

ON DESIGNING BIOLOGICAL NANOSCALE ORGANIZATION

By

Eric J. Young

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology—Doctor of Philosophy

2019

ABSTRACT

ON DESIGNING BIOLOGICAL NANOSCALE ORGANIZATION

By

Eric J. Young

Life at the nanoscale creates a dazzling array machines and structures. Studying these nanoscale creations often requires inter-disciplinary efforts of scientists, along with the support of other personnel. This thesis serves to communicate some personal insights and data captured in studying nanoscale organization of biologically-driven components, as part of such a team. The first chapter addresses spatiotemporal organization of material inside cells, with a focus on scaffolding-type strategies. The second chapter offers a literature perspective on constructing scaffolds with a structurally-characterized protein-domain. The third chapter surveyed functionality of an *in vivo* designer nanoscaffolding system. The fourth chapter, alongside the appendix materials, forms a collection of future-steps and comments on projects I have encountered while working on my thesis project.

To everyone, near and far, who has supported me over this journey, a sincere thank you—this goes out to all of you. Completing this document would not have felt feasible without the love and direction provided by the kind souls this world has to offer.

-Eric

ACKNOWLEDGEMENTS

First order of business: a heartfelt thank you to my thesis advisor Danny Ducat for fostering an environment to explore the depths of what a PhD experience has to offer. You have shown me true kindness and support through the highs and lows. I will forever feel grateful for your support.

To my committee members, and other scientific mentors, all of you have helped shaped my growth as individual. From the collaborative opportunities to heartfelt encouragement and keeping me honest on this path, thank you.

To all of the family, friends, and co-workers, and all the other souls I have met throughout this time, I extend the deepest acknowledgment towards. When things felt overwhelming, life always reminded of how caring people can act.

PREFACE

This thesis document represents my attempt to convey information which I have collected while undertaking my PhD project. One major distinction of this project includes the adoption of describing some information without the use of “to be” language in Chapters 1, 3, 4, and some Appendix materials (at least to the best of my current ability). Known as English-Prime (E-prime), this language style may at first perceptions appear to introduce some level of uncertainty in some statements. But I believe how we communicate with not only our scientific peers, and also the general public (and our loved ones), can improve with having less “is-ness” in our communication. In my opinion, adopting E-prime may help to create statements more reflective of what actually occurred in space-time by uncovering hidden elements and diminishing some perspective bias that can emerge in an “is” statement. Priming our language in such a way helps get “ourselves” out of the picture, so we can hopefully address the real underlying phenomena or issue. With that aside, I welcome you to my thesis document and hope the materials collected give you a glimpse to the broad fields of biochemistry, synthetic biology, and bionanotechnology.

Cheers.

TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
Chapter 1: Spatiotemporal organization within cells.....	1
LITERATURE CITED.....	13
Chapter 2: A literature perspective on designing nanoscaffolds with a self-assembling protein domain.....	18
LITERATURE CITED.....	39
Chapter 3: Visualizing <i>in vivo</i> dynamics of designer nanoscaffolds.....	46
LITERATURE CITED.....	97
Chapter 4: Perspective on future opportunities with engineered components.....	103
LITERATURE CITED.....	112
Chapter 5: Personal remarks on select co-authored publications, example lead-author proposals, and a reference to a co-authored public communication piece.....	116
LITERATURE CITED.....	144

LIST OF TABLES

Table 3.1	Plasmids and gene inserts used in Chapter 3.....	74
-----------	--	----

LIST OF FIGURES

Figure 1.1: Nanoscale spatiotemporal organization strategies inside cells.....	2
Figure 1.2: Colocalizing a pathway via a scaffolding spatiotemporal strategy.....	7
Figure 2.1: BMC-H attributes and potential as modular building blocks.....	24
Figure 2.2: Could molecular-level simulations contribute toward predictive assembly of diverse BMC-H scaffolds?	33
Figure 3.1: Functionalizing HO BMC-H with a Synthetic Zipper protein-protein interaction domain.....	52
Figure 3.2: Cargo recruitment to designer intracellular protein scaffolds.....	56
Figure 3.3: Evaluating scaffold behavior by visualizing intracellular cargo dynamics.....	60
Figure 3.4: Diverse maturation fates of nucleated ScaFS.....	63
Figure S3.5: Analysis of potential scaffold assembly domains.....	75
Figure S3.6: Thin section overexpression of modules lacking nanostructures.....	77
Figure S3.7: PFam0936-protein naturally exhibit diversity in structural features of the C-terminus.....	79
Figure S3.8: Predicted structure of SZ-functionalized ScaFS.....	80
Figure S3.9: Higher-order intracellular architectures formed by ScaFS bearing a C-terminal SZ5 domain attached via a flexible linker.....	81
Figure S3.10: Fluorescent cargo reporter localizes on or near intracellular diffractions produced by overexpressed ScaFS.....	82
Figure S3.11: Comparison of high-resolution approaches for imaging intracellular ScaFS.....	84
Figure S3.12: <i>In vitro</i> co-immunoprecipitation of purified ScaFS and reporter cargo proteins.....	86
Figure S3.13: Expression level of ScaFS correlates with characteristics in intracellular localization phenotype.....	87

Figure S3.14: Co-localization of alternative cargo molecules to compatible ScaFS-induced intracellular protein assemblies.....	88
Figure S3.15: High-resolution image processing pipeline.....	89
Figure S3.16: Scaffold assembly as viewed by time-course imaging of co-expressed ScaFS and cargo in live-cells.....	91
Figure S3.17: Representative time-lapse imaging of nucleation events.....	93
Figure S3.18: Visualization of intracellular cargo dynamics by SRRF-processing.....	95
Figure 4.1: Broad follow up projects with engineered modules.....	105
Figure 4.2: Structural snapshots of individual frames from atomistic molecular dynamic simulations.....	111
Figure 5.1: Engineering and characterizing assemblies for molecular scaffolding.....	122
Figure 5.2: Engineering functional nanostructures.....	129
Figure 5.3: Investigating engineered nanostructures with superresolution microscopy methods.....	133
Figure 5.4: Applying neutron scattering and computational simulation in nanoscale self-assembly.....	137
Figure 5.5: Detailing cross-application of neutron spin echo and molecular dynamic simulation to investigate inter-domain dynamics.....	140

Chapter 1: Spatiotemporal organization within cells

The scientific understanding of the dynamics and specific spatiotemporal ordering of the intracellular environment has steadily evolved over-time, and contrasts remarkably with earliest definitions of the cytosol as: “devoid of any structural organization smaller than the cell itself (Ligrone 2019; Popkin 2016; Lane 2015). Our current descriptions of biological systems now view even the simplest cells (or viruses) as vehicles containing intricate spatiotemporal organization (Govindarajan 2016; Karsenti 2008; Surovtsev 2018; Diekmann 2013). Visualization technologies, ever improving in the resolution of structural and dynamical information collected, have helped advance our understanding. This thesis attempts to cover some considerations for engineering biologically-based nanoscale spatiotemporal organization, principally on the creation of nanoscaffolds within cells.

Beginning to describe spatiotemporal engineering within cells owes some consideration to the forces at play at the nanoscale, namely: Brownian motion and crowding (Schavemaker 2018). Inside cells, particles experience collisions with other molecules stochastically changing their direction and velocity. The sheer concentration of cytosolic material (nearing 300-400 g/L in some cases) further complicates matters (Alberti 2017; Ellis 2001). Small-molecule diffusion coefficients may change nearly 100-fold when comparing behavior in dilute aqueous solutions to the cytosol (Schavemaker 2018; Alberti 2017; Ellis 2001). Practically, this means substrates take longer periods of time to encounter a specific function (Feig 2017). Cellular functions overcome these forces by placing related-function(s) in a

shared space-time. Doing so, appears to increase efficacy and programmability of higher-order behavior(s) (Castellana 2014).

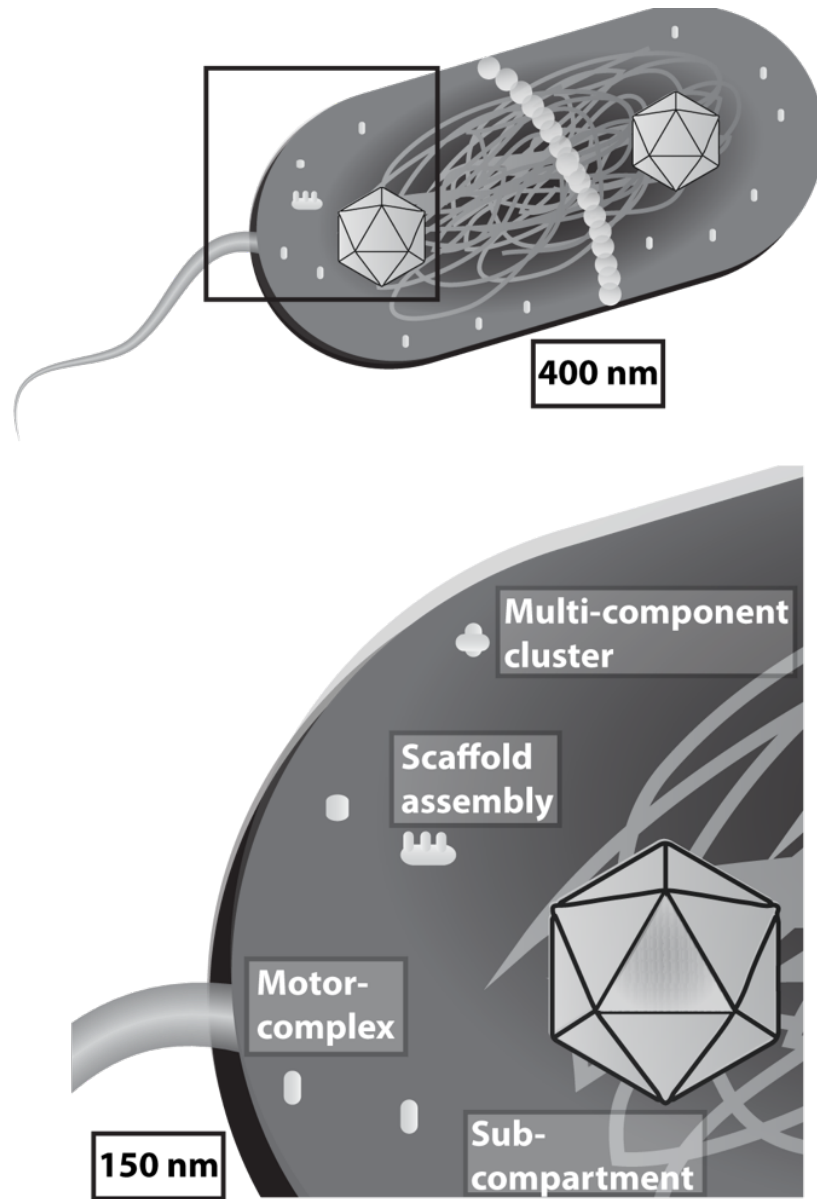


Figure 1.1: Nanoscale spatiotemporal organization strategies inside cells. A variety of approaches accomplish specific spatiotemporal location of nanoscale function leading to benefits to observed output.

Scientists have identified some common themes to the spatiotemporal organization of intracellular material **(Figure 1.1)** (Murat 2010). These organization strategies broadly classify into three categories of differing complexity, although co-localization of inter-related function in shared space-time unites them all. The aggregation of enzymes into particular multimeric complexes— known as metabolons—represents one basic spatiotemporal organization strategy. Structural scaffolds, whereby a distinct structural element tethers auxiliary function(s), another. Finally, cells produce nano-to-micro compartments that define distinct boundaries for unique environments for particular cargos. Each category has individual subtleties and benefits they bring to their respectively organized biochemical functions (Jakobson 2018; Küchler 2016). For example, micro-compartments restrict diffusion of toxic metabolites, scaffolds serve as critical design elements for signaling pathways, and multimeric complexes produce circadian rhythms and efficient metabolic transfer.

Many spatiotemporal organization strategies rely on proteins as their building material, making it pivotal to understand their structure, function, and dynamics inside cells (Luo 2016). To begin, forming any protein-based structural/functional unit requires collective interactions between the amino acids in a poly-peptide sequence and the surrounding environment (Huang 2016). Conditions such as pH, ions, solute, crowding, and hydrophobics intimately influence how folding proceeds to a tertiary-structure state. Protein subunits then can oligomerize further to higher-order quaternary states. Some individual examples include: oligomeric complexes (*e.g.* multi-subunit enzymes), structural geometries (*e.g.* 1D filaments, 2D lattices, 3D compartments) or patterns in spatiotemporal dynamics (*e.g.*

oscillatory waves of particles) (Rudner 2010). Here, we broadly define the process of folding through any higher-order collective either as “self-assembly” or “self-organization.”

Regardless of differences in definition, “self”-driven processes form the heart of protein-based spatiotemporal organization. Often, proteins rely on compatible interfaces, leading to energetically stable conformations and higher-order structures/functions (Keskin 2008; Ahnert 2015). To example the case of divisome ‘machinery’ coordination in some bacteria, a dynamically assembling scaffolding system coupled to a spatiotemporal organizing system leads to midpoint cell bisection (Szwedziak 2014, Loose 2011). The protein-protein interactions and compatible interfaces involved in assembling the scaffold have several outcomes. Firstly, it appears that individual filaments form with a tail-to-head polymerization with active treadmilling behavior. Individual filaments then can laterally bundle together, forming an extended patch work. Bundling appears necessary to help generate a constriction force to invaginate the membrane, while treadmilling behaviors encourages directional movement for associated cell wall synthesis machinery (Bisson-Filho 2017; Yang 2017).

Without these specific properties in spatiotemporal organization driven by the collective ordering of individual subunits, cells would lack the ability to specifically divide at a particular time/space (*i.e.* at the mid-point of the cell). In this context, the filaments of the scaffold acts as a platform to coordinate multiple other cellular components that each contribute towards the broader goal of bisecting a cell. For example, components involved in inserted new lipids and producing cell wall precursors colocalize with the enzymes

responsible for polymerizing new cell wall though binding the same filamentous scaffold. Colocalizing all these functions upon a scaffold produces synergy between the functions, whereas if components lacked this spatiotemporal organization, functions may compete with each other, ultimately leading to lowered efficacy, or even abolishment of the process (Haeusser 2016). These elegantly observed behaviors arising out of self-driven processes has consequently inspired nanoscale engineers to harness proteins (along with other biological materials) to create systems capable of rational spatiotemporal organization of their own desired functions (Yeates 2017; King 2013; Bai 2016).

Designer Scaffolds as Agents for Spatiotemporal Organization

Truly ‘designer’ spatiotemporal organization within cells holds promise for ushering in a new era for technology, energy, and health (Ulijn 2018; Yang 2016; Howorka 2011; Polka; Zhang 2003; Schwille 2011; Rice 2014). Finely controlling the spatiotemporal properties (position and dynamics) of designer components in living cells appears analogous to explosion in understanding and engineerability accomplished by other disciplines (*i.e.* how precision-built components at the microscale enabled computing to develop exponentially). Fully realizing this dream relies on adopting predictive strategies. Using protein-derived building blocks to form structural scaffolds represents one area of development. Below will briefly example some design principles and target applications for scaffolding systems as spatiotemporal coordinators inside cells.

Customized cellular signaling networks comprises one rich area for applying designer protein scaffolds (Good 2011). In some eukaryotic organisms, signaling networks make up

~10% of the proteome and provide indispensable regulators to many key processes. Scaffolds act as a key design element by organizing signaling modules on a common surface, helping to insulate potentially promiscuously interacting signal modules, thereby increasing specificity **(Figure 1.2)** (Good 2011). As an example: in *Saccharomyces cerevisiae* mating, proper co-localization of multiple kinase modules upon a structural scaffold leads to the correct signal cascade (Good 2011). In this system, scaffolding helps dictate a specific local environment for correct transfer of phosphoryl groups between modules. This feature serves to insulate processes from each other that rely on potentially promiscuous partners. This property also makes it so that from a relatively few-number of signaling modules, novel inputs/outputs emerge simply by exchanging modules upon the scaffolding surface (Good 2011; Bashor 2010) Modifying binding site availability or altering topology in some other fashion acts as another mechanism scaffolds selectively sculpt signal cascades (Good 2011; Bashor 2010).

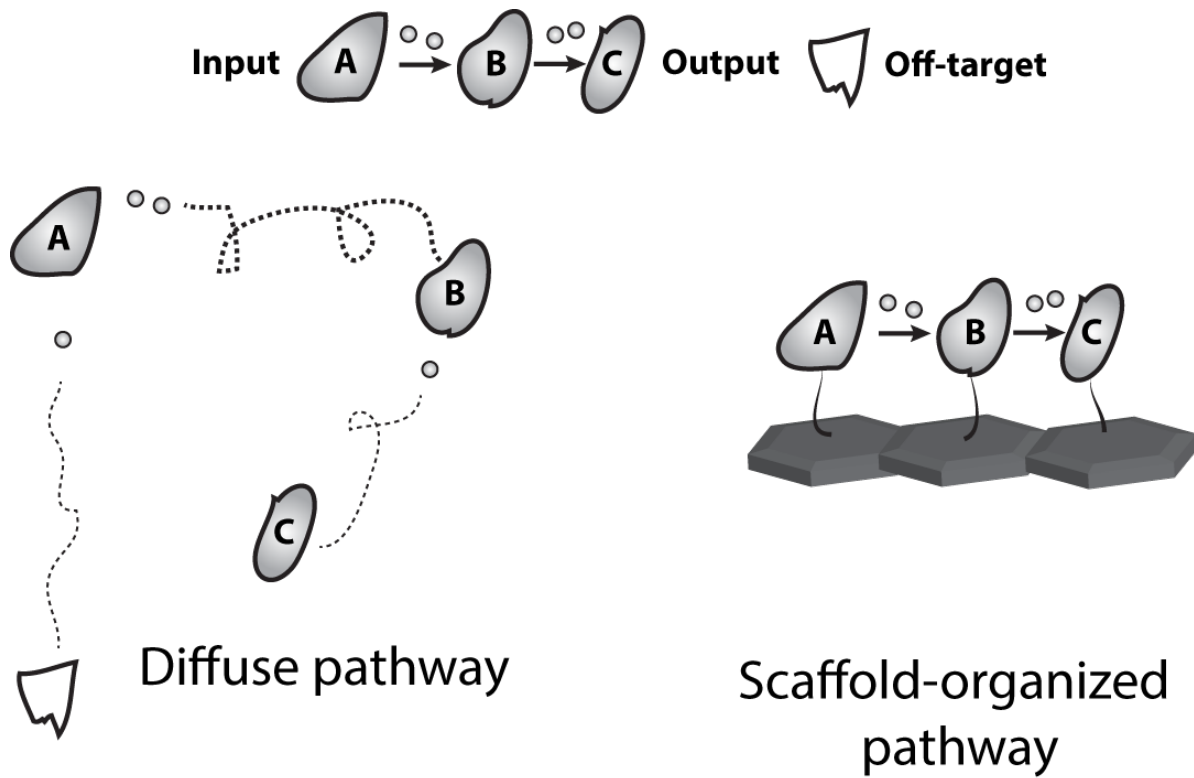


Figure 1.2: Colocalizing a pathway via a scaffolding spatiotemporal strategy. By bringing together common pathway elements on a scaffold, intermediate channeling may occur, boosting output and specificity.

Another ripe area for applying designer protein scaffolds involves co-localizing components of a shared metabolic pathway (Conrado 2008; Lee 2012). Colocalizing enzymatic pathway components upon a scaffold can lead to increases in overall output of a metabolic pathway (Idan 2013). Coordination may increase intermediate “channeling” behaviors that increases local concentration of metabolites and decreases the access of intermediates to other pathways. A seminal application of a designer metabolic scaffold tried to capture on this principle by creating a scaffold composed of a string of protein-protein adaptor domains (Dueber 2009). This *de novo* scaffold possessed adaptor domains designed to interact with other cytosolic proteins tagged with a corresponding ligand domain. IN this manner, th scaffold intended to coordinate a set of otherwise separate proteins by recruiting them all to the same local microenvironment within the cell (**Figure 1.2**). Researchers sought to change the number adaptor domains within the scaffold to change the stoichiometry of the enzymes bound in hopes of alleviating bottle-necks in a rate-limiting enzyme. Scanning this stoichiometry eventually produced a 77-fold product enhancement. Many other research projects support the efficacy of scaffold-organized enzymatic pathways, yet noted variability in output achieved occurs (Horn 2015; Chen 2014; Siu 2015).

Still, scientists overall have reported a variety of other biological scaffolding systems, each with their own advantages and drawbacks. In some cases, scaffold proteins that existed in the published literature (*e.g.*, the signaling scaffold in the yeast mating response pathway discussed above) had additional recruitment domains appended to them so that they would act as a scaffold for new regulatory proteins (Good 2011; Bashor 2010; Won 2011). Results from these studies demonstrated how critical a role scaffolding plays in the control of

information flow through signaling pathways, however such materials do not readily translate to alternative signaling pathways outside of the endogenous pathway. Rapid advances in the predictive assembly of DNA or RNA based oligomers has spawned a mature field of “DNA origami,” which permits a very precise level of control over the types of higher-order structures formed (Wang 2017). Results from such nucleic acid-based metabolic pathway scaffolds support the concept that precise structural arrangements between enzymes can play a critical role in realizing maximum efficiency of a pathway (Fu 2012), yet technical limitations prevent the assembly of most DNA origami designs within living cells (see Chapter 2 for a more complete discussion of alternative scaffold designs). By comparison, early protein-based scaffolds possessed utility for *in vivo* applications, but suffered in structural precision because of inherent design limitations (i.e. scaffolds likely form unspecific aggregates inside cells from an absence of structural integrity) (Whitaker 2011).

So, even though progress has progressed in designing protein scaffolds, some noted additional areas stand out for improvement. Collecting structural and dynamical information of scaffolds and their cargo inside cells identifies as one major area. Producing scaffold-building materials that predictively assemble and organize cargo, another. Reliable translation of scaffolding systems from one chassis organism to another, or to different pathways, has also hampered usage. Ideally models will reach sub-nm levels of structural information and predictive understanding, poising functions for maximum efficiency.

Building Next-generation Designer Scaffolds

Next-generation designer protein scaffolding systems could benefit from improvements that would enable wider-spread adoption for a variety of application. For example, we imagine structural definition, predictive organization, modularity, and tunability as playing integral roles. One way to build this system, whether for signaling, metabolic, or some other scaffolding application, relies on using a bottom-up fabrication mentality. This involves selecting a desired building material for the scaffold, along with a mode of interaction for cargo.

Relying on structurally-characterized protein-domains as the building material for a scaffolding system represent a valuable resource to inspire bottom-up fabrication—those with known higher-order assembly behavior appear even more worthwhile for some projects. Some notable examples include: viral coat proteins, bacterial S-layer proteins, and bacterial microcompartment shell proteins (Selivanovitch 2019; Pum 2014; Planamente 2019). In this PhD, we selected a protein domain (pfam00936 “bacterial microcompartment shell protein domain”) and addressed suitability for the bottom-up design of a scaffolding system. This domain meets many idealized criteria for a scaffold building material, but questions remain before wide-spread application (discussed further in Chapter 2). This project sought to supplement the field by establishing design principles in the self-assembly and higher-order organization of pfam0936-domain containing proteins, while also working towards designing rational scaffolding behaviors. Principally, we explored fundamental questions in protein structure/function and how engineering in a cellular environment may factor in to designs.

Another critical aspect in progressing next-generation scaffolding systems—besides relying on structurally characterized building blocks—involves exploring the dynamic properties of scaffolds and cargos inside cells. This comes to the forefront because the vast majority of cellular processes appear dependent on coordinating partners against the tide of intracellular forces (*e.g.* Brownian motion). Additionally, because cells typically function in a non-equilibrium state, by definition the cellular interior itself dynamically changes over-time in response to internal and environmental cues. Therefore, any engineered cellular process must perform in the context of these dynamics by considering different objectives dependent on the current environment and/or cell state. Current scaffolding systems lack any visualizing information on intrinsic dynamics to scaffold and cargo behavior within living cells. This proves critical because any truly predictive system will require this information. Looking forward to the next-generation of designer scaffolds, it feels ideal to not only have information in scaffold-cargo dynamics across intracellular space-time, but also to utilize dynamically responsive features as part of the design. For example, one could consider triggering scaffold organization only within a particular growth phase to promote heterologous metabolism only at high-cell density. Similarly, timing scaffold organization to sequester particular signaling modules or transcription factors to change the cell cycle could act as another. These types of dynamic scaffolds could greatly expand the utility of designer “logic gates” built from research projects by activating heterologous systems with greater precision when a certain set of defined conditions exist.

In hopes of addressing structural and dynamic properties, robustly assessing designs via inter-disciplinary techniques forms a cornerstone for successful adoption of any next-

generation scaffolding system. Philosophy from synthetic biology, molecular biology, biophysics, computational biology, and bioinformatics came together to help shape the experiments and theory behind my project. Visualizing approaches—such as light microscopy, atomic force microscopy, particle scattering techniques, and electron microscopy—danced together to produce data addressing self-assembly and scaffolding behavior of pfam0396-domain containing proteins. Additionally, computational simulations provided insight into potential inter-domain dynamics. In this chapter, we exemplified spatiotemporal organization systems within cells to help communicate underlying design principles, alongside some target areas for improving next-generation scaffolding systems. The second chapter represents a peer-reviewed perspective article on engineering pfam0936-domain containing proteins for scaffolding function. The third a collection of data and interpretations of *in vivo* dynamics in a scaffolding system designed from pfam0936-domain proteins. The fourth and appendix materials, brief remarks on paths forward and sample funding proposals. I truly hope these materials serve as a conduit for inspiring the next-generation of nanoscale spatiotemporal organization in living systems.

LITERATURE CITED

LITERATURE CITED

- Agapakis, C. M., Boyle, P. M. & Silver, P. A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology* 8, 527–535 (2012).
- Ahnert, S. E., Marsh, J. A., Hernandez, H., Robinson, C. V. & Teichmann, S. A. Principles of assembly reveal a periodic table of protein complexes. *Science* (2015).
- Alberti, S. Phase separation in biology. *Curr. Biol.* 27, R1097–R1102 (2017).
- Bai, Y., Luo, Q. & Liu, J. Protein self-assembly via supramolecular strategies. *Chem Soc Rev* (2016).
- Bashor, C. J., Horwitz, A. A., Peisajovich, S. G. & Lim, W. A. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* 39, 515–537 (2010).
- Bisson-Filho, A. W. *et al.* Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science* 355, 739–743 (2017).
- Castellana, M. *et al.* Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat Biotechnol* 32, 1011–1018 (2014).
- Chen, R. *et al.* Biomolecular scaffolds for enhanced signaling and catalytic efficiency. *Current Opinion in Biotechnology* 28, 59–68 (2014).
- Conrado, R. J., Varner, J. D. & DeLisa, M. P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current Opinion in Biotechnology* 19, 492–499 (2008).
- Diekmann, Y. & Pereira-Leal, J. B. Evolution of intracellular compartmentalization. *Biochem. J.* 449, 319–331 (2013).
- Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* 27, 753–759 (2009).
- Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* 26, 597–604 (2001).
- Feig, M., Yu, I., Wang, P.-H., Nawrocki, G. & Sugita, Y. Crowding in cellular environments at an atomistic level from computer simulations. *J Phys Chem B* 121, 8009–8025 (2017).
- Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme substrate diffusion for an

- enzyme cascade organized on spatially addressable DNA nanostructures. *J. Am. Chem. Soc.* 134, 5516–5519 (2012).
- Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332, 680–686 (2011).
- Govindarajan, S. & Amster-Choder, O. Where are things inside a bacterial cell? *Current Opinion in Microbiology* (2016).
- Haeusser, Daniel P., and William Margolin. Splitsville: structural and functional insights into the dynamic bacterial Z ring. *Nature Reviews Microbiology* 14.5 (2016).
- Horn, A. & Sticht, H. Synthetic protein scaffolds based on peptide motifs and cognate adaptor domains for improving metabolic productivity. *Frontiers in Bioengineering and Biotechnology* (2015).
- Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Current Opinion in Biotechnology* 22, 485–491 (2011).
- Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* 537, 320–327 (2016).
- Idan, O. & Hess, H. Origins of activity enhancement in enzyme cascades on scaffolds. *ACS Nano* 7, 8658–8665 (2013).
- Jakobson, C. M., Tullman-Ercek, D. & Mangan, N. M. Spatially organizing biochemistry: choosing a strategy to translate synthetic biology to the factory. *Sci Rep* 8, 8196 (2018).
- Karsenti, E. Self-organization in cell biology: a brief history. *Nat. Rev. Mol. Cell Biol.* 9, 255–262 (2008).
- Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein–protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* 108, 1225–1244 (2008).
- King, N. P. & Lai, Y.-T. Practical approaches to designing novel protein assemblies. *Curr. Opin. Struct. Biol.* 23, 632–638 (2013).
- Küchler, A., Yoshimoto, M., Luginbühl, S. & Mavelli, F. Enzymatic reactions in confined environments. *Nature* (2016).
- Lane, N. The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 370, (2015).
- Lee, H., DeLoache, W. C. & Dueber, J. E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14, 242–251 (2012).

