the dodecameric structure of Sp1m, these mutations define the chemical anisotropy across the protein surface with amine residues only on the axial face and cysteines located only on the equatorial face. (b) Proposed assembly schemes for building blocks containing strong or weak surface interactions at their axial or equatorial positions. Strong interactions direct the first stage of assembly, leading to multivalency among weak interactions that direct the second stage of assembly.

4.2 Design and Synthesis of Sp1m-DNA Building Blocks

To explore our hypothesis, we selected stable protein 1 (Sp1, PDB: 1TR0), a symmetric homododecameric protein with pseudo hexagonal-prism geometry.²¹³ Sp1 was chosen as a model system due to its well-defined, anisotropic shape and high symmetry. To align the chemical anisotropy of the protein's surface to the shape anisotropy of the protein (Figure 4.1a), we recombinantly expressed a mutant (Sp1m) with 24 surface accessible primary amines and 12 thiols located axially and equatorially, respectively (Materials and Methods, Table 4.1). Importantly, this mutant retains the geometry of the native protein as characterized by transmission electron microscopy (TEM, Figure 4.2b). The designed chemical anisotropy was then exploited to introduce orthogonal DNA ligands to the axial and equatorial faces (Figure 4.2a). In a typical synthesis, the equatorial cysteine residues were first modified with a thiol-reactive hetero-bifunctional crosslinker (Linker 1, Figure 4.2c and Materials and Methods, Figure 4.8) to install azide functional groups. Near-complete (>95%) modification of the cysteine residues was confirmed using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS, Figure 4.2d). The axial primary amines were subsequently reacted with an amine-reactive hetero-bifunctional crosslinker (Linker 2, Figure 4.2c) to install tetrazine functional groups. Although there are two primary amines per monomeric subunit (lysine K74 and *N*-terminus), MALDI-TOF MS analysis indicated high yield (>90%) modification of only a single

primary amine per subunit. Hence only 12 of the 24 surface accessible primary amines are available for DNA attachment. High resolution, top-down proteomic evaluation of this species revealed that the *N*-terminal primary amine was modified, with marginal to no functionalization of K74 (Materials and Methods, Figure 4.12, and Table 4.3). We attribute the low reactivity of K74 to its involvement in hydrogen bonding with an adjacent subunit (Materials and Methods, Figure 4.13).

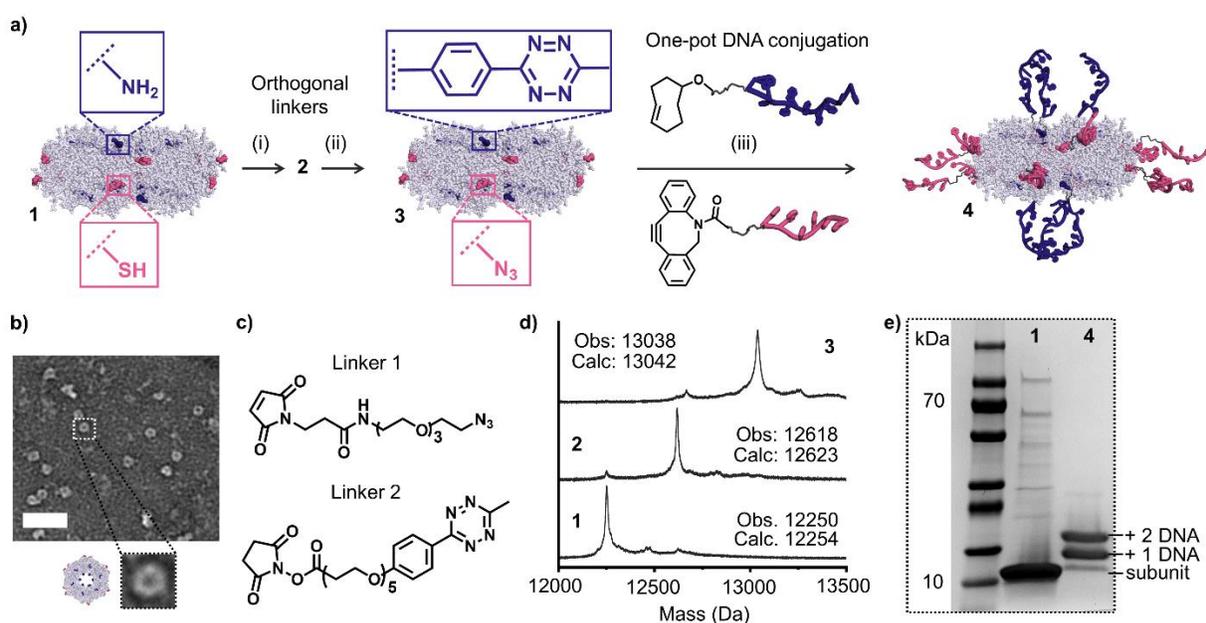


Figure 4.2 – Synthesis and characterization of Sp1m-DNA conjugates. (a) Sp1m (**1**) was modified with DNA in three steps: (i) cysteines were first modified with Linker 1 (c) through a thiol-maleimide Michael addition click reaction to give Sp1m-N₃ (**2**); (ii) primary amines were then modified with Linker 2 (c) to generate **3** through reaction with an NHS-activated ester; (iii) TCO- and DBCO-modified DNA were reacted with **3** in one-pot to generate a Sp1m-DNA building block (**4**). (b) Negative stain TEM of (**1**). Scale bar is 50 nm. Lower image: comparison of a model of Sp1m with a magnified region from the TEM image. (c) Chemical structures of heterobifunctional Linkers 1 and 2. (d) MALDI-TOF MS confirming the consecutive addition of a single molecule of each linker to each subunit of **1**. (e) Denaturing PAGE (left to right) protein ladder, unreacted Sp1m (**1**), and purified Sp1m-DNA conjugate (**4**). The presence of two bands of approximately equal intensity, at higher molecular weight compared to **1**, correspond to a roughly equal mixture of protein subunits with 1 and 2 DNA strands.

Having established a synthetic route to prepare Sp1m with two orthogonal functional groups for click chemistry (tetrazines and azides), we attached DNA to the protein surface. It has

been shown that the inverse electron demand Diels-Alder (IEDDA) reaction between tetrazines and *trans*-cyclooctene (TCO)^{214, 215} is sufficiently orthogonal to the copper-free strain-promoted alkyne-azide cycloaddition (SPAAC) reaction between azides and dibenzocyclooctyne (DBCO),¹⁷⁸ such that these reactants may be used simultaneously to achieve selective, multi-target functionalization.²¹⁶ Therefore, a one-pot reaction was employed to simultaneously conjugate orthogonal TCO- and DBCO-terminated DNA ligands to the linker-modified protein. Denaturing polyacrylamide gel electrophoresis (PAGE) confirmed successful modification of the protein and revealed the attachment of 1 or 2 DNA ligands per protein subunit (Figure 4.2e, and Materials and Methods, Figures 4.9, and 4.10). To understand this distribution and to confirm the orthogonal reactivity of the two DNA conjugation reactions, the reactions were conducted separately and analyzed via denaturing PAGE (Materials and Methods, Figure 4.11). This confirmed that DBCO-DNA ligands react exclusively with the equatorial azides with high conversion (87%, calculated by gel densitometry), resulting in ~10 DNA ligands (of a possible 12) in the equatorial plane. The TCO-DNA ligands react with lower conversion (59%), but good selectivity, suggesting that ~3–4 DNA ligands (of a possible 6) occupy each axial face of the protein, for a total of 6–8 axial DNA (of a possible 12) per building block. We attribute the lower conversion to the proximity of the *N*-termini to each other in the inner portion of the structure, which may lead to steric and electrostatic congestion with the bulky, negatively charged DNA. Given that as few as two closely placed DNA ligands on a protein's surface can act cooperatively to form interface interactions between proteins,⁹⁹ we expected that 3–4 DNA ligands per face would be sufficient to define the axial interaction. Furthermore, the strength of axial or equatorial interactions can be tuned via DNA sequence design, irrespective of the number of DNA ligands. Overall, this conjugation strategy is

highly effective and enabled the preparation of 19 unique Sp1m-DNA building blocks explored in this work.

4.3 Directional Assembly Encoded by Strong Axial or Equatorial DNA Interactions

While the above conjugation strategy controls the spatial distribution of DNA ligands on the protein surface, DNA sequence design allows for the specificity and strength of the resulting DNA-DNA interactions to be programmed. We designed DNA sequences that interact orthogonally, in different directions and at distinct stages, to define a multi-step hierarchical assembly pathway driven by the hybridization of complementary DNA (Figure 4.1b). To this end, we synthesized building blocks where the axial and equatorial DNA sequences have disparate melting temperatures (T_m), such that directionally specific interactions occur at different temperatures (Materials and Methods, Table 4.2). Specifically, interactions were designed to be either “strong” ($T_m \gg$ room temperature, RT) or “weak” ($T_m \ll$ RT). We hypothesized that, upon cooling, the strong interactions would hybridize first and building blocks would undergo a first stage of assembly. This assembled structure would display weakly-interacting DNA ligands in a multivalent fashion, resulting in an emergent interaction with enhanced cooperativity and increased T_m relative to the isolated weak interactions. The emergent interaction would then drive a second stage of assembly and the formation of a complex assembled structure.

To test if our DNA design strategy imparted directionality on the interactions (axial vs equatorial), we initially characterized the assembly outcomes of systems where only strong interactions are present. Temperature-dependent association of Sp1m-DNA conjugates was probed using a donor-quenching Förster resonance energy transfer (FRET) based technique (Figure 4.3a, b).^{99, 217} In a typical experiment, a pair of complementary Sp1m-DNA conjugates was

functionalized with cyanine 3 (Cy3) and cyanine 5 (Cy5) modified DNA, respectively. As the proteins assemble, the efficiency of FRET from excited Cy3 to Cy5 increases, leading to quenching of Cy3 fluorescence. Therefore, FRET efficiency monitored via the change in Cy3 fluorescence upon cooling from 65 to 20 °C, provides a measure of the degree of assembly (Materials and Methods, section 4.7.7). Initially, strong axial interactions (denoted A_S) were studied using two complementary conjugates, Sp1m- $A_S E_{NC}$ and Sp1m- $A'_S E_{NC}$, with Cy5- and Cy3-modified axial DNA, respectively, and non-complementary equatorial (E_{NC}) interactions that will not assemble equatorially. Their temperature-dependent association profile displayed a single transition with a T_m of 57.3 °C and full width half-maximum (FWHM, Materials and Methods, section 4.7.8) of 10.8 °C, compared to $T_m = 43.4$ °C and FWHM = 16.4 °C for the free DNA duplex (Figure 4.3b). The increased T_m and decreased FWHM observed for the Sp1m-DNA conjugates, relative to the free DNA duplex, are suggestive of a multivalent and cooperative interaction between proteins.

Sp1m- $A_S E_{NC}$ and Sp1m- $A'_S E_{NC}$ were then slow cooled (0.1 °C/10 min) and the assembly products were characterized in the dried and native states using negative stain and cryogenic TEM, respectively (Figure 4.3c, d). These micrographs revealed the formation of polymeric, 1-dimensional (1D) protein chains, connected through axial interfaces. Remarkably, in the dried state we can resolve polymeric structures containing tens of proteins (Figure 4.3c and Materials and Methods Figure 4.17) and we observe chains measuring several hundred nm long in the native state (Figure 4.3d and Materials and Methods Figure 4.22), with negligible off-target, non-axial interactions. Negative stain TEM of a control sample where only one building block is present (i.e. Sp1m- $A_S E_{NC}$) shows no evidence of assembly (Materials and Methods, Figure 4.18). Taken

together, these data support our hypothesis that a strong DNA interaction (defined via sequence design) and the axial functionalization of Sp1m (defined via mutant design and specific functionalization) encodes highly directional interactions between proteins.

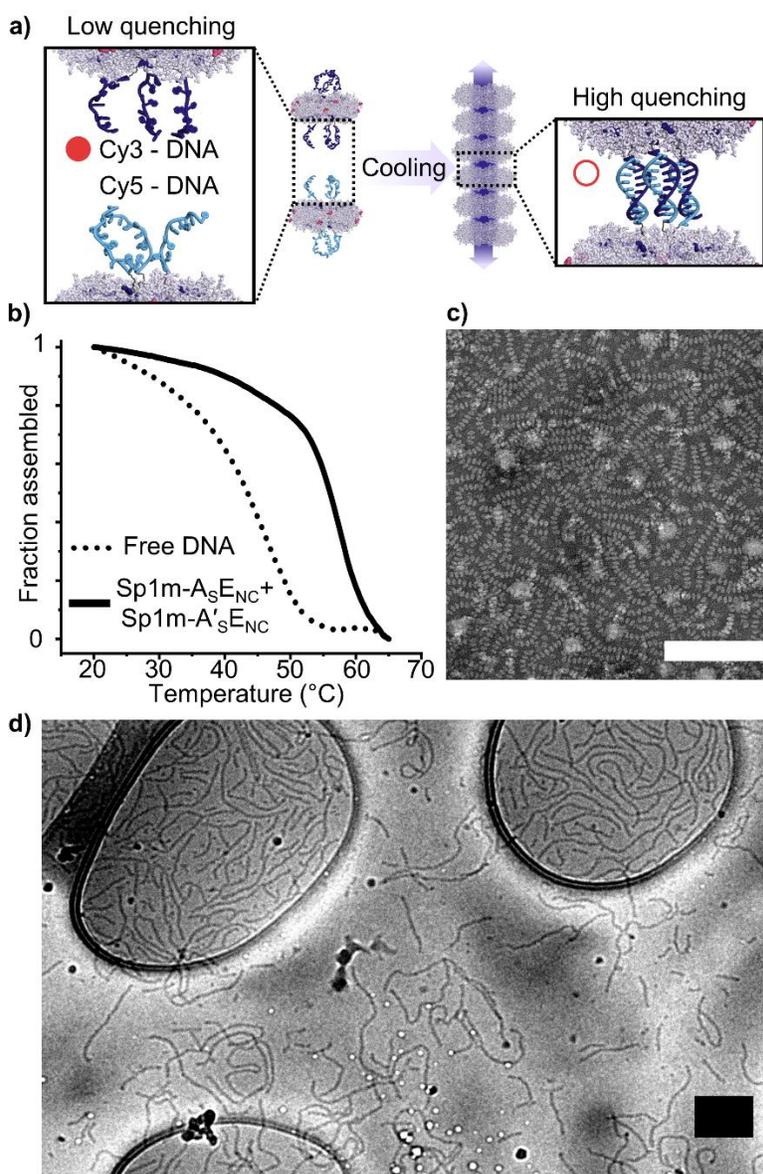


Figure 4.3 – Characterization of the assembly of Sp1m with strong axial (A_S/A'_S) interactions. (a) Scheme showing the donor-quenching FRET experiment. In a typical experiment, a pair of complementary Sp1m-DNA conjugates were functionalized with Cy3- or Cy5-modified axial DNA, respectively. When well separated, excitation of Cy3 results in fluorescence from Cy3 (filled red circle).

However, when Cy3 and Cy5 are in close proximity, FRET from excited Cy3 to Cy5 quenches the fluorescence of Cy3 leading to reduced fluorescent signal (empty red circle). (b) Temperature-dependent association of Sp1m-A_SE_{NC} and Sp1m-A'_SE_{NC} represented as fraction assembled vs temperature, where the fluorescence intensities at 65 and 20 °C correspond to a fraction assembled of 0 and 1, respectively. (c) Negative stain and (d) cryogenic TEM micrographs of slow cooled Sp1m-A_SE_{NC} and Sp1m-A'_SE_{NC}. Scale bars are 150 nm.

Next, the designed strong equatorial interactions (denoted E_S) were interrogated using an identical donor-quenching FRET technique with a pair of complementary Sp1m-DNA conjugates, Sp1m-E_S and Sp1m-E'_S, functionalized with Cy3- and Cy5-modified DNA, respectively (Figure 4.4a). As anticipated, the temperature-dependent association profile for Sp1m-E_S and Sp1m-E'_S displayed a single, sharp transition (Figure 4.4b). Analogous to the strong axial interactions, this transition has a higher T_m (57.3 °C) and lower FWHM (4.1 °C) compared to the free DNA duplex (35.9 and 14.0 °C, respectively), again suggestive of a multivalent and cooperative interaction between proteins. To assess the directionality of these interactions and characterize the assembly products, Sp1m-E_S and Sp1m-E'_S were slow cooled (0.1 °C/10 min) and observed in the dried state using negative-stain TEM (Figure 4.4c and Materials and Methods Figure 4.19) and in their native environment using liquid atomic force microscopy (AFM, Figure 4.4d, e), which enabled quantification of assembly height. Both techniques revealed 2-dimensional (2D) arrays of assembled proteins, connected through equatorial interfaces, suggesting directional interactions in the equatorial plane. Importantly, negative stain TEM of a control sample comprising only one building block (i.e., Sp1m-E_S) shows no evidence of assembly (Materials and Methods, Figure 4.20). Moreover, we confirmed the formation of monolayer assemblies using AFM (Figure 4.4d and Materials and Methods, Figure 4.23), which further supports that favorable interactions only exist in the equatorial plane.

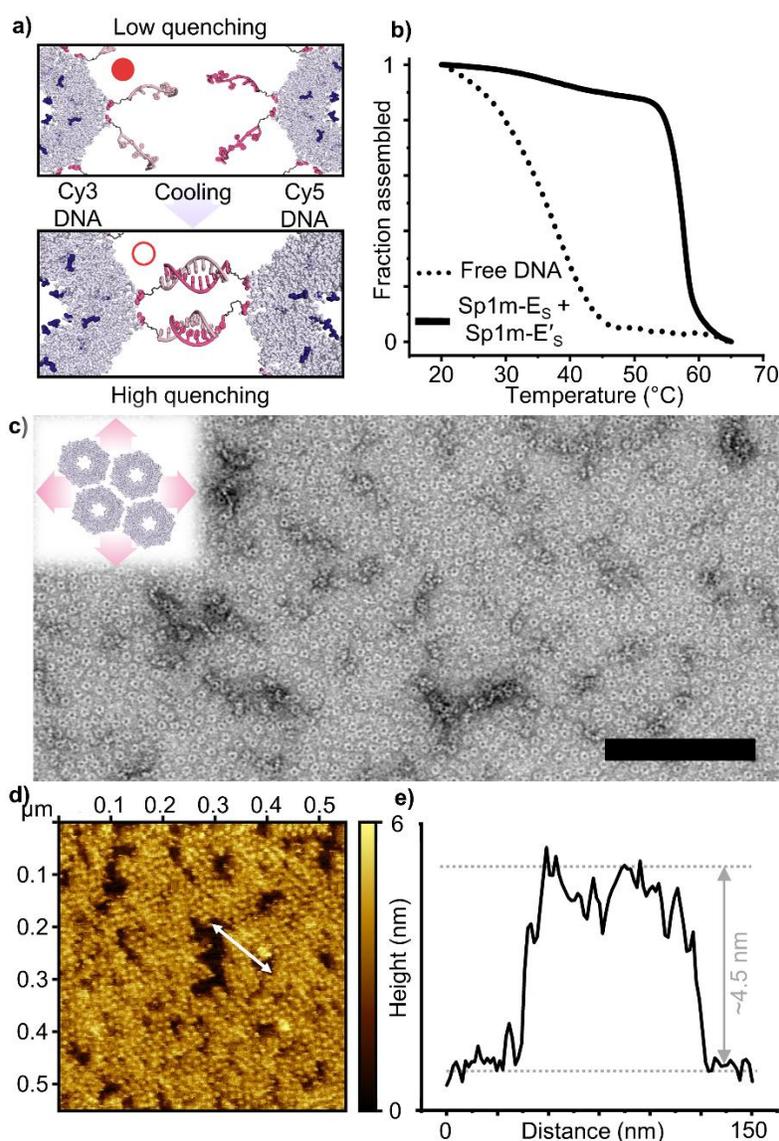


Figure 4.4 – Characterization of the assembly of Sp1m with strong equatorial (E_S/E'_S) interactions. (a) Schematic of the donor-quenching FRET experiment. (b) Temperature-dependent association of Sp1m-E_S and Sp1m-E'_S represented by plot of fraction assembled vs temperature. (c) Negative stain TEM micrograph of slow-cooled Sp1m-E_S and Sp1m-E'_S. Scale bar is 150 nm. (d) Liquid AFM micrograph of slow-cooled Sp1m-E_S and Sp1m-E'_S. White arrow denotes line used for height profile in (e).