- Ligrone, R. Biological Innovations that Built the World. *Springer* (2019).
- Loose, M., Kruse, K. & Schwille, P. Protein Self-organization: lessons from the min system. *Annu Rev Biophys* 40, 315–336 (2011).
- Luo, Q., Hou, C., Bai, Y., Wang, R. & Liu, J. Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* 116, 13571–13632 (2016).
- Murat, D., Byrne, M. & Komeili, A. Cell biology of prokaryotic organelles. *Cold Spring Harbor Perspectives in Biology* 2, a000422 (2010).
- Planamente, S. & Frank, S. Bio-engineering of bacterial microcompartments: a mini review. *Biochem. Soc. Trans.* 47, 765–777 (2019).
- Polka, J. K., Hays, S. G. & Silver, P. A. Building spatial synthetic biology with compartments, scaffolds, and communities. *Cold Spring Harbor Perspectives in Biology* (2016).
- Popkin, G. The physics of life. *Nature* 529, 16–18 (2016).
- Pum, D. & Sleytr, U. B. Reassembly of S-layer proteins. *Nanotechnology* 25, 312001 (2014).
- Rice, M. K. & Ruder, W. C. Creating biological nanomaterials using synthetic biology. *Sci Technol Adv Mater* 15, 014401 (2014).
- Rudner, D. Z. & Losick, R. Protein Subcellular Localization in Bacteria. *Cold Spring Harbor Perspectives in Biology* (2010).
- Schavemaker, P. E., Boersma, A. J. & Poolman, B. How important is protein diffusion in prokaryotes? *Front Mol Biosci* 5, 93 (2018).
- Schwille, P. Bottom-up synthetic biology: engineering in a tinkerer's world. *Science* 333, 1248–1252 (2011).
- Selivanovitch, E. & Douglas, T. Virus capsid assembly across different length scales inspire the development of virus-based biomaterials. *Curr Opin Virol* 36, 38–46 (2019).
- Siu, K.-H. *et al.* Synthetic scaffolds for pathway enhancement. *Current Opinion in Biotechnology* 36, 98–106 (2015).
- Surovtsev, I. V. & Jacobs-Wagner, C. Subcellular organization: a critical feature of bacterial cell replication. *Cell* 172, 1271–1293 (2018).
- Szwedziak, P., Wang, Q., Bharat, T. A. M., Tsim, M. & Löwe, J. Architecture of the ring formed by the tubulin homologue FtsZ in bacterial cell division. *eLife* (2014).

- Ulijn, R. V. & Jerala, R. Peptide and protein nanotechnology into the 2020s: beyond biology. *Chem Soc Rev* 47, 3391–3394 (2018).
- Wang, Pengfei, et al. The beauty and utility of DNA origami. *Chem* (2017).
- Whitaker, W. R. & Dueber, J. E. Metabolic pathway flux enhancement by synthetic protein scaffolding. *Methods Enzymol* 497, 447–468 (2011).
- Won, Angela P., Joan E. Garbarino, and Wendell A. Lim. Recruitment interactions can override catalytic interactions in determining the functional identity of a protein kinase. *Proceedings of the National Academy of Sciences* (2011).
- Yang, L. *et al.* Self-assembly of proteins: towards supramolecular materials. *Chemistry* (2016).
- Yang, X. *et al.* GTPase activity–coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. *Science* 355, 744–747 (2017).
- Yeates, T. O. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu Rev Biophys* 46, 23–42 (2017).
- Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nat Biotechnol* 21, 1171–1178 (2003).

Chapter 2: A literature perspective on designing nanoscaffolds with a self-assembling protein domain

This text exists repurposed with permission from Frontiers in Microbiology under a Creative Commons Attribution License. The original published text exists at: <https://doi.org/10.3389/fmicb.2017.01441>.

Engineering the Bacterial Microcompartment Domain for Molecular Scaffolding Applications

Eric J. Young^{1,2}, Rodney Burton², J.P. Mahalik^{3,4}, Bobby G. Sumpter^{3,4}, Miguel Fuentes-Cabrera^{3,4}, Cheryl A. Kerfeld^{1,2,5} and Daniel C. Ducat^{1,2}

1 Michigan State University, Biochemistry and Molecular Biology; **2** MSU-DOE Plant Research Lab, (East Lansing, Michigan, United States); **3** Oak Ridge National Laboratory, Computational Sciences and Engineering (Oak Ridge, Tennessee, United States); **4** Oak Ridge National Laboratory, Center for Nanophase Material Sciences (Oak Ridge, Tennessee, United States); **5** Berkeley National Lab, Molecular Biophysics and Integrated Bioimaging Division (Berkeley, California, United States)

Abstract

As Synthetic Biology advances the intricacy of engineered biological systems, the importance of spatial organization within the cellular environment must not be marginalized. Increasingly, biological engineers are investigating means to control spatial organization within the cell, mimicking strategies used by natural pathways to increase flux and reduce cross-talk. A modular platform for constructing a diverse set of defined, programmable architectures would greatly assist in improving yields from introduced metabolic pathways and increasing insulation of other heterologous systems. Here, we review recent research on the shell proteins of bacterial microcompartments (BMCs) and discuss their potential application for the tile-like assembly of a range of intracellular scaffolds. We summarize the state of knowledge on the self-assembly of BMC shell proteins and discuss future avenues of research that will be important to realize the potential of BMC shell proteins as predictive and programmable biological materials for bioengineering.

Introduction

With the advent of Synthetic Biology and recent advances in protein engineering, designing, constructing, and controlling biomolecule-based materials at the nanoscale is a rapidly developing field. Currently, there is a lack of modular building blocks for predictably fabricating custom sub-cellular architectures that can be subsequently programmed with precise functions (**Figure 2.1A**). Because of their self-assembly properties, proteins containing the pfamdomain 00936 (pfam0936)—known as the bacterial microcompartment (BMC) domain—offer one unique approach towards the predictive

engineering of biomaterials *in vivo*. In this perspective, we discuss current research in BMC-domain proteins in the context of their potential for the construction of custom nano-architectures and intracellular scaffolds.

Organization of interrelated cellular components in time and space is crucial to increase efficiency of diverse cellular processes, including processes in metabolism, signaling, and division (Agapakis et al. 2012; Pawson & Scott 1997). Typically, cells colocalize components within a common pathway, conferring a host of benefits that include increased enzymatic intermediate flux and limited pathway cross-talk (Good et al. 2011; Agapakis et al. 2012). Biological engineers have increasingly explored a variety of rational colocalization strategies to capitalize on such benefits. These engineered systems range in complexity from simple fusion proteins to dynamic artificial scaffolds (Horn & Sticht 2015; Myhrvold & Silver 2015; Conrado et al. 2008) or compartments (Giessen & Silver 2016). As biologists move towards increasingly complex cellular engineering goals (Bashor et al. 2010), one challenge is designing sophisticated subcellular colocalization approaches that recapitulate the elegance of natural systems (Good et al. 2011). We focus here on molecular scaffold construction.

Polymerizing biomolecules represent ideal building blocks because they can self-assemble into higher-order arrangements *in vivo*. To date, DNA hybridization nanotechnology (e.g. DNA origami) is perhaps the best developed molecular building platform (Pinheiro et al. 2011). DNA architectures are especially flexible under non-physiological conditions where a nearly limitless array of architectures can be predictively constructed and controlled at scales approaching the sub-nanometer (Fu et al. 2012; Funke & Dietz 2016; Wilner et al.

2009). Yet translating this technology to intracellular application has been partially constrained because the concentration of single-stranded nucleic acid building blocks and environmental properties important for nucleic acid folding (e.g. temperature, ions) are not easily manipulated *in vivo* (Pinheiro et al. 2011). While recent studies continue to advance the capability of nucleic acid assemblies achieved within the cell (Conrado et al. 2012; Siu et al. 2015; Elbaz et al. 2016; Myhrvold & Silver 2015), proteins may offer another viable, naturally-inspired solution. One early example of a synthetically designed scaffold was comprised of a string of protein-protein interaction domains that were used to recruit three cognate enzymes involved in the conversion of acetyl-CoA to melavonate (Dueber et al. 2009). Co-recruitment of these enzymes substantially increased the melavonate yield *in vivo*, yet only marginal improvements were reported when this approach was used for other metabolic pathways (Horn & Sticht 2015). One proposed reason that this strategy is not widely successful is that this design lacks an inherent organized structure and may aggregate in unpredictable ways, hindering the rational design process (Lee et al. 2012). A protein-based scaffold of predictable modularity would be composed of defined subunits which self-assemble into a well-defined structure, but can also then be tuned to produce a variety of useful architectures.

Towards this goal, engineering naturally found proteins which self-assemble into defined, nano to macromolecular architectures offers a powerful base to approach artificial scaffold construction (Howorka 2011); the components of bacterial microcompartments (BMCs) are particularly promising in this regard (Kerfeld & Erbilgin 2015). In their native context, BMCs encapsulate related metabolic enzymes within a unique self-assembled protein shell (Axen et al. 2014; Yeates et al. 2010; Kerfeld et al. 2010; Kerfeld & Erbilgin 2015). While the

compliment of BMC shell proteins physiologically form polyhedral compartments, many BMC-domain containing proteins self-assemble into a variety of higher-order structures when expressed in isolation (Havemann et al. 2002; Pang et al. 2014; Sutter et al. 2015; Dryden et al. 2009; Pitts et al. 2012; Parsons et al. 2010; Noël et al. 2015; Lassila et al. 2014; Held et al. 2016; Parsons et al. 2008; Kerfeld et al. 2005). Loci encoding BMC-domain proteins are found in at least 23 bacterial phyla, while each instance having a minimal of three unique pfam0936 domain containing proteins (Axen et al. 2014). This diversity likely includes many new “building blocks” for constructing a multitude of novel, programmable architectures, but unlocking the true potential of the pfam0936 domain will require a deeper understanding of the fundamentals governing self-assembly. We propose that the establishment of design principles—rules which result in a defined, predictable assembly—for the pfam0936 domain will provide the foundation for creating an array of nano to macromolecular structures. These designer structures may then be functionalized to cater to their individual application (**Figure 2.1A**). We discuss the promise and potential limitations of this strategy below.

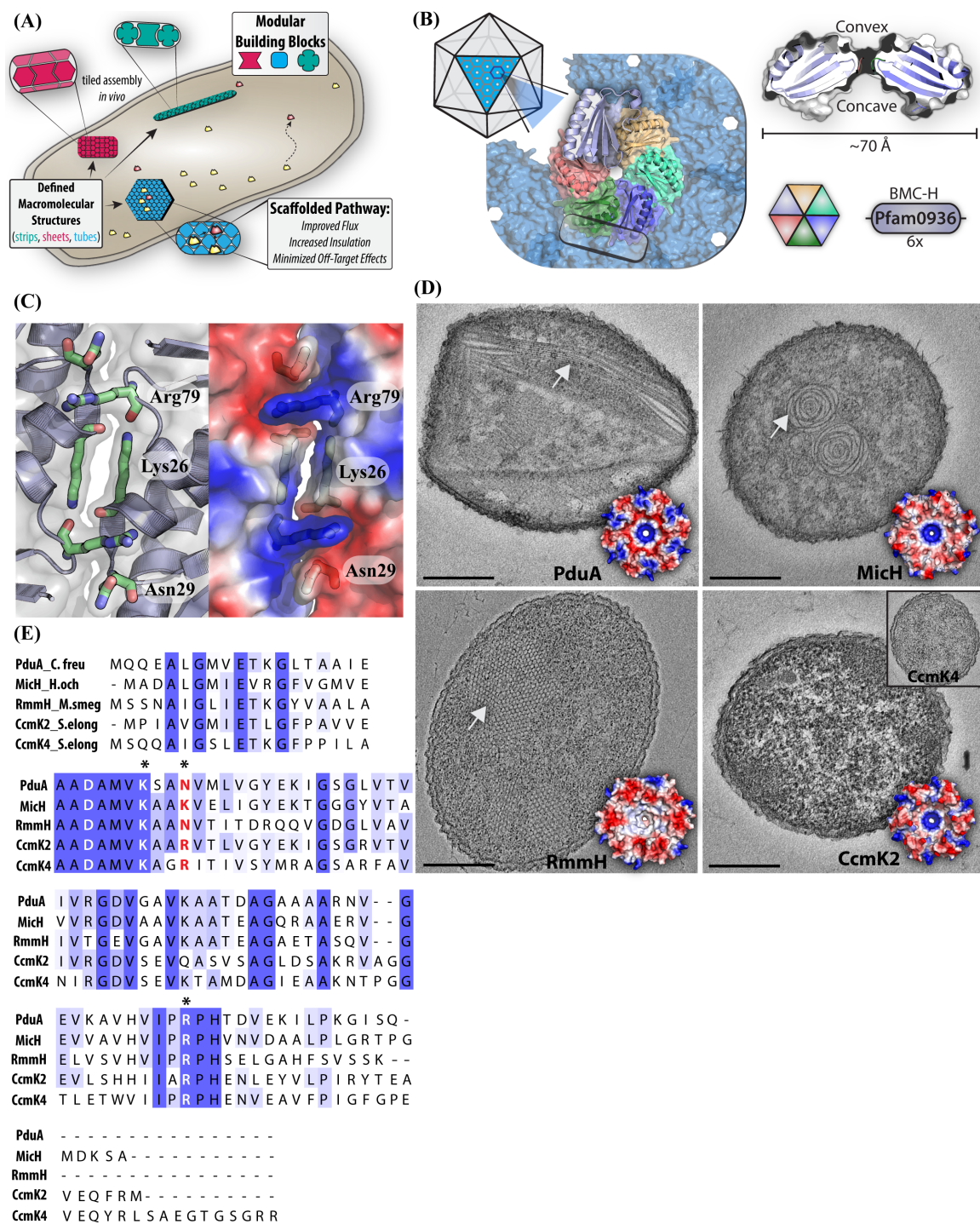


Figure 2.1: BMC-H attributes and potential as modular building blocks. A.) Cartoon schematic depicting distinct BMC hexamers (red, blue, and green) assembling into modular intracellular architectures that can recruit and concentrate cytosolic proteins (yellow and

Figure 2.1 (cont'd) orange). **B.)** General features of BMC hexamers are highlighted through the example protein, PduA. A cross section of a hexamer (right) illustrates the conserved shape and pore, while a hexamer is shown as part of a larger facet (left) that assembles through hexamer-hexamer contacts (box). **C.)** An expanded view of the interface between two BMC-H proteins (PduA, PDB: 3NGK), highlighting electrostatic interactions mediated by key residues (blue = positive). **D.)** Transmission electron microscopy of different assembly of heterologously expressed BMC-H protein in *E. coli* [PduA: Nanotubes, MicH: Rosettes, RmmH: Nanotubes, CcmK2: Lack of Structure, Inset: CcmK4. Scale bar 250 nm. **E.)** Multiple sequence alignment of representative BMC-H proteins. Asterisks indicate key residues positioned at the hexamer-hexamer interface.

Characteristics and Self-Assembly of BMC-H Shell Proteins

Numerous crystal structures of pfam0936-containing proteins have contributed to a detailed structural understanding of BMC shell proteins and models of how they “tile” into the facets of BMCs (**Figure 2.1B**) (summarized in (Kerfeld & Erbilgin 2015)). The signature domain of BMCs has little structural variation across the multitude of functionally distinct and distantly related BMCs, indicating a pivotal role in assembling the BMC shell (Crowley et al. 2010; Kinney et al. 2011). The main constituent of BMC shells are typically small (~100 amino acids) proteins containing the BMC domain (BMC-H) which form a ~70 Å hexagonal disc with distinct faces and a circular pore in the center (**Figure 2.1B**) (Yeates et al. 2010; Kerfeld et al. 2010); other components of BMC shells are BMC-T (containing a tandem fused copy of pfam0936) and BMC-P (pfam03319) proteins (Kerfeld & Erbilgin

2015), but are not a focus of this perspective. The concave side of BMC-H proteins features a surface depression that can harbor both flexible extensions of the N and C protein termini, whereas the convex side has varied electrostatic properties across homologs **(Figure 2.1B)** (Yeates et al. 2010; Kerfeld et al. 2010).

In many crystal structures, a subset of residues found along the edge periphery mediate an inter-hexamer hydrogen bond network, permitting tiled assembly of conjoined arrays **(Figure 2.1C)** (Kinney et al. 2011; Kerfeld et al. 2005; Pitts et al. 2012; Tanaka et al. 2008; Klein et al. 2009; Crowley et al. 2010; Cai et al. 2015; Takenoya et al. 2010). These edge residues (DxxK, RPH) are widely conserved throughout BMC-H proteins and thus, imply a crucial role in maintaining hexamer-hexamer interactions (Kerfeld & Erbilgin 2015). For example, in the crystal structure of a PduA lattice—a canonical example—the antiparallel association of two adjacent lysine residues mediates the bulk of the inter-hexamer association **(Figure 2.1C)** (Crowley et al. 2010; Sinha et al. 2014; Pang et al. 2014). Although, the buried interaction surface area at the hexamer-hexamer interface is typically less than other protein-protein interfaces, it is likely that the multiplicative nature of the interaction (1 hexamer surrounded by six others) provides sufficient cooperativity to permit higher-order arrays (Crowley et al. 2010). Since tiling behavior with consistent inter-hexamer distances has been observed in BMC-H sheets by high-resolution microscopy techniques (Dryden et al. 2009; Sutter et al. 2015), the interface observed in crystals is likely physiologically relevant.

The flexibility of high-order formation of BMC-H homologs begins to take shape outside the confining context of a crystalline array. Many distinct architectures can be formed by

purified BMC-H proteins *in vitro*, including: 100 nm spheroids (Keeling et al. 2014; Kerfeld et al. 2005), extended nanotubes (Noël et al. 2015), and honeycombed tiles (Lassila et al. 2014). The macromolecular assembly behavior and the formation of such high-order structures are influenced by pH and ionic strength (Dryden et al. 2009; Noël et al. 2015; Jorda et al. 2016). Similarly, overexpression of BMC-H proteins *in vivo* leads to the self-assembly of a myriad of higher-order structures inside the cells, including: tubes (Parsons et al. 2010; Noël et al. 2015; Pang et al. 2014), filaments (Pang et al. 2014; Heldt et al. 2009; Parsons et al. 2008; Havemann et al. 2002), and other structures (Parsons et al. 2010; Pitts et al. 2012; Sutter et al. 2015; Parsons et al. 2008; Held et al. 2016; Lin et al. 2014).

Because the methodology used to express BMC-H proteins varies among labs and studies (e.g. host, promoter strength, protein concentration, growth condition, sample preparation), it is not always clear if the distinct intracellular structures generated by BMC-H homologs in separate reports are due to intrinsic self-assembly properties, or the specific experimental conditions. Nonetheless, multiple lines of evidence suggest that properties of BMC-H proteins predispose them towards specific higher-order architectures (Sinha et al. 2014; Pang et al. 2014). To illustrate this point, we heterologously expressed a panel of BMC-H homologs from distinct BMCs under identical conditions in *E. coli*. We find that expression of BMC-H homologs PduA, MicH, RmmH, and CcmK2 form varied macromolecular assemblies *in vivo* (**Figure 2.1D**), generally in agreement with prior reports. PduA and RmmH form nanotube-like structures (Noël et al. 2015; Pang et al. 2014) and MicH (5815 BMC-H) forms “swiss roll” rosettes thought to be an extended sheet of rolled up protein (Sutter et al. 2015). Despite orderly tiling in CcmK2 and CcmK4 crystal structures (*Synechococcus elongatus* PCC 7942), over-expressing these homologs in *E. coli*

does not lead to the formation of prominent macromolecular structures (**Figure 2.1D**). It is unclear if the absence of visible structures via transmission electron microscopy (TEM) thin section represents a lack of higher-order self-assembly, or if smaller assemblies are formed in the cytoplasm which are insufficiently discriminated from other cytoplasmic elements; as was previously proposed for other smaller BMC assemblies (Lassila et al. 2014). Collectively, it appears heterologously expressed BMC-H proteins form an assortment of *in vivo* assemblies, but exactly how the intrinsic features of each homolog (differences in primary structure) contribute to differences in self-assembly are currently unknown.

Towards Understanding the Relationship Between Primary Structure and Macromolecular Assembly of BMC-H Proteins

As evident by the diversity of structures formed by BMC-H proteins (**Figure 2.1D**), there must be subtle primary structure differences that dictate changes in higher-order assembly. One region anticipated to influence assembly dynamics surrounds residues at the inter-hexamer junction; although there is strict conservation of some sidechains at this interface, some positions exhibit variance across homologs (**Figure 2.1E**) (Cai et al. 2015). For example, both CcmK2 and CcmK4 contain arginine in comparison to the asparagine residues of PduA, RmmH or lysine of MicH (**Figure 2.1E, red text**). Supporting this hypothesis, experimental evidence generated by targeted amino acid substitutions at interface residues of PduA indicate they influence macromolecular assembly *in vivo* (Pang et al. 2014). Other studies of BMC-H proteins with modified hexamer-hexamer interface

residues show they alter the formation of isolated BMCs (Sinha et al. 2014), size of tiled arrays *in vitro* (Sutter et al. 2015), or disrupt crystal packing contacts and orientation (Sinha et al. 2014; Pang et al. 2014).

In addition to residues at the hexamer-hexamer interface, other hexamer features may dictate the self-assembly behavior of BMC-H homologs. It has been well documented that the overall electrostatic surface profiles varies significantly among homologs (**Figure 2.1D**) (Kinney et al. 2011; Kerfeld et al. 2010). Electrostatic differences—known to affect the self-assembly of proteins (Keskin et al. 2008)—could influence the preferred interaction orientation between BMC-H proteins, predisposing them to a particular assembly architecture. Besides the overall electrostatic profile, other unique regions in primary structure of BMC-H homologs could manipulate self-assembly; one such region is the variable C-terminal region (**Figure 2.1E, boxed**). Longer C-terminal extensions (**Figure 2.1E, CcmK4**) originally were hypothesized to interfere with lateral molecular tiling through steric clash, a hypothesis partially supported by the observation of that truncation mutants form hexamers which pack more tightly in crystal lattices (Tanaka et al. 2009). Some crystal forms of CcmK2 orthologs appear to have hexamers which are stacked upon each other (dodecamer) interacting through the C-terminal extensions (Tanaka et al. 2009; Samborska & Kimber 2012). Although the physiological significance of these crystal contacts is uncertain, *in vitro* analysis of the molecular weight (Tanaka et al. 2009) and FRET interaction (Samborska & Kimber 2012) of CcmK2 with modified C-termini provide some supporting evidence of a functional role. However, it should be noted that C-terminal truncation does not disrupt the formation of heterologous BMC shells in the presence of other BMC shell components (Cai et al. 2016).

To establish detailed models for oligomeric BMC-H self-assembly, dynamic techniques that can interrogate the differences in the nucleation and expansion of shell protein arrays are required. High-speed atomic force microscopy (HS-AFM) is one emergent technique because of the high spatial and temporal resolution it affords. HS-AFM was used to capture the individual changes in BMC-H proteins association/dissociation rates into larger sheets (Sutter et al. 2015). Dynamic light scattering, and complimentary biophysical techniques can also quickly assess particle size (from single molecules to large assemblies) based on changes in the optical properties creating a more high-throughput pipeline to evaluate factors controlling assembly; as recently employed in a reengineered PduA variant (Jorda et al. 2016).

Computational frameworks and molecular dynamics offer another potentially powerful tool for predicting and understanding the behavior of BMC shell proteins. One recent study simulated the steps of BMC assembly by utilizing a computational model that varied the strength of hexamer-hexamer and hexamer-cargo affinity (Perlmutter et al. 2016). From this, two classes of BMC assembly emerged that proceeded through distinct hierarchies (Perlmutter et al. 2016). Another recent example used Monte Carlo simulations with a coarse-grained potential to study the 2D self-assembly of CcmK2 (*Synechocystis* sp PCC 6803) (Mahalik et al. 2016). In these simulations, 2D sheets were found to form rapidly after the association of an initial clustering of four hexamers, suggesting that self-assembly is rate-limited by a nucleation event (Mahalik et al. 2016). In turn, nucleation rates strongly depended on the concentration of hexamers and their relative 2D orientation upon collision (Mahalik et al. 2016).

To illustrate of how subtle differences in primary structure influences hexamers' self-assembly, we performed preliminary simulations of the initial steps of the 3D self-assembly with two different BMC-H homologs (**Figure 2.2A-C**). Employing the Thomas-Dill (Thomas & Dill 1996) coarse-grained potential—found to best approximate the fully atomic potential in (Mahalik et al. 2016)—to compute the angular dependence in the potential of mean force (PMF) between two RmmH hexamers (*Mycobacterium smegmatis*) or two CcmK2 hexamers (*S. elongatus* 7942). PMF was calculated as a function of the distance between the center of mass of each hexamer tile and the angle θ (defined in Figure 2.2A). CcmK2 and RmmH structures were obtained by relaxing the corresponding crystal structure in an aqueous environment. The angular dependence of PMF for CcmK2 and RmmH pairs is shown in Figure 2.2B; the PMF of RmmH has a clear minimum at $\theta \sim -45^\circ$, whereas CcmK2 is, by comparison, flatter. This association angle is consistent with the formation of a regularly repeating curved surface akin to models created of *in vivo/in vitro* RmmH nanotubes (Noël et al. 2015). In contrast, *in vitro* macromolecular assemblies of CcmK2 orthologs (Keeling et al. 2014), depict flexibility in the overall macromolecular structure supporting a lack of defined interaction orientations. The differences in angular dependence of PMF is further demonstrated if one uses different initial structures of the same BMC-H protein. This is illustrated in Figure 2.2C for CcmK2. In this figure, the PMFs of CcmK2 pairs formed from either the crystal or the relaxed structure are shown. The PMF of the crystalline pair has a clear angular minimum ($\theta \sim -80^\circ$), where the PMF of the relaxed pair lacks any clear minimum. This is notable because the backbone root mean-square deviation between the two structures only differs by 0.67 Å. The distinct PMFs are caused by differences in side-chain rotamers altered in the relaxation procedure. These small

structural changes are enough to cause a difference of almost 1 $K_B T$, indicating that subtle structural changes which may occur between crystallization and *in situ* conditions can have a profound impact on self-assembly; also the overall lower PMF values for CcmK2 suggests a net weaker hexamer-hexamer interaction, consistent with other predictions (Perlmutter et al. 2016).

Although these observations of hexamer behavior in computational frameworks are preliminary, they are illustrative of the potential for *in silico* techniques to aid in understanding and predicting behavior across several scales (from single molecules to large arrays). Combining computational approaches with experimental methods could help define general design rules for predicting the architecture formed by BMC-H proteins.

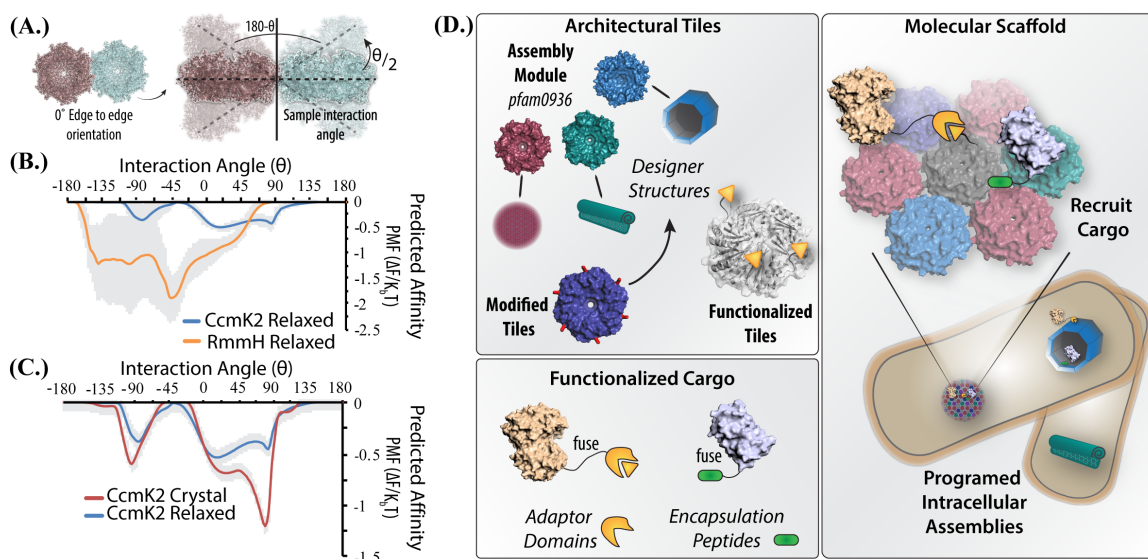


Figure 2.2: Could molecular-level simulations contribute toward predictive assembly of diverse BMC-H scaffolds? **A.)** Illustration of design of molecular dynamics simulations where the potential of mean force (PMF) is calculated from two adjacent hexamers. Keeping the relative orientation fixed, the hexamers are systematically rotate out of the plane by an angle $\theta/2$ and the change in PFM is recalculated **B.)** Differences in the predicted PMF depending on the inter-hexamer angle are shown for solvated crystals structures of RmmH and CcmK2 (standard deviation is depicted in gray). **C.)** Differences in the angular PMF profile can solely arise by comparing the crystal structure versus solvated structure **D.)** Illustration of pipeline for constructing BMC-H based programmable nanostructures. BMC-H proteins with different assembly characteristics can be selected from existing homologs (magenta, green, and blue) or created by modification of key residues (red) and modified to encode protein interaction domain (orange). Enzymes and other cargo can be directed to BMC-H assemblies by fusing corresponding ligand domain, or the use of native encapsulation peptides (green). In this manner, it is feasible to envision

Figure 2.2 (cont'd) a diversity of subcellular protein architectures that can be functionalized to scaffold many distinct metabolic or signaling pathways.

Functionalization of Existing Macromolecular Structures

The engineering of scaffolding platforms based on self-assembling protein modules can be considered a two-faced challenge: one side is predictably making a discrete structure, and the other is functionalizing the structure for a specific purpose. Ideally, functionalizing the surface of BMC-H protein architectures should be modular in itself, so that scaffold structures could be “repurposed” for new enzymes and pathways with minimal redesign. There are two ways in which functional proteins could be organized to BMC-H assemblies: through natural or synthetic motifs.

One approach to functionalizing heterologous assemblies is to use binding motifs BMCs natively employed to recruit cargo (Fan et al. 2012; Aussignargues et al. 2015). Frequently, BMC core proteins contain small peptides (~20 amino acids) as extensions of the N- or C-termini that are necessary for encapsulation (Kim & Tullman-Ercek 2014; Choudhary et al. 2012; Cai et al. 2016; Jakobson et al. 2015; Kinney et al. 2012; Lawrence et al. 2014; Wagner et al. 2017; Quin et al. 2016; Held et al. 2016; Lin et al. 2014; Gonzalez-Esquer et al. 2015). Collectively known as “encapsulation peptides” (EPs), modeling (Kinney et al. 2012; Fan et al. 2012) and solution structures (Lawrence et al. 2014) adopt an alpha-helical conformation with an amphipathic charge distribution (Aussignargues et al. 2015). Although EPs vary widely in primary structure, it has been demonstrated that non-native EPs can interact with non-cognate BMCs (Jakobson et al. 2015) and amino acid substitutions can alter the affinity for BMC encapsulation (Kim & Tullman-Ercek 2014), likely through conservation of specific amphipathic characteristics. While promising, general use of EP motifs for predictive recruitment is currently impaired by uncertainty in

the interface location and affinity EP-BMC component binding (Lawrence et al. 2014; Fan et al. 2012; Aussignargues et al. 2015).

An alternative strategy is appending natural or synthetically derived protein-protein interaction domains—known as adaptor domains—to BMC-H proteins. In this way, virtually any protein encoded with the cognate adaptor domain could be post-translationally concentrated, conferring specified enzymatic functions to the designer architecture. While this approach is potentially powerful, it must be determined if fusion of adaptor domains to BMC-H proteins will alter higher-order assembly. In published work, fusions to BMC-H proteins have ranged from small affinity tags (Havemann et al. 2002; Samborska & Kimber 2012; Dryden et al. 2009; Kerfeld et al. 2005) to fluorescent proteins (~26 kDa: ~2X the size of a single pfam0936 domain), and these modified BMC-H proteins still incorporate into BMCs (Savage et al. 2010; Sun et al. 2016; Cameron et al. 2013; Cai et al. 2013; Parsons et al. 2010; Cai et al. 2015; Parsons et al. 2008). Yet, fusions of certain characteristics (e.g. size, charge) could disrupt association through steric clash or compromised electrostatics, thereby changing higher-order behavior. Indeed, although a fluorescent protein fusion to major shell protein CcmK2 incorporates into functional BMCs (Cameron et al. 2013), an unmodified copy of the protein must also be present, as the fusion is unable to solely complement a full $\Delta CcmK2$ background; this is supported also by fluorescent protein fusions to CcmK2 aggregates and does not interact with other shell protein components (Lin et al. 2014). In contrast, it is notable that some shell proteins can still assemble functional BMCs without a native copy present (Parsons et al. 2010; Sun et al. 2016), suggesting that some fusions can be tolerated or other shell protein components can relieve the compromised function of the fusion (Chowdhury et al. 2016). So, it remains to

be fully elucidated how specific fusions and how certain properties (e.g. fusion orientation — N- vs. C-termini — fusion size/charge) will alter self-assembly behavior.

Future Applications and Closing Remarks

The capacity to form a spectrum of defined architectures that can be tailored through functionalization is a powerful tool for the future of Synthetic Biology. For instance, it can be anticipated that the suitability of a given architecture for scaffolding pathway components would be dependent on the overall macromolecular geometry of the assembly **(Figure 2.2D)**. Multiple lines of evidence indicate that for metabolic pathway scaffolds to reach peak efficacy close proximity of active sites and a barrier preventing metabolic intermediate diffusion are ideal—widely known as substrate channeling—so structures which accomplish this (i.e. nanotubes) would be highly desirable (Castellana et al. 2014; Whitaker & Dueber 2011; Bauler et al. 2010; Idan & Hess 2013; Wheeldon et al. 2016). Alternatively, structures with wider accessibility to the cytosol, for instance 1D or 2D geometries like filaments or large planar sheets, could be more appropriate for the colocalization of signal-transduction or redox pathways, akin to a macromolecular switchboard (Good et al. 2011). Understanding the details on BMC shell protein assembly and orientation of facets would open additional exciting possibilities for scaffolding orthogonal pathways on distinct sides of the structure.

Broadly speaking, we have outlined the promise and hurdles inherent to the use of BMC shell protein hexamers as building blocks for designer scaffold assemblies. Currently, most instances of heterologous BMC shell protein higher-order assembly are in prokaryotes,

though demonstration that BMC shell proteins can assemble in plant chloroplasts (Lin et al. 2014) indicates that this tool has potential across a wide diversity of organisms. As the influence of primary structure on higher-order structure formation is further disentangled we envision the use of the pfam0936 domain as a simple molecular building block for construction of diverse architectures that can be employed as synthetic scaffolds. Further elucidation of this and general principles of multiprotein complex formation (Murugan et al. 2015; Glover & Clark 2016; Ahnert et al. 2015) will usher in an era of true *in vivo* nanometer scale molecular engineering for the design of programmable synthetic subcellular architectures.

Acknowledgements

The authors would like to thank Dr. Alicia Withrow of the MSU Center for Advanced Microscopy for support with TEM. A portion of this work, i.e. simulations and writing of the manuscript, was conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility. The rest of the work was performed at MSU-DOE PRL, funded through the Department of Energy (Grant: DE-FG02-91ER20021).

LITERATURE CITED

LITERATURE CITED

- Agapakis, C.M., Boyle, P.M. & Silver, P.A., Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology*, 8(6), pp.527–535. (2012).
- Ahnert, S.E. et al., Principles of assembly reveal a periodic table of protein complexes. *Science* (2015).
- Aussignargues, C. et al., Bacterial microcompartment assembly: The key role of encapsulation peptides. *Communicative & Integrative Biology*, 8(3) (2015).
- Axen, S.D., Erbilgin, O. & Kerfeld, C.A., A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Computational Biology*, 10(10) (2014).
- Bashor, C.J. et al., Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annual review of biophysics*, 39, pp.515–537. (2010).
- Bauler, P. et al., Channeling by proximity: the catalytic advantages of active site colocalization using brownian dynamics. *The journal of physical chemistry letters*, 1(9), pp.1332–1335. (2010).
- Cai, F. et al., Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. *ACS synthetic biology*, 4(4), pp.444–453. (2015).
- Cai, F. et al., Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiology*, 170(3), pp.1868–1877. (2016).
- Cai, F. et al., 2013. The structure of CcmP, a tandem bacterial microcompartment domain protein from the β -carboxysome, forms a subcompartment within a microcompartment. *The Journal of biological chemistry*, 288(22), pp.16055–16063. (2013).
- Cameron, J.C. et al., 2013. Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* 155(5), pp.1131–1140. (2013).
- Castellana, M. et al., Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology*, 32(10), pp.1011–1018. (2014).
- Choudhary, S. et al., 2012. Engineered protein nano-compartments for targeted enzyme localization. *PloS one*, 7(3) (2012).
- Chowdhury, C., Chun, S. & Sawaya, M.R., The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. *Molecular*

Microbiology (2016).

Conrado, R.J. et al., DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic acids research*, 40(4), pp.1879–1889. (2012).

Conrado, R.J., Varner, J.D. & DeLisa, M.P., Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current opinion in biotechnology*, 19(5), pp.492–499. (2008).

Crowley, C.S. et al., Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. *The Journal of biological chemistry*, 285(48), pp.37838–37846. (2010).

Dryden, K.A. et al., Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein science : a publication of the Protein Society*, 18(12), pp.2629–2635. (2009).

Dueber, J.E. et al., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology*, 27(8), pp.753–759. (2009).

Elbaz, J., Yin, P. & Voigt, C.A., Genetic encoding of DNA nanostructures and their self assembly in living bacteria. *Nature communications*, 7, p.11179. (2016).

Fan, C., Cheng, S. & Sinha, S., Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proceedings of the National Academy of Sciences* (2012).

Fu, J. et al., 2012. Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *Journal of the American Chemical Society*, 134(12), pp.5516–5519 (2012).

Funke, J.J. & Dietz, H., Placing molecules with Bohr radius resolution using DNA origami. *Nature Nanotechnology*, 11(1), pp.47–52. (2016).

Giessen, T.W. & Silver, P.A., Encapsulation as a strategy for the design of biological compartmentalization. *Journal of Molecular Biology*, 428(5), pp.916–927. (2016).

Glover, D.J. & Clark, D.S., Protein Calligraphy: A New Concept Begins To Take Shape. *ACS Central Science*. (2016).

Gonzalez-Esquer, C.R., Shubitowski, T.B. & Kerfeld, C.A., Streamlined Construction of the Cyanobacterial CO₂-Fixing Organelle via Protein Domain Fusions for Use in Plant Synthetic Biology. *The Plant Cell*. (2015).

Good, M.C., Zalatan, J.G. & Lim, W.A., 2011. Scaffold proteins: hubs for controlling the flow of cellular information. *Science*, 332(6030), pp.680–686. (2011).

- Havemann, G.D., Sampson, E.M. & Bobik, T.A., PduA is a shell protein of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in *Salmonella enterica* serovar Typhimurium. *Journal of bacteriology*. (2002).
- Held, M. et al., Engineering formation of multiple recombinant Eut protein nanocompartments in *E. coli*. *Scientific Reports* (2016).
- Heldt, D. et al., Structure of a trimeric bacterial microcompartment shell protein, EutB, associated with ethanol utilization in *Clostridium kluyveri*. *The Biochemical Journal*, 423(2), pp.199–207. (2009).
- Horn, A. & Sticht, H., Synthetic protein scaffolds based on peptide motifs and cognate adaptor domains for improving metabolic productivity. *Frontiers in Bioengineering and Biotechnology* (2015).
- Howorka, S., Rationally engineering natural protein assemblies in nanobiotechnology. *Current opinion in biotechnology*, 22(4), pp.485–491. (2011).
- Idan, O. & Hess, H., Engineering enzymatic cascades on nanoscale scaffolds. *Current opinion in biotechnology*, 24(4), pp.606–611. (2013).
- Jakobson, C.M. et al., Localization of Proteins to the 1, 2-Propanediol Utilization Microcompartment by Non-native Signal Sequences Is Mediated by a Common Hydrophobic Motif. *Journal of Biological Chemistry* (2015).
- Jorda, J. et al., Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chemical Communications*, 52(28), pp.5041–5044. (2016).
- Keeling, T.J. et al., Interactions and structural variability of β -carboxysomal shell protein CcmL. *Photosynthesis research*, 121(2), pp.125–133. (2014).
- Kerfeld, C.A. et al., Protein structures forming the shell of primitive bacterial organelles. *Science*, 309(5736), pp.936–938. (2005).
- Kerfeld, C.A. & Erbilgin, O., Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in Microbiology*, 23(1), pp.22–34. (2015).
- Kerfeld, C.A., Heinhorst, S. & Cannon, G.C., Bacterial microcompartments. *Annual review of microbiology*, 64, pp.391–408. (2010).
- Keskin, O. et al., Principles of Protein–Protein Interactions: What are the Preferred Ways For Proteins To Interact? *Chemical reviews*, 108(4), pp.1225–1244. (2008).
- Kim, E.Y. & Tullman-Ercek, D., A rapid flow cytometry assay for the relative quantification

- of protein encapsulation into bacterial microcompartments. *Biotechnology journal*, 9(3), pp.348–354. (2014).
- Kinney, J.N. et al., Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *The Journal of Biological Chemistry*, 287(21), pp.17729–17736. (2012).
- Kinney, J.N., Axen, S.D. & Kerfeld, C.A., Comparative analysis of carboxysome shell proteins. *Photosynthesis research*, 109(1), pp.21–32. (2011).
- Klein, M.G. et al., Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *Journal of Molecular Biology*, 392(2), pp.319–333. (2009).
- Lassila, J.K. et al., Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *Journal of Molecular Biology*, 426(11), pp.2217–2228. (2014).
- Lawrence, A.D. et al., Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS synthetic biology*, 3(7), pp.454–465. (2014).
- Lee, H., DeLoache, W.C. & Dueber, J.E., Spatial organization of enzymes for metabolic engineering. *Metabolic engineering*, 14(3), pp.242–251. (2012).
- Lin, M.T. et al., β -Carboxysomal proteins assemble into highly organized structures in Nicotianachloroplasts. *The Plant Journal*, 79(1), pp.1–12. (2014).
- Mahalik, J.P. et al., Theoretical study of the initial stages of self-assembly of a carboxysome's facet. *ACS Nano*. (2016).
- Murugan, A., Zou, J. & Brenner, M.P., Undesired usage and the robust self-assembly of heterogeneous structures. *Nature Communications*. (2015).
- Myhrvold, C. & Silver, P.A., Using synthetic RNAs as scaffolds and regulators. *Nature Publishing Group*, 22(1), pp.8–10. (2015).
- Noël, C.R., Cai, F. & Kerfeld, C.A., Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit. *Advanced Materials Interfaces*. (2015).
- Pang, A. et al., Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *The Journal of biological chemistry*, 289(32), pp.22377–22384. (2014).

- Parsons, J.B. et al., Biochemical and structural insights into bacterial organelle form and biogenesis. *The Journal of biological chemistry*, 283(21), pp.14366–14375 (2008).
- Parsons, J.B. et al., Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular Cell*, 38(2), pp.305–315. (2010).
- Pawson, T. & Scott, J.D., Signaling through scaffold, anchoring, and adaptor proteins. *Science*, 278(5346), pp.2075–2080. (1997).
- Perlmutter, J.D., Mohajerani, F. & Hagan, M.F., Many-molecule encapsulation by an icosahedral shell. *eLife*, 5, p.5342. (2016).
- Pinheiro, A.V. et al., Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology*, 6(12), pp.763–772. (2011).
- Pitts, A.C. et al., Structural insight into the *Clostridium difficile* ethanolamine utilisation microcompartment. *PloS One*, 7(10) (2012).
- Quin, M.B. et al., Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. *Applied Microbiology and Biotechnology* (2016).
- Samborska, B. & Kimber, M.S. A dodecameric CcmK2 structure suggests β -carboxysomal shell facets have a double-layered organization. *Structure* (2012).
- Savage, D.F. et al., Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science*, 327(5970), pp.1258–1261. (2010).
- Sinha, S. et al., Alanine scanning mutagenesis identifies an asparagine–arginine–lysine triad essential to assembly of the shell of the Pdu Microcompartment. *Journal of Molecular Biology*, 426(12), pp.2328–2345. (2014).
- Siu, K.-H. et al., Synthetic scaffolds for pathway enhancement. *Current opinion in biotechnology*, 36, pp.98–106. (2015).
- Sun, Y. et al., Light modulates the biosynthesis and organization of cyanobacterial carbon fixation machinery through photosynthetic electron flow. *Plant Physiology*. (2016).
- Sutter, M. et al., Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Letters* (2015).
- Takenoya, M., Nikolakakis, K. & Sagermann, M., Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of *Escherichia coli*. *Journal of Bacteriology*, 192(22), pp.6056–6063. (2010).

- Tanaka, S. et al., Atomic-Level Models of the Bacterial Carboxysome Shell. *Science*, 319(5866), pp.1083–1086. (2008).
- Tanaka, S. et al., Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Science* 18(1), pp.108–120. (2009).
- Thomas, P.D. & Dill, K.A., An iterative method for extracting energy-like quantities from protein structures. *Proceedings of the National Academy of Sciences*, 93(21), pp.11628–11633. (1996).
- Wagner, H.J., Capitain, C.C. & Richter, K., Engineering bacterial microcompartments with heterologous enzyme cargos. *Engineering in Life Sciences*. (2017).
- Wheeldon, I. et al., Substrate channelling as an approach to cascade reactions. *Nature Chemistry*, 8(4), pp.299–309. (2016).
- Whitaker, W.R. & Dueber, J.E., Metabolic pathway flux enhancement by synthetic protein scaffolding. *Methods in Enzymology*, 497, pp.447–468. (2011).
- Wilner, O.I. et al., Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature Nanotechnology*, 4(4), pp.249–254. (2009).
- Yeates, T.O., Crowley, C.S. & Tanaka, S., Bacterial microcompartment organelles: protein shell structure and evolution. *Annual Review of Biophysics*, 39, pp.185–205. (2010).

Chapter 3: Visualizing *in vivo* dynamics of designer nanoscaffolds

Material in this text exists in part to an accepted publication in the journal Nano Letters
<http://dx.doi.org/10.1021/acs.nanolett.9b03651>

Eric J. Young^{1,2}, Jonathan K. Sakkos², Jingcheng Huang^{1,2}, Jacob K. Wright^{1,2}, Benjamin Kachel³, Miguel Fuentes-Cabrera^{4,5}, Cheryl A. Kerfeld^{1,2,6}, Daniel C. Ducat^{1,2}

¹ MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan, 48824 USA; ² Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan, 48824 USA; ³ Institute for Technical Microbiology, Mannheim University of Applied Sciences, Mannheim, Germany; ⁴ Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830 USA; ⁵ Center for Nanophase Material Sciences Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830 USA; ⁶ Environmental Genomics and Systems Biology and Molecular Biophysics and Integrated Bioimaging Divisions, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Abstract

Biochemistry routinely appears organized spatiotemporally for apparent improvements to pathway efficacy and control. Research towards nano-scale scaffolding platforms which rationally organize heterologous functions form one broad approach for engineering spatiotemporal organization within cells. Herein, we evaluate bacterial microcompartment shell (pfam0936-domain) proteins as modules for constructing well-defined nanometer scale scaffolds *in vivo*. We use a suite of visualization techniques to evaluate scaffold assembly and dynamics. We demonstrate recruitment of targeted cargo molecules onto assembled scaffolds by appending reciprocally-interacting adaptor domains, and refine these interactions by tuning scaffold expression level. Real-time observation of this system reveals a nucleation-limited step where multiple scaffolds initially form within a cell. Over time, nucleated scaffolds reorganize into a single intracellular assembly, likely due to inter-scaffold competition for protein subunits. Our results suggest design considerations for using self-assembling proteins as building blocks to construct nanoscaffolds, while also providing a platform to visualize scaffold-cargo dynamics *in vivo*.

Introduction

Once considered as an amorphous mix of freely-diffusing components, it has become increasingly evident that all cells use strategies to organize discrete subsets of signaling and metabolic proteins across space and time. One recurring theme of this organization at the nanoscale is the concentration of functionally related proteins into larger complexes and micro-domains—when physically coordinated by binding to a common surface, we define this process as scaffolding. Scaffolding proteins can act as mechanisms to control information flow of signaling and metabolic pathways (Good *et al.*, 2011). For example, colocalizing enzymes upon a common scaffolding surface increases their effective local concentration, potentially improving a pathway's fidelity and flux, while also lowering cross-reactivity and toxicity (Agapakis *et al.*, 2012; Castellana *et al.*, 2014; Wheeldon *et al.*, 2016). Within the past decade, bioengineers have sought to develop designer scaffolds to confer similar benefits upon heterologous pathways (Dueber *et al.*, 2009; Good *et al.*, 2011; Agapakis *et al.*, 2012; Chen *et al.*, 2014; Siu *et al.*, 2015; Wheeldon *et al.*, 2016).

Constructing scaffolds presents a number of scientific challenges and requires the development of biomaterials for predictively design at the nano-to-micron scale (Pugh *et al.*, 2018; Lee *et al.*, 2012; Glover and Clark, 2016; Luo *et al.*, 2016). To date, designer scaffolds have been built from a variety of biomaterials including: proteins, nucleic acids, and lipids (Good *et al.*, 2011; Chen *et al.*, 2014; Horn and Sticht, 2015; Siu *et al.*, 2015; Myhrvold *et al.*, 2016). Multiple studies have indicated that designer scaffolds can enhance metabolic flux by co-localizing enzymes of heterologous metabolic pathways as well as modulate signal transduction networks (Dueber *et al.*, 2009; Good *et al.*, 2011; Agapakis *et al.*, 2012; Chen *et*

al., 2014; Siu *et al.*, 2015; Horn *et al.*, 2015 Wheeldon *et al.*, 2016). However, it appears difficult to translate designs from one scaffolding application to another, mainly because previous designs have relied on building blocks lacking structural definition, leading to unpredictable cytosolic aggregation (Whitaker and Dueber, 2011). While these agglomerates may improve pathway flux via “proximity channeling” of metabolites (Lee *et al.*, 2012, 2016; Castellana *et al.*, 2014), it seems difficult to reproduce these benefits in other contexts. Even less information exists on dynamics of designer scaffolds over-time. Ultimately, scaffolding systems should feature structural information nearing Angstrom-level precision and information in spatiotemporal position over-time (Luo *et al.*, 2016).

Recently, researchers have explored scaffolding systems based on self-assembling protein building blocks that appear structurally well-defined. The protein domain family pfam00936 (“bacterial microcompartment shell protein domain”) has promise in this regard (Lee, Mantell, Hodgson, *et al.*, 2018; Schmidt-Dannert *et al.*, 2018; Zhang *et al.*, 2018). This ~90 amino acid domain oligomerizes into a modular, “tile” appearing protein with a hexagonal shape ~7 nm in diameter and ~3.5 nm thick (Young *et al.*, 2017). Oligomers then can self-assemble into a variety of higher-order architectures by “inter-tile” lateral interactions governed by electrostatics and shape complementarity (Sutter *et al.* Science 2017, Young *et al.*, 2017). In a native context, pfam00936 proteins form the architecture of the exterior shell of bacterial microcompartments (BMCs) that encapsulate specialized metabolic processes within many prokaryotic species (Kerfeld *et al.*, 2018; Lee *et al.*, 2019). When heterologously expressed outside of their native context, many pfam00936-proteins retain the capacity to form higher-order nanoarchitectures, including structures potentially attractive to serve as protein-based nanoscaffolds (*e.g.*, sheets, strips, spheroids, or tubes; (Young *et al.*, 2017;

Hagen, A., Kerfeld, 2018; Lee, Mantell, Hodgson, *et al.*, 2018; Schmidt-Dannert *et al.*, 2018; Zhang *et al.*, 2018). Indeed, recent studies have reported the use of functionalized pfam00936-domain proteins to organize target cargo molecules (Lee, Mantell, Hodgson, *et al.*, 2018; Schmidt-Dannert *et al.*, 2018; Zhang *et al.*, 2018), and in one instance this co-localization strategy was shown to increase the ethanol titer of *Escherichia coli* (*E. coli*) engineered with pfam00936-based nanotubes (Lee, Mantell, Hodgson, *et al.*, 2018). Additional development and analysis of this promising biomaterial may enable a wide range of increasingly sophisticated designer scaffolding applications.

In this study, we used pfam00936-domain containing proteins as assembly modules to form higher-order architectures suitable as intracellular scaffolds (**Figure 3.1A**) and developed a platform for high-resolution visualization of real-time dynamics *in vivo*. We expressed a small library of pfam00936 proteins to explore the nanoarchitectures formed in *E. coli* under different expression conditions. We appended adaptor domains to selected pfam00936 proteins (**Figure 3.1A**) and sought conditions that allowed for the specific recruitment of reciprocally-tagged cargo proteins, while preserving their capacity to form higher-order nanoarchitectures. Our results suggest design principles for creating pfam00936-based nanoscaffolds and revealed dynamics to consider in engineering scaffolding platforms from self-assembling proteins. We propose further refinement of visualization techniques as an important component for the development of nanoscale scaffolding platforms within living systems.

Results

We began by creating a small library of pfam00936 domain-containing proteins to generate a toolbox of components suitable for use as “**S**caffolds **F**ormed by BMC-**S**hell proteins” (hereafter; ScaFS). We individually expressed nine different ScaFS under a strong promoter (P_{T7}) and visualized intracellular assembly by transmission electron microscopy (TEM) of cellular thin sections (**Figure 3.1B and S3.5 & S3.6**). The selected ScaFS have differences in primary structure changing their surface electrostatics, lateral interface residues, and C-terminal extensions (**Figure S3.5A**) (Young *et al.*, 2017). Previous literature indicate that these differences influence self-assembly properties, although it remains difficult to predict how modifying primary sequence (e.g. single amino acid substitutions or domain extensions) will translate into which type(s) of higher-order architectures that form (Young *et al.*, 2017). Seven candidate ScaFs formed discrete structures that could be visualized by TEM (including tubes, sheets, and “rosettes”; **Figure S3.5**) while two ScaFS from *Halothecae* sp. PCC 7418, did not form discernable nanoarchitectures in the cytosol of *E. coli*, although they appeared to express correctly (**Figure S3.6**).

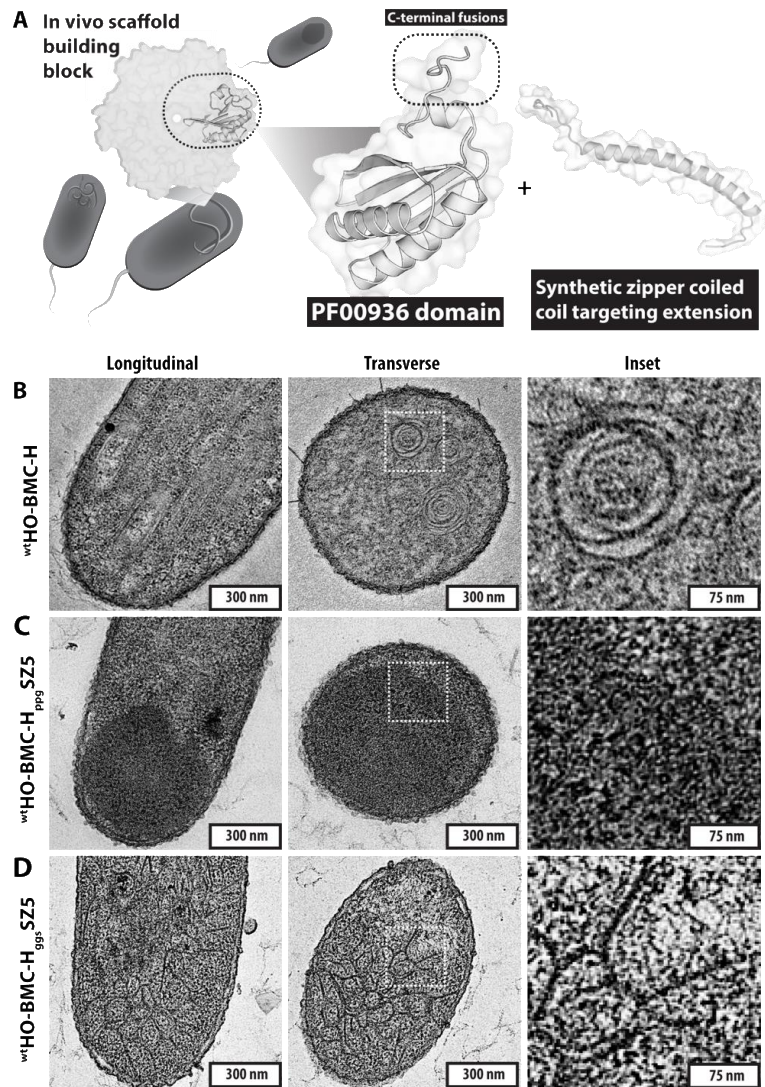


Figure 3.1: Functionalizing HO BMC-H with a Synthetic Zipper protein-protein interaction domain. A.) Cartoon schematic of ScaFS building block and the attachment of a C-terminal SZ coiled-coil. **B.)** Representative TEM images of intracellular assemblies formed by overexpression of unmodified HO BMC-H (^{wt}HO 5815 BMC-H). **C.)** Representative TEM images of the HO BMC-H ScaFS bearing a C-terminal Synthetic Zipper (SZ5) domain attached via a proline-rich (rigid; ppg) linker. **D.)** Representative TEM images as in (C) except where the linker region is composed of glycine-serine residues (flexible; ggs). Scale bar (white) size as indicated.