4.4 Multi-Stage Assembly Encoded by Strong and Weak DNA Interactions

Having validated our design for encoding strong, directional interactions between proteins and characterized the assembly behavior resulting from these single-step assembly processes, we proceeded to investigate systems that could undergo defined, multi-step assembly. Guided by our

hypothesis that building blocks with both sufficiently strong and weak surface interactions would be able to traverse a hierarchical assembly pathway that relies on emergent multivalency to induce the second stage of assembly, we designed building blocks displaying axial and equatorial DNA with vastly different interaction strengths, as characterized by T_m (Materials and Methods, Table 4.2). In all cases, the weak interaction comprises self-complementary DNA sequences with a theoretical $T_m < 10$ °C, to ensure negligible association at ambient temperature prior to undergoing the first stage of assembly. To characterize these assembly steps, we again turned to a donor-quenching FRET based technique to capture their assembly profiles as a function of temperature (Figure 4.5a, d).

A pair of Sp1m building blocks, Sp1m-A_SE_{W1} and Sp1m-A'_SE_{W1}, were synthesized in which the proteins were functionalized at the axial positions with the previously discussed strong DNA sequences (A_S and A'_S) and at the equatorial positions with a self-complementary weak DNA sequence (E_{W1}). We modified the equatorial DNA sequences of Sp1m-A_SE_{W1} and Sp1m-A'_SE_{W1} with Cy3 and Cy5 dyes, respectively, such that upon the formation of 1D protein chains, driven by the strong axial interactions, the proximity of equatorial DNA would increase and thus partial quenching of the Cy3 fluorescence would occur. Further quenching would take place when the 1D structures associate through hybridization of equatorial DNA stands, indicating a second stage of assembly. As a control, an additional pair of building blocks, Sp1m-A_SE_{NC} and Sp1m-A'_SE_{NC}, was synthesized whereby the equatorial DNA ligands of Sp1m-A_SE_{NC} and Sp1m-A'_SE_{NC} were modified with Cy3 and Cy5 dyes, respectively. The degree of assembly for both systems was determined by measuring the fluorescence of Cy3 upon cooling from 65 to 20 °C (Figure 4.5b, c). The assembly profiles of both sets of building blocks revealed a sharp transition at $T_m = 54$ °C,

consistent with the T_m measured for the assembly of axial-only system (57.3 °C), that can be attributed to the association of proteins through axial interactions. The discrepancy in T_m is due to the difference in salt concentration between experiments. Additionally, for the building blocks modified with self-complementary equatorial DNA (Sp1m-A_SE_{W1} and Sp1m-A'_SE_{W1}) a second transition occurs. This transition has a T_m of 32.7 °C, which is greater than expected for the free six base-pair (bp) E_{W1} duplex (theoretical $T_m < 5$ °C), indicating a highly cooperative assembly event.

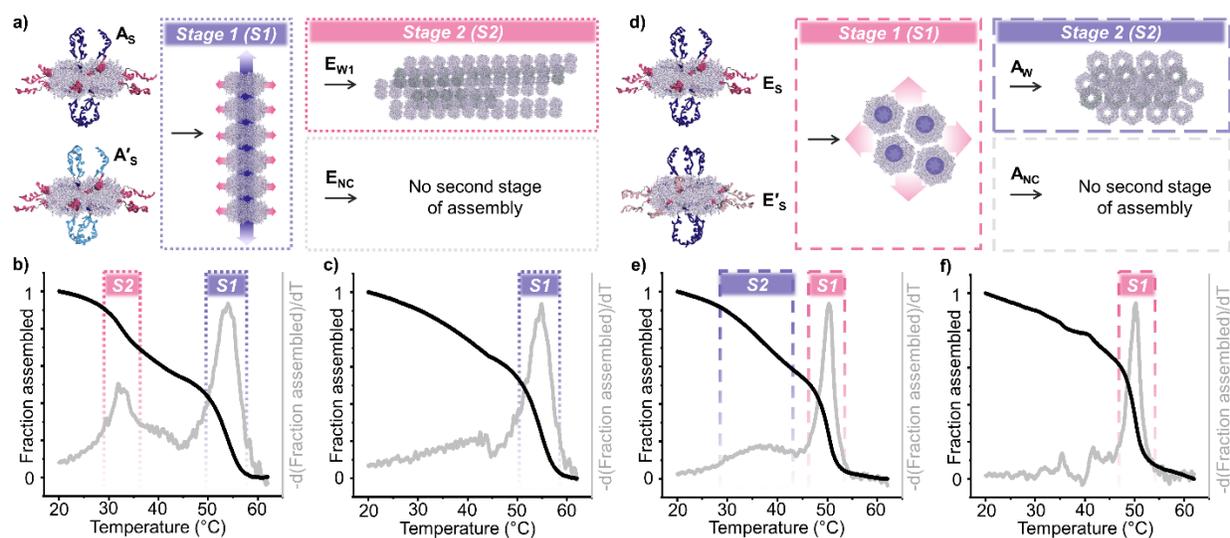


Figure 4.5 – FRET-based characterization of temperature-dependent hierarchical assembly processes. (a–c) Hierarchical assembly mediated by strong axial (A_S/A'_S) interactions. (a) Scheme showing the hypothesized assembly outcomes for two pairs of A_S/A'_S building blocks: Sp1m- $A_S E_{W1}$ with Sp1m- $A'_S E_{W1}$; and Sp1m- $A_S E_{NC}$ with Sp1m- $A'_S E_{NC}$. Temperature-dependent association of (b) Sp1m- $A_S E_{W1}$ and Sp1m- $A'_S E_{W1}$ and (c) Sp1m- $A_S E_{NC}$ and Sp1m- $A'_S E_{NC}$ represented by plots of fraction assembled vs temperature. Both pairs show the first stage of assembly mediated by A_S/A'_S interactions, but only with E_{W1} is a second stage of assembly observed. (d–f) Hierarchical assembly mediated by strong equatorial (E_S/E'_S) interactions. (d) Scheme showing hypothesized assembly outcomes for two pairs of E_S/E'_S building blocks: Sp1m- $A_W E_S$ with Sp1m- $A_W E'_S$; and Sp1m- $A_{NC} E_S$ with Sp1m- $A_{NC} E'_S$. Temperature-dependent association of (e) Sp1m- $A_W E_S$ and Sp1m- $A_W E'_S$ and (f) Sp1m- $A_{NC} E_S$ and Sp1m- $A_{NC} E'_S$ represented by plots of fraction assembled vs temperature. Both pairs show the first stage of assembly mediated by E_S/E'_S interactions, but only with A_W is a second stage of assembly observed.

DNA interactions are greatly influenced by their ionic environment,²¹⁸ and thus we investigated how this two-step assembly profile would change under different salt conditions. We repeated the cooling experiment at a higher and lower salt concentration (20 mM and 5 mM vs 10 mM MgCl₂, Materials and Methods, Figure 4.25a). Interestingly, in both 5 and 20 mM MgCl₂, the transition at 32.7 °C disappeared and the assembly profiles display a single transition at 52.0 and 55.2 °C, respectively, but these conditions result in significantly different relative fractions assembled (Materials and Methods, Figure 4.25b). Assembly driven by axial interactions results in a much greater fraction assembled in 20 mM MgCl₂ compared to lower salt concentrations, suggesting that at high salt concentration, the two assembly steps become concerted and cannot be resolved. At the lowest salt concentration (5 mM), the assembly profile suggests that only the first (axial) stage of assembly occurs and that a salt concentration between 5 and 20 mM is required for both assembly stages to occur and be resolvable. These trends are consistent with the influence of ionic environment on the hybridization of DNA; however, it is notable that the two stages of assembly differ substantially in the extent to which they are influenced by changes in salt concentration, therefore pointing to additional methods to fine tune hierarchical assembly pathways. Overall, this set of experiments provides evidence for a temperature-dependent, programmed, multi-step assembly pathway defined by DNA interactions and supports the hypothesis that Sp1m-DNA conjugates assemble first through axial interactions and then through equatorial interactions. Importantly, this second stage of assembly relies on an emergent interaction that is encoded by DNA sequences in the initial building block but is only activated after the first assembly step. This process is akin to the hierarchical generation of tertiary and quaternary protein structures defined exclusively by the information present in the primary amino acid sequence.

Next, we investigated whether a reversed assembly pathway could be programmed by simply switching the relative strengths of DNA interactions at the axial and equatorial positions. Accordingly, a new set of building blocks, Sp1m-A_wE_s and Sp1m-A_wE'_s, was synthesized employing the previously discussed strong equatorial complementary DNA sequences (E_s and E'_s) as well as weak self-complementary DNA sequences at the axial positions (A_w). We modified the axial DNA sequences of Sp1m-A_wE_s and Sp1m-A_wE'_s with Cy3 and Cy5 dyes, respectively, where we expected to observe partial quenching for the first stage of assembly (formation of 2D structures through strong equatorial interactions), and further quenching upon subsequent axial interactions during cooling from 65 to 20 °C. To provide a comparison where axial interactions are inhibited, Sp1m-A_{nc}E_s and Sp1m-A_{nc}E'_s were synthesized with non-complementary axial DNA ligands (A_{nc}) modified with Cy3 and Cy5 dyes, respectively. When comparing the temperature-dependent assembly profiles for these two sets of building blocks, the system containing both interaction types (Sp1m-A_wE_s and Sp1m-A_wE'_s) displays two distinct transitions ($T_m = 50.4$ and 38.1 °C) whereas the system with A_{nc} interactions displays only a single transition (50.4 °C; Figure 4.5e, f). We therefore attribute the common transition at 50.4 °C to the initial association of proteins in the equatorial plane to form 2D structures and the unique transition at 38.1 °C to the subsequent onset of axial interactions between these 2D structures. The transition at 38.1 °C is relatively broad, compared to the first assembly step, which may be due to polydispersity in the domain sizes of the 2D structures that associate in this step (Materials and Methods, Figure 4.24). Together, these experiments support the hypothesis that Sp1m-A_wE_s and Sp1m-A_wE'_s undergo a reversed, thermally controlled, multi-step assembly pathway, first associating through equatorial interactions and then via axial interactions.

4.5 Programming Structural Outcomes via DNA Design

We have shown that designing the relative strength of DNA ligands and their spatial arrangement on the protein surface directs assembly along different pathways with distinct assembly outcomes. We next explored whether the assembly outcome could be changed while maintaining the same pathway, via DNA sequence design. To that end, we focused on characterizing the structures that arise from an axial-first, equatorial-second assembly pathway. In addition to the previously described system, Sp1m-A_SE_{W1} and Sp1m-A'_SE_{W1} (Figure 4.5), we designed an additional building block, Sp1m-A'_SE_{W2}, where the equatorial sites of the second building block were modified with a weak self-complementary sequence (E_{W2}) orthogonal to E_{W1}. The E_{W1} and E_{W2} DNA sequences (Materials and Methods, Table 4.2) are identical in length and bp composition to ensure that differences in the assembly outcome result from differences in the presentation of the emergent second interaction, rather than inherent differences in the interaction strength between the two self-complementary DNA designs. The building blocks were slow cooled (0.1 °C/10 min) in two combinations: Sp1m-A_SE_{W1} with Sp1m-A'_SE_{W1} (Figure 4.6a), and Sp1m-A_SE_{W1} with Sp1m-A'_SE_{W2} (Figure 4.6c). The complementarity of A_S and A'_S ensures that, in the latter system, E_{W1} and E_{W2} are presented alternately (Figure 4.6c). TEM characterization of both samples reveals the formation of 1D protein chains, formed via A_S interactions, that interact with each other, suggesting that these two systems traverse the same assembly pathway. However, the two sets of building blocks give significantly different structural outcomes (Figure 4.6b, d). For the system containing only E_{W1}-based building blocks, the 1D protein chains have a high propensity to form bundles and fold up on themselves via intra-chain interactions (Figure 4.6a). However, when one of the building blocks is modified with E_{W2}, the 1D protein chains instead

interact to form elongated filaments. Moreover, TEM suggests that registry between the proteins in each chain is better enforced in this sample (Materials and Methods, Figure 4.21). The presence of alternating, orthogonally self-complementary interaction areas, spatially encoded on the surface of the 1D chain by DNA, favors inter-chain association by maximizing DNA hybridization. This highlights how two, orthogonal, self-complementary E_w sequences decrease the propensity for the 1D protein chains to fold and bundle and is a key demonstration of how DNA design can not only define a specific assembly pathway but also direct the final structural outcome. Additional control over the final structural outcome, such as achieving higher order or enforcing registry between particles, can in principle be achieved through further exploration of DNA design parameters, including length, flexibility, absolute and relative interaction strength, position of DNA attachment, and number of DNA strands, among many others.

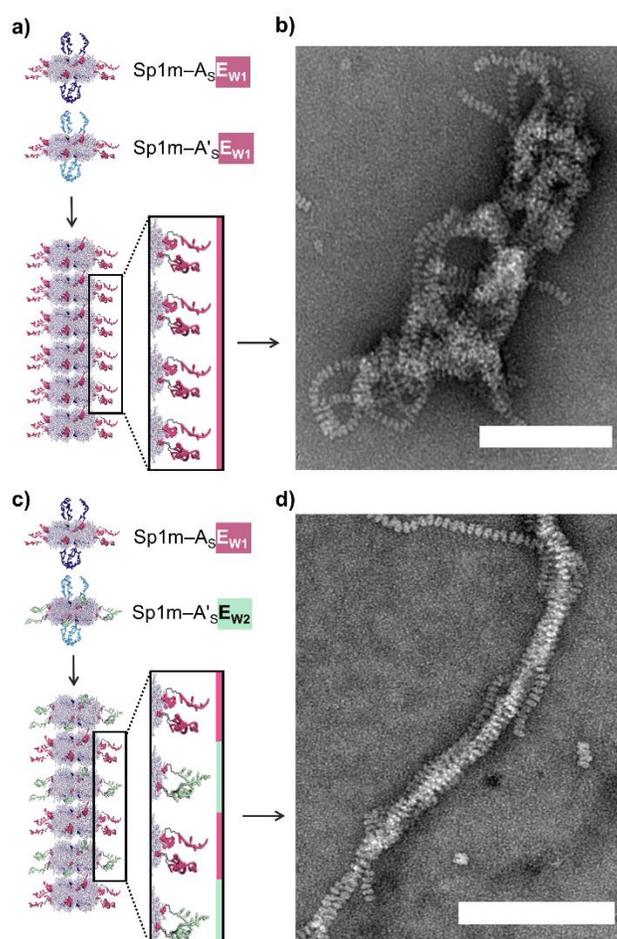


Figure 4.6 – Characterization of assembly outcomes from axial-first, equatorial-second hierarchical assembly processes. (a) Scheme showing 1D protein chains displaying equatorial E_{W1} DNA homogenously. (b) Negative-stain TEM micrograph of slow-cooled assembly of $Sp1m-A_S E_{W1}$ and $Sp1m-A'_S E_{W1}$. (c) Scheme showing 1D protein chains displaying alternating equatorial E_{W1} and E_{W2} DNA. (d) Negative-stain TEM micrograph of slow-cooled assembly of $Sp1m-A_S E_{W1}$ and $Sp1m-A'_S E_{W2}$. Scale bars are 150 nm.

4.6 Conclusions

This work harnesses the programmability of DNA and the chemical addressability of protein surfaces to control the hierarchical, multi-step assembly of protein building blocks mediated by multiple, distinct DNA hybridization events. Through functionalization of a protein's surface with DNA ligands at axial and equatorial positions, we introduced highly directional

interactions between specific geometric interfaces. We programmed multi-step assembly profiles by defining disparate recognition properties at different locations within discrete protein building blocks, which allows us to control the assembly pathways and structural outcomes. Furthermore, we used DNA to define multiple orthogonal interactions within a single assembly pathway, thereby realizing distinct protein-based materials as a function of both the type of pathway traversed and the DNA design employed. This principle, in which all information required for hierarchical assembly is encoded into an initial primary structure, has long been exploited by Nature to realize sophisticated architectures from amino acid sequences, but seldom by using nucleic acids. In contrast to canonical uses of nucleic acids in Nature—primarily information storage and sometimes as a template to organize structures—DNA is rarely, if ever, employed as a programmable “bond” to direct complex assembly pathways. These findings show that, through judicious design, one can use DNA to build structures on demand with a degree of hierarchical control atypical for synthetic nanoscale programmable matter but reminiscent of complex structures in Nature. In principle, because DNA is encoding the information for hierarchical assembly, this approach can be applied to any protein, or other nanoscale building block, where the surface can be appropriately modified with DNA. These insights reveal how to go beyond a single-step assembly pathway for the bottom-up assembly of nanomaterials and will enable the synthesis of hierarchically structured materials by design.

4.7 Materials and Methods

4.7.1 Protein Design and Expression

Owing to the relatively orthogonal chemistries of thiols and primary amines, we chose to create a mutant where the axial and equatorial faces of the protein each displayed one type of these

functional groups. Sp1m was selected, not only because of its desirable geometry, but also because it possesses no native, surface-exposed cysteine residues. Therefore, due to symmetry, a single cysteine mutation (E20C) would be sufficient to introduce thiols to the equatorial face of the protein. Two additional mutations (K18Q and K44Q) were required to remove excess lysine residues from the protein surface to localize primary amines to the axial faces (Table 4.1).

Protein sequence^a	MATRTPKLVKHTLLTRFQDCITREQIDNYINDYTNLLDLIPSMQSFNWTDLGME SAELNRGYTHAFESTFESKSGLQEYLDAAALAAFAEGFLPTLSQRLVIDYFLY-
Gene sequence	ATG GCG ACC CGC ACC CCG AAA CTG GTT AAA CAC ACC CTG CTG ACC CGC TTC CAG GAT TGC ATT ACC CGC GAA CAG ATC GAC AAC TAC ATC AAC GAC TAC ACC AAC CTG CTG GAT CTG ATT CCG AGC ATG CAG AGC TTC AAC TGG GGC ACC GAC CTG GGT ATG GAG AGC GCG GAA CTG AAC CGT GGT TAC ACC CAC GCG TTC GAG AGC ACC TTT GAA AGC AAA AGC GGC CTG CAG GAG TAT CTG GAT AGC GCG GCG CTG GCG GCG TTT GCG GAA GGT TTT CTG CCG ACC CTG AGC CAA CGC CTG GTT ATT GAT TAC TTT CTG TAT TAA

^a Mutations relative to native protein sequence are highlighted: deletion of native lysines [K18Q and K44Q] in gray; addition of cysteine [E20C] in red.

Table 4.1 – Protein and Gene Sequences for Sp1m.

The gene for Sp1m (Table S1) was purchased from Integrated DNA Technologies (IDT) and cloned into a pET-28a(+) vector using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The plasmid was transformed into electrocompetent BL21*(DE3) *E. coli* cells (New England Biolabs) by electroporation, and cells were grown overnight on LB agar plates with 50 µg/mL kanamycin.

Single colonies were picked and used to inoculate 8 mL cultures which were grown overnight at 37 °C in LB broth with 50 µg/mL kanamycin. These cultures were added to 750 mL of YTP Broth (2×) with 50 µg/mL kanamycin. Cells were grown at 37 °C to an optical density of

0.6, then induced with 1 mM IPTG and allowed to grow overnight at 25 °C. Cells were spun down ($6000 \times g$, 4 °C, 25 min) and resuspended in 100 mL of 20 mM HEPES buffer (pH 7.4) containing 2 mM DTT and 1 mM EDTA, then lysed using a high-pressure Avestin Emulsiflex C5 homogenizer. The cell lysate was clarified by centrifugation ($15000 \times g$, 4 °C, 25 min). The clarified lysate was then digested with 50 $\mu\text{g}/\text{mL}$ proteinase K and 2 $\mu\text{L}/\text{mL}$ DNase I at 37 °C for 45 min, followed by the addition of 1 mM phenylmethylsulfonyl fluoride (PMSF) and incubation at 25 °C for 30 min. Subsequently, the digested lysate was clarified by centrifugation ($15000 \times g$, 4 °C, 15 minutes). The soluble protein was concentrated to a volume of ~ 3 mL (50 K MWCO centrifugal filter, 8 °C, 7800 rpm, 15 min cycles) and filtered through a 0.2 μm filter (Amicon). The sample was then loaded in 1 mL volumes onto a Q anion exchange column (GE Healthcare), and the column was washed with a gradient of 100% buffer A (20 mM HEPES, pH 7.4) to 100% buffer B (20 mM HEPES, pH 7.4 + 1 M NaCl) at 3 mL/min over 100 min. The desired protein eluted at an approximate conductivity of 45 mS/cm. The eluted protein was diluted with an equal volume of 70% glycerol and stored in 1 mL aliquots at -20 °C.

4.7.2 Oligonucleotide Design and Synthesis

A complete list of oligonucleotides used in the study can be found in Table 4.2. Extinction coefficients (ϵ_{260}) listed in Table 4.2 were used to determine oligonucleotide concentration based on their absorbance at 260 nm (Cary 5000 UV-vis spectrophotometer). Values of ϵ_{260} were calculated using the IDT Oligo Analyzer tool, with corrections made for absorbance by Cy3, Cy5, and DBCO using values from Glen Research. Melting temperatures (T_m) were calculated using the IDT Oligo Analyzer tool, using $[\text{DNA}] = 1 \mu\text{M}$ and $[\text{Mg}^{2+}] = 10 \text{ mM}$.

Name	Sequence (5' to 3') ^a	T_m (°C) ^b	ϵ_{260} (M ⁻¹ cm ⁻¹)	Calcd MW (Da)	Found MW (Da)
A _S	TCO-CTGGAAGTGT	44	93700	3231	3227
A _S -Cy5	TCO-Cy5- CTGGAAGTGT	44	103700	3760	3760
A' _S	TCO-ACAGTTCCAG	44	99300	3200	3197
A' _S -Cy3	TCO-Cy3- ACAGTTCCAG	44	104230	3703	3702
A _W	TCO-AATATATT	8	87100	2596	2593
A _W -Cy5	TCO-Cy5-AATATATT	8	97100	3129	3125
A _W -Cy3	TCO-AATATATT-Cy3	8	92030	3103	3100
A _{NC}	TCO-TTTTTT	nc	49200	1952	1950
A _{NC} -Cy5	TCO-Cy5-TTTTTT	nc	59200	2485	2482
A _{NC} -Cy3	TCO-TTTTTT-Cy3	nc	54130	2459	2456
E _S	DBCO-CTACAAATCT	35	104200	3542	3536
E _S -Cy3	DBCO-Cy3- CTACAAATCT	35	109130	4049	4042
E' _S	DBCO-AGATTTGTAG	35	113400	3653	3647
E' _S -Cy5	DBCO-Cy5- AGATTTGTAG	35	123400	4186	4179
E _{W1}	DBCO-AATATT	≤ 0	73000	2361	2361
E _{W1} -Cy3	DBCO-Cy3-AATATT	≤ 0	77930	2868	2865
E _{W1} -Cy5	DBCO-Cy5-AATATT	≤ 0	83000	2894	2890
E _{W2}	DBCO-TAATTA	≤ 0	73600	2361	2362
E _{NC}	DBCO-TTTTTTTT	nc	73400	2942	2940
E _{NC} -Cy3	DBCO-Cy3-TTTTTTTT	nc	78330	3450	3443
E _{NC} -Cy5	DBCO-Cy5-TTTTTTTT	nc	83400	3476	3468

^a Non-standard nucleotides:

- TCO (*trans*-cyclooctene) - 2-cyanoethyl (*E*)-cyclooct-4-enyl *N,N*-diisopropyl phosphoramidite, synthesized as described in section 4.7.3. $\epsilon_{260} = 0 \text{ M}^{-1} \text{ cm}^{-1}$ (i.e., no correction applied).
- DBCO (dibenzocyclooctyne) – 5'-DBCO-TEG phosphoramidite (Glen Research #10-1941). $\epsilon_{260} = 8000 \text{ M}^{-1} \text{ cm}^{-1}$.

- Cy3 – cyanine 3 phosphoramidite (Glen Research #10-5913). $\epsilon_{260} = 4930 \text{ M}^{-1} \text{ cm}^{-1}$.
- Cy5 – cyanine 5 phosphoramidite (Glen Research #10-5915). $\epsilon_{260} = 10000 \text{ M}^{-1} \text{ cm}^{-1}$.

^b Melting temperatures (T_m , rounded to nearest °C) were calculated for complementary and self-complementary sequences using the IDT Oligo Analyzer tool, using $[\text{DNA}] = 1 \text{ }\mu\text{M}$ and $[\text{Mg}^{2+}] = 10 \text{ mM}$. Sequences used as non-complementary interactions are indicated by nc.

Table 4.2 – Oligonucleotide Sequences, Extinction Coefficients, and Calculated and Observed Masses.

All materials for oligonucleotide synthesis were obtained from Glen Research or Fisher, unless otherwise specified. Oligonucleotides were synthesized using standard conditions on solid controlled pore glass (CPG) supports, with acetonitrilic 4,5-dicyanoimidazole solution as activator and either iodine solution or 0.5 M camphorsulfonyloxaziridine (CSO) solution as oxidizer. Typically, oligonucleotides containing cyanine-based phosphoramidites were synthesized using an Applied Biosystems ABI 391 synthesizer using CSO as oxidant whereas those without cyanine-based phosphoramidites were synthesized using a Bio Automation MerMade 12 oligonucleotide synthesizer using iodine as oxidant. Oligonucleotides containing TCO or DBCO modifications were synthesized as above and then coupled using syringes, by hand, to install the 5'-functional moiety. Hand-coupling procedures utilized 0.5 M CSO for DBCO and 0.1 M CSO for TCO.

Oligonucleotides were cleaved from CPG using 30% NH_3 (aq) for 16–22 h at room temperature and purified using reverse-phase HPLC (Agilent 1260 Infinity II LC system) with a gradient of acetonitrile in triethylammonium acetate buffer that varied depending on the oligonucleotide design. The identities of purified oligonucleotides were confirmed by matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Bruker MALDI Rapiflex Tissue Typer instrument in linear mode using negative ion detection

using 2',6'-dihydroxyacetophenone as matrix and diammonium hydrogen citrate as co-matrix. All observed masses were within 10 Da of the calculated mass.

4.7.3 Synthesis of TCO Phosphoramidite

2-Cyanoethyl (*E*)-cyclooct-4-enyl *N,N*-diisopropyl phosphoramidite (**7**, “TCO phosphoramidite”) was synthesized according to a modified literature procedure (Figure 4.7).²¹⁹ NMR were recorded on a Bruker Avance III HD system equipped with a TXO Prodigy probe at 298 K and referenced to the internal TMS standard at $\delta = 0$ ppm. High resolution mass spectrometry (HRMS) data were recorded using a Bruker Impact II Q-TOF High Resolution Time of Flight mass spectrometer connected to a Bruker Elute UHPLC. Samples were injected via flow analysis using a flow of 0.3 mL/min of 5:95 water:acetonitrile with 0.1% formic acid.

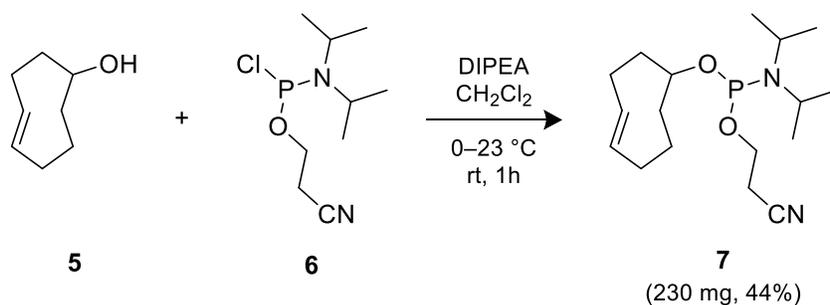


Figure 4.7 –Synthesis of TCO Phosphoramidite (**7**)

(*E*)-cyclooct-4-enol (**5**, 200 mg, 1.58 mmol, 1 equiv; Broadpharm) was suspended in dry CH_2Cl_2 (6 mL, Acros) under a N_2 atmosphere. Diisopropylethylamine (1.38 mL, 7.92 mmol, 5 equiv; Acros) was added to the suspension, which was then cooled to 0 °C in an ice-water bath. 2-Cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (**6**, 530 μL , 2.38 mmol, 1.5 equiv; Sigma-Aldrich) was added dropwise. The mixture was allowed to warm to room temperature whereupon it was stirred for 1 h. The solution was dried onto Celite (Sigma-Aldrich) and purified by flash

column chromatography, eluting with a gradient of 0 to 30% ethyl acetate in hexanes, to give pure product as a colorless oil (230 mg, 0.70 mmol, 44%) that was stored in the dark at -20°C .

4.7.4 Synthesis and Characterization of Sp1m-DNA Conjugates

The functionalization of Sp1m (**1**) with small molecule linkers is summarized in Scheme S2. Sp1m (**1**, stored in 1 mL aliquots in glycerol/HEPES at -20°C , see SI Section 1) was allowed to warm to 4°C and glycerol was removed by size exclusion chromatography (SEC, using a Bio-Rad ENrich SEC 650 10×300 column on a Bio-Rad NGC 10 medium-pressure chromatography system), eluting with 20 mM HEPES (pH 7.4). The fractions containing Sp1m (verified by the absorbance at 280 nm) were combined. An aqueous solution of tris(2-carboxyethyl)phosphine (TCEP, 200 mM) was added to the protein to a final concentration of 2 mM TCEP. The protein was optionally stored at 4°C overnight before proceeding to the next step.

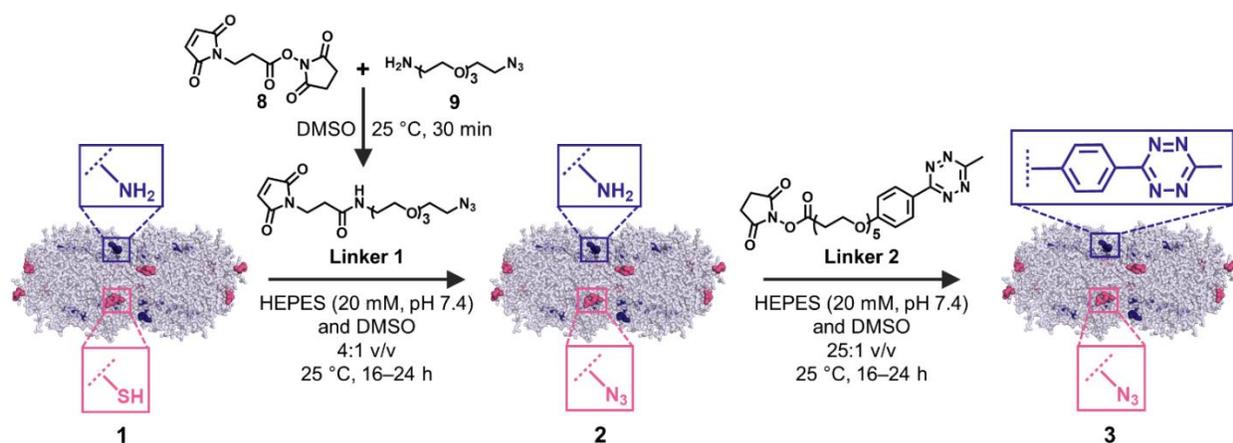


Figure 4.8 – Functionalization of Sp1m with Azide and Tetrazine Linkers

The concentration of Sp1m was determined from its UV absorbance at 280 nm ($\epsilon_{280} = 173,280 \text{ M}^{-1} \text{ cm}^{-1}$) and adjusted to $5 \mu\text{M}$ by addition of 20 mM HEPES (pH 7.4) or concentration via

centrifugal filter (30 K MWCO, $3000 \times g$, $4\text{ }^{\circ}\text{C}$, 3 min), as necessary. The protein solution was portioned into 1.5 mL Eppendorf tubes in 400 μL aliquots.

To prepare the maleimide-azide linker (**Linker 1**), a solution of azido-PEG3-amine (**8**, 2 μL ; Broadpharm) in DMSO (48 μL) was added to solution of 3-maleimido-propionic NHS ester (**9**, 2.5 mg; Broadpharm) in DMSO (50 μL). The mixture was shaken at 650 rpm at $25\text{ }^{\circ}\text{C}$ for 30 min. The reaction was quenched by addition of Tris (1 M, pH 7, 10 μL) and shaken for a further 5 min. The mixture (110 μL) was added to an aliquot of Sp1m (**1**, 400 μL , 5 μM) and shaken at 650 rpm overnight at $25\text{ }^{\circ}\text{C}$. The reaction mixture was purified by SEC (as described above, eluting with 20 mM HEPES) and fractions containing Sp1m- N_3 (**2**) were pooled and concentrated to 5 μM via centrifugal filter (30 K MWCO, $3000 \times g$, $4\text{ }^{\circ}\text{C}$, 4 min). Successful addition of the azide linker was verified by MALDI-TOF MS (Figure 4.2d).

The azide linker-modified Sp1m (**2**) was immediately portioned into 1.5 mL Eppendorf tubes in 500 μL aliquots. To each aliquot was added a solution of methyltetrazine-PEG5-NHS ester (**Linker 2**, 0.6 μL ; Broadpharm) in DMSO (20 μL) with thorough mixing by pipette aspiration. The solution was shaken at 650 rpm for 20 h at $25\text{ }^{\circ}\text{C}$. The reaction mixture was purified by SEC (as described above, eluting with 20 mM HEPES) and fractions containing protein were pooled and concentrated via centrifugal filter (30 K MWCO, $3000 \times g$, $4\text{ }^{\circ}\text{C}$, 4 min). The concentration was determined using $\epsilon_{280} = 173,280\text{ M}^{-1}\text{ cm}^{-1}$ and the absorbance by tetrazine was deemed negligible. Successful addition of the tetrazine linker was verified by MALDI-TOF MS (Figure 4.2d). Sp1m with both azide and tetrazine linkers (hereafter, Sp1m-2L or **3**) was typically reacted with DNA immediately (see next subsection), although Sp1m-2L (**3**) could be stored at $4\text{ }^{\circ}\text{C}$ for 24 h without loss in reactivity.

DNA conjugation reactions were typically performed on the 0.5, 0.7, or 1 nmol scale with respect to Sp1m-2L (**3**). For a 1 nmol reaction, to a 1.5 mL Eppendorf tube were added:

- 150 μL of Sp1m-2L (**3**) solution in 20 mM HEPES, pH 7.4 (6.66 μM , 1 nmol, 1 equiv);
- 100 μL of TCO-DNA solution in water (1800 μM , 180 nmol, 15 equiv per tetrazine);
- 50 μL of DBCO-DNA solution in water (3000 μM , 150 nmol, 12.5 equiv per azide);
- 34 μL of NaCl solution in water (5 M, to bring reaction concentration to 500 mM); and
- 4 μL of HEPES, pH 7.4 (1 M, to bring reaction concentration to 20 mM).

The solution was shaken at 650 rpm for 20 h at 37 °C. Most unreacted DNA was removed by washing the reaction mixture three times in a 4 mL centrifugal filter with 20 mM HEPES (30 K MWCO, 3000 $\times g$, 4 °C, 3 min cycles). The remaining unreacted DNA was removed from the reaction mixture by SEC (as described in the previous subsection, eluting with 20 mM HEPES) and fractions containing protein were pooled, concentrated to 300–500 μL via centrifugal filter (50 K MWCO, 3000 $\times g$, 4 °C, 2 min cycles), and stored at 4 °C. Addition of DNA was verified by denaturing polyacrylamide gel electrophoresis (PAGE, Figures 4.9 and 4.10). Samples were used within 7 days for subsequent experiments.

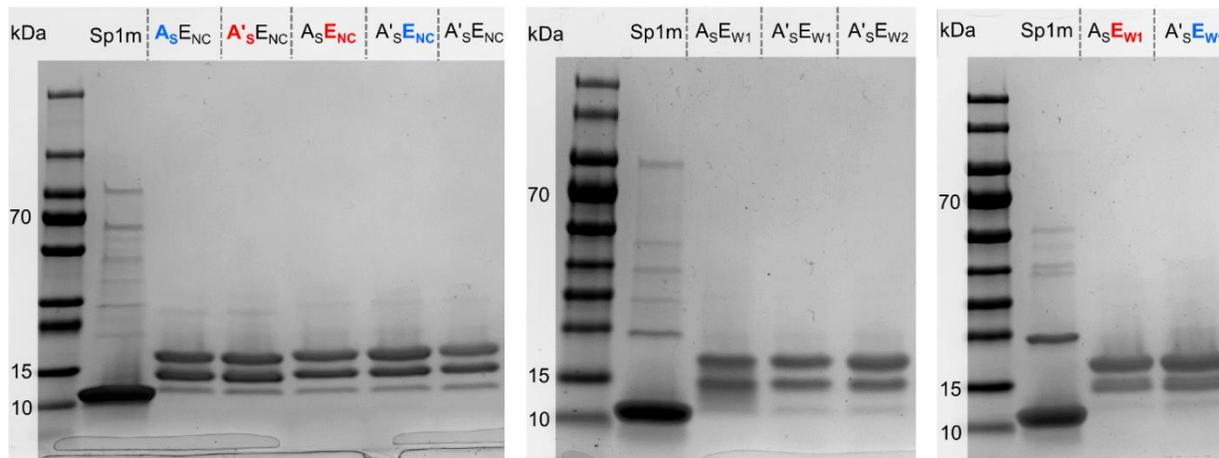


Figure 4.9 – Denaturing 4–15% PAGE of Sp1m-DNA conjugates containing As or A's DNA. Color coding to reflect Cy3- (red) or Cy5- (blue) labelled DNA. Leftmost lane in each panel is a protein ladder; molecular weight (in kDa) of select bands are indicated. Protein bands visualized by staining with SimplyBlue (Thermo Fisher).

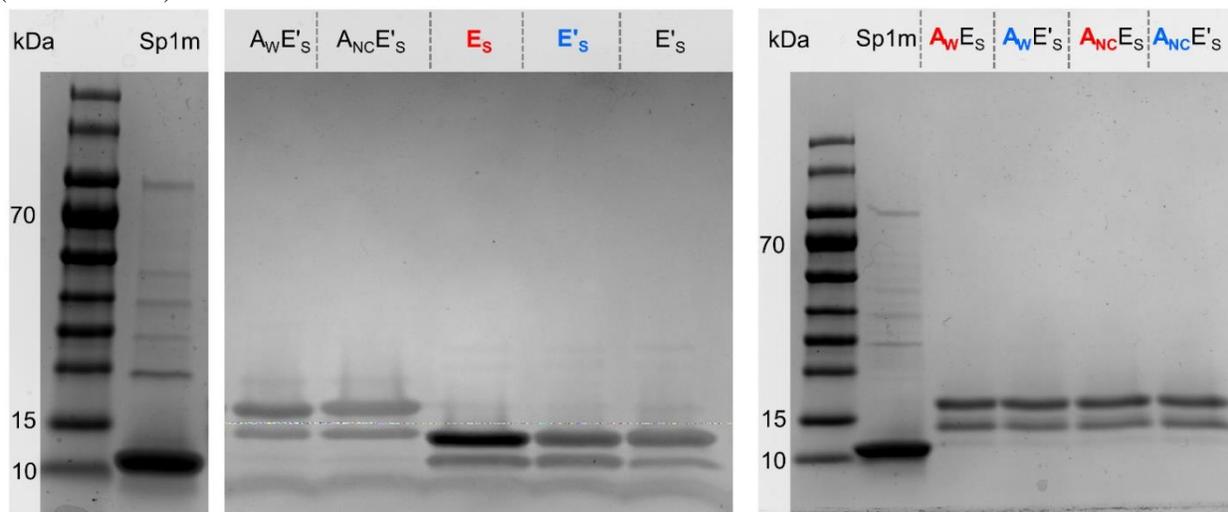


Figure 4.10 – Denaturing 4–15% PAGE of Sp1m-DNA conjugates containing Es or E's DNA. Color coding to reflect Cy3- (red) or Cy5- (blue) labelled DNA. Leftmost lane in each panel is a protein ladder; molecular weight (in kDa) of select bands are indicated. Protein bands visualized by staining with SimplyBlue (Thermo Fisher).

4.7.5 Cross-Reactivity of TCO/Tetrazine and DBCO/Azide Conjugation Reactions

Previous reports have shown that the inverse electron demand Diels-Alder (IEDDA) reaction between tetrazines and TCO is sufficiently orthogonal to the copper-free strain-promoted

alkyne-azide cycloaddition (SPAAC) reaction between azides and DBCO, such that these reactants may be used simultaneously to achieve selective, multi-target functionalization. We chose to employ a similar strategy to conjugate two, orthogonal DNA strands to azide and tetrazine functional groups on the surface of Sp1m simultaneously in a one-pot reaction. We probed the orthogonality of these reactions in the context of our system and determined the yield of each reaction in a series of control experiments.

Using the protocol described in section 4.7.4, we synthesized azide-modified Sp1m (Sp1m-N₃, **2**), and azide/tetrazine modified Sp1m (Sp1m-2L, **3**). We performed DNA conjugation reactions (similar to the protocol described in SI Section 3.2) with **2** and **3** using a single type of DNA, modified with either TCO or DBCO, to determine yield and cross-reactivity assessed by denaturing PAGE and densitometry analysis using ImageJ software (Figure 4.11).