We selected HO BMC-H 5815 (Lassila *et al.*, 2014; Young *et al.*, 2017) from *Haliangium ochraceum* to evaluate the effects of appending an additional adaptor domain useful for mediating protein-protein interactions. We selected heterodimeric, coiled-coil protein domains from a published toolbox, termed “Synthetic Zippers (SZ; Thompson *et al.*, 2012). Since pfam00936-domain containing proteins naturally exhibit considerable diversity in the size and composition in extensions at the C-terminus (Young *et al.* 2017) (**Figure S3.7**), SZs were fused to the C-terminus with one of two differing linker sequences. The linker sequences were designed to be either “rigid” (proline-rich; ppg) or “flexible” (composed of glycine and serine; ggs) (Chen *et al.*, 2013). The unmodified HO BMC-H (^{WTHO} BMC-H) formed protein sheets that frequently rolled in upon themselves to create characteristic “rosettes” (**Figure 3.1B**). However, we observed only amorphous electron dense regions with characteristics similar to protein inclusion bodies following expression of HO BMC-H tagged with a SZ via a rigid linker (^{WTHO} BMC-H_{ppgSZ5}), (**Figure 3.1C**). Conversely, HO BMC-H SZ fusions with a flexible linker design (^{WTHO} BMC-H_{ggsSZ5}) formed curved, sheet-like structures within the cytosol (**Figure 3.1D**), apparently maintaining a higher-order arrangement similar to ^{WTHO} BMC-H (**Figure 3.1B**). The sheets formed by ^{WTHO} BMC-H_{ggsSZ5} tended to pack in a less dense arrangement, appearing as a webwork of curls (**Figure 3.1B,D**). Protein structure prediction software suggested a high degree of steric clash was predicted between the SZ extensions when they were attached via a rigid linker (**Figure S3.8**). We therefore tested additional ScaFs derived from ^{WTHO} BMC-H tagged solely with a SZ via the flexible-linker design and found that each of these formed higher-order assemblies in the cytosol of *E. coli* (**Figure S3.9**) that exhibited features similar to the parental ScaFs (*i.e.*, not modified with the appended SZ; **Figure S3.5**).

We next examined if SZ-appended ScaFS were capable of specifically binding and concentrating targeted cargo proteins tagged with a compatible SZ domain. Towards this, we constructed fluorescent protein fusions tagged with partner SZ domain (SZ6); which has been reported to form heterodimers with SZ5 with nanomolar affinity ($K_d < 15$ nM)(Thompson *et al.*, 2012). Upon co-expression of the reporter and a compatibly-tagged ScaF, fluorescence signal would be expected to relocate to the vicinity of higher-order assemblies (**Figure 3.2A**). As expected, a SZ6-tagged fluorescent reporter (SZ6-mNeonGreen) displayed a diffuse localization pattern in the cytosol of *E. coli* when expressed alone and visualized by widefield light microscopy (**Figure S3.10A**). However, when co-expressed with a compatible ScaFS, the reporter strongly concentrated to punctate and filamentous structures within the cell (**Figure 3.2B-D** and **Figures S3.10B, D and S3.11**). The fluorescence signal overlaid with cytosolic diffracting bodies observed in brightfield and which are consistent in shape and size to the protein sheets visualized by TEM (**Figures 3.1-3.2B,D** and **Figures S3.10-S3.11**). The higher resolution afforded by 3D-Structured Illumination Microscopy (SIM) further indicated that SZ6-tagged reporters form higher-order organizations similar to ScaFS assemblies viewed by TEM (**Figure 3.2CD, Figure S3.10D, S3.11** and **Movies 1-3**).

Unexpectedly, we also observed that a negative control fluorescent reporter (*i.e.* a fluorescent protein lacking a cognate SZ domain) would often concentrate near the diffracting bodies formed by overexpressed ScaFS (**Figure 3.2E** and **Figure S3.10C,E**), suggesting the presence of ScaFS-cargo interactions not mediated by SZ binding. In these instances, there was often a significant pool of cytosolic reporter (**Figure 3.2E** and **Figure S3.10C,E**), while the concentrated reporter pool appeared localized in the vicinity of, but not

directly overlaid upon, the finer structural features of the underlying body. These subtle features were further highlighted by deconvolution of widefield images (**Figure 3.2E and Figures S3.10E**). In a co-immunoprecipitation assay only SZ6-tagged cargo was co-precipitated with purified SZ5-tagged ScaFS: untagged reporters did not demonstrate binding affinity (**Figure S3.12**). Taken together, the data suggested that untagged fluorescent cargo might be concentrated as a result of an artifact *in vivo*, rather than direct binding. We observed other artifacts associated with the expression of ScaFS from the strong T7 promoter, including decreased growth rate and distorted cellular morphologies (**Figure 3.2B-E and Figures S3.10-S3.11**), which have also been reported by other groups (Liang *et al.*, 2017; Lee, Mantell, Brown, *et al.*, 2018).

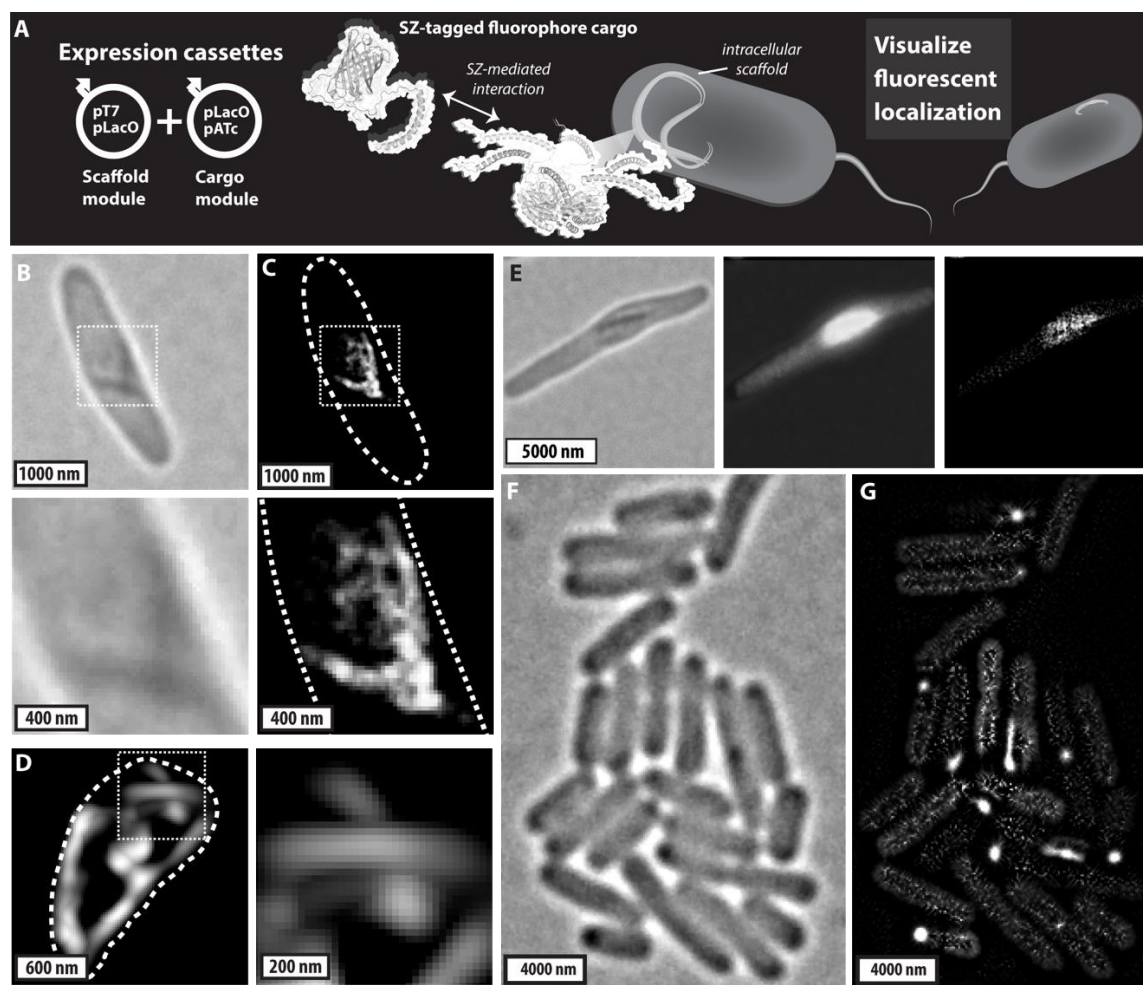


Figure 3.2: Cargo recruitment to designer intracellular protein scaffolds. A.) Cartoon schematic of fluorophore cargo recruitment to SZ-tagged ScaFS. **B.)** (top) Representative filamentous diffracting bodies observed 2 h following expression of $P_{T7}::K28AHO$ BMC-H -SZ5 (50 μ M IPTG); (bottom) magnified view of inset. **C.)** Deconvolved localization pattern of compatibly-tagged, co-expressed cargo protein $P_{LacO}::SZ6$ -mNG, as in (B), demonstrating cargo co-localization to fine filamentous features. **D.)** Representative 3D-SIM image of cargo-scaffold localization as in (C); highlighting fluorescent structures (*i.e.*, sheets/ribbons) similar to those observed by TEM. **E.)** Representative localization of an untagged cargo

Figure 3.2 (cont'd) ($P_{LacO}::mNG$; 50 μ M IPTG) in a cell exhibiting a $K^{28A}HO$ BMC-H -SZ5 diffracting body (left). Unprocessed (center), and deconvolved (right) fluorescence images of the cargo. Representative brightfield (**F**) and SRRF-processed fluorescence (**G**) images 3 h following expression of $P_{LacO}::K^{28A}HO$ BMC-H -SZ5 (100 μ M IPTG) and $P_{aTc}::SZ6$ -mScarlet-I (5nM aTc). Scale bars as indicated.

The localization artifacts and cytotoxic effects observed when using a strong promoter to express ScaFS encouraged us to try reducing size through the use of alternative, tunable promoters (Lee *et al.*, 2011). *E. coli* cells expressing ScaFS via a tunable promoter retained normal morphology, and typically exhibited either no obvious internal diffracting bodies (low inducer concentration), or small punctate or filament-like diffractions (intermediate-to-high inducer concentrations; **Figure 3.2F-G** and **Figure S3.13**). When co-expressed, two different SZ6-tagged fluorescent reporters concentrated strongly onto intracellular puncta and filaments (**Figures S3.13-S3.14**). Furthermore, untagged fluorescent cargo (lacking the cognate SZ6 binding domain) remained delocalized throughout the cytosol when co-expressed with ScaFS under a low or moderate level of expression (**Figure S3.13-S3.14**). Image error-mapping and super-resolution image processing (Gustafsson *et al.*, 2016; Culley *et al.*, 2018) (**see Figure S3.15** and **Materials and Methods**) refined descriptions of fluorescent location. After processing, SZ-tagged fluorescent signal appeared in specific subcellular locations, while untagged cargo appeared diffuse (**Figure 3.2F** and **Figure S3.15**).

We next visualized the dynamics of scaffold nucleation and maturation over time in live-cells. To accomplish this, we induced expression of SZ6-mNG for ~30 min to build a cytosolic pool of fluorescent cargo, then induced assembly by expressing a compatible ScaFS (^{K28A}HO BMC-H_{ggs}SZ5). Live-cell imaging revealed that cargo fluorescence initially appeared diffuse, but rapidly relocates into intracellular puncta following ScaFS expression (**Figure S3.16**). At least one fluorescent focus was evident in ~90% of cells within 60 minutes of ScaFS expression (n=31). Over the time-course, cargo continued to concentrate to subcellular

domains in the cell, although the localization pattern at later time points increasingly resembled filaments rather than small puncta (2-18 hours; **Figure S3.16**). By contrast, untagged cargo exhibited a primarily diffuse localization throughout the cytosol, regardless of the length of time following induction (**Figure S3.16**). 3D-total internal reflectance microscopy (3D-TIRFM) tracked cargo location in individual cells with an improved spatio-temporal resolution. This technique allowed us to observe cargo clustering in small concentrated regions as early as ~10 minutes following ScaFS induction, which we interpret as nucleation events of scaffold assembly (**Figure 3.3A; white arrowheads, Figure S3.17, Movie 4**). In the minutes immediately following a nucleation event, we frequently observed a rapid rise in the relative fluorescence of the puncta (**Figure 3.3A**), although not all puncta exhibited identical kinetics. In some instances, the “maturation” appeared to stall even as a sister-foci within the cell continued to grow in size and/or intensity (**Figure 3.3C**).

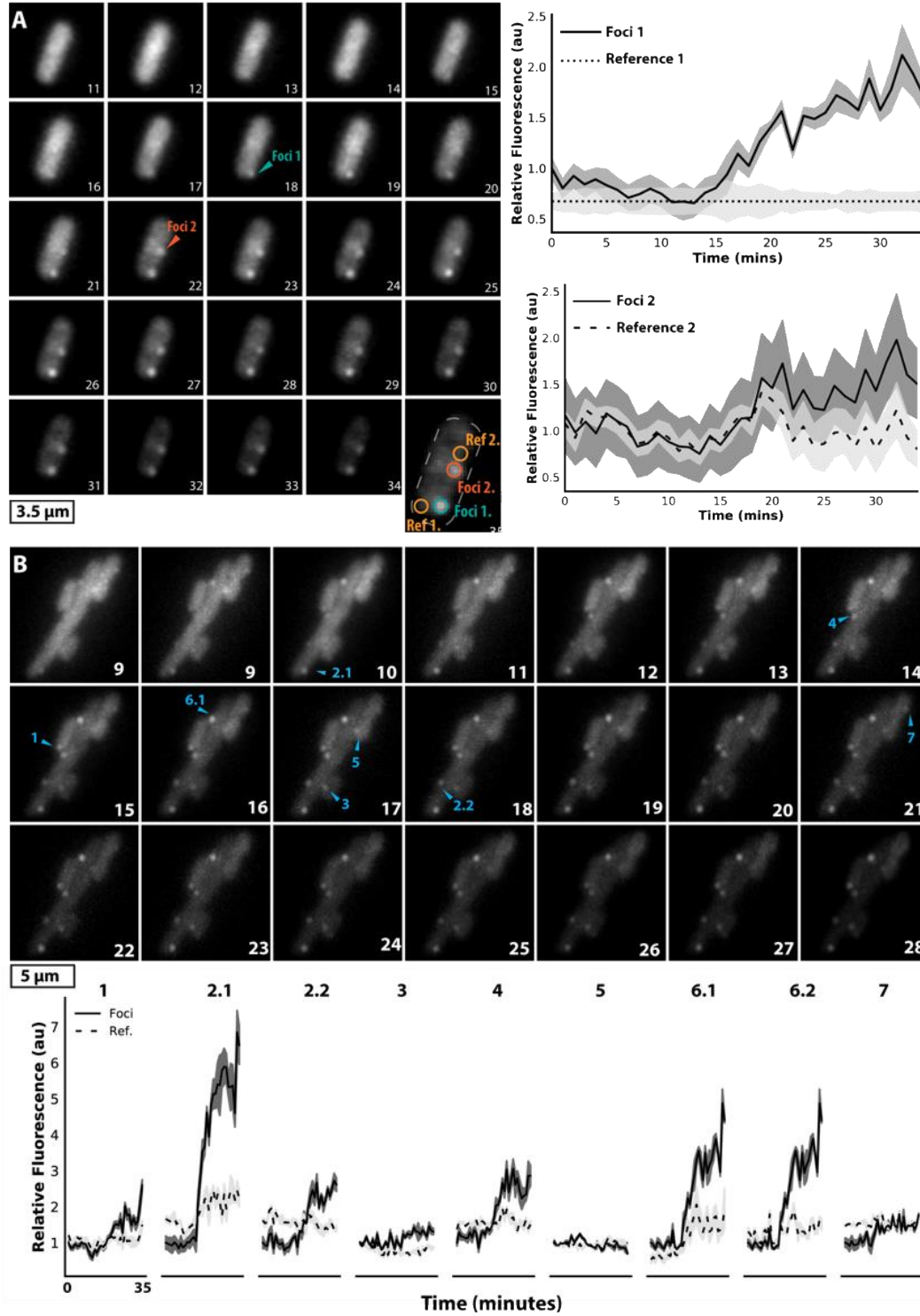


Figure 3.3: Evaluating scaffold behavior by visualizing intracellular cargo dynamics.

A.) 3D-TIRF microscopy max intensity projections of cargo ($P_{aTc}::SZ6\text{-mNG}$) following

Figure 3.3 (cont'd) induction of a compatible ScaFS ($P_{LacO}::K28AHO$ BMC- $H_{ggs}SZ5$). Representative frames are displayed for each minute following induction (at time = 0); appearance of cargo puncta is indicated by arrowheads. Enlarged image: highlighting two fluorescent foci and reference regions within the cell (circles) that are used to track the fluorescence kinetics associated with the recruitment of cargo at newly nucleated scaffold assemblies. **B.)** Additional 3D-TIRF microscopy of targeted cargo ($P_{ATc}::SZ6$ -mNeonGreen) in a field of cells following induction of a compatible ScaFS ($P_{LacO}::K28AHO$ BMC- $H_{ggs}SZ5$). Individual scaffold puncta were identified upon first indication of nucleation (red arrowheads; numbers) and the associated fluorescence intensity was tracked over time (dark grey traces). Reference areas of background fluorescence are indicated as light grey traces. [IPTG = 250 μ M, ATc = 5 nM]

SRRF analysis of the early stages of nucleation suggested additional subtleties in dynamics. For instance, we could often detect persistent fluorescent foci via SRRF at time points earlier than they became resolved by conventional widefield imaging (**Figure 3.4A and Figure S3.18**). Newly nucleated foci exhibited local movements while we observed relatively little motion with the larger “mature” cellular foci (**Figure 3.4B and Movies 5-6**). We captured multiple examples where one focus decreased in size and intensity while another maintained size or became more prominent (**Figure 3.3D**). Similarly, we observed instances where a small cluster of cargo could be detected at a given cellular position for minutes, but eventually diminish (**Figure 3.4A and Figure S3.18**). We interpret these events as nucleated ScaFS, but ones that fail to grow and instead become disassembled over time (**Figure 3.4A**). Indeed, the nucleation of one or more assemblies at early time points typically was accompanied by a decrease in the background level of cargo (**Figures 3.3-3.4 and Figures S3.17-S3.18**), suggesting that the level of freely-diffusible ScaFS appears depleted rapidly as nucleated scaffolds grow. Depletion of this free pool of ScaF subunits has implications for “competitive” behaviors between self-assembling systems that share the cytosol, with potential limitations on number and size of scaffolds that can be supported by a single cell (**see Discussion**).

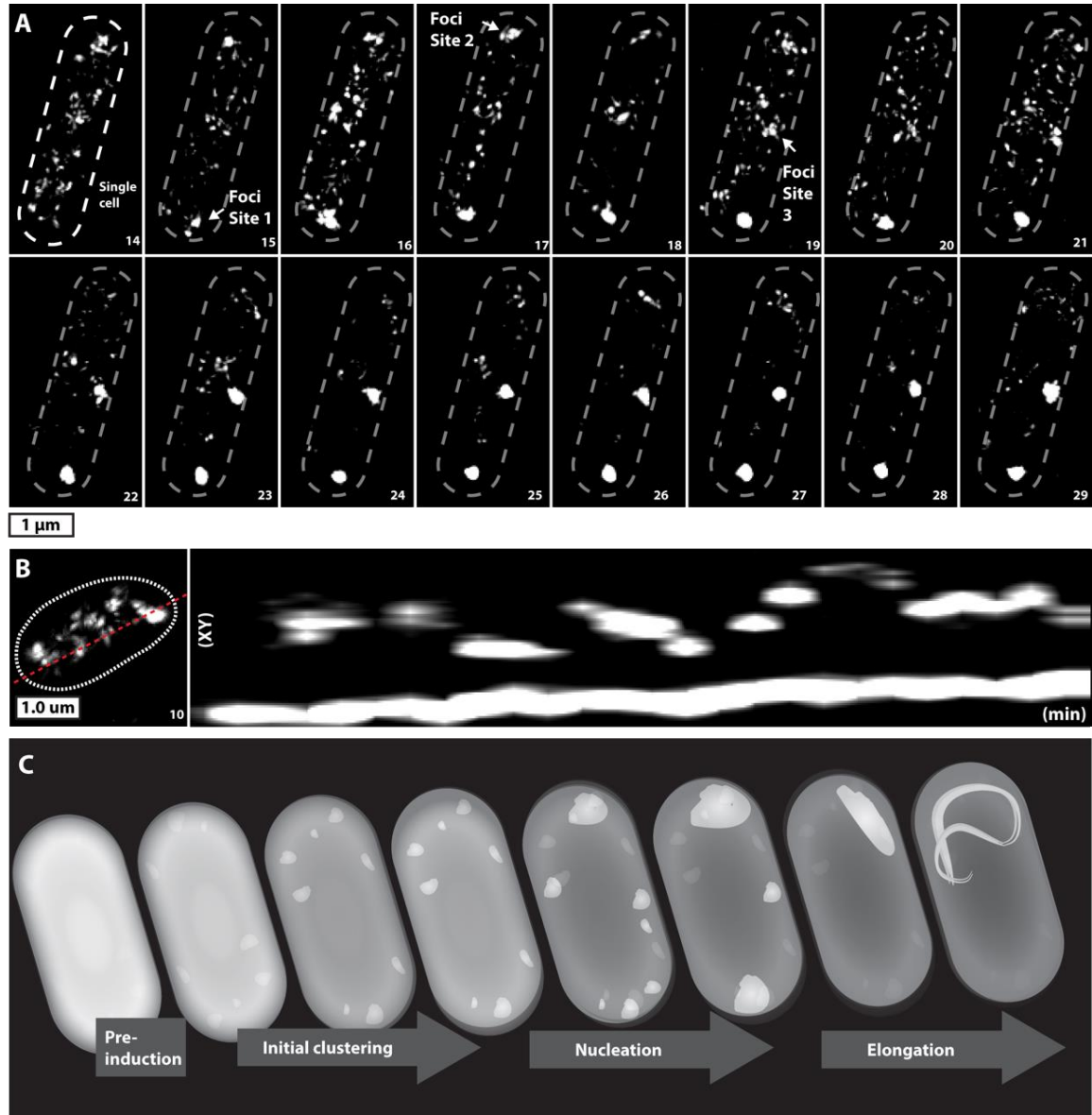


Figure 3.4: Diverse maturation fates of nucleated ScaFS. A.) Representative time series of nucleation events in live *E. coli* as in Figure 3A, but analyzed via SRRF. Three foci appear nucleated during the observation period, although only 2 persist throughout the time-course. **B.)** Kymograph of cargo foci position over time. Larger foci remain relatively fixed in position, while small foci exhibit more dynamic repositioning. **C.)** Cartoon schematic of scaffold-cargo reorganization *in vivo*. Expression leads to the buildup subunits in the cytosol,

Figure 3.4 (cont'd) eventually triggering the formation of numerous assemblies. Nucleated scaffolds expand rapidly and deplete the cytosolic pool of subunits. Over time, inter-scaffold competition leads to the domination of 1-2 large assemblies and the disassembly of other scaffolds within the cell.

Discussion

Recently literature exploring pfam0936-domain proteins as a biomaterial for constructing structurally-defined scaffolds (and compartment-like architectures) indicate potential for predictive positioning of cargo nearing angstrom resolution (Young *et al.*, 2017; Bari *et al.*, 2018; Lee, Mantell, Brown, *et al.*, 2018; Lee, Mantell, Hodgson, *et al.*, 2018; Schmidt-Dannert *et al.*, 2018; Plegaria *et al.*, 2018). We extend upon these efforts in the data collected here, expanding the toolbox of functionalized modules useful for spatiotemporal organization. We observed cargo interaction specificity as partially dependent upon size of scaffolds produced within cells, with large intracellular assemblies leading to artifacts in localization. We also found that the method used for attaching functionalization domains to pfam00936 proteins can impact their self-assembling properties. Using fluorescent reporter cargo proteins, we developed a platform for visualizing real-time assembly and reorganization. Our observations appear consistent with a nucleation-growth model of assembly, and the dynamic reorganization of nucleated scaffolds over time has broader implications for the use of self-assembling proteins as the building blocks for scaffolding applications.

In contrast to the relatively amorphous cellular “agglomerates” reported with other protein scaffolding materials expressed in the cell (Castellana *et al.*, 2014; Lee *et al.*, 2016), pfam00936-domain scaffolding proteins may form higher-order assemblies with predictable nanoscale features. Our results reinforce and expand upon recent results published using functionalized pfam00936-domain proteins as scaffolds in order to concentrate cargo in both *in vitro* (Bari *et al.*, 2018; Schmidt-Dannert *et al.*, 2018) and *in vivo* (Lee, Mantell, Brown, *et al.*, 2018; Lee, Mantell, Hodgson, *et al.*, 2018), while retaining higher-order structural organization (**Figures 3.1-3.2**). However, domain fusion to pfam0936 proteins may have a

influence on higher-order assembly properties (Young *et al.*, 2017; Huber *et al.*, 2017; Lee, Mantell, Hodgson, *et al.*, 2018; Hagen *et al.*, 2018). Our results support defined higher-order assembly when appending adaptor domains through a flexible linker sequence (*i.e.*, glycine-rich; **Figure 3.1D**)—resolvable higher-order features were lost when the same module was attached with a more rigid linker sequences (**Figure 3.1C**). While inconclusive, protein structural models of the rigid linker highlight how adaptor orientation, size, and degrees of motion allowed could influence the ability of pfam00936-domain proteins to hexamerize and/or assemble to a higher-order organization (**Figure S3.8**).

Our results reinforce the importance of regulating the size and number of scaffolds within the cell, yet also reveal current limitations in our capacity to control these processes. Previously, most publications on designer scaffolds have utilized highly active promoters which make correspondingly large intracellular bodies (Delebecque *et al.*, 2011; Lee *et al.*, 2012; Giessen and Silver, 2016; Myhrvold *et al.*, 2016; Wang *et al.*, 2017; Lee, Mantell, Brown, *et al.*, 2018; Lee, Mantell, Hodgson, *et al.*, 2018). Large scaffolds may impose significant cellular burden by being physically incompatible with other cellular processes (*e.g.*, constriction of the divisosome). Indeed, we see morphological defects in cells overexpressing ScaFS (**Figure 3.2B-D** and **Figures S3.10-S3.11**) that appear ameliorated at lower expression levels. Furthermore, scaffolds should present a number of binding sites roughly equivalent to cargo molecules; an overabundance of binding sites could act to locally tether cargo proteins at positions effectively isolated from one another (Lim, 2010). Finally, we see that control (*i.e.*, untagged) fluorophores with no detectable affinity *in vitro* concentrate in a delocalized fashion around the vicinity of large ScaFS (**Figures S3.5-S3.6**). This suggests unexpected artifacts could complicate rational design when large scaffolds form

intracellularly (*e.g.*, restricting diffusion, creating localized areas of protein aggregation). The smaller scaffolds produced by lower expression allowed for specific recruitment intended cargo proteins and minimized cellular distortion (**Figures 3.3-3.4** and **Figures S3.13-S.3.15**).

Through time-lapse imaging, we observe dynamic remodeling behaviors important to consider when using self-assembling proteins to produce smaller and/or more numerous scaffolds within cells (**Figures 3.3-3.4, Figures S3.16-S3.18**). These observations of early-assembly dynamics support a nucleation-growth model proposed by recent computer simulations of the BMC-shell protein component CcmK2, where nucleation of a small cluster of interacting hexamers (≥ 4 hexamers) appeared rate-limiting (Mahalik *et al.*, 2016). On longer timescales (>15 min), we observe nucleated assemblies undergo dynamic reorganization that tend to lead to one of two fates; 1) enlargement and elongation, or; 2) dissolution. At later time points (*i.e.*, >3 hours), typically only a single-focus persisted in the cytosol of bacteria (**Figure S3.16**); although we cannot exclude the possibility of smaller assemblies existing below our resolution limits. This dynamic appears consistent with a phenomenon known as Ostwald Ripening (Streets and Quake, 2010), where in a closed system (like that of the cytosol), each separate assembly pulls from a pool of diffuse building blocks, setting up a “competition” which may lead to one site dominating at the expense of all others (**Figure 3.4C**). This illustrates a potential common design feature inherent to self-assembling protein systems. In some scaffolding applications, 1-2 large scaffold clusters per cell may feel adequate or desired (*e.g.*, to promote proximity channeling in metabolic reactions). In these instances, partitioning pathways that ensure equitable inheritance to both daughter cells upon division may benefit cells. Towards this, positioning systems, such

as those of the ParA family, may work towards this goal—especially ones already used for segregating BMCs reported to interact with pfam0936-domain proteins (MacCready *et al.*, 2018). For scaffolding applications that would benefit most from numerous small assemblies, it seems necessary to consider other methods to prevent Ostwald Ripening behaviors, such as pfam0936-proteins engineered to act as assembly terminators through domain-fusion or amino acid substitution (Sutter *et al.*, 2019).

Regardless of current design limitations, nanoscaffolds built from pfam0936-domain proteins appear poised to form a powerful bioengineering tool, but fully realizing their potential requires deep examination of their dynamic structure inside of living cells. Finer-resolution multi-dimensional (3D/4D) imaging coupled to machine learning image analysis could help in this regard (Xu *et al.*, 2019; Ziatdinov *et al.*, 2017; Falk *et al.*, 2019). Such datasets could also form a basis to inform computational simulations, from coarse grained to fully atomistic (Frederix *et al.*, 2018, Hafner *et al.*, 2019). Development of such tools will act as stepping-stones towards ultimately realizing truly predictive engineered scaffolds at the nanoscale.