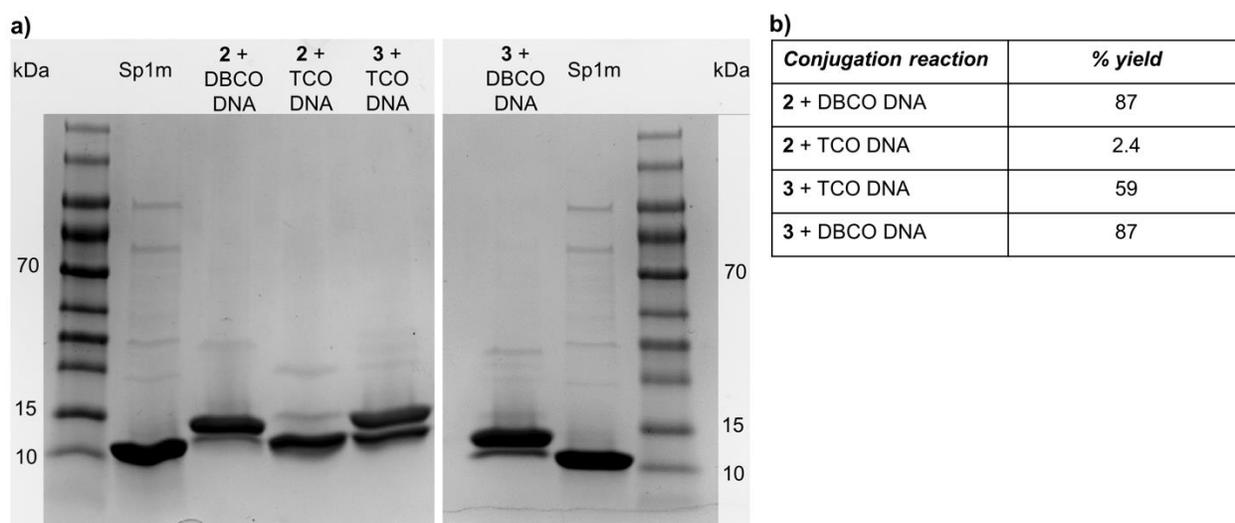


Figure 4.11 – Reactivity of TCO- and DBCO-modified DNA with Sp1m-N₃ (2**) and Sp1m-2L (**3**).** (a) Denaturing 4–15% PAGE showing conjugation products of TCO- and DBCO-DNA with **2** and **3**. Leftmost and rightmost lanes are identical protein ladders; molecular weight (in kDa) of select bands are indicated. Protein bands visualized by staining with SimplyBlue (Thermo Fisher). (b) Reaction yields determined by densitometry analysis of (a).

We observe high yielding (87%) conjugation upon reaction of azide-modified Sp1m (**2**) with DBCO-DNA. An identical degree of conjugation (87%) is observed for the reaction between **3** and DBCO-DNA, suggesting minimal cross-reactivity between DBCO and tetrazine. A moderate degree of conjugation (59%) is observed for the reaction between **3** and TCO-DNA. A negligible degree of cross-reactivity (2.4%) is observed for the reaction between TCO-DNA and azide-modified Sp1m (**2**).

4.7.6 Proteomics Analysis

To determine the extent of functionalization of Sp1m with Linkers 1 and 2, and the amino acid site of attachment of Linker 2, high resolution MS analysis of Sp1m-2L (**3**) was conducted by the Northwestern Proteomics Core Facility.

Protein sample was diluted in solvent A (95% H₂O, 4.8% ACN, 0.2% formic acid) and subjected to nanocapillary reverse-phase liquid chromatography performed on a Bionex Ultimate 3000 chromatographic system (Thermo Scientific, San Jose, CA) using a combination of an in-house prepared trap column (150 μ m diameter, 25 mm length) and analytical column (75 μ m diameter, 250 mm length) packed with PLRP-S resin (5 μ m particle size; Agilent, Santa Clara, CA). Separations were carried out using a gradient of solvent B from 5 to 50% in 50 min followed by re-equilibration at 5% B (solvent B consisted of 4.8% H₂O in ACN and 0.2% formic acid). The column outlet was coupled to a nanoelectrospray ionization (nanoESI) source built in-house. All MS experiments were carried out on an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific, San Jose, CA) operated in “protein mode.” Top-down MS acquisition parameters were set as follows: broadband MS (MS¹) used 120,000 resolving power, with an automatic gain control (AGC) target value of 5×10^5 ; MS² spectra were acquired at 60,000 resolving power, with target

AGC values of 1×10^6 . Four microscans were averaged for both MS¹ and MS². Precursor ions were selected using a 3 m/z isolation window. The dynamic exclusion option was used with a 60 s duration. Data acquisition was performed by applying a data-dependent top-2 method, with the two most abundant precursors selected for fragmentation.

Top-down MS revealed that the major functionalization product of Sp1m-2L (**3**) has a subunit mass of 13033.40 Da (Figure 4.12) which corresponds well with conjugation sites of C20 and *N*-terminus for linker 1 and linker 2, respectively (Table 4.3). A very minor product with a mass of 13451.57 Da was also observed, correlating to conjugation of linker 2 at both the *N*-terminus and K74, in addition to linker 1 at C20. However, a product with a mass suggesting modification of solely C20 and K74 was never observed. These data support that Sp1m is modified by both linker 1 and linker 2, almost exclusively, at the C20 and *N*-terminus sites, respectively.

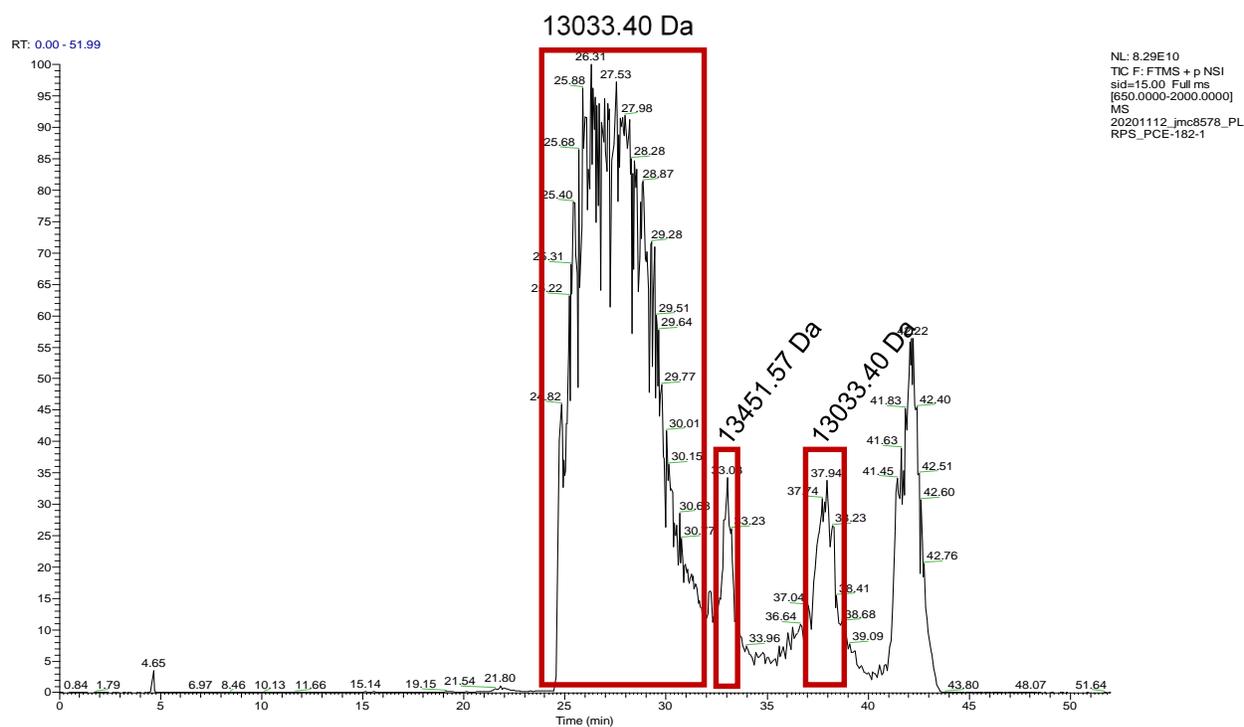


Figure 4.12 – Chromatogram from nanocapillary reverse-phase liquid chromatography of **3.** Masses determined by MS-MS are indicated, and tabulated in Table 4.3.

Retention time (min)	Observed mass (Da)	Linker attachment position ^a		Theoretical mass (Da) ^b
		Linker 1	Linker 2	
27 ^c	13033.40	C20	<i>N</i> -terminus	13033.86
33	13451.57	C20	<i>N</i> -terminus and K74	13451.59
38 ^c	13033.40	C20	<i>N</i> -terminus	13033.86

^a Attachment positions of linkers 1 and 2 on Sp1m-2L (**3**) that are consistent with observed mass. ^b Theoretical mass consistent with Sp1m-2L with linker attachment positions listed. In all cases, the *N*-terminal methionine residue has not been included. ^c An additional peak at 42 min has identical observed and theoretical masses to those at 27 and 38 min. The observation of multiple peaks corresponding to this species likely arises from overloading of the column, supported also by the large size of the peak at 27 min.

Table 4.3 – Observed and Theoretical Masses and Corresponding Linker Attachment Positions on Sp1m-2L.

The low reactivity of K74 may be explained by its local chemical environment (Figure 4.13). Closer inspection of the Sp1 crystal structure (PDB: 1TR0) (1) reveals a hydrogen bond network connecting K74 with multiple sidechains through water molecules. We hypothesize that these hydrogen bonds significantly reduce the nucleophilicity of the primary amine moiety on K74, resulting in negligible reactivity with NHS-activated esters.

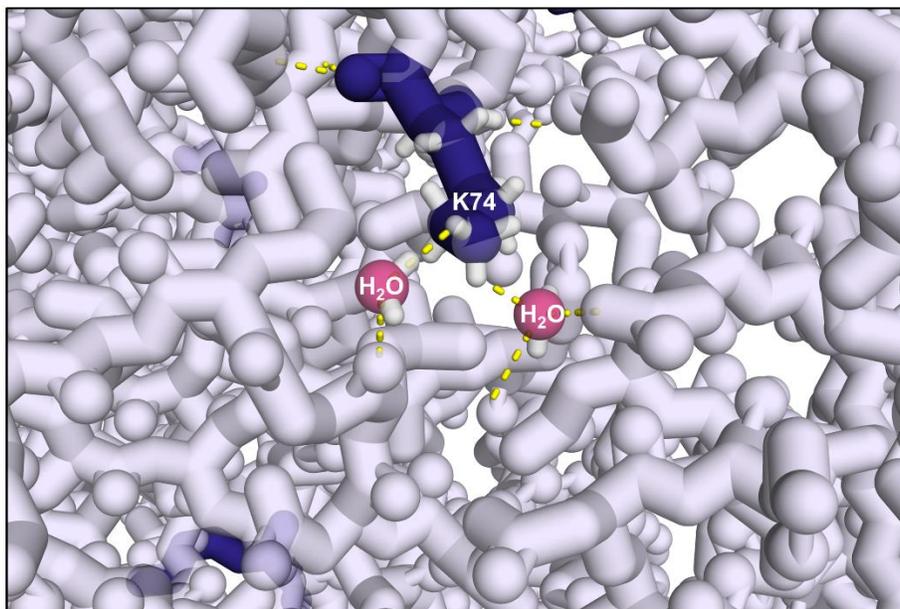


Figure 4.13 – Local chemical environment of K74 (purple) showing hydrogen bonds (depicted as yellow dashed lines) with adjacent amino acid residues and water molecules (pink). Protein coordinates taken from PDB: 1TR0 (1).

4.7.7 Donor-Quenching FRET Studies

Combinations of Sp1m-DNA conjugates at a total Cy3 concentration of 300 nM were mixed (1:1 ratio, 50 μ L) and placed in a 96-well plate, heated at 65 $^{\circ}$ C for 5 min and then cooled from 65 to 20 $^{\circ}$ C at 0.1 $^{\circ}$ C/0.5 min using a Bio-Rad CFX96 Touch real time PCR system. All samples were measured in triplicate, and the data reported represent the average of the three runs. Cy3 fluorescence was measured at 0.1 $^{\circ}$ C intervals.

Plots of fraction assembled vs temperature were obtained by measuring the fluorescence intensity (I) of two samples: a sample where the donor fluorophore (Cy3) is in the presence of a FRET acceptor (Cy5) (I_{DA}) and a sample where only the donor fluorophore (Cy3) is present (I_D). Comparing the fluorescence of both systems allows for the assembly-dependent FRET quenching of the donor to be distinguished from the inherent temperature-dependent change in fluorescence

of the donor. From the raw intensity profiles (Figures 4.14 and 4.15), the temperature-dependent FRET efficiency was determined as:

$$\text{FRET efficiency} = 1 - I_{DA}/I_D$$

Using the FRET efficiency, we defined “fraction assembled” by taking the maximum FRET ratio as fraction assembled = 1 and the minimum FRET ratio as fraction assembled = 0.²¹⁷ This method was used to generate all plots in Figures 4.3 to 4.5. Since each system has different assembly outcomes/end-points, the fraction assembled is defined independently for each system.

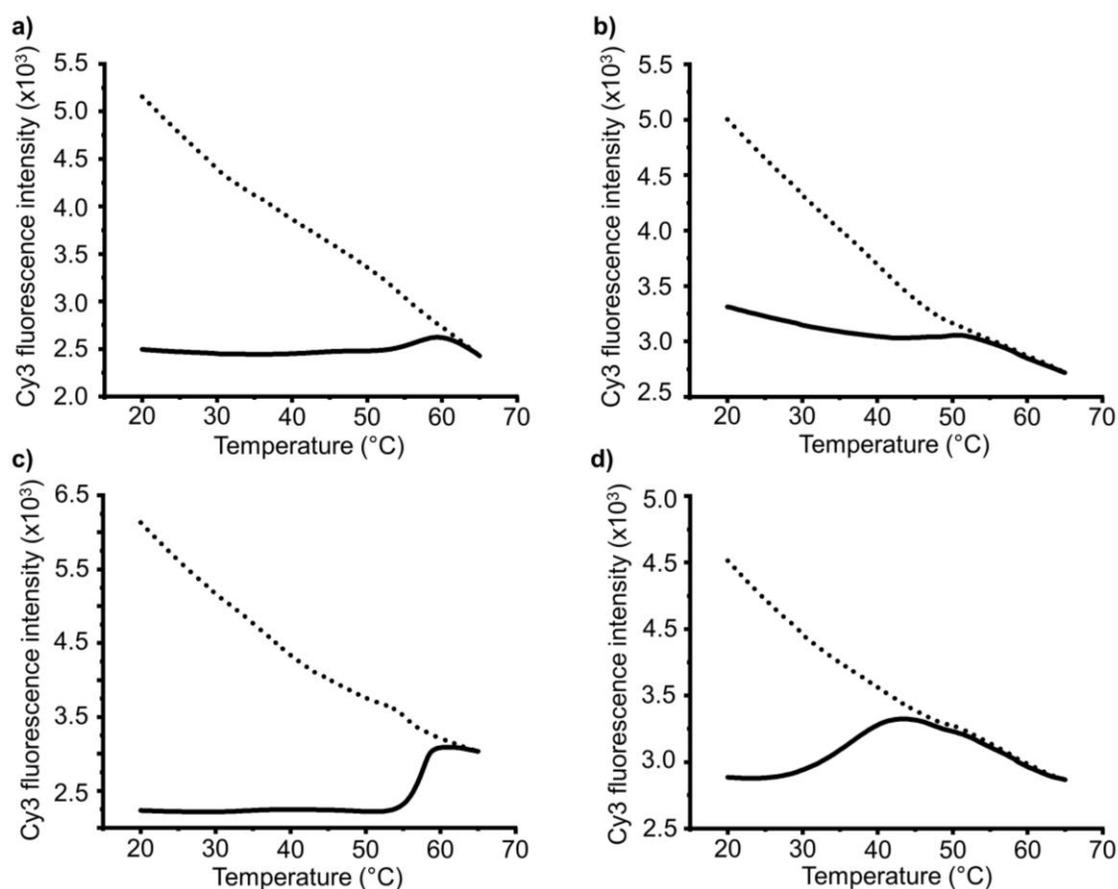


Figure 4.14 – Raw Cy3 intensity data used to determine fraction assembled vs temperature curves in Figures 4.3 and 4.4. (a) Sp1m-A_SENC + Sp1m-A'_SENC, (b) A_S + A'_S (DNA only), (c) Sp1m-E_S + Sp1m-E'_S, and (d) E_S + E'_S (DNA only). The solid lines show Cy3 intensities for Cy3 + Cy5 systems and the dotted lines show Cy3 intensities for Cy3-only systems.

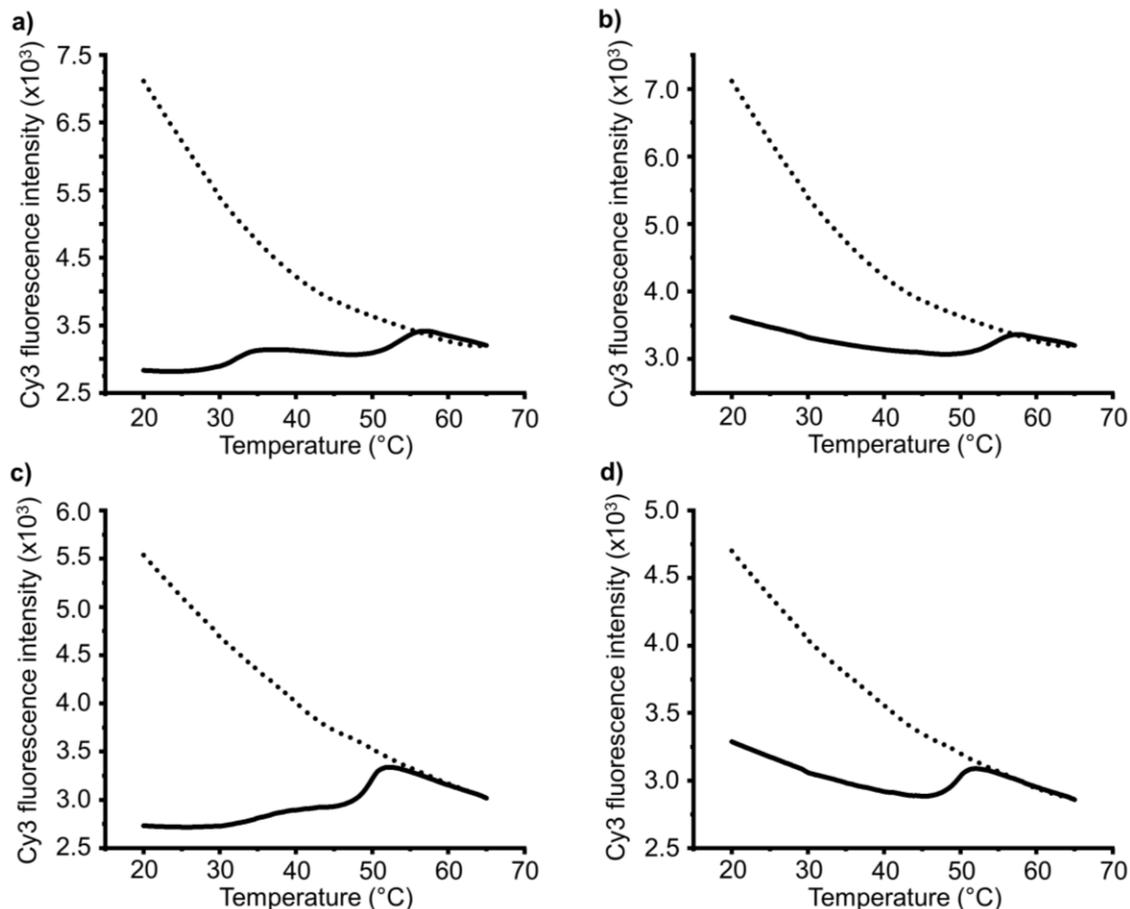


Figure 4.15 – Raw Cy3 intensity data used to determine fraction assembled vs temperature curves in Figure 4.5. (a) Sp1m-AsEw₁ + Sp1m-A'sEw₁, (b) Sp1m-AsE_{NC} + Sp1m-A'sE_{NC}, (c) Sp1m-AwE_s + Sp1m-AwE's, and (d) Sp1m-A_{NC}E_s + Sp1m-A_{NC}E's. The solid lines show Cy3 intensities for Cy3 + Cy5 systems and the dotted lines show Cy3 intensities for Cy3-only systems.

4.7.8 Melting Temperature and Full-Width Half Maximum of Assembly Transitions

The data from fraction assembled vs temperature plots (Figures 4.3 and 4.4) were fit with a sigmoidal curve using the “Sigmoidal Fit” function in OriginPro from which the 1st derivative was calculated (Figure 4.16, solid lines). The derivatized data were subsequently fit with a gaussian curve using the “Single Peak Fit” function in OriginPro (Figure 4.16, dotted lines). Melting temperatures (T_m) were taken as the peak of the fitted gaussian and the full-width half-maximum (FWHM) was also measured.

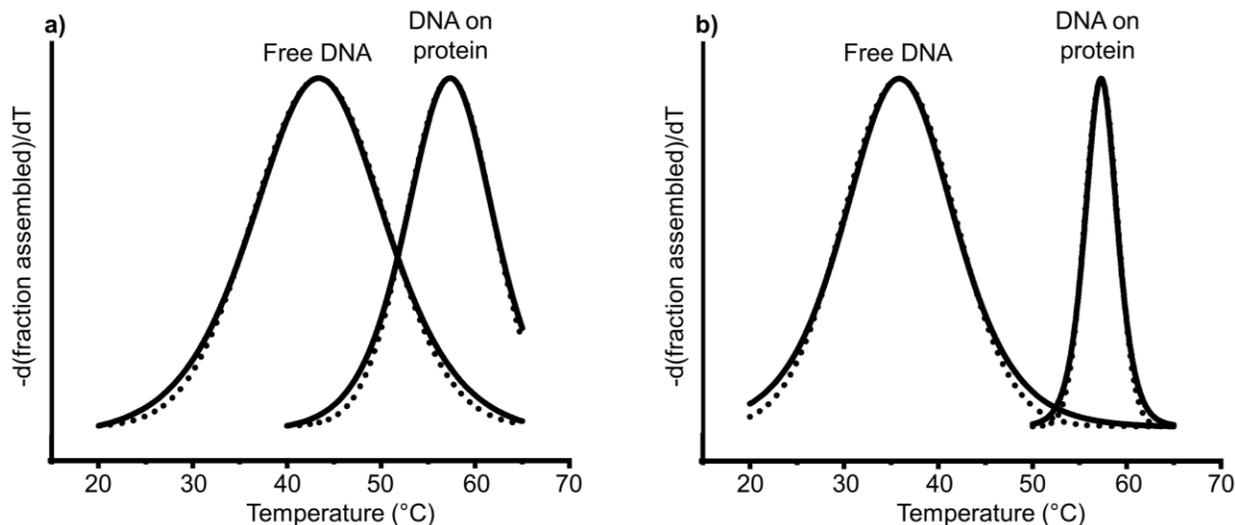


Figure 4.16 – Graphs showing first derivative of fraction assembled vs temperature (solid line) and fitted gaussian curves (dotted line) for (a) strong axial (A_S/A'_S) DNA system and (b) strong equatorial (E_S/E'_S) DNA system.

4.7.9 Transmission Electron Microscopy (TEM)

4 μL of slow-cooled sample (diluted to 100 nM if necessary) were adsorbed onto a glow-discharged carbon-coated Cu grid (Ted Pella) for 2 min. Excess liquid was wicked away by applying filter paper to the underside of the grid. A solution (4 μL) of either 2% uranyl acetate or 0.75% uranyl formate stain (Electron Microscopy Solutions) was applied for 1 min. The sample was allowed to air dry for 10 min after wicking away excess stain. Images were collected on a JEOL 1230 transmission electron microscope at 100 or 120 kV accelerating voltage.

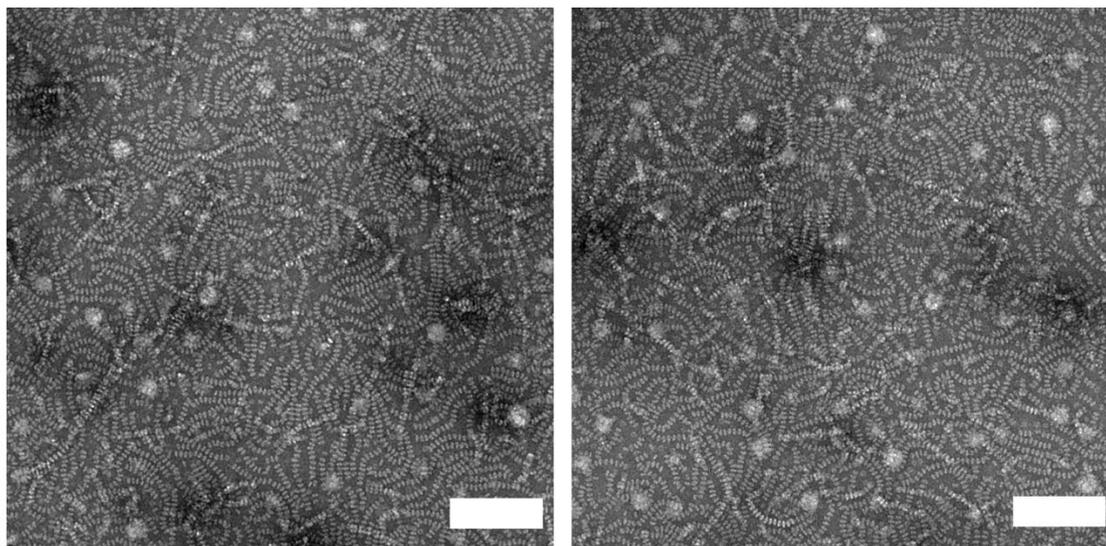


Figure 4.17 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-As_{ENC} and Sp1m-A's_{ENC}. Scale bars are 150 nm.

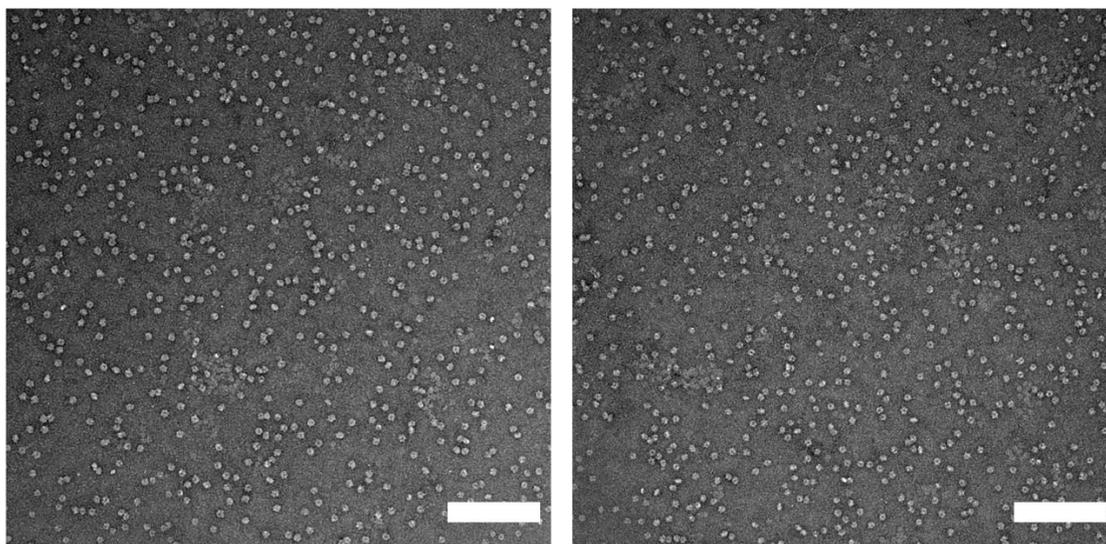


Figure 4.18 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-As_{ENC}. Scale bars are 150 nm.

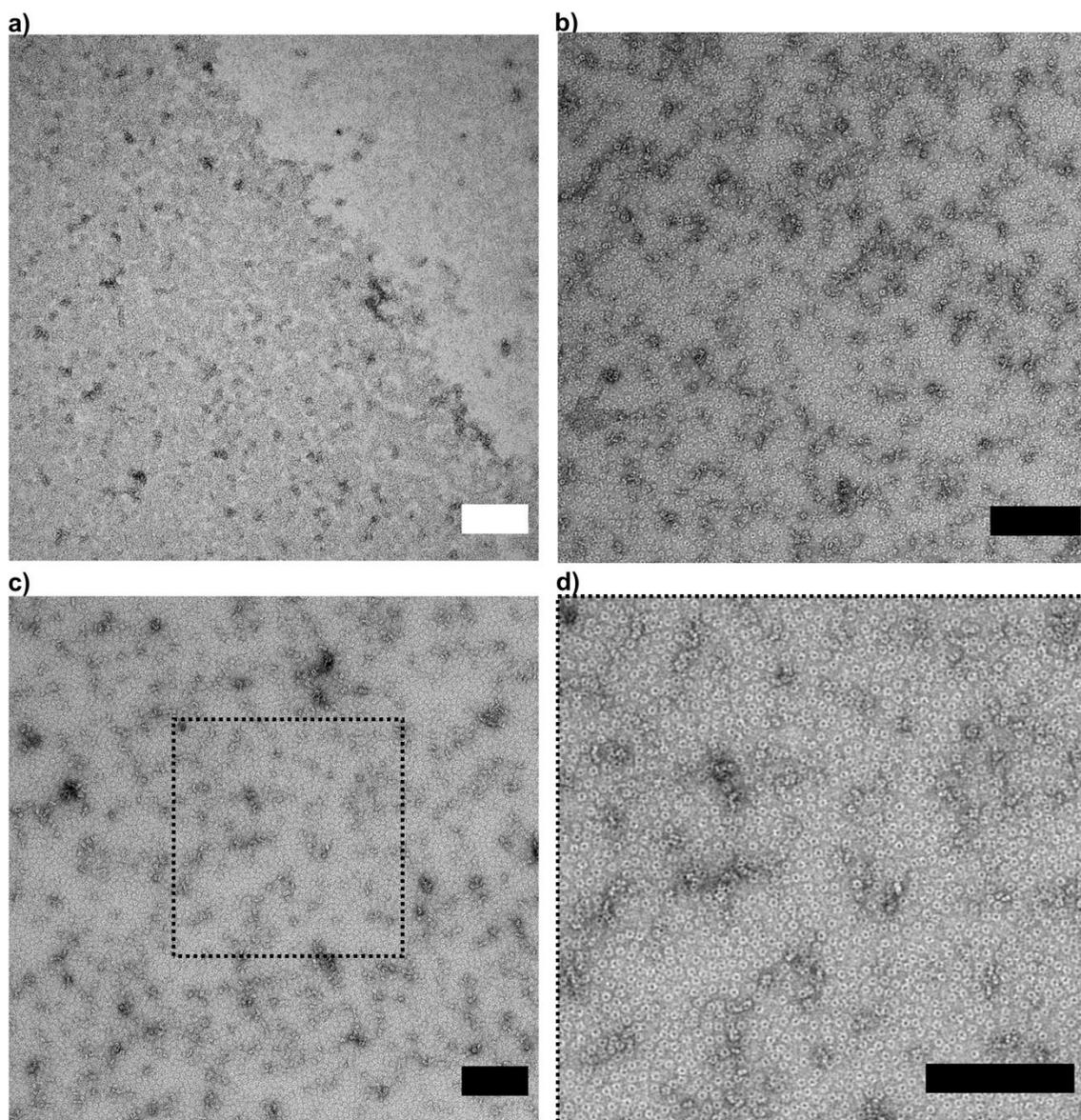


Figure 4.19 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-Es and Sp1m-E's. (a–c) Micrographs showing the sample imaged at different locations and magnifications. (d) An expanded view of the outlined area in (c). Scale bars are 150 nm.

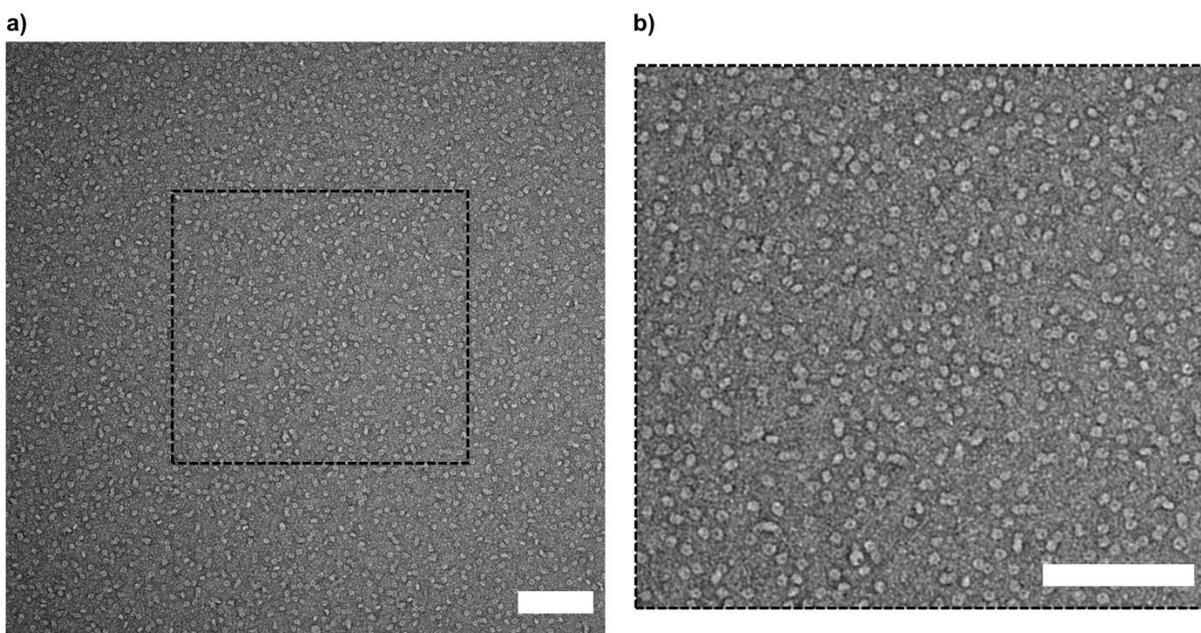


Figure 4.20 – Representative negative-stain TEM micrographs of slow-cooled Sp1m-Es. (a) Micrograph showing a wide-field image of the sample. (b) An expanded view of outlined area in (a). Scale bars are 150 nm.

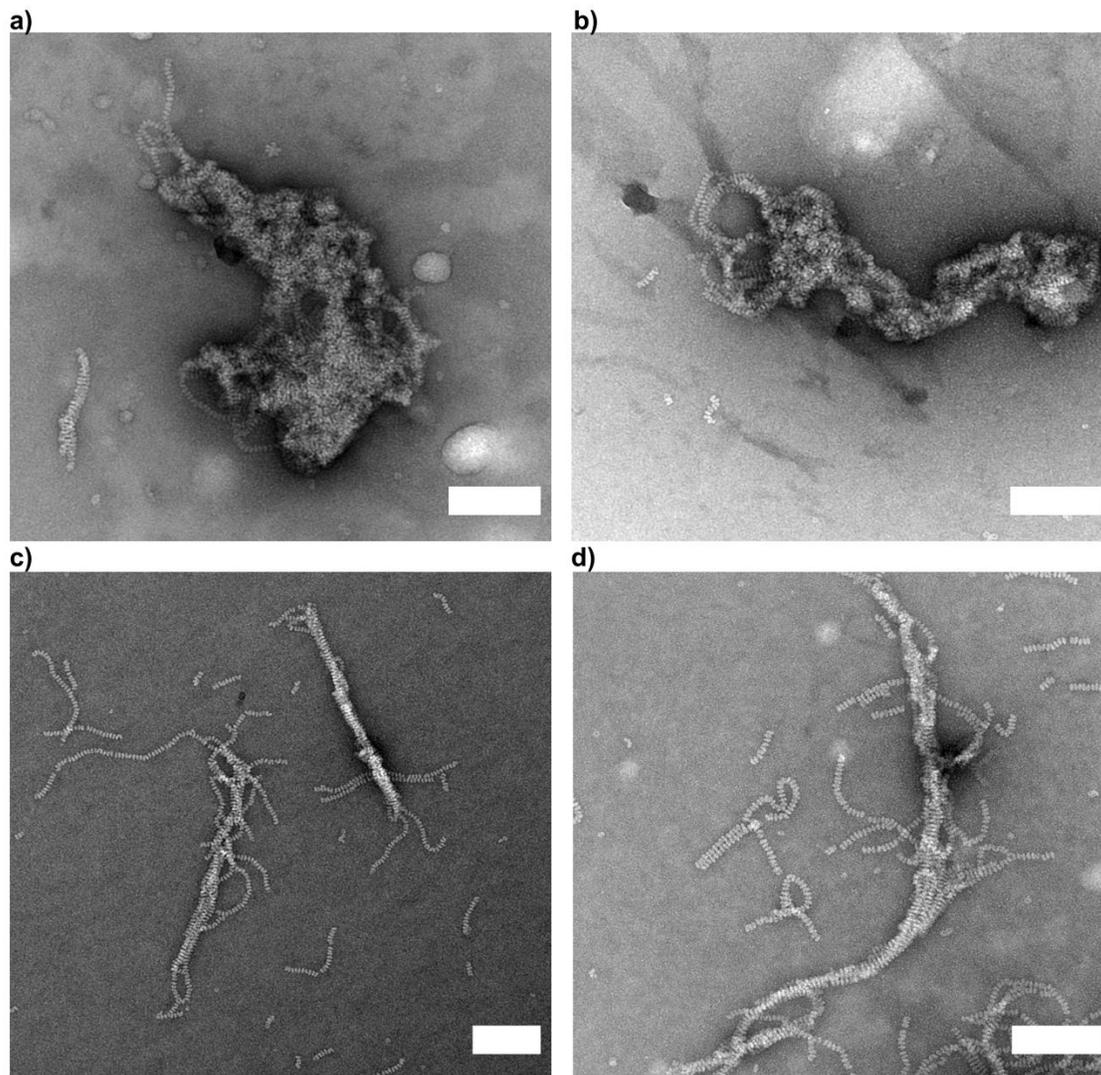


Figure 4.21 – TEM micrographs of structures tuned by equatorial DNA design. (a and b) Representative negative-stain TEM micrographs of slow-cooled Sp1m-A_SE_W1 and Sp1m-A'_SE_W1. (c and d) Representative negative-stain TEM micrographs of slow-cooled Sp1m-A_SE_W1 and Sp1m-A'_SE_W2. Scale bars are 150 nm.

For cryogenic-EM, 4 μ L of 500 nM sample were deposited on a glow-discharged lacey carbon-coated grid (Ted Pella) and plunge-frozen using a FEI Vitrobot Mark IV using a blot time of 5 s at 10 °C and high humidity. Images were collected on a Hitachi HT-7700 Biological S/TEM at 100 kV accelerating voltage.

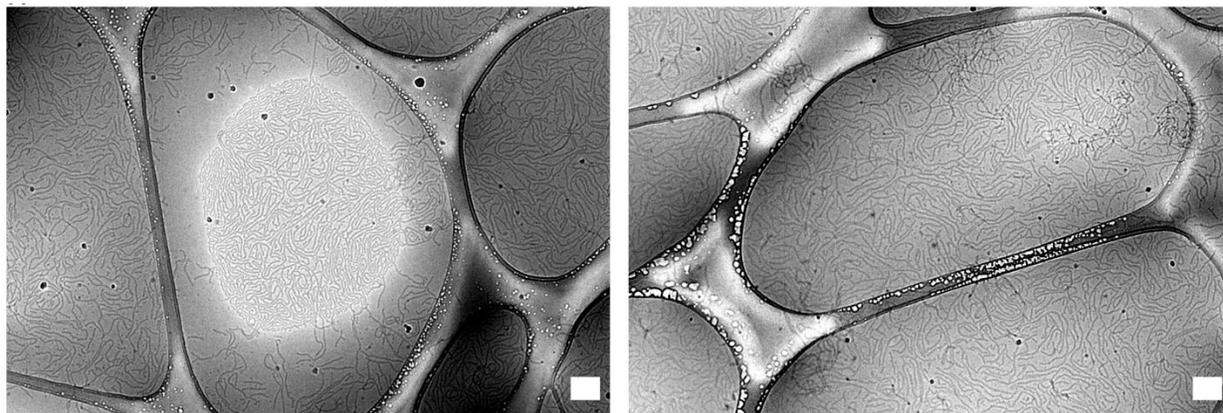


Figure 4.22 – Representative cryo-TEM micrographs of slow cooled Sp1m-AsENC and Sp1m-A'sENC. Scale bars are 150 nm.

4.7.10 Atomic Force Microscopy

5 μL of 500 nM sample were deposited on a freshly cleaved mica substrate. 10 μL of buffer (10 mM MgCl_2 , 20 mM HEPES pH 7.4) was added to the substrate and the sample was left to incubate overnight in a high humidity environment to minimize evaporation. All AFM images were captured in ScanAsyst PeakForce Tapping mode on a BioScope Resolve AFM (Bruker) using a SCANASYST-FLUID+ probe. The effective imaging force ranged from 100 to 200 pN, within the typical force range for AFM imaging of biomolecules.