Acknowledgements

This work was supported by Department of Energy grant (DE-FG02-91ER20021) at MSU-DOE PRL. A portion of this work (e.g. SIM and TIRF microscopy), via a DOE-SCGSR fellowship, was also conducted at the Center for Nanophase Material Sciences, which is a DOE Office of Science User Facility at Oak Ridge National Lab. The authors graciously thank Jennifer Morrell-Falvey and Alicia Withrow for invaluable assistance in light and electron microscopy, respectively. Additionally, we would like to thank Clement Aussignargues in providing HO-BMC-H plasmids, and members of both the Ducat and Kerfeld labs for guidance and fruitful discussion.

Author Contribution

EJY conceived the project ideas, conducted experiments, analyzed data, and wrote/edited the manuscript. JS, JH, JW, BK, MF-C also provided experimental support and data analysis, while also editing of the manuscript. CK and DD conceived project ideas and wrote/edited the manuscript.

Materials and Methods

Cloning of Assembly and Cargo Module Constructs

Genes were synthesized by IDT. Isothermal assembly was used to clone sequences into their respective destination plasmids (pET11b, pBbB6k, pBbA2a). *E. coli* DH5 α strains were used for plasmid construction and propagation (see Table 3.1 for strains and plasmid list)

Expression of Assembly and Cargo Modules

E. coli BL21 ArcticExpress (DE3) competent cells (Agilent) were used for all cellular expression studies. Competent cells were transformed with constructs containing the gene(s) of interest and plated on LB agar plates with the appropriate antibiotic(s) to provide selection. Individual colonies were picked into 5 mL LB cultures with the appropriate antibiotics(s) and incubated at 30°C, with a rotational stirring of 250 RPM. These overnight cultures were inoculated 1:100 into fresh liquid cultures with conditions dependent upon the experimental plan. For TEM analysis, cells were inoculated in 50 mL LB, grown until ~OD₆₀₀ 0.8, induced with 100 μ M IPTG, and incubated for an additional 6 hours, before 2 mL of the parent culture was fixed overnight at 4°C. Cells used to purify protein were cultured in a similar manner, the whole culture was pelleted, and cell pellets were stored at -20°C for subsequent protein isolation. Cultures analyzed via light microscopy were inoculated into 1 mL LB or SOB media in culture tubes and incubated the indicated amount of time before imaging. Appropriate inducer(s) (IPTG or aTc) were added as indicated for each experimental design.

Purification of Assembly and Cargo Proteins

Pellets were solubilized in 30 mL resuspension buffer (50 mM Tris pH 7.8, 100 mM NaCl, 10 mM MgCl₂) on ice, lysed in a cell disruptor at 20k psi, and centrifuged at 20k *g* to separate soluble and insoluble fractions. For assembly module isolation, pellets were washed ~3 times with resuspension buffer + 1% Triton X-100, followed by washes with 500 mM NaCl resuspension buffer. ScaF protein was stored at 4°C until further analysis. Cargo modules were purified from the soluble cellular lysis fraction using Strep-tactin resin (IBA) according to the manufacture's protocol.

Magnetic Bead Precipitation

Purified cargo and ScaF modules (~20 µg each) were added together and allowed to mix gently for 2 hours on a rotator at room temperature. Washed Anti-HA Magnetic Beads (Pierce) were added to the solution and incubated for 30 minutes in order to allow binding to the HA tags appended to the C-terminus of ScaF constructs. Beads were then magnetically collected, washed three times, and then eluted with 0.5 M NaOH and analyzed via SDS-PAGE gel electrophoresis.

Transmission Electron Microscopy Thin Section Analysis

Expression of shell proteins was driven from T7 promoters as described above, then 2 mL cell aliquots were fixed in 2.5% glutaraldehyde/paraformaldehyde in sodium cacodylate buffer overnight at 4°C. Cells were pelleted by centrifugation at 4k rpm for 2 minutes and washed with sodium cacodylate buffer three times. Samples then were processed with a microwave assisted protocol beginning with 1% osmium tetroxide. Cells were washed with

HPLC-grade water until clear, and stained with 2% uranyl acetate, followed by another wash cycle. Samples were dehydrated with a gradient acetone series, then infiltrated with Spurr resin and cured at 60°C for ~3 days. Blocks were trimmed to highlight areas of cellular concentration and then ultra-thin sectioned on an RMC MX ultra-microtome with a diamond knife (Diatome 45°). Sections (~50 nm) were collected on copper mesh grids and stained subsequently with 2% uranyl acetate, washed, and then incubated with Reynolds lead citrate for 5 minutes each. Imaging was performed on a JEM 100CX II transmission electron microscope (JEOL) with a Prius SC200-830 CCD camera (Gatan). Raw data files were processed with FIJI-ImageJ software.

Light Microscopy

Cells were imaged on either a Zeiss Observer D1 or Zeiss Elyra P1 microscope. Suspended cells (1-3 μ L) were loaded onto an agarose pad and covered with a coverslip to minimize cell movement. These agarose pads were composed of M9 medium + 1% agarose (Thermo Fisher). Samples imaged with Structured Illumination Microscopy were collected as Z-stacks (5% power, 5 grid). Frames were collected in total internal reflectance mode (15% power, ~200 nm Zeiss software depth) as a Z-stack through each field every single minute. All raw data was processed with FIJI-ImageJ software.

Image Processing and Analysis

FIJI-ImageJ software was used in the processing and analysis of all raw data files (Schindelin, J. *et al.* 2012). Transmission electron microscopy thin section examples were all equally processed with a 'Enhance Local Contrast' plugin function. Deconvolution examples of

widefield images used plugin 'Iterative Deconvolve 3D' utilizing an estimated measured PSF of the Zeiss Observer D1. Total internal reflectance and structured illumination microscopy Z-stack frames were compiled using a 'Max Intensity Stack Projection.' In time lapse examples, frames were registered with 'Correct 3D Drift' plugin before stack formation. Values for foci development were tracked by measuring parameters within a circular region of interest in comparison to a reference area within a cell that lacked discernable features. Widefield microscopy images further processed by NanoJ-SRRF, and subsequent NanoJ-Squirrel error minimization, followed parameters specified in an experiment. Images were further processed with median, unsharp mask, or tophat filters as indicated.

Supporting Information:

Table 3.1 Plasmids and gene inserts used in Chapter 3			
Plasmid Gene			Resistance
1	pET-11b	MicH (HO BMC-H 5815)	AMP
2	pET-11b	K28A_MicH (HO BMC-H 5815)	AMP
3	pET-11b	K28P_MicH (HO BMC-H 5815)	AMP
4	pET-11b	R78A_MicH (HO BMC-H 5815)	AMP
5	pET-11b	CcmK2 (7418)	AMP
6	pET-11b	CcmK1 (7418)	AMP
7	pET-11b	PduA	AMP
8	pET-11b	RmmH	AMP
9	pET-11b	MicHppgSZ5-HA	AMP
10	pET-11b	MicHggsSZ5-HA	AMP
11	pET-11b	K28A_MicHggsSZ5-HA	AMP
12	pET-11b	K28P_MicHggsSZ5-HA	AMP
13	pET-11b	R78A_MicHggsSZ5-HA	AMP
14	pET-11b	MicHggsSZ19-HA	AMP
15	pBbB6k	MicHggsSZ5-HA	KAN
16	pBbB6k	K28A_MicHggsSZ5-HA	KAN
17	pBbB6k	STREP-mNeonGreen	KAN
18	pBbB6k	SZ6-ggs-mNeonGreen	KAN
19	pBbA2a	STREP-mNeonGreen	AMP
20	pBbA2a	SZ6-ggs-mNeonGreen	AMP
21	pBbA2a	STREP-mScarlet-I	AMP
22	pBbA2a	SZ6-ggs-mScarlet-I	AMP

Figure S3.5 (cont'd) (CcmO), *Halotheca* sp. PCC 7418 (CcmK1, CcmK2), *Mycobacterium smegmatis* (RmmH), *Citrobacter freundii* (PduA), and *Haliangium ochraceum* (MicH). Alignment of CcmO has been split into its two respective pfam0936 domains **B.)** Higher-order protein assemblies formed within the cytosol of *E. coli* strains overexpressing the indicated ScaFs. Representative longitudinal and transverse sections of assemblies formed by HO-BMC-H and amino acid point substitution (K28A, K28P, and R78A) are displayed. [IPTG = 100 μ M]

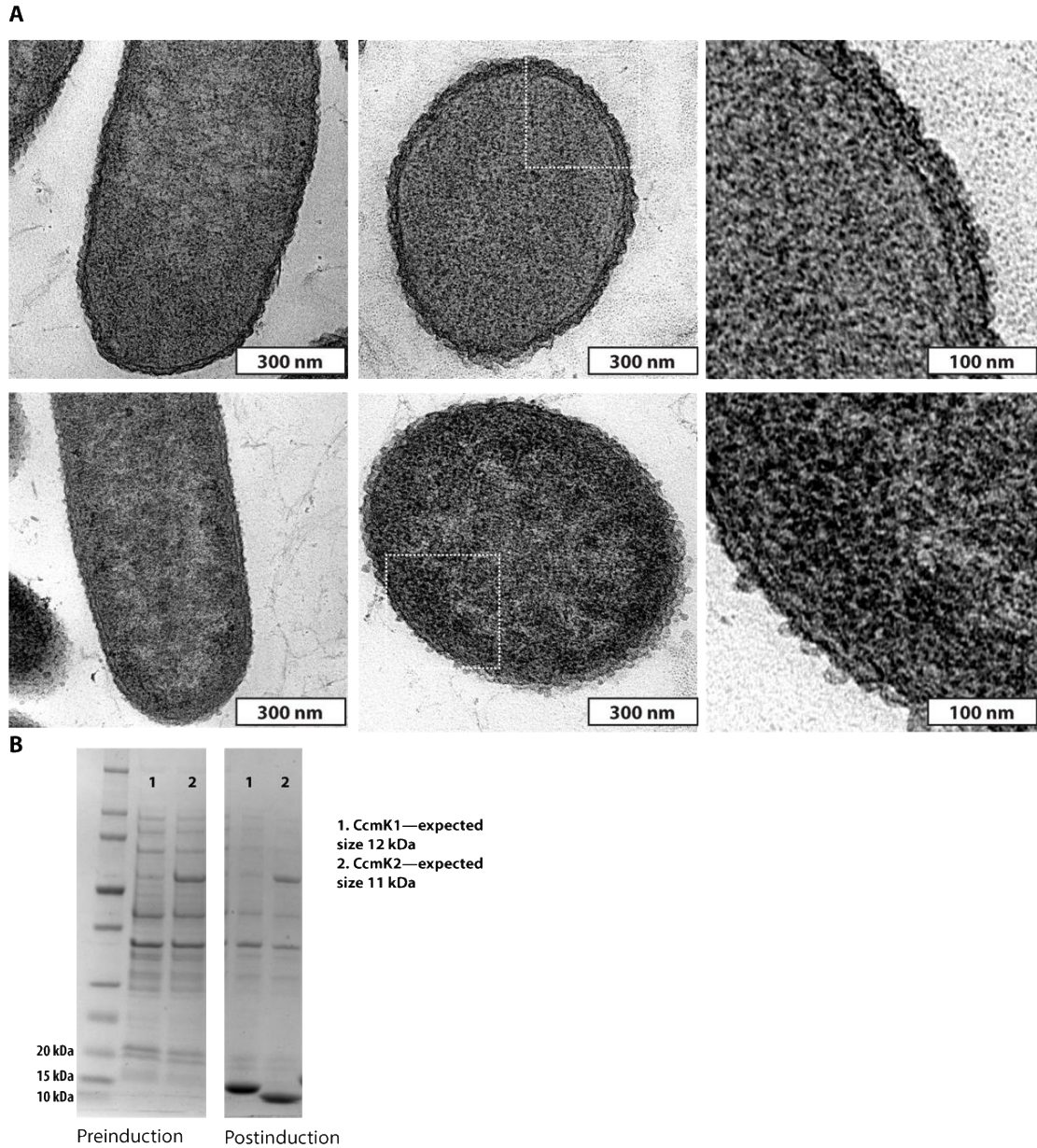


Figure S3.6: Thin section overexpression of modules lacking nanostructures. Two ScaFs from *Halotheca* spp 7418 do not form resolvable higher-order structures when heterologously expressed in *E. coli*. **A.)** TEM images of thin sections of bacteria expressing CcmK1 (top) and CcmK2 (bottom) from a T7 promoter. **B.)** SDS-PAGE analysis of whole cell lysates of *E. coli* obtained before (left) and after (right) induction of shell proteins (100 μ M

Figure S3.6 (cont'd) IPTG). Despite no **(Fig. S3.6 cont.)** visible higher-order structures, protein bands are visible in the cells induced to express CcmK1 (1; expected size = 12 kDa) and CcmK2 (2; expected size = 11 kDa), indicating successful expression.

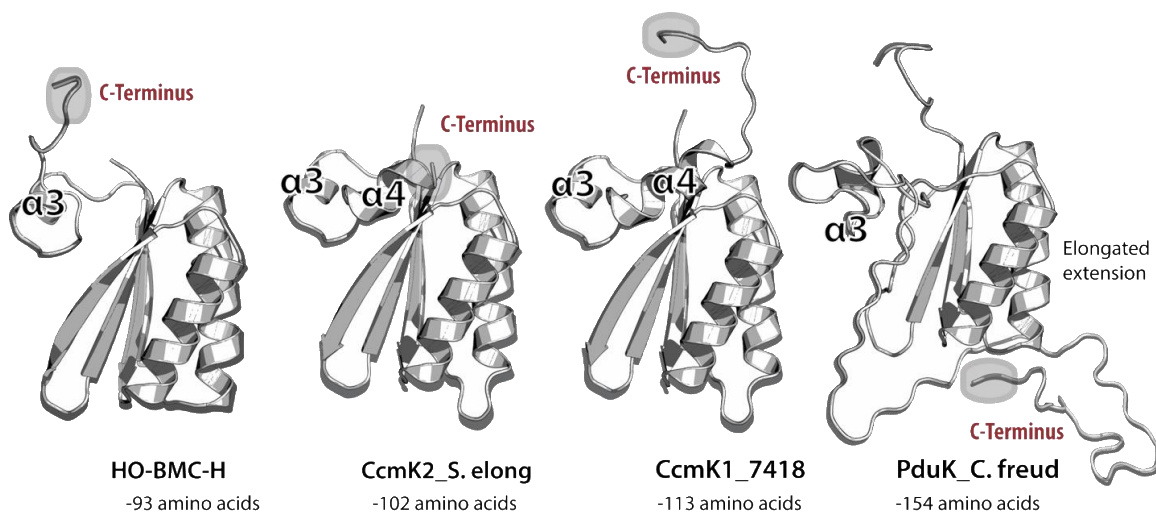


Figure S3.7: Pfam0936-proteins naturally exhibit diversity in the structural features of the C-terminus. The predicted structure of four modules used in this study illustrates the conservation of the N-terminal β -sheets and the first 2 α -helices in this protein family. By contrast, the C-terminus (red) varies widely, including differences in length as well as the presence/composition of additional α -helices ($\alpha 3$ and $\alpha 4$).

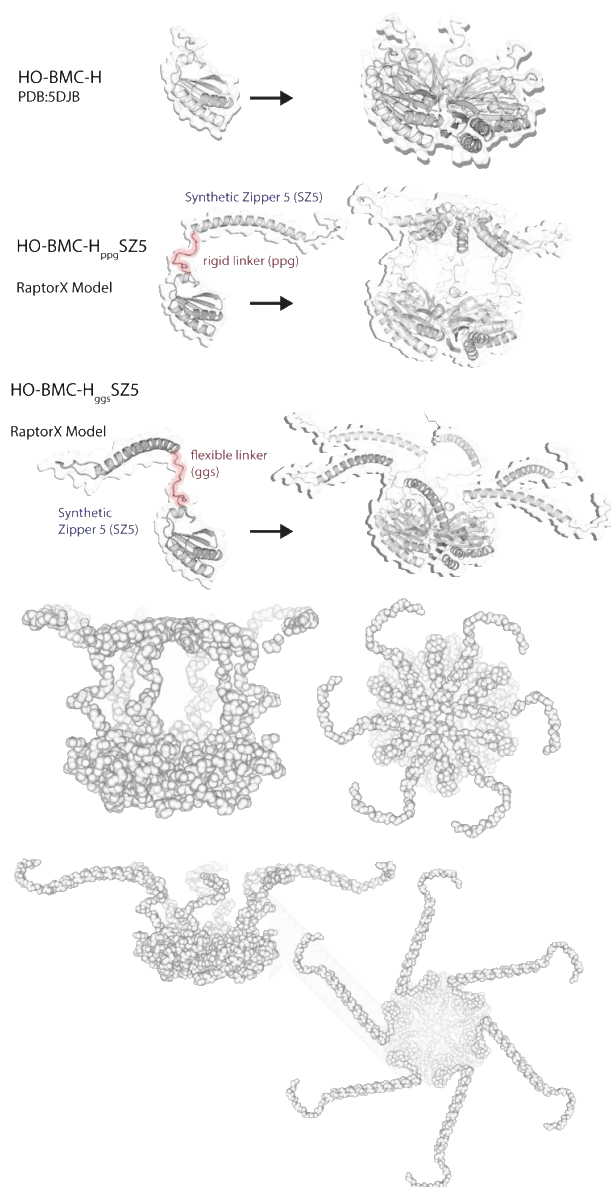


Figure S3.8: Predicted structure of SZ-functionalized ScaFS. Raptor-X (Källberg, M. *et al.* 2012) monomer models were constructed of HO-BMC-H either attached to a Synthetic Zipper domain (SZ5) through either a “rigid” proline-rich linker sequence (ppg: red middle) or a “flexible” glycine-serine linker (ggs: red bottom). Modeling software GalaxyHomomer (Baek et al. 2017) then built hexamerized models; while inconclusive, the predicted model of the hexamerized HO-BMC-H_{ppg}SZ5 illustrates how steric clash may arise between adaptor domains.

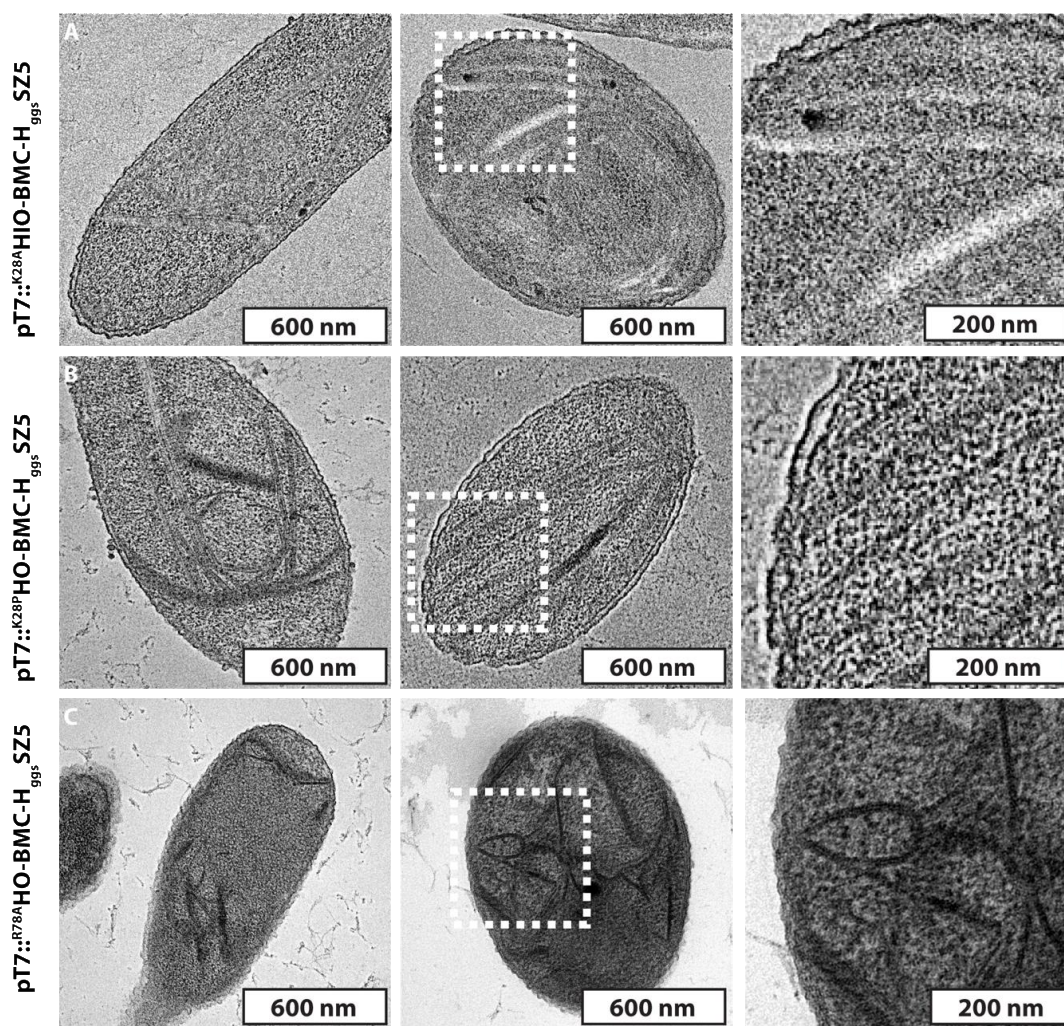


Figure S3.9: Higher-order intracellular architectures formed by ScaFS bearing a C-terminal SZ5 domain attached via a flexible linker. Representative TEM images of *E. coli* cells overexpressing derivatives of HO-BMC-H (K28A, K28P, and R78A) that are tagged with a C-terminal SZ5 domain via a glycine-serine linker. These ScaFS retain similar higher-order assembly features as those observed when their parental (*i.e.*, untagged) modules appear overexpressed.

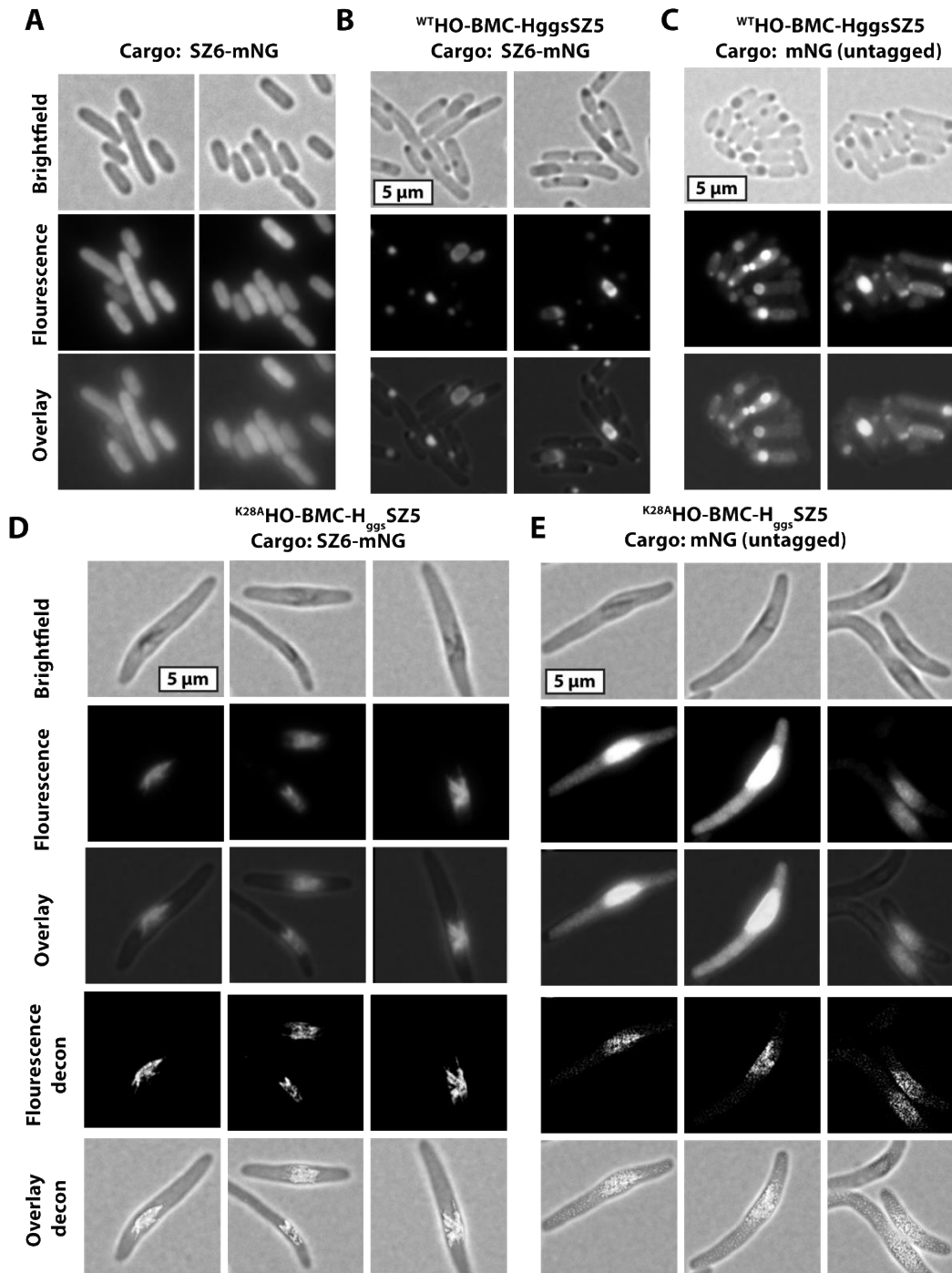


Figure S3.10: Fluorescent cargo reporter localizes on or near intracellular diffractions produced by overexpressed ScaFS. Representative *E. coli* cells co-expressing both a SZ5-tagged ScaF and fluorescent cargo with (A, B, D) or without (C, E) a N-terminally fused

Figure S3.10 (cont'd) Synthetic Zipper (SZ6). The length of induction and the concentration of inducer (50 μ M IPTG) are identical for all examples. Reporter cargo are expressed from P_{LacO} and ScaFs are expressed from a P_{T7} promoter. Scale bars as indicated.

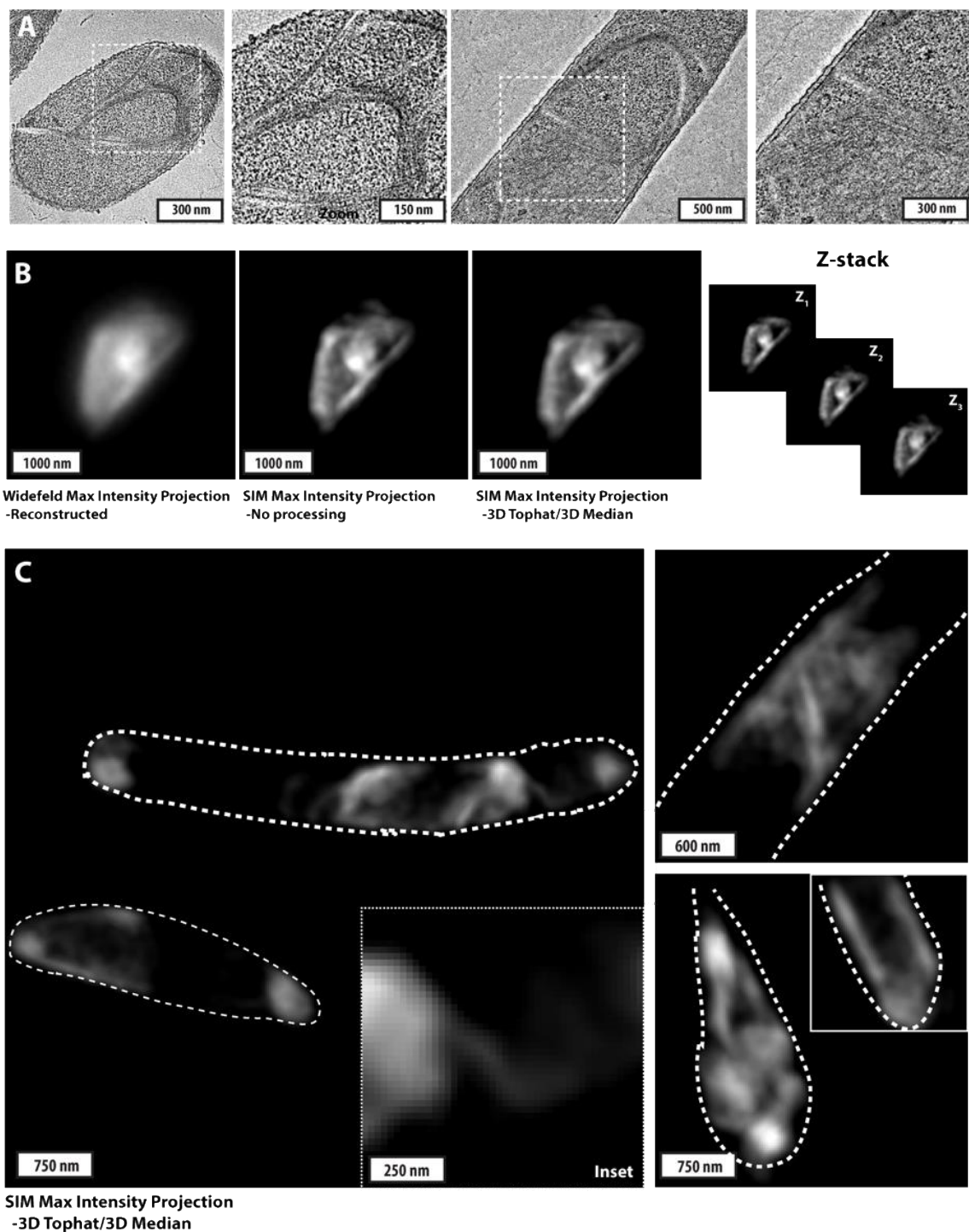


Figure S3.11: Comparison of high-resolution approaches for imaging intracellular ScaFS. **A.)** Representative TEM images of filament/sheet-like bodies observed in *E. coli* expressing a ScaF (K^{28A} HO-BMC-H_{ggs}SZ5) from a T7 promoter: insets are magnified to

Figure S3.11 (cont'd) highlight fine details. **B.)** Comparison of imaging techniques and processing of fluorescence data. Representative images of SZ6-mNG cargo localization in a cell expressing (P_{T7}::K^{28A}HO-BMC-H_{ggs}SZ5) as imaged by maximum projection of widefield fluorescence images (left) in comparison to a 3D-SIM maximum projection (center), and images processed via a 3D-Tophat/Median filter (right). See **Materials and Methods** for additional information on processing **C.)** Representative fluorescence images processed via 3D-SIM and Tophat/Median of SZ6-mNG cargo in cells expressing K^{28A}HO-BMC-H_{ggs}SZ5 that highlight higher-order organization of targeted cargo. Scale bars as indicated.