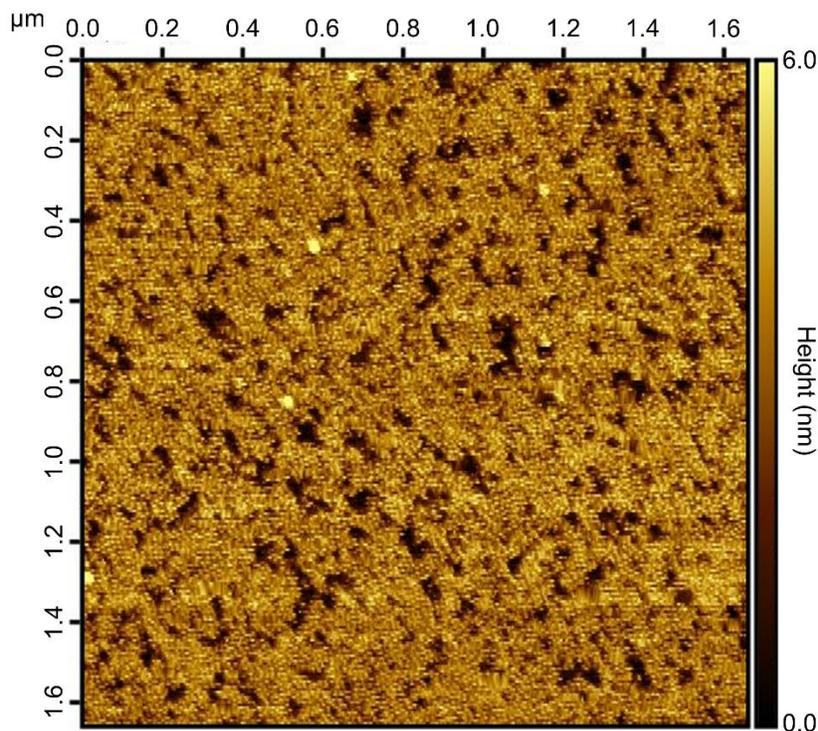


Figure 4.23 – AFM micrograph of slow-cooled Sp1m-ES and Sp1m-E'S. Reveals a monolayer two-dimensional protein assembly over a large ($1.6 \times 1.6 \mu\text{m}$) area.

AFM was used to investigate the structures resulting from the hierarchical assembly of Sp1m-AwE_S and Sp1m-AwE'_S (Figure 4.5 d to f). Having shown that the strong equatorial interactions drive the assembly of 2D protein sheets, we hypothesized that subsequent axial interactions between sheets would lead to multi-layered terraced structures. The measurement of height profiles of AFM micrographs show quantized height increments of 6 nm (Figure 4.24), suggesting a mixture of one- and two-layered materials.

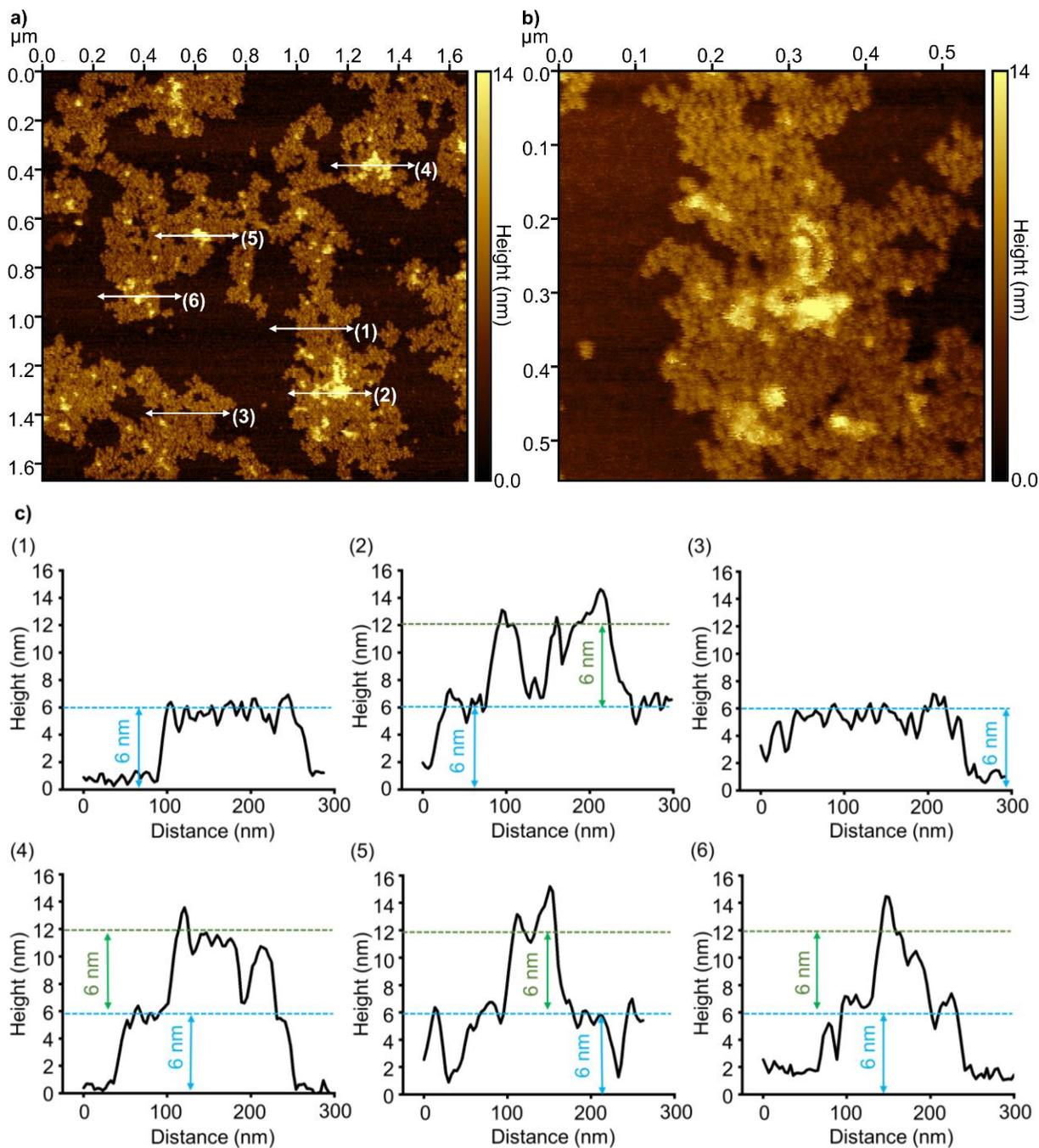


Figure 4.24 – AFM characterization of hierarchical structures. (a and b) AFM micrographs of slow-cooled Sp1m-AwEs and Sp1m-AwE's reveal large area two-dimensional protein assembly containing areas of different heights. (c) Height profiles measured along white lines indicated in (a) revealing quantized layer heights, with increments measuring 6 nm.

4.7.11 Effect of Salt Concentration on Hierarchical Assembly of Sp1m-AsEw and Sp1m-A'sEw

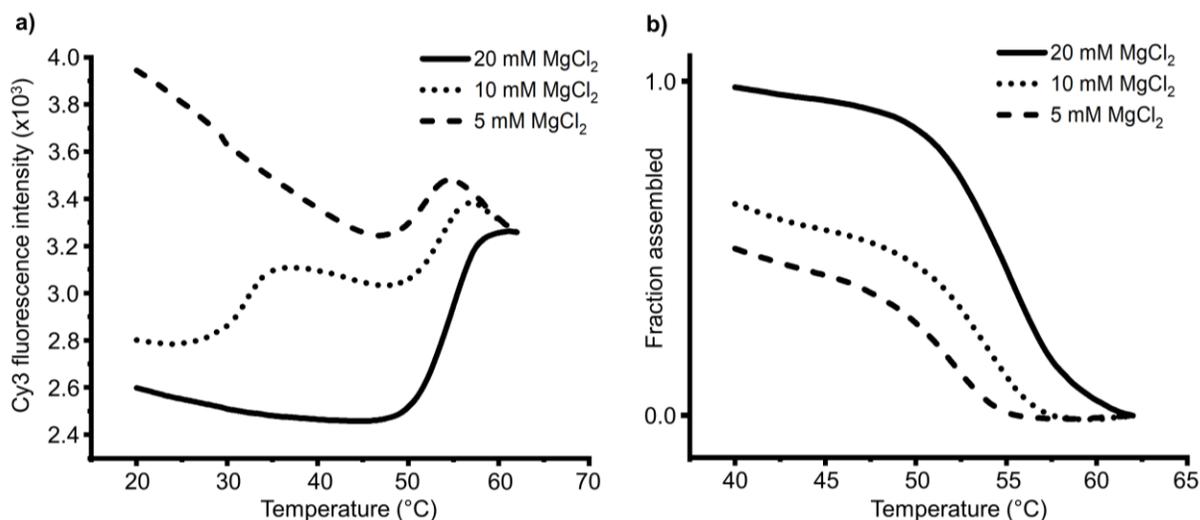


Figure 4.25 – Graphs showing the influence of salt concentration on the assembly of Sp1m-AsEw and Sp1m-A'sEw. (a) Raw Cy3 intensity data of dye-labelled building blocks measured at different salt concentrations. (b) Fraction assembled vs temperature data normalized to greatest fraction assembled (20 mM MgCl₂) to show relative fraction assembled as a function of salt concentration.

As noted in the main text (section 4.4), changing the concentration of MgCl₂ from 10 mM (Figure 4.5) to 20 mM or 5 mM eliminates the observation of a transition at 32.7 °C and yield assembly profiles with a single transition at 52.0 and 55.2 °C, respectively (Figure 4.25a). However, these profiles represent two distinct assembly processes, whereby either both axial and equatorial assembly steps become concerted (20 mM) or only the first (axial) stage occurs (5 mM), supported by the relative fractions assembled (Figure 4.25b). This behavior arises because varying the concentration of MgCl₂ leads to changes in both the *absolute* and *relative* melting temperatures of the two stages of assembly, such that the melting temperature associated with the second (equatorial) assembly step is affected by changes in [MgCl₂] than the first (axial) step. We hypothesize that this may be because in the second stage of assembly, larger structures possessing

a higher density of negative charge (DNA) are influenced more by the charge screening effects of salt than the smaller Sp1-DNA monomers that assemble in the first assembly step. These results suggest that ionic strength/salt identity could provide an additional design parameter to direct hierarchical assembly pathways in DNA-mediated systems.

CHAPTER FIVE

Conclusions and Outlook

5.1 Conclusions

This thesis teaches new methods to program the assembly pathways of proteins using DNA through judicious design of DNA sequences and protein-DNA conjugates. The lessons include: (1) how energy barriers to association can be tuned by DNA sequence and secondary structure design; (2) how multiple, orthogonal DNA interactions can be programmed through the design of a protein's amino acid sequence; and (3) how hierarchical assembly pathways can be encoded through a combination of both DNA and amino acid sequence design. This work takes advantage of the vast design space of oligonucleotides as programmable bonding material and our findings emphasize the impact of controlling the assembly pathway of building blocks to specify desired architectural outcomes.

In chapter two we chose to study a simple system: a monomeric protein with a single oligonucleotide surface ligand, where the oligo can drive the polymerization of proteins. The aim was to manipulate the secondary structure of the oligonucleotide through sequence design to modulate the energy barriers to protein association. We successfully demonstrated that a specific polymerization pathway, step-growth or chain-growth, could be designed for a given system as a function of DNA sequence and secondary structure. These distinct pathways enabled the synthesis of different product distributions: a mixture of cyclic and linear polymers or exclusively linear polymers for step-growth and chain-growth pathways, respectively. Moreover, we were able to confirm the "living" nature of protein polymers formed via the chain-growth pathway through chain-extension experiments. Importantly, "living" chain-ends will allow for the synthesis of polymer architectures with greater complexity and structural diversity, for example: brush, block

and star. This work demonstrates a level of control currently not possible with other chemical techniques.

Chapter three provides important insight into how proteins displaying multiple, orthogonal DNA interactions can be designed and the types of materials that arise from the assembly of such building blocks. We wanted to explore the idea of encoding multiple interaction “patches” on the surface of proteins, akin native protein surfaces, to direct the assembly of multicomponent systems. Methods to synthesize protein-based Janus particles were developed and the assembly of these building blocks were interrogated in the context of colloidal crystallization. We found that the unique, asymmetric DNA ligand design of these Janus particles led to the assembly of unique, layered colloidal crystals, previously inaccessible using PAEs with isotropic DNA functionalization. These results provide a blueprint for designing novel protein-based PAEs, with highly designed and specific ligand shells encoded with multiple interactions, that will enable the synthesis of materials that rival and exceed the complexity of those found in Nature.

In chapter four, the concept of programming hierarchical assembly pathways of proteins through DNA sequence design and defining interaction geometry (axial and equatorial) is explored. This work represents an important conceptual advance in the field, since only single-step assembly pathways of protein-DNA conjugates had been previously investigated. Here we demonstrate the pathway can be controlled and designed using DNA interactions encoded on the surface of a protein building block. Importantly, all the information required to determine the assembly multi-step pathway is contained within the building blocks. Through DNA design, we can change the directionality of protein assembly and pathway by which protein-DNA conjugates will assemble, as well as realize distinct structures by directing assembly along different pathways. These findings will facilitate

the assembly of an entirely new class of protein-DNA materials with structures possessing defined levels of hierarchy.

The field of DNA-mediated assembly continues to evolve methods that enable the elegant and sophisticated supramolecular, bottom-up assembly of structures and systems from nanoscale building blocks. However, for the field to enter a new paradigm and to realize functional materials with complex architectures, it is not simply just interactions that need to be programmed, but additionally the assembly pathway that building blocks traverse to access specific architectures must also be designed and controlled. Overall, the body of work presented here describes some key steps towards this challenge and will provide blueprints that will map the development of new strategies to control the assembly pathways of proteins, and other nanoscale building blocks.

5.2 Outlook

5.2.1 Summary and future steps

The assembly pathway of supramolecular materials is critical to determining their architectures and functions. This is intrinsic to the way protein-based material and systems are formed in Nature and is the major theme that underscores this thesis. While we have established key tools for programming assembly pathways of protein-DNA conjugates, future work is needed to fully understand and finely control these assembly mechanisms. This will greatly enhance the precision of materials produced and will likely widen the generalizability and scope of these methods.

An important class of materials not directly discussed in the body of work presented here, but one that presents many opportunities for future study, is protein single crystals. Protein single crystals represent materials with the greatest precision in positional, orientational, and translational

order and are amenable to X-ray analysis to provide important angstrom-level structural information. The large size and conformational flexibility of most proteins makes their crystallization challenging. Crystallization relies on a protein finding the optimal combination of competing interactions with other proteins, solvent, and salts, in a symmetry-related way, to reach an equilibrium structure. The sheer complexity and chemical inhomogeneity of a protein surface means that this process is slow or, for particularly large and flexible proteins, may not occur at all. Having demonstrated the power of high-fidelity, relatively simplistic, and programmable DNA interactions in directing protein association, throughout this thesis, a natural next step is to apply such methods to solve challenges in the field of protein crystallography.

5.2.2 Leveraging symmetry and valency control to enhance protein crystallization

Protein crystallography provides valuable, angstrom-level-resolution, structural insight into the macromolecules that engender the infrastructure of life. The information gained through solving protein crystal structures has provided fundamental understanding of their functions, informed the design of therapeutics and has been recognized by multiple Nobel prizes.²²⁰ Among the many properties of proteins that render routine preparation of diffraction-quality crystals a challenge, their inherent enantiopurity is particularly difficult to overcome. As chiral molecules, naturally occurring proteins can only crystallize in the 65 non-centrosymmetric space groups; the other 165 space groups are centrosymmetric and require either that the protein itself is centrosymmetric or that the protein and its enantiomer exist in equal numbers in the crystal.²²¹ This is impossible or impractical, respectively, for all but the smallest oligopeptides. Fortunately, the likelihood for a protein to crystallize may be improved by controlling and promoting the formation of intermolecular contacts that are compatible with crystal symmetry. A promising

future direction is the development of symmetric DNA-based templates control protein oligomerization and enhance their propensity for crystallization through synthetic symmetrisation.

Pioneered by the Yeates lab, synthetic symmetrisation utilizes surface-associated motifs that drive symmetric interactions between proteins to promote their crystallization due to a reduction in the number of unique fortuitous crystal contacts.²²² This concept is supported by the observation that protein oligomers typically crystallize in space groups that support the point group of the oligomer.²²² To this end, several approaches have been taken, including biosynthetic protein fusion,²²³ metal-mediated coordination chemistry,²²⁴ cross-linking,²²⁵ and protein-scaffold synthetic symmetrisation.²²⁶ However, these methods are not widely generalizable and lack the ability to systematically program a wide variety of symmetries and valences. In contrast, DNA surface ligands provide a powerful way to direct the oligomerization and assembly of proteins with programmable interaction strength and length. In future research, DNA could be utilized to control the symmetry and valency of protein clusters to enhance their propensity towards crystallization by reducing the number of unique fortuitous crystal contacts and be able to systematically explore the effects of valency and symmetry. This proposed approach will address the shortcomings in the field and dramatically expand synthetic control over the crystallization process and outcome.

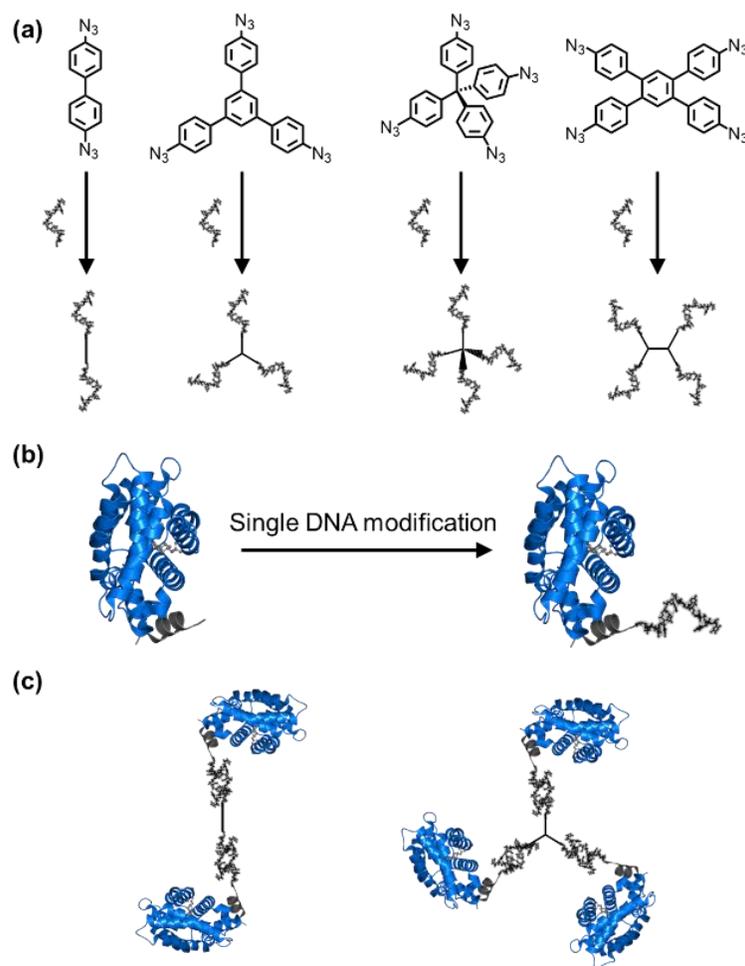


Figure 5.1 – Preparation of protein clusters with defined valency. (a) Functionalization of small molecule templates with DNA. (b) Modification of a protein of interest with DNA. (c) Assembly of protein-DNA conjugates with DNA templates will yield discrete, symmetric protein oligomers.

Utilizing the recognition and programmability of DNA, in concert with the defined symmetry of small molecules, it is possible to synthesize a series of DNA templates (**Figure 5.1a**). These templates would serve as directing structures that induce a specific oligomerization state of protein-DNA conjugates (e.g., dimeric, trimeric, tetrameric). The chemistry required for the synthesis of such DNA templates has previously been developed in our group²²⁷ and will provide access to a library of templates. Briefly, azide-containing small molecules will be reacted with dibenzocyclooctyne (DBCO)-terminated DNA on a solid-phase support and the products will be

purified using HPLC or PAGE. DNA templates will be characterized by PAGE and MALDI-TOF-MS. Protein-DNA conjugates (**Figure 5.1b**) will be synthesized using established bioconjugation techniques that target a specific amino acid residue such as cysteine, lysine, or N- and C-termini.²²⁸ Purification of these protein-DNA conjugates will be achieved using size exclusion chromatography (SEC), PAGE, and liquid chromatography time of flight (LC-TOF) mass spectrometry. Upon hybridization of the protein-DNA conjugates with a given DNA template, a protein cluster will be synthesized with defined valency and symmetry (**Figure 5.1c**). These clusters will be thoroughly characterized by PAGE, multi-angle light scattering (SEC-MALS), and transmission electron microscopy (TEM).