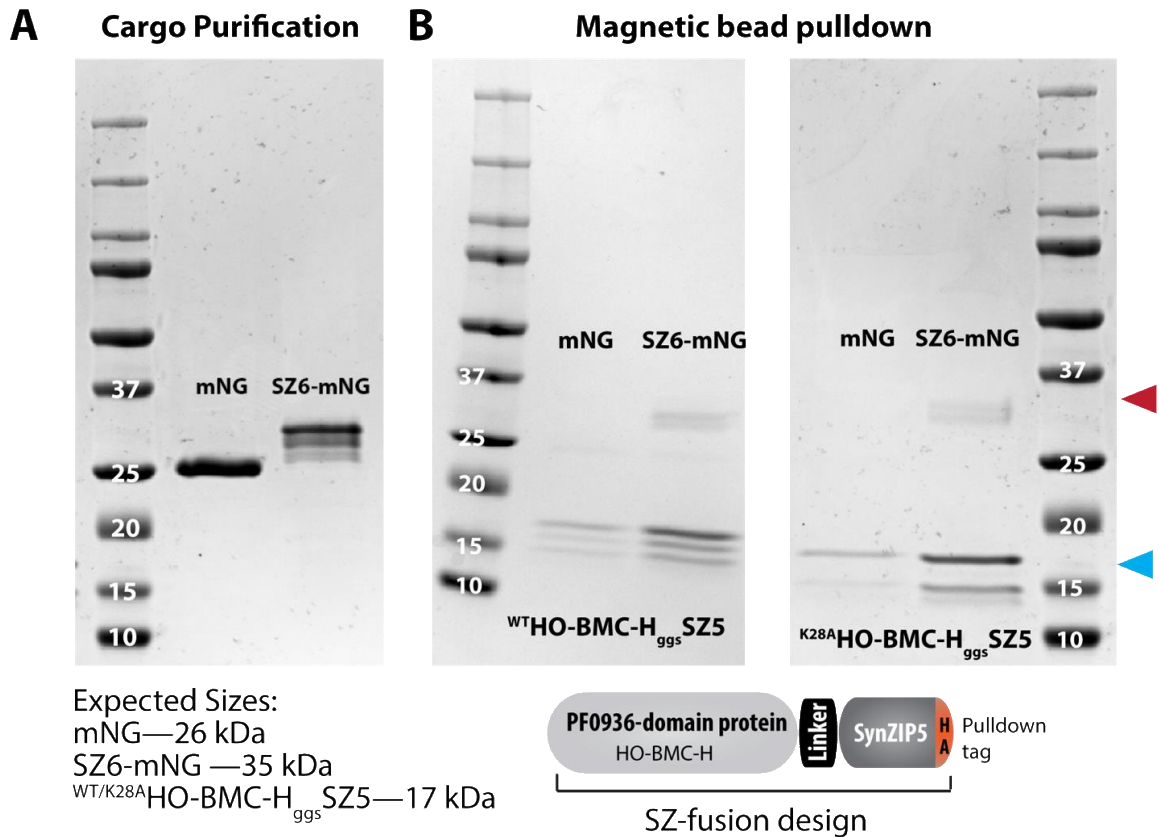


Figure S3.12: *In vitro* co-immunoprecipitation of purified ScaFS and reporter cargo proteins. **A.)** SDS-PAGE analysis of strep-column purified fluorophore cargo. Proteins purified with a SZ-domain contain a multiple band pattern likely attributable to partial protein breakdown products. **B.)** Co-immuno precipitation of either untagged (mNG) or reciprocally-tagged (SZ6-mNG) with isolated ScaFs (^{WT}HO-BMC-H_{ggs}SZ5; left or ^{K28A}HO-BMC-H_{ggs}SZ5;right). Precipitation targeted the HA tag appended to the C-terminus of ScaF constructs. Expected molecular weight for SZ-tagged ScaF = 17 kDa (blue arrowhead). Expected molecular weight for SZ-tagged fluorophore = 35 kDa (red arrowhead).

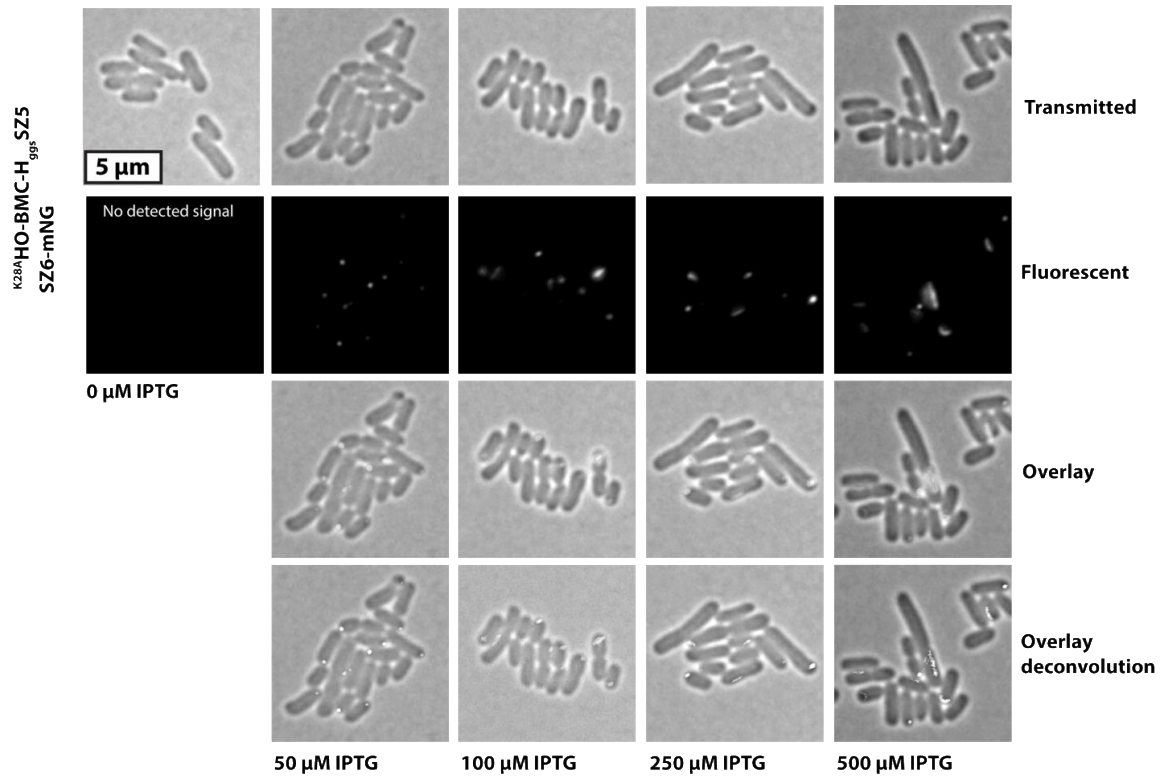


Figure S3.13: Expression level of ScaFS correlates with characteristics in intracellular localization phenotype. Representative fields of *E. coli* co-expressing $K^{28A}HO-BMC-H_{ggs}SZ5$ and SZ6-mNeonGreen (SZ6-mNG) at various concentrations of ScaF inducer (IPTG). Increased IPTG regulates expression of $K^{28A}HO-BMC-H_{ggs}SZ5$ from P_{LacO} . With increasing inducer concentrations, punctate, then filamentous diffracting bodies are observed within the cytosol of *E. coli*. SZ6-mNG co-localizes with these diffracting bodies as observed by fluorescence microscopy.

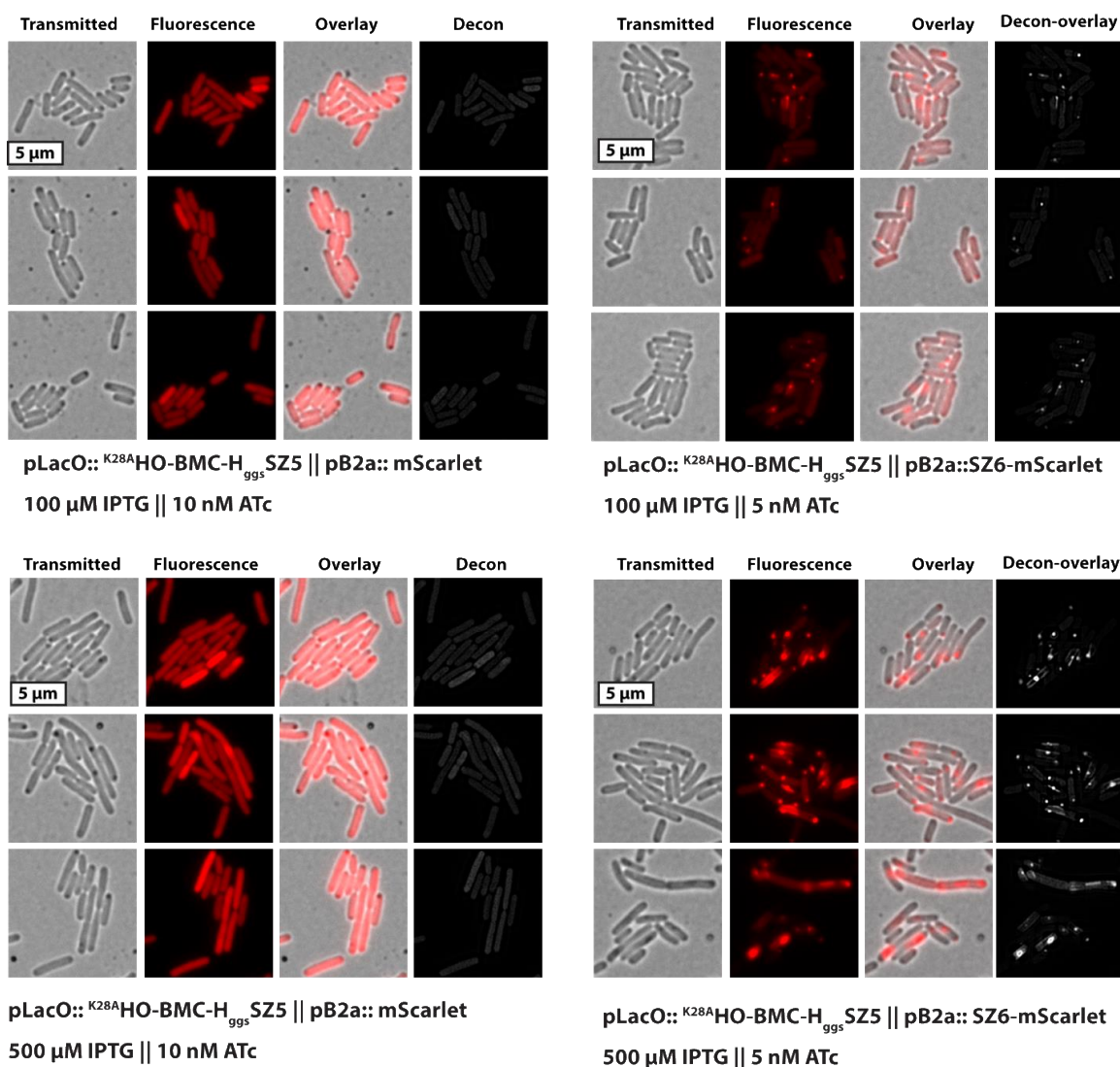


Figure S3.14: Co-localization of alternative cargo molecules to compatible ScaFS-induced intracellular protein assemblies. Representative panels of the localization pattern for untagged cargo (mScarlet-I) and SZ-tagged cargo (SZ6-mScarlet-I) when co-expressed in the same cells as K^{28A} HO-BMC-H_{ggs}SZ5. IPTG concentration is modulated to poise K^{28A} HO-BMC-H_{ggs}SZ5 expression at low (100 μ M IPTG; top) or high induction (500 μ M IPTG; bottom) levels. Untagged cargo remains delocalized throughout the cytosol, while compatibly-tagged cargo co-localizes with the K^{28A} MHO-BMC-H_{ggs}SZ5-induced diffracting bodies. Scale bars as indicated.

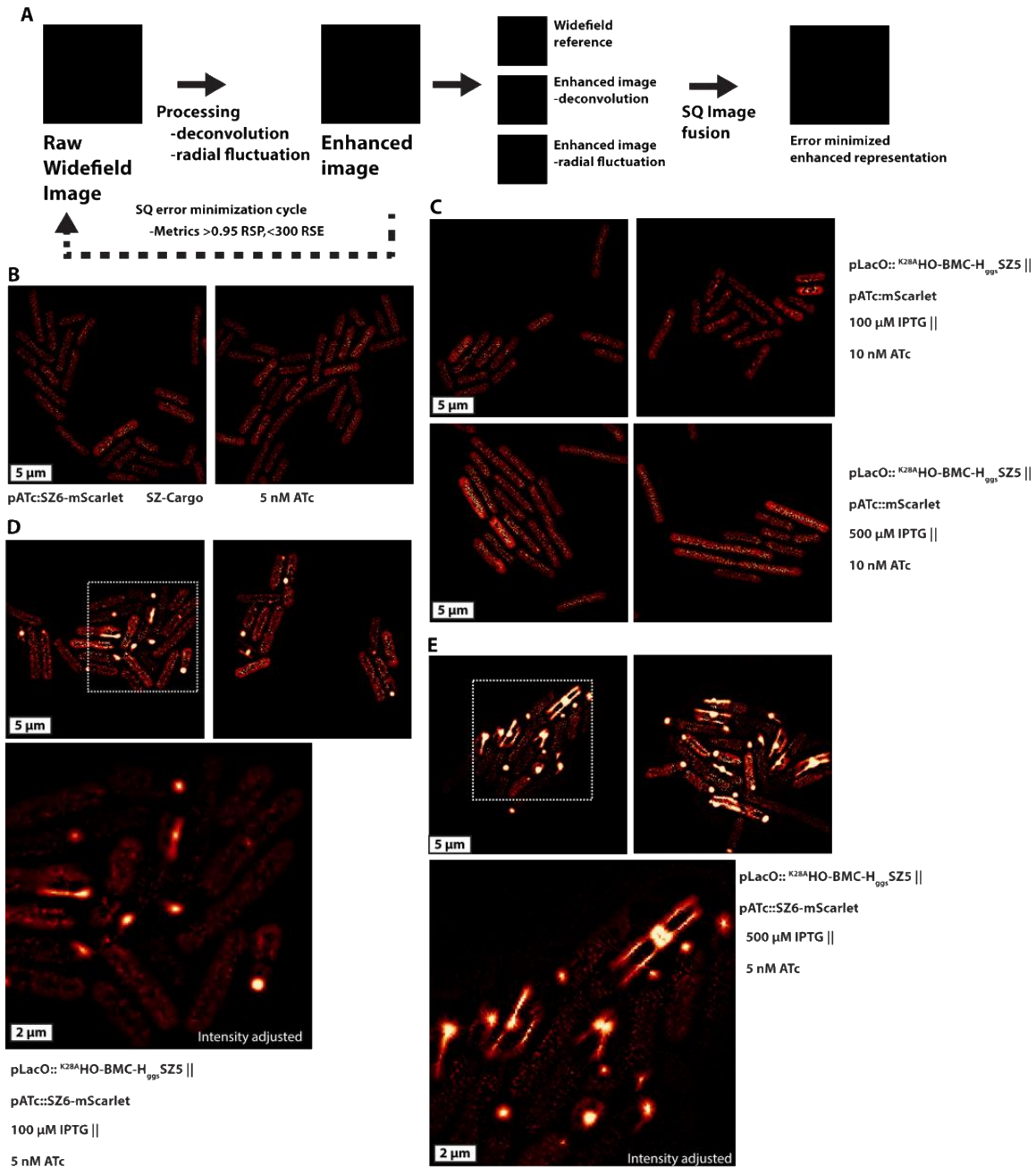


Figure S3.15: High-resolution image processing pipeline. A.) A widefield frame was processed by Iterative Deconvolve 3D, or by SRRF, with indicated NanoJ-SQUIRREL metrics. Widefield, deconvolved, and SRRF images were fused with NanoJ-SQUIRREL for further reduction of image artifacts. B.) Representative error minimized images of diffuse SZ6-mScarlet-I cargo expressed in *E. coli* without a cognate ScaF. C.) Representative images of an

Figure S3.15 (cont'd) untagged cargo, mScarlet-I, in a cell expressing ^{K28A}HO-BMC-H_{ggs}SZ5, indicating no specific cargo localization features. **D.)** Representative error minimized images of cells expressing a ScaF (^{K28A}HO-BMC-H_{ggs}SZ5; 100 μM IPTG) and cognate cargo (SZ6-mScarlet-I; 5 nM aTc). Cargo molecules localize to defined puncta within the cell. **E.)** Same as (D) except with a stronger induction of the ScaF (500 μM IPTG), showing the development of larger puncta and filament-like structures within the cell. Scale bars as indicated.

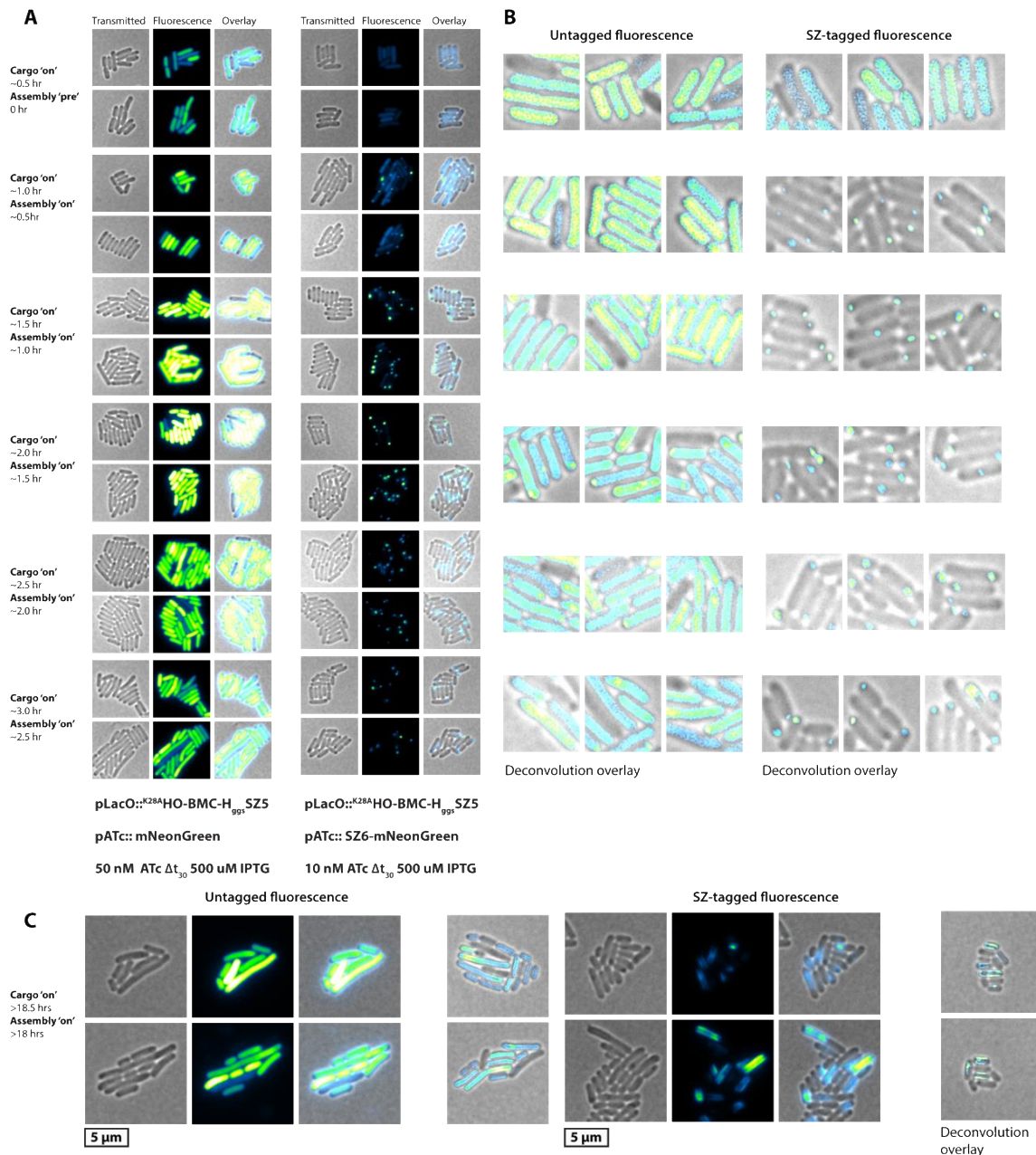


Figure S3.16 (cont'd) image deconvolution applied. **C.)** Representative cells as in (A), except viewed after an overnight induction; Scale bars as indicated.

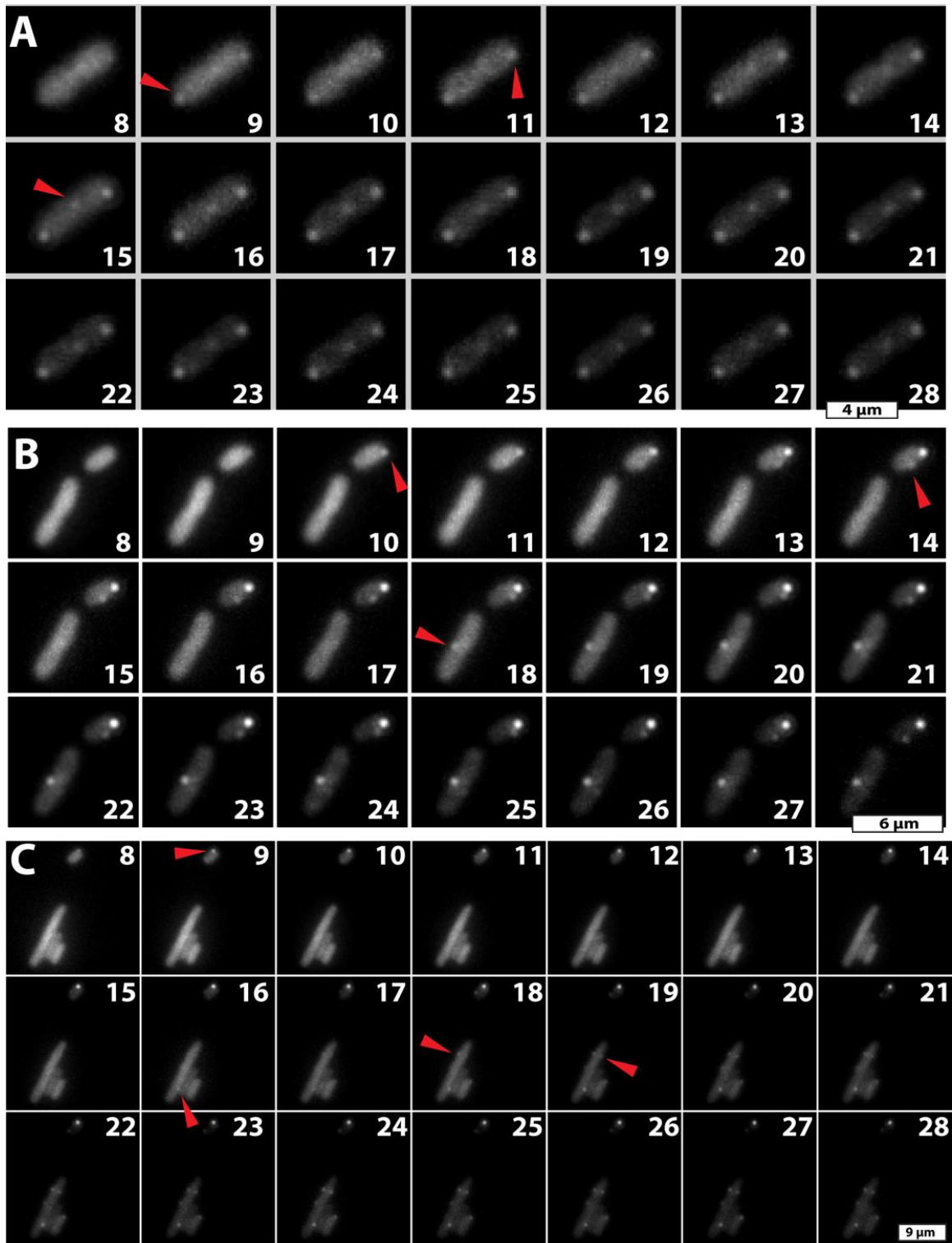


Figure S3.17: Representative time-lapse imaging of nucleation events. Individual *E. coli* cells (A-B) and a field (C) induced to express SZ6-mNeonGreen for 30 minutes prior to the

Figure S3.17 (cont'd) induction of compatible SZ-tagged ScaFS (5 nM aTc induction of cargo from P_{aTc}). $K28A$ HO-BMC-H_{ggs}SZ5 was then induced with 250 μ M IPTG (P_{LacO}): the number of minutes following addition of IPTG is indicated in each frame. Images represent max intensity projections with scale bar sizes as indicated.

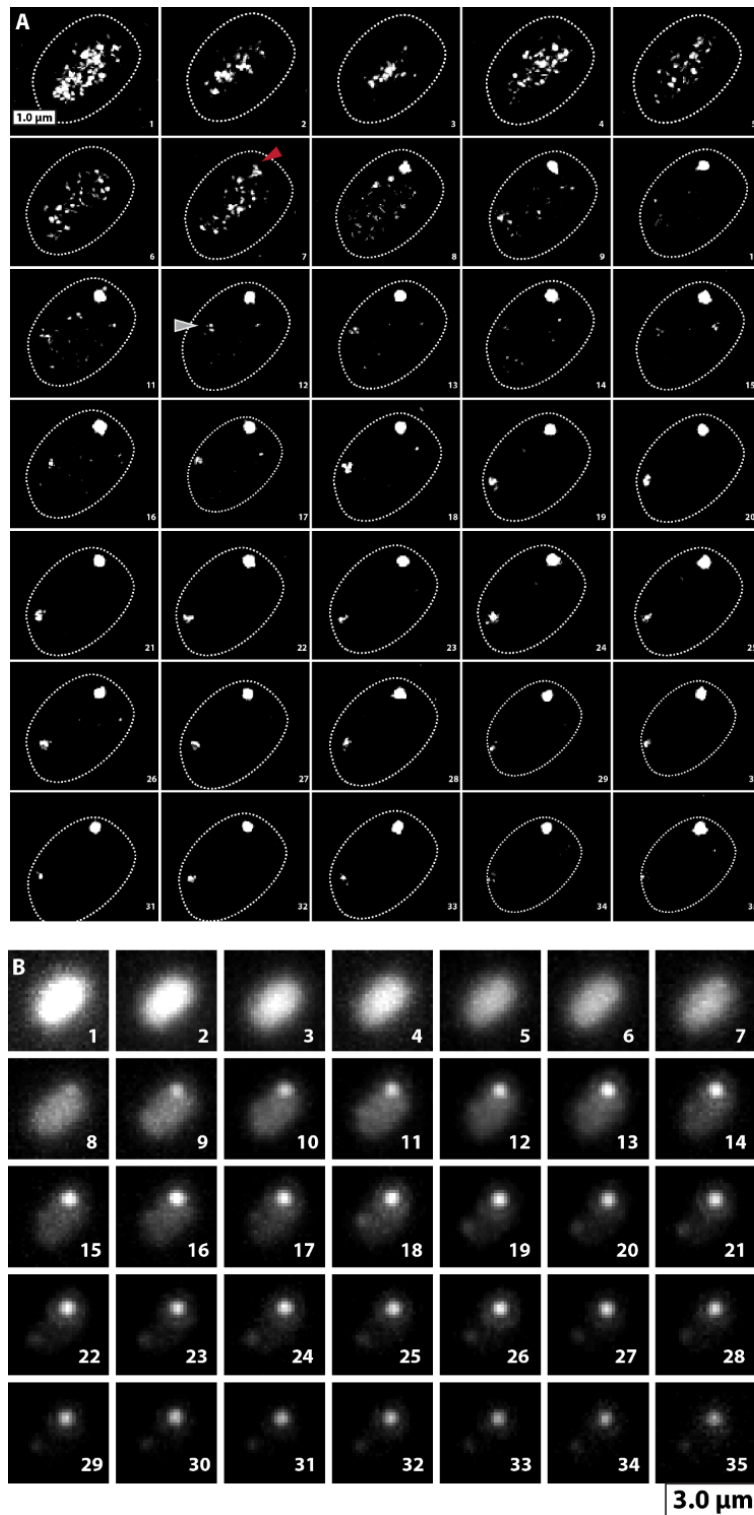


Figure S3.18: Visualization of intracellular cargo dynamics by SRRF-processing. A.)

Representative time lapse of a cell expressing SZ6-mNeonGreen and induced to express

Figure S3.18 (cont'd) ^{K28A}HO-BMC-H_{ggs}SZ5 at t = 0. Two clusters of cargo appear to form (red and grey arrowheads). The first focus forms at 7 minutes (red arrowhead) and rapidly increases in size and intensity. A second focus forms at 12 minutes (grey arrowhead) and transiently becomes brighter, before becoming increasingly difficult to resolve. **B.)** 3D-TIRF max intensity projection of unprocessed fluorescence images in timeseries **(A)**. Scale bars as indicated.

LITERATURE CITED

LITERATURE CITED

- Agapakis, C.M., Boyle, P.M., and Silver, P.A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* 8: 527–535. (2012).
- Bari, N.K., Kumar, G., Bhatt, A., Hazra, J.P., Garg, A., Ali, M.E., and Sinha, S. Nanoparticle fabrication on bacterial microcompartment surface for the development of hybrid enzyme-inorganic catalyst. *ACS Catal.* 8: 7742–7748. (2018).
- Baek, M., Park, T., Heo, L., Park, C. & Seok, C. GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res.* 45, W320–W324 (2017).
- Bienick, M.S., Young, K.W., Klesmith, J.R., Detwiler, E.E., Tomek, K.J., and Whitehead, T.A. The interrelationship between promoter strength, gene expression, and growth rate. *PLoS One* (2014).
- Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature Publishing Group* 14, 53–56 (2017).
- Castellana, M., Wilson, M.Z., Xu, Y., Joshi, P., Cristea, I.M., Rabinowitz, J.D., et al. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* 32: 1011–1018. (2014).
- Chen, R., Chen, Q., Kim, H., Siu, K.H., Sun, Q., Tsai, S.L., and Chen, W. Biomolecular scaffolds for enhanced signaling and catalytic efficiency. *Curr. Opin. Biotechnol.* 28: 59–68. (2014).
- Chen, X., Zaro, J.L., and Shen, W.C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* 65: 1357–1369. (2013).
- Culley, S., Albrecht, D., Jacobs, C., Pereira, P.M., Leterrier, C., Mercer, J., and Henriques, R. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* 15: 263–266. (2018).
- Delebecque, C.J., Lindner, A.B., Silver, P.A., and Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science (80-.)*. 333: 470–474. (2011).
- Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A. V, et al. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27: 753–759. (2009).
- Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16: 67–70. (2019).

- Frederix, P. W. J. M., Patmanidis, I. & Marrink, S. J. Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem Soc Rev* 47, 3470–3489 (2018).
- Giessen, T.W. and Silver, P.A. Encapsulation as a strategy for the design of biological compartmentalization. *J. Mol. Biol.* 428: 916–927. (2016).
- Glover, D.J. and Clark, D.S. Protein calligraphy: A new concept begins to take shape. *ACS Cent. Sci.* 2: 438–444. (2016).
- Good, M.C., Zalatan, J.G., and Lim, W.A. Scaffold Proteins: Hubs for controlling the flow of cellular information. *Science* (80). 332: 680–686. (2011).
- Gustafsson, N., Ashdown, G., Owen, D., Pereira, P., Henriques, R., and Culley, S. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications* (2016).
- Hagen, A., Kerfeld, C. In vitro assembly of engineered bacterial microcompartment shells enables encapsulation of diverse cargo. *Nano Letters* (2019).
- Horn, A.H.C. and Sticht, H. Synthetic Protein scaffolds based on peptide motifs and cognate adaptor domains for improving metabolic productivity. *Front. Bioeng. Biotechnol.* (2015).
- Huber, I. *et al.* Construction of recombinant pdu metabolosome shells for small molecule production in *Corynebacterium glutamicum*. *ACS Synth Biol* (2017).
- Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. The cost of protein production. *Cell Rep.* 14: 22–31. (2016).
- Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7, 1511–1522 (2012).
- Kerfeld, C.A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* (2018).
- Lassila, J. K., Bernstein, S. L., Kinney, J. N., Axen, S. D. & Kerfeld, C. A. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *Journal of Molecular Biology* 426, 2217–2228 (2014).
- Lee, H., DeLoache, W.C., and Dueber, J.E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14: 242–251. (2012).

- Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S., and Warren, M.J. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. *Metab. Eng.* 36: 48–56. (2016).
- Lee, M.J., Mantell, J., Brown, I.R., Fletcher, J.M., Verkade, P., Pickersgill, R.W., et al. De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. *Nat. Commun.* (2018).
- Lee, M.J., Mantell, J., Hodgson, L., Alibhai, D., Fletcher, J.M., Brown, I.R., et al. Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Biol.* 14: 142–147. (2018).
- Lee, M.J., Palmer, D.J., and Warren, M.J. Biotechnological advances in bacterial microcompartment technology. *Trends Biotechnol.* 37: 325–336. (2019).
- Lee, T., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., et al. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* 5: 12. (2011).
- Liang, M., Frank, S., Lünsdorf, H., Warren, M.J., and Prentice, M.B. Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in *E. coli*. *Biotechnol. J.* 12 (2011).
- Lim, W.A. Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.* 11: 393–403. (2011).
- Luo, Q., Hou, C., Bai, Y., Wang, R., and Liu, J. Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* 116: 13571–13632. (2016).
- MacCready, J.S., Hakim, P., Young, E.J., Hu, L., Liu, J., Osteryoung, K.W., et al. Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* (2018).
- Mahalik, J.P., Brown, K.A., Cheng, X., and Fuentes-Cabrera, M. Theoretical study of the initial stages of self-assembly of a carboxysome's facet. *ACS Nano* 10: 5751–5758. (2016).
- Myhrvold, C., Polka, J.K., and Silver, P.A. Synthetic Lipid-containing scaffolds enhance production by colocalizing enzymes. *ACS Synth. Biol.* 5: 1396–1403. (2016).
- Pugh, G. C., Burns, J. R. & Howorka, S. Comparing proteins and nucleic acids for next-generation biomolecular engineering. *Nature Reviews Chemistry* 2018 2:7 2, 113–130 (2018).
- Plegaria, J.S and Kerfeld, C.A. Engineering nanoreactors using bacterial microcompartment architectures. *Current Opinion in Biotechnology* 51:1-7, (2018).
- Savage, D.F., Afonso, B., Chen, A.H., and Silver, P. A Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* 327: 1258–61. (2010).

- Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Publishing Group* 9, 676–682 (2012).
- Schmidt-Dannert, S., Zhang, G., Johnston, T., Quin, M.B., and Schmidt-Dannert, C. Building a toolbox of protein scaffolds for future immobilization of biocatalysts. *Appl. Microbiol. Biotechnol.* 102: 8373–8388. (2018).
- Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat Meth* 10, 407–409 (2013).
- Siu, K.H., Chen, R.P., Sun, Q., Chen, L., Tsai, S.L., and Chen, W. Synthetic scaffolds for pathway enhancement. *Curr. Opin. Biotechnol.* 36: 98–106. (2015).
- Streets, A.M. and Quake, S.R. Ostwald ripening of clusters during protein crystallization. *Phys. Rev. Lett.* 104 (2010).
- Sutter, M., McGuire, S., Ferlez, B., and Kerfeld, C.A. Structural characterization of a synthetic tandem-domain bacterial microcompartment shell protein capable of forming icosahedral shell assemblies. *ACS Synth. Biol.* 8: 668–674. (2019).
- Thompson, K.E., Bashor, C.J., Lim, W.A., and Keating, A.E. Synzip protein interaction toolbox: In vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* 1: 118–129. (2012).
- Wang, Y., Heermann, R., and Jung, K. CipA and CipB as scaffolds to organize proteins into crystalline inclusions. *ACS Synth. Biol.* 6: 826–836. (2017).
- Wheeldon, I., Minter, S.D., Banta, S., Barton, S.C., Atanassov, P., and Sigman, M. Substrate channelling as an approach to cascade reactions. *Nat. Chem.* 8: 299–309. (2016).
- Whitaker, W.R. and Dueber, J.E. Metabolic pathway flux enhancement by synthetic protein scaffolding. *Methods Enzymol.* 497: 447–468. (2011).
- Xu, M., Singla, J., Tocheva, E.I., Chang, Y.W., Stevens, R.C., Jensen, G.J., and Alber, F. De Novo structural pattern mining in cellular electron cryotomograms. *Structure* 27: 679–691 (2019).
- Young, E.J., Burton, R., Mahalik, J.P., Sumpter, B.G., Fuentes-Cabrera, M., Kerfeld, C.A., and Ducat, D.C. Engineering the bacterial microcompartment domain for molecular scaffolding applications. *Front. Microbiol.* (2017).
- Zhang, G., Quin, M.B., and Schmidt-Dannert, C. Self-assembling protein scaffold system for easy in vitro coimmobilization of biocatalytic cascade enzymes. *ACS Catal.* 8: 5611–5620. (2018).

Ziatdinov, M., Dyck, O., Maksov, A., Li, X., Sang, X., Xiao, K., et al. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS Nano* 11: 12742–12752. (2017).

Chapter 4: Perspective on future opportunities with engineered components

This doctoral education experience began with a broad goal: build a new system for spatially organizing biochemical function from modules with the pfam0936-domain. Over this ~6-year timeline, I dived into a variety of techniques, instrumentation, and concepts. Below accounts some “loose-ends” and personal takes on this project, along with some broader thoughts as well.

A Parts Library

An integral aspect to any synthetic biology endeavor consists of a well-documented library of genetic parts. Work from this project has created an *Escherichia coli*-based glycerol stock, alongside a corresponding plasmid bank, totaling near 200 individual sequences of novel assembly and cargo modules (**Figure 4.1A, Ducat Lab Benchling Server**). Notable uncharacterized examples include: HO-BMC-H-based (MicH in some documentation) modules with amino acid substitutions and SZ-adaptor domain fusions, specialized SuperResolution microscopy fluorescent cargo modules, and contrast enhancing cargo for electron microscopy (Wang 2014; Lam 2014). Most sequences currently reside in expression vectors of differing promoters (*i.e.*, some in high over-expression T7-based vectors, others with tunable features) (Lee 2011). The “ease” of modern molecular biology cloning techniques makes it so these sequences could readily transfer to other chassis organisms or expression platforms.

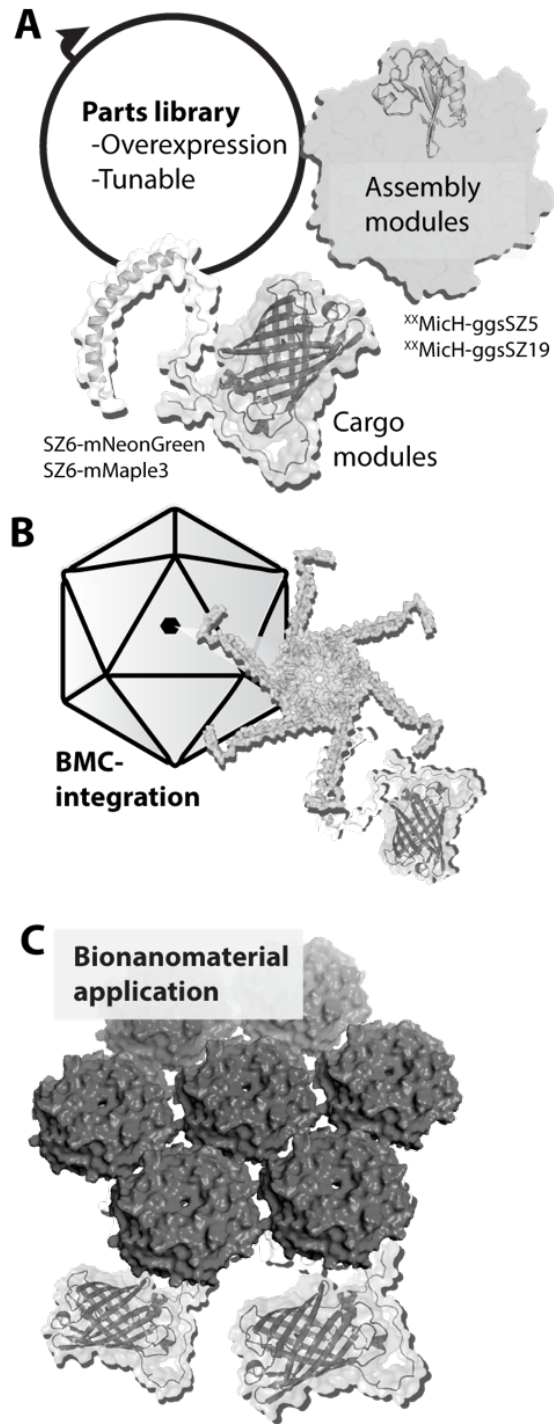


Figure 4.1: Broad follow up projects with engineered modules. A.) A molecular ‘parts’ library **B.)** Integrating modules into BMC-architectures **C.)** Applying modules towards more bionanomaterial driven-applications

Opportunities for integrating designed assembly and cargo modules into frameworks which produce bacterial microcompartment architectures also exists (**Figure 4.1B**). In light of structural models supporting C-terminal extensions protruding on the outside of BMCs (Plegaria 2017; Kerfeld 2018), SZ-tagged assembly modules may serve as unique functional adaptors to BMC-architectures. Modules developed could find use in disentangling native behaviors or creating new designer function(s) (Lee 2019). For instance, SZ-mediated interactions could tether cargo to the surface of a BMC, or circular permutation of termini may allow specific organization to the interior—even other fusion designs may create new unexpected properties (Sutter 2019).

This parts library not only forms a resource for projects using assembly and cargo modules within living systems, but also as an *in vitro* bionanomaterial toolkit (**Figure 4.1C**). Some properties of pfam0936-domain proteins (*e.g.*, spontaneous assembly, wide-temperature stability, defined structure) make them attractive candidates as supports for biological catalysts (*e.g.*, enzymes) and inorganic catalysts (Hagen 2018; Schmidt-Dannert 2018; Bari 2018). Whether as fully formed compartments or as structural scaffolds, these building blocks could serve as the next molecular “circuit board” for bionanomaterial engineering.

Improving *in situ* Imaging and Analysis

The bulk of investigative efforts explored visualizing the behavior of higher-order organization of modules with microscopic imaging techniques. Improving spatial-temporal resolution forms a paramount aspect in creating predictive models for nanoscale assembly. While some groundwork has begun, building a truly predictive nanoscaffolding system calls for use of state-of-the-art techniques and extensive inter-disciplinary action. We will briefly touch on a few technologies I believe important for advancement.

Transmission electron microscopy ultrathin section excelled at identifying general ultrastructure features of intracellular higher-order assemblies produced, but suffered in producing any details beyond that—especially those in 3D organization. Serial imaging of sections could help visualize 3D ultrastructure organization. Tomographic tilt-series imaging represents another potentially fruitful approach. Using a cryo-electron microscopy based for investigation may also reduce potential artifacts from standard transmission electron microscopy sample preparation, while also bringing additional opportunity for class averaging methods (Milne 2013; Okonomou 2016).

Continued to apply light-microscopy based techniques will also serve as another powerful tool for imaging *in situ* behavior. Information captured in Chapter 3 barely dipped into the plethora of platforms created for visualizing samples with light-based methodology (Liu 2014; Görlitz 2017; Pujals 2017). One route forward would work to combine multiple super-resolution imaging suites (*e.g.*, structured illumination, stimulated emission depletion, single-molecule localization methods) to improve models of cargo and assembly

organization, even though “simpler” widefield type set-ups could still yield worthwhile data sets. Regardless of the specific imaging platform, designing experiments with an eye towards collecting 4D information—those that include changes over-time in XYZ position—will illuminate the most productive datasets (—see Chapter 5 material for sample proposal).

Any data captured with a microscopic imaging suite presents additional opportunity for deeper-level analysis by external processing packages. ImageJ (more specially the FIJI “batteries included” version) worked as one of most integral parts of this PhD project (Schindelin 2012). Whether enhancing the contrast of electron microscopy images or compiling frames from a light-microscopy time-series, without this tool, producing much of the data reported would have presented much more difficulty. I also personally feel that I have just scratched the surface on what appears offered as a part of FIJI. One notable plugin found within the FIJI framework that could directly aid in light-microscopy image analysis and build more quantitative models comprises MicrobeJ (Ducret 2016). There also exists a host of other plugins developed for super-resolution imaging suites that can improve reliable data collection and analysis (this would include the previously utilized plugins part of the NanoJ suite) (Ball 2015; Sage 2015). Machine-learning algorithms—such as the U-net plugin within FIJI (Falk 2019)—represent another developing way to diminish user-bias in microscopic image analysis and uncover any hidden patterns in complex data sets (Brent 2018; Arganda-Carreras 2017).

It will also take applying other imaging technologies to gain the capacity to comprehensively model cargo and assembly behavior at the nanoscale. X-ray crystallography continues to

serve as one benchmark technique to inform researchers of the nearly-atomic structures of native and designed modules. This approach offers powerful “structural snapshots” of particle conformations that arise during an *in vitro* crystallization process. Small-angle scattering techniques (X-ray or neutron-based) serves as another method for visualizing nanoscale properties (Jacques 2010; Hong 2014). Using neutron-based scattering approaches has an added benefit of specific-labeling (resulting from deuterium-labeling of desired components) for creating increased contrast of samples (Callaway 2015). Studying modules with scattering techniques in isolated systems (*in vitro*)—paired with other biophysical techniques (e.g. atomic force microscopy (Nievergelt 2018))—would work as one starting point (—see Chapter 5 material for sample proposal). Eventually, making a gradual move towards *in vivo* data sets within more complex systems (*i.e.*, intracellular behavior inside a chassis organism) could compliment to other data sets (Gundlach 2016).

Computationally Modeling Behavior

Simulations, or data collected from computationally driven methods, could also improve our models of assembly and cargo behavior. In this project, atomistic molecular dynamic simulations acted as one resource for visualizing potential dynamics of three pfam0936-domain proteins (**Figure 4.2, Movies 8-10**). Briefly, atomistic molecular dynamics can simulate a trajectory of atoms in a protein-structure within a solvated-water box (Frederix 2018). At this point, rather than draw significant conclusions (where starting structures and time-scales strongly dictate the simulation landscape) only some general themes appear noted.

Each module experienced a length of simulation-time (>100 ns) until reaching an equilibrium state. In the case of an unmodified HO-BMC-H—assembled as a hexameric subunit—the C-termini have the greatest freedom in motion (**Figure 4.2a, Movie 8**). Over the course of the simulation, tails appear to be gently sampling different conformations, sometimes forming brief bonds with neighboring tails or the hexamer beneath. This behavior contrasts with the simulation of HO-BMC-H fused to SZ5 with a rigid linker design (ppg) (**Figure 4.2b, Movie 9**). Already in a particular crisscrossed conformation from the starting structural-model, extensions continue to stay locked and frustrated together throughout the simulation. In comparison, HO-BMC-H fused to SZ5 via a flexible linker design (ggs) starts from a more extended conformation (**Figure 4.2c, Movie 10**). The extensions then sample varying positions over-time, sometimes interacting with each-other, while other times bonding momentarily to the hexamer below. Deeper-level analysis in software—such as Visual Molecular Dynamics (Humphrey 1996)—will help to note any trends in local changes in structure. Expanding to more complex systems (*i.e.*, such as those with several hexamers, cargos, or crowders), iterative trajectories, and longer simulation time-scales will create even-more worthwhile models—coarse-graining methods will prove indispensable too (Frederix 2018; Feig 2017; Shoemark 2018; Perkett 2014).

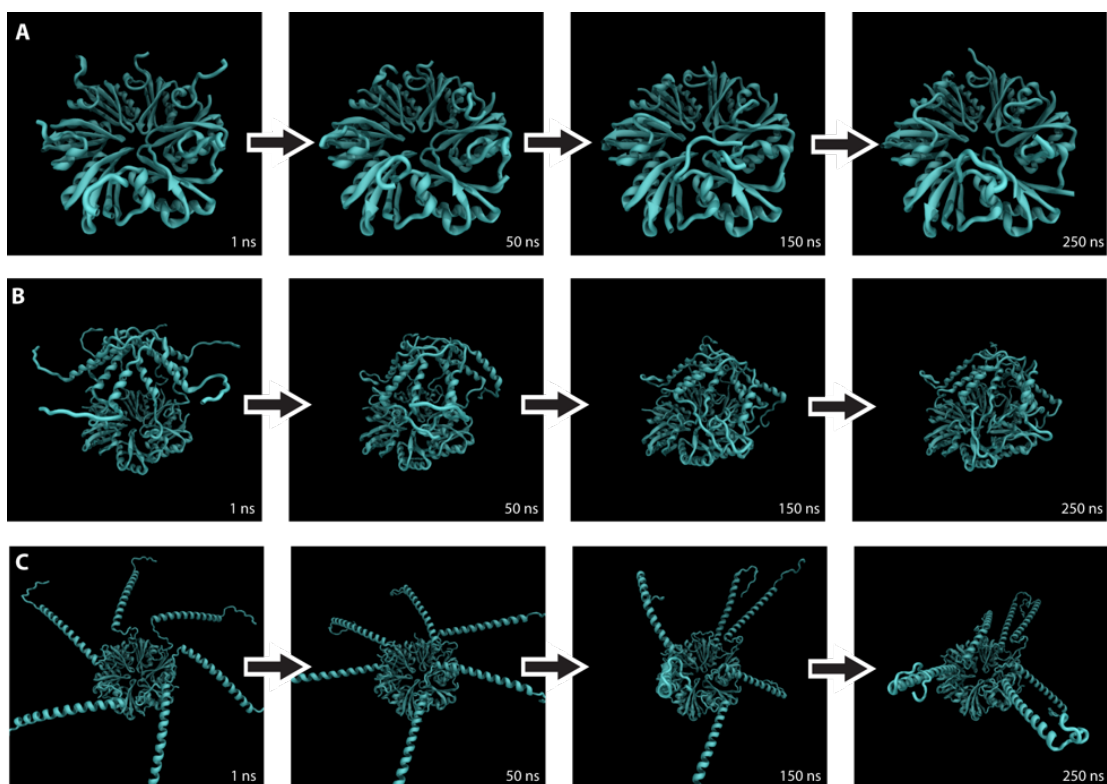


Figure 4.2: Structural snapshots of individual frames from atomistic molecular dynamic simulations. A.) $^{WT}HO-BMC-H$ B.) $^{WT}HO-BMC-H_{ppgSZ5}$ C.) $^{WT}HO-BMC-H_{ggsSZ5}$

*Special thanks to Miguel Fuentes-Cabrera, Xiaolin Cheng, and NERSC facilities for running calculations

LITERATURE CITED

LITERATURE CITED

- Arganda-Carreras, I. et al. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33, 2424–2426 (2017).
- Ball, G. et al. SIMcheck: A toolbox for successful super-resolution structured illumination microscopy. *Scientific Reports* 5, 15915 (2015).
- Brent, R., et al. Deep learning to predict microscope images. *Nature Methods*. (2018).
- Callaway, D. J. E. & Bu, Z. Nanoscale protein domain motion and long-range allostery in signaling proteins—a view from neutron spin echo spectroscopy. *Biophys Rev* 7, 165–174 (2015).
- Ducret, A., Quardokus, E. M. & Brun, Y. V. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology* 1, 1–7 (2016).
- Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nature Publishing Group* 16, 67–70 (2019).
- Feig, M., Yu, I., Wang, P.-H., Nawrocki, G. & Sugita, Y. Crowding in cellular environments at an atomistic level from computer simulations. *J Phys Chem B* 121, 8009–825 (2017).
- Frederix, P. W. J. M., Patmanidis, I. & Marrink, S. J. Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem Soc Rev* 47, 3470–3489 (2018).
- Görlitz, F. et al. Mapping Molecular Function to Biological Nanostructure: Combining Structured Illumination Microscopy with Fluorescence Lifetime Imaging (SIM + FLIM). *Photonics* (2017).
- Greber, B. J., Sutter, M. & Kerfeld, C. A. The plasticity of molecular interactions governs bacterial microcompartment shell assembly. *Structure* 27, 749–763 (2019).
- Gundlach, von, A. R. et al. Use of small-angle X-ray scattering to resolve intracellular structure changes of Escherichia coli cells induced by antibiotic treatment. *J Appl Crystallogr* 49, 2210–2216 (2016).
- Hagen, A. R. et al. In vitro assembly of diverse bacterial microcompartment shell architectures. *Nano Letters* (2018).
- Hong, L., Smolin, N. & Smith, J. C. de Gennes narrowing describes the relative motion of protein domains. *Phys. Rev. Lett.* 112, 158102 (2014).

- Humphrey, W., Dalke, A. and Schulten, K., VMD - Visual Molecular Dynamics, *J. Molec. Graphics*, vol. 14, pp. 33-38. (1996).
- Jacques, D. A. & Trewthella, J. Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci.* 19, 642–657 (2010).
- Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* 16, 277–290 (2018).
- Lam, S. S. et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Meth* 12, 51–54 (2014).
- Lee, M. J., Palmer, D. J. & Warren, M. J. Biotechnological advances in bacterial microcompartment technology. *Trends in Biotechnology* 37, 325–336 (2019).
- Lee, T. S. et al. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* 5, 12 (2011).
- Liu, Z. et al. Super-resolution imaging and tracking of protein-protein interactions in sub-diffraction cellular space. *Nature Communications* 5, 4443 (2014).
- Milne, J. L. S. et al. Cryo-electron microscopy--a primer for the non-microscopist. *FEBS J.* 280, 28–45 (2013).
- Nievergelt, A. P., Banterle, N., Andany, S. H., Gönczy, P. & Fantner, G. E. High-speed photothermal off-resonance atomic force microscopy reveals assembly routes of centriolar scaffold protein SAS-6. *Nature Nanotechnology* 33, 302 (2018).
- Oikonomou, C. M. & Jensen, G. J. A new view into prokaryotic cell biology from electron cryotomography. *Nature* (2016).
- Perkett, M. R. Studying protein self-assembly and conformational dynamics using rare event simulation methods. PhD Thesis: Brandeis University (2014).
- Plegaria, J. S. & Kerfeld, C. A. Engineering nanoreactors using bacterial microcompartment architectures. *Current Opinion in Biotechnology* 51, 1–7 (2017).
- Pujals, S. X. L., Tao, K., Terradellas, A. X., Gazit, E. & Albertazzi, L. Studying structure and dynamics of self-assembled peptide nanostructures using fluorescence and super resolution microscopy. *Chem. Commun.* 53, 7294–7297 (2017).
- Sage, D. et al. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat Meth* 12, 717–724 (2015).
- Schindelin, J., Arganda-Carreras, I., Frise, E., Nature, V. K. Fiji: an open-source platform for biological-image analysis. *Nat Meth* (2012).

- Schmidt-Dannert, S., Zhang, G., Johnston, T., Quin, M. B. & Schmidt-Dannert, C. Building a toolbox of protein scaffolds for future immobilization of biocatalysts. *Appl. Microbiol Biotechnol.* 102, 8373–8388 (2018).
- Shoemark, D. K. et al. The dynamical interplay between a megadalton peptide nanocage and solutes probed by microsecond atomistic MD; implications for design. *Phys Chem Chem Phys* (2018).
- Sutter, M., McGuire, S., Ferlez, B. & Kerfeld, C. A. Structural characterization of a synthetic tandem-domain bacterial microcompartment shell protein capable of forming icosahedral shell assemblies. *ACS Synth Biol* 8, 668–674 (2019).
- Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8452–8457 (2014).

Chapter 5: Personal remarks on select co-authored publications, example lead-author proposals, and a reference to a co-authored public communication piece

This chapter represents a collection of information in hopes to inspire ‘that next step’ in some of the projects I had the opportunity of conducting research in outside of my primary thesis project. Additionally, I have included 3 lead-author scientific research proposals, to help advance the broader fields of synthetic biology/bionanomaterials and aid others looking to propose their own research concepts. Finally, I share a link to a co-authored public-orientated scientific communication piece because I believe it represent a developing aspect important to influencing public perception positively about the hard-work and dedication that can go in to publicly-funded science.