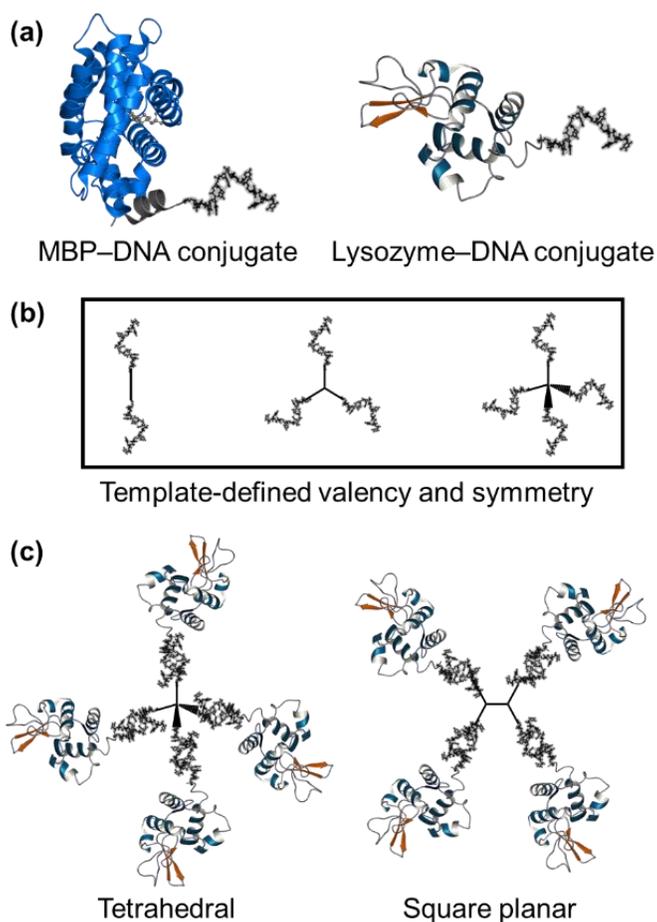


Figure 5.2 – Preparation of protein clusters from monomeric proteins. (a) Protein-DNA conjugates. (b) DNA templates. (c) Protein clusters with identical valency but distinct symmetries.

Next, crystallization of two naturally monomeric proteins, for example, maltose binding protein (MBP) and lysozyme, could be investigated, as their well-known crystal structures provide a point of comparison to determine the influence of oligomerization on crystallization. Crystallization trials can be performed using sitting drop, vapor diffusion method to screen for conditions that lead to formation of high-quality protein crystals. Commercially available screens that assess a broad scope of conditions organized in a 96-well format (100 μ l reservoir) along with hanging drops (1 μ l, 50:50 protein:reservoir) can be used and readily prepared by a Mosquito pipetting robot. It should be possible to determine the optimal protein cluster crystallization conditions by investigating the impact of forming defined protein clusters prior to crystallization or allowing protein-DNA conjugates to associate with the template transiently during crystallization. The programmability of DNA allows for the design of the template DNA to be modulated to tune the interaction strength and distance between protein and template. If different crystal structures are observed from these two initial states, we may obtain valuable insights into the favourability of crystallization in different oligomerization states.

While it has been shown that dimerization of naturally monomeric proteins leads to new crystal structure forms,²²⁵ a systematic study of valency has not been conducted. The hypothesis that the number of unique fortuitous crystal contacts will decrease as valency increases, owing to the increased symmetry with each increase in valence can be systematically explored. Comparing the space group and crystal contacts of known crystal structures of monomeric proteins (MBP and lysozyme, **Figure 5.2a**) with the novel structures of higher-order oligomeric, symmetric proteins will reveal how subsequent increase in valence (**Figure 5.2b**) influences crystallization. This

project could also investigate if the symmetry of a protein cluster influences crystallization, independent of valency. For this, two tetrameric protein structures using either a tetrahedral or square-planar DNA template (**Figure 5.2c**) should be synthesized. These two geometries differ with respect to their point groups (T_d and D_{4h} , respectively) which could impact both the packing and crystal contacts that arise in a given crystal structure. To ensure the fidelity of the programmed geometry, short, double-stranded DNA can be used and, if necessary, rigidified further by incorporating locked nucleic acids.²²⁹ Molecular cores based on larger constructs (e.g. polysiloxane cages)²³⁰ may also be used to emphasize the desired geometry. If clusters with both geometries give rise to the formation of single crystals, the interfaces that form in these crystals can be compared using PDBEPIA. These experiments will reveal how geometry and symmetry impact crystallization.

Once an understanding of how programming valency and symmetry of protein clusters impacts crystallization is established, these principles can be applied to enhance crystal quality and a protein's propensity to crystallize by choosing a difficult-to-crystallize protein or a protein with an unknown structure and applying the DNA-templated approach to induce oligomerization and crystallization. For example, PTOV1 is an approx. 400 amino acid protein overexpressed in prostate cancer that has eluded crystal structure determination for almost 20 years.²³¹ These types of proteins serve as excellent candidates for structure elucidation. Performing a qualitative and quantitative analysis of crystals grown using these methods versus those grown without intervention will be important to determine differences in average size, number, and crystal habit. These parameters are critical indicators of a protein's propensity to crystallize when assessed over a broad range of conditions.

This future research area harnesses the exquisite programmability of DNA interactions to organize protein's into advantageous geometries for crystallization and enables the systematic exploration of synthetic symmetrisation affords. Importantly, this work builds on fundamental observations and early demonstrations that symmetry is an advantage in protein crystallization and pushes this concept into a regime where we can assess precisely how valency, symmetry and, to some extent, geometry impacts protein crystallization. The findings will produce key lessons which can be applied to unknown or difficult-to-crystallize proteins to to widen the bottleneck that currently exists in protein crystallography and expand our fundamental understanding of crystallization processes.

REFERENCES

- (1) Crick, F. H. *Symp Soc Exp Biol.* **1958**, *12*, 138-163.
- (2) Berg, P. *Annu. Rev. Biochem.* **1961**, *30*, 293–324.
- (3) Baker, D. A. *Nature.* **2000**, *405*, 39–42.
- (4) Gabor Miklos, G. L.; Maleszka, R. *Proteomics.* **2001**, *1*, 169–178.
- (5) Bernal, J. D.; Crowfoot, D. *Nature.* **1934**, *133*, 794–795.
- (6) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature.* **1958**, *181*, 662-666.
- (7) Brooks-Bartlett, J. C.; Garman, E. F. *Interdiscip. Sci. Rev.* **2015**, *40*, 244–264.
- (8) Mast, N.; White, M. A.; Bjorkhem, I.; Johnson E. F.; Stout, D; Pikuleva, I. A. *Proc. Natl Acad. Sci. USA.* **2008**, *105*, 9546–9551.
- (9) Sun, F.; Huo X.; Zhai, Y.; Wang, A.; Xu, J.; Su, D.; Bartlam, M.; Rao, Z. *Cell.* **2005**, *121*, 1043–1057.
- (10) Aldaz, H.; Rice, L. M.; Stearns, T.; Agard, D. A. *Nature.* **2005**, *435*, 523–527.
- (11) Waugh, D. F. *Adv. Protein Chem.* **1954**, *9*, 325–437.
- (12) Sun, T. L.; Zhou, B.; Lai, L. H.; Pei, J. F. *Bmc. Bioinformatics.* **2017**, *18*, 277.
- (13) Ding, Z.; Kihara, D. *Sci. Rep.* **2019**, *9*, 1–13.
- (14) Berggård, T.; Linse, S.; James, P. *Proteomics.* **2007**, *7*, 2833–2842.
- (15) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. *Sig. Transduct. Target. Ther.* **2020**, *5*, 213.
- (16) Wagstaff, J.; Löwe, J. *Nat. Rev. Microbiol.* **2018**, *16*, 187–201.
- (17) Pieters, B. J. G. E.; van Eldijk, M. B.; Nolte, R. J. M.; Mecinovic, J. *Chem. Soc. Rev.* **2016**, *45*, 24-39.
- (18) Xu, G., Gong, L., Yang, Z. & Liu, X. Y. *Soft Matter.* **2014**, *10*, 2116–2123.

- (19) Prasad, B. V.; Schmid, M. F. *Adv. Exp. Med. Biol.* **2012**, 726, 17–47.
- (20) Rodis, P.; Hoff, J. E. *Plant Physiol.* **1984**, 74, 907–911.
- (21) Domingo, R.; Holmes, K. C. *Annu. Rev. Biophys.* **2011**, 40, 169–186.
- (22) Bearer, E. L. *Am. J. Respir. Cell Mol. Biol.* **1993**, 8, 582–591.
- (23) Pollard, T. D.; Cooper, J. A. *Science.* **2009**, 326, 1208–1212.
- (24) Rayment, I.; Holden, H. M.; Whittaker, M.; Yohn, C. B.; Lorenz, M.; Holmes, K. C.; Milligan, R. A. *Science.* **1993**, 261, 58–65.
- (25) Chou, S. Z.; Pollard, T. D. *Proc. Natl Acad. Sci. USA.* **2019**, 116, 4265–4274.
- (26) Cooper, G. M. *Structure and Organization of Actin Filaments* [Online]. In *The Cell: A Molecular Approach*, 2nd ed.; Sinauer Associates: Sunderland; **2000**; <https://www.ncbi.nlm.nih.gov/books/NBK9908/>.
- (27) Sara, M.; Sleytr, U. B. *J. Bacteriol.* **2000**, 182, 859–868.
- (28) Sleytr, U. B.; Beveridge, T. J. *Trends Microbiol.* **1995**, 7, 253–260.
- (29) Fagan, R.; Fairweather, N. *Nat. Rev. Microbiol.* **2014**, 12, 211–222.
- (30) Aumiller, Jr. W. M.; Uchida, M.; Douglas, T. *Chem. Soc. Rev.* **2018**, 47, 3433–3469.
- (31) Rochat, R. H; Liu, X.; Murata, K.; Nagayama, K.; Rixon, F. J.; Chiu, W. *J. Virol.* **2011**, 85, 1871–1874.
- (32) Suttle, C. *Nature.* **2005**, 437, 356–361.
- (33) Harrison, P. M.; Arosio, P. *Biochim. Biophys. Acta-Bioenerg.* **1996**, 1275, 161–203.
- (34) Li, H.; DeRosier, D.; Nicholson, W.; Nogales, E.; Downing, K. *Structure.* **2002**, 10, 1317–1328.
- (35) Hess, H. & Ross, J. L. *Chem. Soc. Rev.* **2017**, 46, 5570–5587.
- (36) Nogales, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, 30, 397–420.
- (37) Roll-Mecak, A.; *Dev. Cell.* **2020**, 54, 7–20.
- (38) Zwetsloot, A. J.; Tut, G.; Straube, A. *Essays Biochem.* **2018**, 62, 725–735

- (39) Schek, H. T.; Gardner, M. K.; Cheng, J.; Odde, D. J.; Hunt, A. J. *Curr. Biol.* **2007**, *17*, 1445–1455.
- (40) Heim, M.; Romer, L.; Scheibel, T.; *Chem. Soc. Rev.*, **2010**, *39*, 156–164.
- (41) Fratzl, P.; Weinkamer, R. *Prog. Mater. Sci.* **2007**, *52*, 1263–1334.
- (42) Goelet, P.; Lomonosoff, G. P.; Butler, P. J. G.; Akam, M. E.; Gait, M. J.; Karn, J. *Proc. Natl Acad. Sci. U. S. A.* **1982**, *79*, 5818 —5822
- (43) Fraenkel-Conrat, H.; Williams, R. C.; *Proc. Natl Acad. Sci. USA.* **1955**, *41*, 690–698.
- (44) Eber, F. J.; Eiben, S.; Jeske, H.; Wege, C. *Nanoscale.* **2015**, *7*, 344–355.
- (45) Greising, S. M.; Gransee, H. M.; Mantilla, C. B.; Sieck, G. C. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2012**, *4*, 457–473.
- (46) Frontera, W. R.; Ochala, J. *Calcified Tissue Int.* **2015**, *96*, 183–195.
- (47) Lite, T.-L. V.; Grant, R. A.; Necedal, I.; Littlehale, M. L.; Guo, M. S.; Laub, M. T. *eLife*, **2020**, *9*, e60924.
- (48) Sian-Yang, O.; Dunstan, D. E. *Protein Sci.* **2014**, *23*, 1315–1331.
- (49) Kitagishi, H.; Oohora, K.; Yamaguchi, H.; Sato, H.; Matsuo, T.; Harada, A.; Hayashi, T. *J. Am. Chem. Soc.* **2007**, *129*, 10326–10327.
- (50) Ringler, P.; Schulz, G. E. *Science.* **2003**, *302*, 106–109.
- (51) Dotan, N.; Arad, D.; Frolow, F.; Freeman, A. *Angew. Chem. Int. Ed.* **1999**, *38*, 2363–2366.
- (52) Bai, Y.; Luo, Q.; Liu, J. *Chem. Soc. Rev.* **2016**, *45*, 2756–2767.
- (53) Zhang, L.; Wu, Y-W.; Brunsveld, L. *Angew. Chem., Int. Ed.*, **2007**, *46*, 1798–1802.
- (54) Uhlenheuer, D. A.; Young, J. F.; Nguyen, H. D.; Scheepstra, M.; Brunsveld, L. *Chem. Commun.*, **2011**, *47*, 6798–6800.
- (55) Nguyen, H. D.; Dang, D. T.; van Dongen J. L.; Brunsveld, L. *Angew. Chem., Int. Ed.*, **2010**, *49*, 895–898.
- (56) Uhlenheuer, D. A.; Milroy, L.-G; Neiryneck, P.; Brunsveld, L. *J. Mater. Chem.* **2011**, *21*, 18919–18922.

- (57) Hou, C. X.; Li, J. X.; Zhao, L. L.; Zhang, W.; Luo, Q.; Dong, Z. Y.; Xu, J. Y.; Liu, J. Q. *Angew. Chem., Int. Ed.* **2013**, *52*, 5590–5593.
- (58) Tainer, J. A.; Roberts, V. A.; Getzoff, E. D. *Curr. Opin. Biotech.* **1992**, *3*, 378–387.
- (59) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. *Acc. Chem. Res.* **2010**, *43*, 661–672.
- (60) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. *J. Am. Chem. Soc.* **2015**, *137*, 10468–10471.
- (61) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. D.; Tezcan, F. A. *Nature*. **2016**, *533*, 369–373.
- (62) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A. *J. Am. Chem. Soc.* **2015**, *137*, 11598–11601.
- (63) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. *Nat. Chem.* **2012**, *4*, 375–382.
- (64) Bolon, D. N.; Wah, D. A.; Hersch, G. L.; Baker, T. A.; Sauer, R. T. *Mol. Cell.* **2004**, *13*, 443–449.
- (65) Huang, P. S.; Love, J. J.; Mayo, S. L.; *Protein Sci.* **2007**, *16*, 2770–2774.
- (66) Mou, Y.; Yu, J. Y.; Wannier, T. M.; Guo, C. L.; Mayo, S. L. *Nature*. **2015**, *525*, 230–233.
- (67) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. *Science*. **2012**, *336*, 1171–1174.
- (68) Huang, P. S.; Boyken, S. E.; Baker, D. *Nature*. **2016**, *537*, 320–327.
- (69) Marcos, E.; Adriano-Silva, D. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1374.
- (70) Ben-Sasson, A. J.; Watson, J. L.; Sheffler, W.; Johnson, M. C.; Bittleston, A.; Somasundaram, L.; Decarreau, J.; Jiao, F.; Chen, J.; Mela, J.; Drabek, A. A.; Jarrett, S. M.; Blacklow, S. C.; Kaminski, C. F.; Hura, G. L.; De Yoreo, J. J.; Kollman, J. M.; Ruohola-Baker, H.; Derivery, E.; Baker, D. *Nature*. **2021**, *589*, 468–473.
- (71) Huang, P.-S.; Feldmeier, K.; Parmeggiani, F.; Velasco, D. A. F.; Hocker, B.; Baker, D. *Nat. Chem. Biol.* **2016**, *12*, 29–34.
- (72) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D. *Nature* **2016**, *535*, 136–139.

- (73) Xu, C.; Lu, P.; Gamal El-Din, T. M.; Pei, X. Y.; Johnson, M. C.; Uyeda, A.; Bick, M. J.; Xu, Q.; Jiang, D.; Bai, H.; Reggiano, G.; Hsia, Y.; Brunette, T. J.; Dou, J.; Ma, D.; Lynch, E. M.; Boyken, S. E.; Huang, P.; Stewart, L.; DiMaio, F.; Kollman, J. M.; Luisi, B. F.; Matsuura, T.; Catterall, W. A.; Baker, D. *Nature*. **2020**, *585*, 129–134.
- (74) Shen, H.; Fallas, J. A.; Lynch, E.; Sheffler, W.; Parry, B.; Jannetty, N.; Decarreau, J.; Wagenbach, M.; Vincente, J. J.; Chen, J.; Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; De Yoreo, J.; Jacobs-Wagner, C.; Kollman, J.; Baker, D. *Science*. **2018**, *362*, 705–709.
- (75) Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; Mulligan, V. K.; Duebe, J. E.; Novak, W. R. P.; El-Samad, H.; Baker, D. *Nature*. **2019**, *572*, 205–210.
- (76) Watson, J. D.; Crick, F. H. *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–131.
- (77) Reese, C. B. *Org. Biomol. Chem.* **2005**, *3*, 3851–3868.
- (78) Seeman, N. C.; Sleiman, H. F. *Nat. Rev. Mater.* **2017**, *3*, 17068.
- (79) Chen, J.; Seeman, N. C.; *Nature*. **1991**, *350*, 631–633.
- (80) Winfree, E.; Liu, F.; Wenzler, L.; Seeman, N. C.; *Nature*. 1998, **394**, 539–544.
- (81) Rothmund, P. *Nature*. **2006**, *440*, 297–302.
- (82) Alivisatos, A. P.; Johnsson, K. P.; Peng, X.; Wilson, T. E.; Loweth, C. J.; Bruches Jr, M. P.; Schultz, P. G.; *Nature*. **1996**, *382*, 609–611.
- (83) Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J.; *Nature*. **1996**, *382*, 607–609.
- (84) Laramy, C. R.; O'Brien, M. N.; Mirkin, C. A. *Nat. Rev. Mater.* **2019**, *4*, 201–224.
- (85) Macfarlane, R. J.; Lee, B.; Jones, M. R.; Harris, N.; Schatz, G. C.; Mirkin C. A.; *Science*. **2011**, *334*, 204–208.
- (86) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. *Nucleic. Acids. Res.* **1994**, *22*, 5530–5539.
- (87) Seki, Y.; Ishiyama, T.; Sasaki, D.; Abe, J.; Sohma, Y.; Oisaki, K.; Kanai, M. *J. Am. Chem. Soc.* **2016**, *138*, 10798–10801
- (88) Szijj, P. A.; Kostadinova, K. A.; Spears, R. J.; Chudasama, V. *Org. Biomol. Chem.* **2020**, *18*, 9018-9028.
- (89) Kim, C. H.; Axup, J. Y.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2013**, *17*, 412–419.

- (90) Sørensen, R. S.; Okholm, A. H.; Schaffert, D.; Kodal, A. L. B.; Gothelf, K. V.; Kjems, J. *ACS Nano*. **2013**, *7*, 8098–8104.
- (91) Wilchek, M.; Bayer, E. A. *Anal. Biochem.* **1988**, *171*, 1-32.
- (92) Kim, Y.-Y.; Bang, Y.; Lee, A.-H.; Song, Y.-K. *ACS Nano*. **2019**, *13*, 1183–1194.
- (93) Shimada, J.; Maruyama, T.; Hosogi, T.; Tominaga, J.; Kamiya, N.; Goto, M. *Biotechnol. Lett.* **2008**, *30*, 2001–2006.
- (94) Adler, M.; Langer, M.; Witthohn, K.; Wilhelm-Ogunbiyi, K.; Schoffski, P.; Fumoleau, P.; Niemeyer, C. M. *J. Pharm. Biomed. Anal.* **2005**, *39*, 972–972.
- (95) Kong, G.; Xiong, M.; Liu, L.; Hu, L.; Meng, H. M.; Ke, G.; Tan, W. *Chem. Soc. Rev.* **2021**, *50*, 1846–1873.
- (96) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. *Nature Nanotech.* **2009**, *4*, 249–254.
- (97) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. *Proc. Natl Acad. Sci. USA*. **2015**, *112*, 4564–4569.
- (98) Kashiwagi, D.; Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. *J. Am. Chem. Soc.* **2018**, *140*, 26–29.
- (99) McMillan, J. R.; Mirkin, C. A. *J. Am. Chem. Soc.* **2018**, *140*, 6776–6779.
- (100) Carothers, W. H. *Trans. Faraday Soc.* **1936**, *32*, 39-49.
- (101) Flory, P. J. *Chem. Rev.* **1946**, *39*, 137-197.
- (102) Goto, A.; Fukuda, T. *Prog. Polym. Sci.* **2004**, *29*, 329-385.
- (103) Colombani, D. *Prog. Polym. Sci.* **1997**, *22*, 1649-1720.
- (104) Gregory, A.; Stenzel, M. H. *Prog. Polym. Sci.* **2012**, *37*, 38-105.
- (105) Gunkel, G.; Weinhart, M.; Becherer, T.; Haag, R.; Huck, W. T. S. *Biomacromolecules*. **2001**, *12*, 4169–4172.
- (106) Müller, A. H. E.; Matyjaszewski, K. *Controlled and Living Polymerizations: From Mechanisms to Applications*. Wiley-VCH Verlag GmbH & Co. KGaA, Germany, **2010**.
- (107) Lutz, J. F.; Lehn, J. M.; Meijer, E. W.; Matyjaszewski, K. *Nat. Rev. Mater.* **2016**, *1*, 16024.

- (108) Besenius, P. *J. Polym. Sci. Part A: Polym. Chem.*, **2017**, *55*, 34–78.
- (109) Markovitch, O.; Agmon, N. *J. Phys. Chem. A*. **2007**, *111*, 2253–2256.
- (110) Aida, T.; Meijer, E. W.; Stupp, S. I. *Science*. **2012**, *335*, 813–817.
- (111) Brunsveld, L.; Folmer, B. J. B.; Meijer, E. W.; Sijbesma, R. P. *Chem. Rev.* **2001**, *101*, 4071–4097.
- (112) Yang, L.; Tan, X.; Wang, Z.; Zhang, X. *Chem. Rev.* **2015**, *115*, 7196–7239.
- (113) Kueh, H. Y.; Mitchison, T. J. *Science*. **2009**, *325*, 960–963.
- (114) Bugyi, B.; Carlier, M. F. *Annu. Rev. Biophys.* **2010**, *39*, 449–470.
- (115) Chesarone, M. A.; Goode, B. L. *Curr Opin Cell Biol.* **2009**, *21*, 28–37.
- (116) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311–335.
- (117) Aida, T.; Meijer, E. W. *Isr. J. Chem.* **2020**, *60*, 33–47.
- (118) Desai, A.; Mitchison, T. J. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 83–117.
- (119) Slep, K. C. *Curr. Opin. Cell. Biol.* **2010**, *22*, 88–95.
- (120) Carlson, J. C. T.; Jena, S. S.; Flenniken, M.; Chou, T.-f.; Siegel, R. A.; Wagner, C. R. *J. Am. Chem. Soc.* **2006**, *128*, 7630–7638.
- (121) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. *Nat. Chem.* **2012**, *4*, 375.
- (122) Modica, J. A.; Lin, Y.; Mrksich, M. *J. Am. Chem. Soc.* **2018**, *140*, 6391–6399.
- (123) Liu, K.; Nie, Z.; Li, W.; Rubinstein, M.; Kumacheva, E. *Science*. **2010**, *329*, 197–200.
- (124) Ogi, S.; Sugiyasu, K.; Manna, S.; Samitsu, S.; Takeuchi, M. *Nat. Chem.* **2014**, *6*, 188.
- (125) Kang, J.; Miyajima, D.; Mori, T.; Inoue, Y.; Itoh, Y.; Aida, T. *Science*. **2015**, *347*, 646–651.
- (126) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science*. **2015**, *347*, 1260901.
- (127) Garst, A. D.; Edwards, A. L.; Batey, R. T. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a003533.
- (128) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627–8637.

- (129) Dirks, R. M.; Pierce, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15275–15278
- (130) Danev, R.; Buijsse, B.; Khoshouei, M.; Plitzko, J. M.; Baumeister, W. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15635–15640.
- (131) Khoshouei, M.; Radjainia, M.; Baumeister, W.; Danev, R. *Nat. Commun.* **2017**, *8*, 16099.
- (132) Chua, E. Y. D.; Vogirala, V. K.; Inian, O.; Wong, A. S. W.; Nordenskiöld, L.; Plitzko, J. M.; Danev, R.; Sandin, S. *Nucleic Acids Res.* **2016**, *44*, 8013–8019.
- (133) Usov, I.; Mezzenga, R. *Macromolecules.* **2015**, *48*, 1269–1280.
- (134) Xu, J.; Fogleman, E. A.; Craig, S. L. *Macromolecules.* **2004**, *37*, 1863–1870.
- (135) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6185–6190.
- (136) Marko, J. F.; Siggia, E. D. *Macromolecules.* **1995**, *28*, 8759–8770.
- (137) Bustamante, C.; Marko, J.; Siggia, E.; Smith, S. *Science.* **1994**, *265*, 1599–1600.
- (138) Cloutier, T. E.; Widom, J. *Mol. Cell.* **2004**, *14*, 355–362.
- (139) Vafabakhsh, R.; Ha, T. *Science.* **2012**, *337*, 1097–1101.
- (140) Du, Q.; Kotlyar, A.; Vologodskii, A. *Nucleic Acids Res.* **2008**, *36*, 1120–1128.
- (141) Wolters, M.; Wittig, B. *Nucleic Acids Res.* **1989**, *17*, 5163–5172.
- (142) Yan, J.; Marko, J. F. *Phys. Rev. Lett.* **2004**, *93*, 108108.
- (143) Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **2004**, *342*, 775–785.
- (144) Demurtas, D.; Amzallag, A.; Rawdon, E. J.; Maddocks, J. H.; Dubochet, J.; Stasiak, A. *Nucleic Acids Res.* **2009**, *37*, 2882–2893.
- (145) Chen, R. P.; Blackstock, D.; Sun, Q.; Chen, W. *Nat. Chem.* **2018**, *10*, 474–481.
- (146) Gentekos, D. T.; Dupuis, L. N.; Fors, B. P. *J. Am. Chem. Soc.* **2016**, *138*, 1848–1851.
- (147) Wang, J.-S.; Matyjaszewski, K. *J. Am. Chem. Soc.* **1995**, *117*, 5614–5615.
- (148) Figg, C. A.; Winegar, P. H.; Hayes, O. G.; Mirkin, C. A. *J. Am. Chem. Soc.* **2020**, *142*, 8596–8601.

- (149) Zheng, S. Q.; Palovcak, E.; Armache, J.-P.; Verba, K. A.; Cheng, Y.; Agard, D. A., *Nat. Methods*. **2017**, *14*, 331.
- (150) Rohou, A.; Grigorieff, N. *J. Struct. Biol.* **2015**, *192*, 216-221.
- (151) Kimanius, D.; Forsberg, B. O.; Scheres, S. H. W.; Lindahl, E., *eLife*. **2016**, *5*, e18722.
- (152) Campbell, M. G.; Karbstein, K.; *PLoS One*. **2011**, *6*, e16194.
- (153) Fromont-Racine, M.; Senger, B.; Saveanu, C.; Fasiolo, F. *Gene*. **2003**, *313*, 17–42.
- (154) Henras, A. K.; Soudet, J.; Gerus, M.; Lebaron, S.; Caizergues-Ferrer, M.; Mougin, A.; Henry, Y. *Cell. Mol. Life. Sci.* **2008**, *65*, 2334–2359.
- (155) Strunk, B. S.; Karbstein, K. *RNA*. **2009**, *15*, 2083–2104.
- (156) Ameismeier, M.; Cheng, J.; Berninghausen, O.; Beckmann, R. *Nature*. **2018**, *558*, 249–253.
- (157) Woolls, H.; Lamanna, A. C.; Karbstein, K. *J. Biol. Chem.* **2011**, *286*, 2578–2586.
- (158) Doi, R. H.; Kosugi, A. *Nat. Rev. Microbiol.* **2004**, *2*, 541–551.
- (159) Murata, S.; Yashiroda, H.; Tanaka, K. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 104–115.
- (160) Glotzer, S. C.; Solomon, M. J. *Nat. Mater.* **2007**, *6*, 557–562.
- (161) Jacobs, W. M.; Frenkel, D. *J. Am. Chem. Soc.* **2016**, *138*, 2457–2467.
- (162) Jacobs, W. M., Reinhardt, A. & Frenkel, D. *Proc. Natl Acad. Sci. USA*. **2015**, *112*, 6313–6318.
- (163) Ravaine, S.; Duguet, E. *Curr. Opin. Colloid Interface Sci.* **2017**, *30*, 45–53.
- (164) Perro, A.; Reculosa, S.; Ravaine, S.; Bourgeat-Lami, E. B.; Duguet, E. *J. Mater. Chem.* **2005**, *15*, 3745–3760.
- (165) Chen, Q.; Bae, S. C.; Granick, S. *Nature*. **2011**, *469*, 381–384.
- (166) Gong, Z.; Hueckel, T.; Yi, G. R. Sacanna, S. *Nature*. **2017**, *550*, 234–238.
- (167) Liu, W.; Halverson, J.; Tian, Y.; Tkachenko, A. V.; Gang, O. *Nat. Chem.* **2016**, *8*, 867–873.
- (168) Bianchi, E.; Doppelbauer, G.; Filion, L.; Dijkstra, M.; Kahl, G. *J. Chem. Phys.* **2012**, *136*, 214102.

- (169) McMillan, J. R.; Brodin, J. D.; Millan, J. A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C. A. *J. Am. Chem. Soc.* **2017**, *139*, 1754–1757.
- (170) Lattuada, M.; Hatton, T. A. *Nano Today*. **2011**, *6*, 286–308.
- (171) Chen, T.; Chen, G.; Xing, S.; Wu, T.; Chen, H. *Chem. Mater.* **2010**, *22*, 3826–3828.
- (172) van Ravensteijn, B. G. P.; Kamp, M.; van Blaaderen, A.; Kegel, W. K. *Chem. Mater.* **2013**, *25*, 4348–4353.
- (173) Edwardson, T. G. W.; Lau, K. L.; Bousmail, D.; Serpell, C. J.; Sleiman, H. F. *Nat. Chem.* **2016**, *8*, 162–170.
- (174) Chamberlain, A. K.; Bowie, J. U. *Protein Sci.* **2004**, *13*, 2270–2274.
- (175) Jin, W.; Takada, S. *J. Mol. Biol.* **2008**, *377*, 74–82.
- (176) Kim, Y. E.; Kim, Y.-n.; Kim, J. A.; Kim, H. M.; Jung, Y. *Nat. Commun.* **2015**, *6*, 7134.
- (177) Lapiene, V.; Kukolka, F.; Kiko, K.; Arndt, A.; Niemeyer, C. M. *Bioconj. Chem.* **2010**, *21*, 921–927.
- (178) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- (179) Auyeung, E.; Macfarlane, R. J.; Choi, C. H. J.; Cutler, J. I.; Mirkin, C. A. *Adv. Mater.* **2012**, *24*, 5181–5186.
- (180) Macfarlane, R. J.; O'Brien, M. N.; Petrosko, S. H.; Mirkin, C. A. *Angew. Chem., Int. Ed.* **2013**, *52*, 5688–5698.
- (181) Cutler, J. I.; Auyeung, E.; Mirkin, C. A. *J. Am. Chem. Soc.* **2012**, *134*, 1376–1391.
- (182) Radovic-Moreno, A. F.; Chernyak, N.; Mader, C. C.; Nallagatla, S.; Kang, R. S.; Hao, L.; Walker, D. A.; Halo, T. L.; Merkel, T. J.; Rische, C. H.; Anantatmula, S.; Burkhart, M.; Mirkin, C. A.; Gryaznov, S. M. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3892–3897.
- (183) Zheng, D.; Giljohann, D. A.; Chen, D. L.; Massich, M. D.; Wang, X.-Q.; Iordanov, H.; Mirkin, C. A.; Paller, A. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11975–11980.
- (184) Paramelle, D.; Sadvovoy, A.; Gorelik, S.; Free, P.; Hobley, J.; Fernig, D. G., *Analyst*. **2014**, *139*, 4855–4861.
- (185) Gotti, C.; Sensini, A.; Zucchelli, A.; Carloni, R.; Focarete, M. L. *Appl. Mater. Today*. **2020**, *20*, 100772.

- (186) Fitzpatrick, A. W. P.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. A.; Jaroniec, C. P.; Wang, L. C.; Ladizhansky, V.; Muller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P. J.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V.; Griffin, R. G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5468–5473.
- (187) Gale, M.; Pollanen, M. S.; Markiewicz, P.; Goh, M. C.; *Biophys. J.* **1995**, *68*, 2124–2128.
- (188) Okada, Y.; Ohno, T. *Mol. Gen. Genet. MGG.* **1972**, *114*, 205–213.
- (189) Morphew, D.; Shaw, J.; Avins, C.; Chakrabarti, D. *ACS Nano.* **2018**, *12*, 2355–2364.
- (190) Schnitzer, T.; Vantomme, G. *ACS Cent. Sci.* **2020**, *6*, 2060–2070.
- (191) Haxton, T. K.; Whitelam, S. *Soft Matter.* **2013**, *9*, 6851–6861.
- (192) Rao, A. B.; Shaw, J.; Neophytou, A.; Morphew, D.; Sciortino, F.; Johnston, R. L.; Chakrabarti, D. *ACS Nano.* **2020**, *14*, 5348–5359.
- (193) Sun, H.; Li, Y.; Yu, S.; Liu, J. *Front. Bioeng. Biotechnol.* **2020**, *8*, 295.
- (194) 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. *Science.* **2016**, *353*, 389–394.
- (195) Churchfield, L. A.; Tezcan, F. A. *Acc. Chem. Res.* **2019**, *52*, 345–355.
- (196) Yang, M.; Song, W. J.; *Nat. Commun.* **2019**, *10*, 5545.
- (197) Garcia-Seisdedos, H.; Empereur-Mot, C.; Elad, N.; Levy, E. D. *Nature.* **2017**, *548*, 244–247.
- (198) Si, C.; Li, J.; Luo, Q.; Hou, C.; Pan, T.; Li, H.; Liu, J. *Chem. Commun.* **2016**, *52*, 2924–2927.
- (199) Burazerovic, S.; Gradinaru, J.; Pierron, J.; Ward, T. R. *Angew. Chemie Int. Ed.* **2007**, *46*, 5510–5514.
- (200) Insua, I.; Montenegro, J. *J. Am. Chem. Soc.* **2020**, *142*, 300–307.
- (201) Zhao, Z.; Wang, J. C-Y.; Zhang, M.; Lykтей, N. A.; Jarrold, M. F.; Jacobson, S. C.; Zlotnick, A. *Nat. Commun.* **2021**, *12*, 589.
- (202) Chung, S.; Shin, S. H.; Bertozzi, C. R.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16536–16541.

- (203) Van Driessche, A. E. S.; Van Gerven, N.; Bomans, P. H. H.; Joosten, R. R. M.; Friedrich, H.; Gil-Carton, D.; Sommerdijk, N. A. J. M.; Sleutel, M. *Nature*. **2018**, *556*, 89–94.
- (204) Zhang, S.; Alberstein, R. G.; De Yoreo, J. J.; Tezcan, F. A. *Nat. Commun.* **2020**, *11*, 3770.
- (205) McMillan, J. R.; Hayes, O. G.; Winegar, P. H.; Mirkin, C. A. *Acc. Chem. Res.* **2019**, *52*, 1939–1948.
- (206) Xu, Y.; Jiang, S.; Simmons, C. R.; Narayanan, R. P.; Zhang, F.; Aziz, A-M.; Yan, H.; Stephanopoulos, N. *ACS Nano*. **2019**, *13*, 3545–3554.
- (207) 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.; Tezkan, F. A. *ACS Cent. Sci.* **2018**, *4*, 1578–1586.
- (208) Winegar, P. H.; Hayes, O. G.; McMillan, J. R.; Figg, C. A.; Focia, P. J.; Mirkin, C. A. *Chem.* **2020**, *6*, 1007–1017.
- (209) Partridge, B. E.; Winegar, P. H.; Han, Z.; Mirkin, C. A. *J. Am. Chem. Soc.* **2021**, jacs.1c04191.
- (210) Hayes, O. G.; McMillan, J. R.; Lee, B.; Mirkin, C. A. *J. Am. Chem. Soc.* **2018**, *140*, 9269–9274.
- (211) Kashiwagi, D.; Shen, H. K.; Sim, S.; Sano, Koki, Ishida, Y.; Kimura, A.; Niwa, T.; Taguchi, H.; Aisa, T. *J. Am. Chem. Soc.* **2020**, *142*, 13310–13315.
- (212) McMillan, J. R.; Hayes, O. G.; Remis, J. P.; Mirkin, C. A.; *J. Am. Chem. Soc.* **2018**, *140*, 15950–15956.
- (213) Dgany, O.; Gonzalez, A.; Sofer, O.; Wang, W.; Zolotnitsky, G.; Wolf, A.; Shoham, Y.; Altman, A.; Wolf, A.; Shoseyov, O.; Almog, O. *J. Biol. Chem.* **2004**, *279*, 51516–51523.
- (214) Blackman, M. L.; Royzen, M.; Fox, J. M. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
- (215) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. *Chem. Soc. Rev.* **2017**, *46*, 4895–4950.
- (216) Karver, M. R.; Weissleder, R.; Hilderbrand, S. A. *Angew. Chemie Int. Ed.* **2012**, *51*, 920–922.
- (217) Saccà, B.; Meyer, R.; Niemeyer, C. M.; *Nat. Protoc.* **2009**, *4*, 271–285.
- (218) Samanta, D.; Iscen, A.; Laramy, C. R.; Ebrahimi, S. B.; Bujold, K. E.; Schatz, G. C.; Mirkin, C. A. *J. Am. Chem. Soc.* **2019**, *141*, 19973–19977.

- (219) Wiessler, M.; Lorenz, P.; Fleischhacker, H.; Kliem, C.; Jaschke, A.; Schoch, J. *US9169283B2*, **2012**.
- (220) Galli, S. J. *Chem. Educ.* **2014**, *91*, 2009–2012
- (221) Matthews, B. W. *Protein Sci.* **2009**, *18*, 1135–1138.
- (222) Wukovitz, S. W.; Yeates, T. O. *Nat. Struct. Biol.* **1995**, *2*, 1062–1067.
- (223) Kwon, N. Y.; Kim, Y.; Lee, J. O. *Curr. Opin. Struct. Biol.* **2020**, *60*, 110–116.
- (224) Laganowsky, A.; Zhao, M.; Soriaga, A. B.; Sawaya, M. R.; Cascio, D.; Yeates, T. O. *Protein Sci.* **2011**, *20*, 1876–1890.
- (225) Banatao, D. R.; Cascio, D.; Crowley, C. S.; Fleissner, M. R.; Tienson, H. L.; Yeates, T. O. *Proc. Natl. Acad. Sci.* **2006**, *103*, 16230–16235.
- (226) Leibly, D. J.; Arbing, M. A.; Pashkov, I.; Devore, N.; Waldo, G. S.; Terwilliger, T. C.; Yeates, T. O. *Structure.* **2015**, *23*, 1754–1768.
- (227) Thaner, R. V.; Eryazici, I.; Farha, O. K.; Mirkin, C. A.; Nguyen, S. T. *Chem. Sci.* **2014**, *5*, 1091–1096.
- (228) DeGruyter, J. N.; Malins, L. R.; Baran, P. S. *Biochemistry* **2017**, *56*, 3863–3873.
- (229) Campbell, M. A.; Wengel, J. *Chem. Soc. Rev.* **2011**, *40*, 5680–5689.
- (230) Feng, X.; Zhang, R.; Li, Y.; Hong, Y. L.; Guo, D.; Lang, K.; Wu, K. Y.; Huang, M.; Mao, J.; Wesdemiotis, C.; Nishiyama, Y.; Zhang, W.; Zhang, W.; Miyoshi, T.; Li, T.; Cheng, S. Z. D. *ACS Cent. Sci.* **2017**, *3*, 860–867.
- (231) Bénédict, P.; Paciucci, R.; Thomson, T. M.; Valeri, M.; Nadal, M.; Càceres, C.; De Torres, I.; Estivill, X.; Lozano, J. J.; Morote, J.; Reventós, J. *Oncogene* **2001**, *20*, 1455–1464.