Remarks on co-authored projects—

Giant Viral-particle Fluorescent Organization

(Published originally in Schrad et al. 2017)

Wonders in nanoscale structure/function, viral systems come in many shapes and sizes. Discovery of “giant” viruses (viral-like particles readily observed with light microscopy) has changed the scope of the viral-world even more. By applying electron-microscopy and light microscopy approaches, we imaged attributes of a giant viral system (‘Samba mimivirus’). Comparing results to another characterized giant viral system from the same family (‘*Acanthamoeba polyphaga* mimivirus’) finds conserved, alongside some unique features. My personal contributions to this work involved imaging the two giant viral systems with light microscopy methods. In transmitted light imaging, we found Samba mimivirus formed a conjoined lattice of viral particles, a feature lacked by the *Acanthamoeba polyphaga* mimivirus sample. Fluorescent imaging of biomolecule specific dyes (DNA/protein) corroborated electron microscopy observations in organization, but questions remained addressable with improved light microscopy approaches.

Rendering a cryo-electron microscopy single-particle 3D organization proved difficult due to Samba mimivirus particle heterogeneity/size. 3D imaging with light microscopy may help describe organization. While giant viral systems exist at the cusp of a meaningful resolution with light microscopy, imaging with specifically labeled fluorescent molecules, just as done previously, but with microscope possessing a total-internal-reflectance set-up and motorized Z-sections would create representations with more volume associated information. This general idea could also extend to other light microscopy set-ups which

collect sub-diffraction information (such as 3D-SIM) to bolster models. The ImageJ associated plugin (NanoJ-VirusMapper) would work as a useful resource in classifying imaged particles and generating spatial maps better detailing organization (Gray 2016). Furthermore, processing images with ImageJ spatial filter plugins (e.g. NanoJ-SRRF) and using fluorescent probes more amenable to super-resolution imaging regimes (e.g. Alexafluor-684) could improve sub-diffraction information.

Ultrastructure of Cyanobacteria Encapsulated in a Polymer-matrix

(Published originally in Weiss et al. 2017)

Cyanobacteria form an exciting class of chassis organisms used for synthetic biology engineering efforts in part to their photosynthetic-driven metabolism. By incorporating a heterologous sugar transport module, a cyanobacteria strain (*Synechococcus elongatus* PCC 7942 CscB⁺) can export sucrose-molecules at productivities surpassing other typically-considered carbohydrate sources (e.g. sugar cane, corn). ‘Plug and play’ application of this strain with other chassis organisms creates opportunity for designer “co-culture” systems, driven by “green” photosynthetically-produced carbon. In this work, a bio-plastic producing organism (*Halomonas boliviensis*), along with encapsulating sucrose-secreting cyanobacteria within a bead polymer-matrix formed our co-culture system. I supplemented this project by studying nanoscale organization of both chassis’ organisms with transmission electron microscopy thin section. Imaging dominating electron opaque bodies within *Halomonas boliviensis* taken as produced bio-plastic reserves. Studying encapsulated cyanobacteria revealed an interesting observation: confinement appeared to alter expectations for cells undergoing a routine binary fission. Instead of growing in

exponentially in number, cells in “pockets” within a bead maintained single digit numbers throughout measurement time-points (~150 days).

Given that polymer-encapsulation has emerged as one design approach for spatially organizing chassis organisms, the influence of encapsulation on physiology warrants deeper investigation. Resolving this appears addressable in several ways, but will require some creativity. Special light-microscopy set-ups engineered to track cell development over-time within beads, or other polymer structures, would help visualize this. Again, a more destructive method using fixed sections with either electron microscopy or scanning probe methods could also help, but an keeping an eye towards creating comprehensive data sets (i.e. those with finer time resolution) would benefit projects the most.

Ultrastructure of Spatially Organized Organelles in Cyanobacteria

(Published originally in MacCready et al. 2018)

The faithful segregation and spatial distribution of particular cellular components plays a vital role for all organisms. In the case of cyanobacterial carbon fixation “nanofactories” (called carboxysomes), literature observed instances of non-random distribution within cells, but the underlying mechanisms and partners in the process remained poorly understood. This study worked in genetically identifying partners influencing spatial positioning of carboxysomes (genes named maintenance of carboxysome distribution A/B) and then built a model describing the spatial organization mechanism based on *in vitro*, *in vivo*, and *in silico* methodology. My contributions to this project consisted of analyzing overexpression, or knock-out, strains of Mcd genes with TEM biological thin section in our

model cyanobacteria ('*Synechococcus elongatus* PCC 7942'). Data supported general trends that elements play an important role in even intracellular spatial positioning, but we also observed ultrastructural changes to assembled carboxysomes. This led us to propose that positioning elements may play a role in helping to regulate the actual structure of carboxysomes.

Teasing apart consequences the Mcd system has on carboxysome ultrastructure/composition appears important not only for understanding an exciting nanoscale spatial organization system, but for adopting these modules for designer functionality. Again, I propose imaging microscopy to play a critical role in this. Another attempt in a more comprehensively collected data set in TEM would serve to validate previous observations. Using light-microscopy to monitor fluorescent-reporter lines, but with more advanced microscopy methods capable of capturing sub-diffraction information will also elude to any change in morphology/composition. Finally, studying components *in vitro* and *in silico* with thoughts towards mimicking intracellular conditions may reveal new behaviors obscured by other methodology.

A collection of peer-reviewed lead-author proposals—

2017 Oakridge National Lab CNMS User Proposal

As our understanding and ability to engineer cells at the nanoscale grows, actively considering spatial organization of introduced function is an important design consideration moving forward (Young 2017). Scaffolding is one type of spatial organization strategy commonly used throughout natural biological systems to recruit like-function on an orthogonal component (e.g. membranes, nucleic acids, proteins) (Young 2017). Biological engineers have applied this concept to increase product outputs of metabolic pathways and rewire signaling in cells to some success, yet current synthetic scaffolds developed lack clearly defined surfaces, limiting rational pathway engineering. We aim to engineer a family of self-assembling proteins to construct in vivo architectures with defined and predictable assembly that could be utilized as molecular scaffolds.

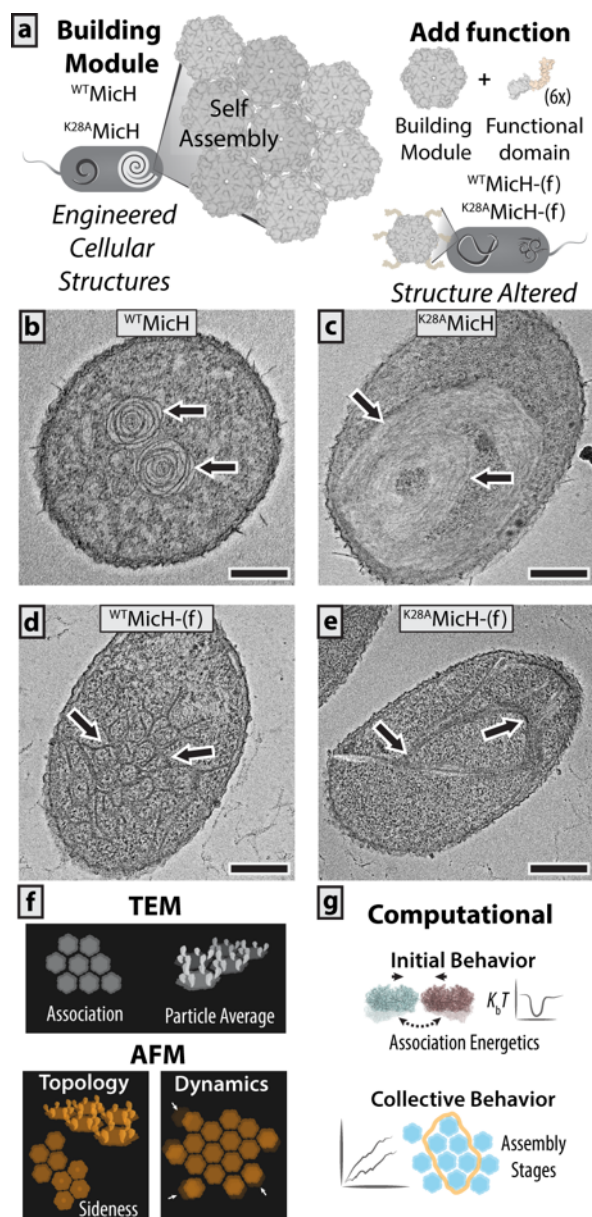


Figure 5.1: Engineering and characterizing assemblies for molecular scaffolding. A.) Schematic of engineering heterologous higher-order structures built from self-assembling proteins **B-E.)** *In vivo structure* formation of indicated building material **F.)** Structurally characterizing building modules via TEM and AFM **G.)** Computational mean to investigate building module association and collective behavior

Our building material is based on a single protein normally found assembling the facets of a bacterial microcompartment (**Fig. 5.1a**). In isolation, this protein—named MicH for this proposal—has been shown to self-assemble into large, macromolecular structures *in vivo* (Young 2017; Lassila 2014) and extended hexagonal arrays *in vitro* (Sutter 2016). Not only does the wild-type module (^{wt}MicH) self-assemble into a distinct *in vivo* structure (**Fig. 5.1b**), modification (^{K28A}MicH) results in a differing *in vivo* higher-order architecture (**Fig. 5.1c**). The *in vivo* higher-order structures formed are described as rosettes; likely a 2D lattice which begins to roll upon itself. We hypothesize differences in self-assembly behavior between the modules leads to the differing macromolecular formation (Young 2017). While these structures are intriguing intracellular assemblies to study self-assembly behavior, they lack a crucial feature to be used as a functional scaffold.

The next step in engineering a molecular scaffold is designing targeted recruitment to the assembly surface; this is known as functionalization. Our approach consists in appending a functional domain—a protein-protein interaction motif which would allow recruitment of a corresponding motif—to either building module. After fusion, self-assembly is maintained (**Fig. 5.1d,e**), but the macromolecular structure differs compared to the parent building modules (**Fig. 5.1b,c**). ^{wt}MicH-(f) maintains the filament characteristics of the parent module, but forms more of a ‘webwork’ versus a rosette; ^{K28A}MicH-(f) also loses the rosette phenotype, forming filaments which appear bundled in some area of the cell. We would like to investigate the structural details and assembly dynamics which differ between the functionalized and parent MicH modules. For this, we propose using the combination of high-resolution imaging techniques and computational modeling. This will begin to elucidate how

functionalized MicH modules form (macro)molecular assemblies, while establishing design principles of assembly for our scaffold building material¹. Knowledge gained from this proposal ultimately will increase our ability to engineer predictively assembling and structurally defined in vivo molecular scaffolds, while also supplementing the broader themes within CNMS Interface Directed Assembly research foci.

Our core questions of this proposal comprise:

- 1.) What is the in vitro (macro)molecular assembly of functionalized MicH modules?
- 2.) What are the differences in assembly dynamics between modules?
- 3.) What can computational modeling tell us about the initial- and mid stages of self-assembly?

AFM and TEM will be used to visualize and measure (macro)molecular assembly characteristics and dynamics of both parent and functionalized MicH models. Purified, recombinant protein will be used in these experiments. Computational resources will also be used to investigate (macro)molecular assembly. This will range from atomistic molecular dynamics to Metropolis Monte Carlo simulations with coarse grained potentials—methods developed in (Young 2017; Mahalik 2016). As a step further, a look up table methodology has been developed by our collaborators at CNMS, which is written in Java and already deployed in CADES (<https://cades.ornl.gov>). This will allow us to visualize even larger states of assembly (i.e. hundreds of modules).

The purification of parent and functionalized MicH modules will be done at the PI's home institution, as this is already routinely done and is straightforward. All four modules are stable proteins and therefore can be transported with minimal difficulty (i.e. shipping on ice). At our institution, we will also perform the initial characterization steps of assembly via DLS and negative stain TEM to narrow down conditions (e.g. buffers, concentrations) amendable for higher-order structuring.

With the technical expertise and equipment at CNMS, we aim to uncover the structural characteristics and dynamic behaviors of the parent and functionalized MicH modules. To accomplish this, we will apply AFM and TEM techniques (**Fig. 5.1f**). Both techniques have been utilized to study the parent modules in terms of their structural and behavioral attributes (Young 2017; Sutter 2016). First, routine TEM negative stain will be used to visualize the macromolecular assembly type of the four modules, while conventional contact-mode AFM will be used to measure assembly patch size and thickness of assemblies. Averaging AFM topographs will also resolve additional features such as respective orientation of the molecular sheets ('sideness'). Previous published data collected with high-speed AFM also demonstrates differences in the association rates of the parent MicH modules (Sutter 2016), suggesting we will also see a change in the dynamics of functionalized modules once measured. If feasible, more advanced methods utilizing EM will be done to further uncover differences in module association and higher-order assembly. Particle class averages could be used to generate a more-detailed picture—moving towards molecular resolution—of module association, while EM tomography would render a clearer 3D picture. Information gathered in the AFM and TEM experiments will then be related back

to what is known/predicted about the atomic structures and eventually supplement models of assembly found in computational avenues.

In parallel utilizing resources at CNMS NTI, we will investigate the assembly of our building modules with computational approaches. We will first create MD-relaxed structural models of the parent (PDB:5DJB) and functionalized MicH modules. Comparisons between the ‘static’ and ‘relaxed’ structures will provide insights into possible structural changes which occur to these modules outside the confines of a crystalline lattice (Young 2017). These structures will be used subsequently to compute a coarse-grained potential of mean force (PMF) using the Thomas-Dill potential⁴ (Fig. A.1g). From this PMF—which quantifies the energetics based on relative orientation and distance—look-up tables will be generated. Armed with the relative orientation coarse-grained PMF and look-up tables, we will simulate the initial and mid-stages of self-assembly, respectively, for each module. This will be quantified by K-means clustering techniques for the presence of short and long-range ordering. In addition to generating a PMF based on relative orientation, we will attempt to generate a PMF based on angular association for each module. Recently, this approach was successfully adopted to investigate preferred interaction angles and lend clues into assembly in 3D (Young 2017). We hypothesize this simulation data will compare well with recent structural knowledge gained on preferred interaction angles of MicH found in the setting of a compartment facet (Sutter 2017).

Ultimately, through the combination of these research tasks at CNMS, we hope to develop a model of assembly from single molecules to higher-order assemblies of the parent and

functionalized building modules. By enhancing our image of assembly, our capabilities to predicatively engineer structures for scaffold utilization based on our building material will increase, furthering the repertoire of materials to construct in vivo molecular scaffolds.

We anticipate the experimental side of the project (AFM, TEM) to require 15 days total to investigate the tasks outlined for the four protein modules. The theoretical tasks we have described will require 30 days' time and 300,000 CPU hours at NERSC and 400,000 CPU hours at NTI.

Our team has extensive experimental experience and background knowledge on the family of proteins utilized as our building material. We recently published an in-depth perspective article in collaboration with the Fuentes-Cabrera (ORNL) and Kerfeld (MSU/LBNL) groups on this subject.

Technology to Investigate Intracellular Assembly of Functional Nanostructures

2017 DOE-SCGSR Awardee

Background

Microbes and other forms of life use a variety of nano-sized structures to organize molecular function (Howorka 2011). Such organization places like-function in closer proximity, increasing specificity and speed (Wheeldon 2016). Consequently, productively redesigning cells requires engineering synthetic nano-sized structures to rationally localize function (Papapostolou 2009). In pursuit of this, we are using a proteinaceous material to build functional nanostructures inside cells.

Our base-material comprises hexagonally shaped proteins formed from the pfam0936 protein family (**Fig. 5.2a**)—bacterial microcompartment domain, reviewed in (Young 2017). Sole expression of different assembly modules lends to the self-assembly of distinct nanostructures inside cells (**Fig. 5.2a**) (Young 2017). To engineer molecular cargo to the surface of assemblies, we incorporated a functionalizing domain and have preliminary data demonstrating self-assembly is maintained after addition (**Fig. 5.2b-d**). Additionally, functionality is supported by the recruitment of a reciprocally tagged cargo protein to presumed nanostructures (**Fig. 5.2e-g**). Current limitations in methodology are hindering detailed characterization of self-assembly behavior and measuring precision of recruitment (**Fig. 5.2h**). Establishing these finer details will enhance the capability of these modules as biological nanomaterials.

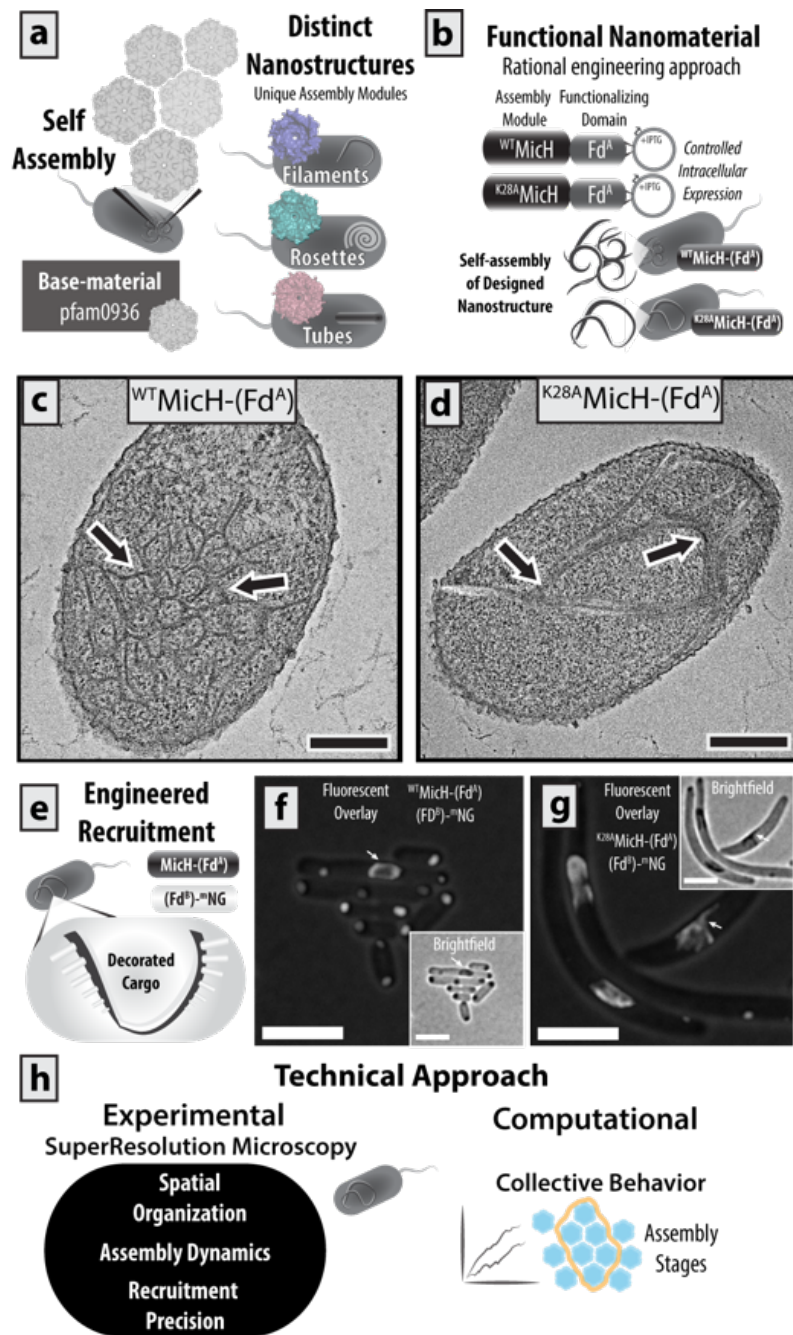


Figure 5.2: Engineering functional nanostructures. A.) Assembly modules with the pfam0936 domain can form nanostructures. **B.)** Rational engineering functional domains to assembly modules. **C-D.)** Self-assembly of distinct nanostructures formed by functionalized assembly modules. **E.)** Engineering cargo recruitment to functionalized nanostructures **F-G.)**

Figure 5.2 (cont'd) Specific targeting of fluorescent cargo protein to functionalized nanostructures **H.)** Technical approach to further investigate self-assembly and functionalization of assembly modules

We propose developing methods for the high-resolution visualization of nanostructure assembly and functionalization: this will aid productive intracellular engineering of user-defined biological materials and ways to investigate natural biological systems. Research will be conducted at Oak Ridge National Laboratory. Mentorship and collaboration will be with CNMS Nanomaterials Theory Institute theorist Miguel Fuentes-Cabrera, with additional support by ORNL Bio-inspired Nanomaterials and Biological and Nanoscale Systems staff (Point of Contact: Mitch Doktycz).

Briefly, state-of-the-art SuperResolution Microscopy (SRM) techniques—SIM, dSTORM, and fPALM reviewed in (Turkowsky 2016)—will be harnessed to track stages of assembly and measure recruitment precision (**Fig. 5.2h**). Overall, this proposal is novel in two regards: 1.) To our knowledge, this is the first instance of using SRM technology to investigate engineered nanostructures inside cells. 2.) Experimental data will inform computational models of self-assembly, working together towards predictive and rational structure design of our base-material

Ultimately, insight gained from the outlined objectives will increase our predictive knowledge and understanding of our assembly modules from single molecules to higher-order structure. This broad goal fits with current themes at ORNL CNMS and will build upon

previous productive collaborations with the Fuentes-Cabrera group (Young 2017). Additionally, by establishing high spatial/temporal imaging methods, addressing the Office of Science priority research area of *Novel In situ Imaging and Measurement Technologies for Biological System Science*. Finally, this proposal will contribute significantly to my thesis project by establishing techniques to collect high-resolution data on the nature of nanostructure assembly and cargo recruitment precision not available at my home institution by utilizing the expertise in biological imaging and theoretical modeling at ORNL.

Research Objectives and Goals

Our proposal can be broken up into two main research objectives (**Fig 5.2h.**). We expect these objectives to work synergistically in developing methods into nanoscale organization. Furthermore, experimental data will benchmark theoretical resources to refine models describing assembly behavior. By refining assembly behavior in this collaborate effort, we strive to establish preliminary design principles—rules enabling predictive understanding (Young 2017). This proposed approach will advance part of my thesis work of investigating self-assembly and functionalization of these modules (Young 2017), while also fitting in the broader theme of studying nanoscale organization.

Objective 1: Establishing SRM Localization Technique to Probe Nanostructure

Assembly

Preliminary data by conventional light microscopy suggests assembly size can increase in relationship to protein concentration **(Fig. 5.3a)**. This is supportive of computational modeling suggesting assembly is driven by a nucleation driven process dependent on local concentration of assembly units (Mahalik 2016). To increase the spatial resolution and through-put analyses of intracellular assembly, SRM localization technique Direct Stochastic Optical Reconstruction Microscopy (dSTORM) will be utilized. By virtue of the precise labeling and high spatial resolution of dSTORM, we anticipate visualizing intracellular assembly from beginning stages (small clusters) to larger nanostructures not feasible with conventional light or electron microscopy **(Fig. 5.3b)** (Turkowsky 2016, Gahlmann 2014).

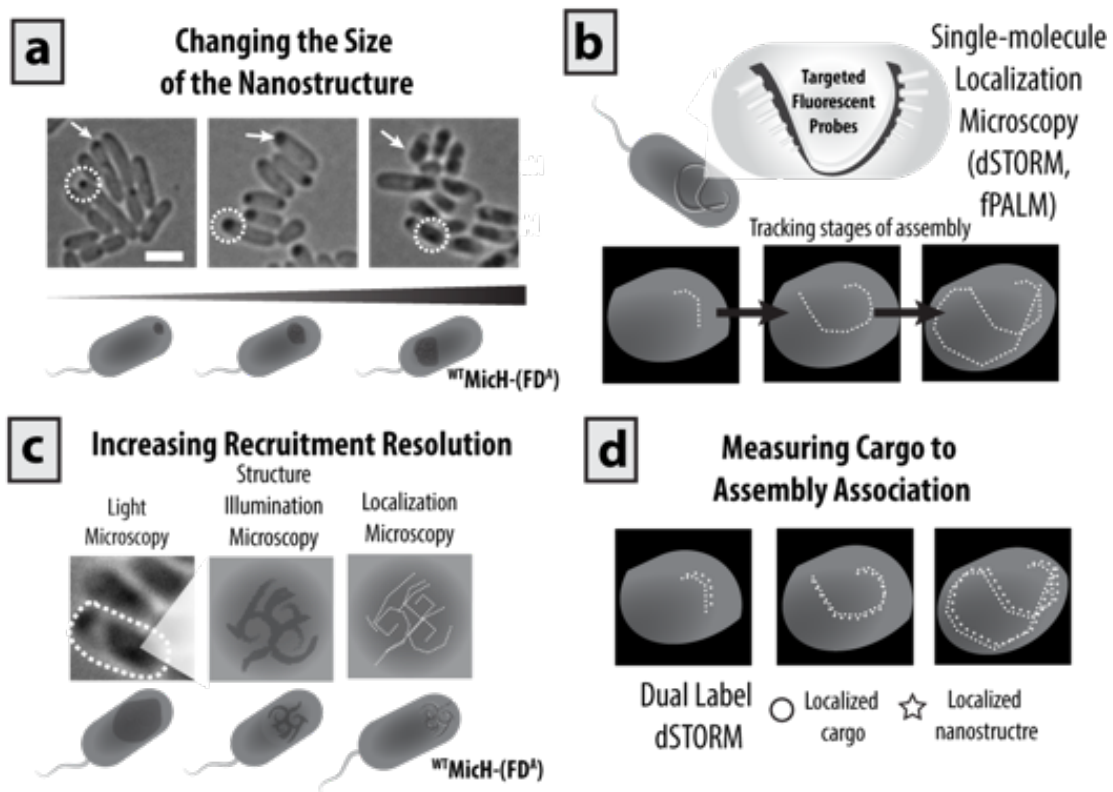


Figure 5.3: Investigating engineered nanostructures with superresolution microscopy methods. **A.)** Nanostructures size can be changed by intracellular concentration of assembly module production **B.)** Single-molecule localization microscopy to visualize assembly dynamics **C.)** Superresolution techniques to increase confidence and precision of recruitment **D.)** Dual-label localization microscopy will spatially visualize the relationship between cargo and engineered nanostructures

To target fluorescent probes for dSTORM imaging we will rely on the specific association of an immunogenic tag (HA) to an antibody in fixed-cells; this tag has been shown to be accessible *in vitro* and *in vivo* methods on our assembly modules (data not shown). Primary antibody conjugation to SR fluorescent probe Alexa-647 will be used to image assembly. Imaging will take place at ORNL Multimodality live-cell imaging lab on a Zeiss Elyra SuperResolution Confocal Microscope. The goal is to investigate assembly by manipulating the intracellular concentration of protein (length and strength of expression) **(Fig. 5.3b)**. Staff expertise at the laboratory will aid in correct analysis of SR localization data to achieve optimal spatial resolution and accuracy. Data will benchmark computational predicted assembly behavior.

Objective 2: Measuring Precision of Recruitment to Engineered Nanostructures

Resolution of engineered cargo recruitment **(Fig. 5.2e-g)** suffers from structural information lost due to the diffraction limitation of light (Turkowsky 2016); this is exemplified by the differences in the details of assembly between light and electron microscopy **(Fig. 5.3c-d, f-g)**. Refining the resolution of cargo recruitment will better reflect the nanoscale organization, increasing the confidence of targeted recruitment. This is an important consideration for rationally designing molecular function to engineered nanostructures (Young 2017). We will use SR techniques Structured Illumination Microscopy (SIM) and Fluorescent Photoactivation Localization Microscopy (fPALM)—along with dSTORM—in concert to refine recruitment resolution, working towards visualizing the precise relationship between cargo and nanostructure organization **(Fig 5.3c)**.

Applying three SRM techniques will serve as an approach to stepwise increase the confidence of recruitment and precision of recruitment. From a technical standpoint, we expect SIM to be the most approachable method to yield immediate improvements in resolution, increasing confidence of recruitment. In pursuit of increasing recruitment resolution to single molecules, both SRM localization techniques—dSTORM and fPALM—are amendable to our system. We will investigate parameters in both approaches, such as fluorescent protein identity (mNeonGreen, mMaple3) and probe labeling density/method, to yield the experimental set up most conducive to reproducible data at the highest molecular resolution. Finally, measuring the precision between the spatial relationship between nanostructure assembly location and cargo position will be attempted (**Fig. 5.3d**). This will be accomplished by dual-label dSTORM (Turkowsky 2016). Conditions will be systematically assessed again yields the most reproducible data at the highest molecular resolution. All experimental data in Objective 2 will be collected on the equipment specified in Objective 1, with guidance and collaboration by ORNL imaging support staff.

2018 Oakridge National Lab SNS-CNMS User Proposal

Proteins are one of nature's building blocks used for constructing nano to micro-scale multimeric structures and machines. Synthetic biology, material science, and other related disciplines hope to harness the amendable nanoscale properties found by proteins (self-assembly, defined molecular interactions) to develop modules for bottom-up nanostructure fabrication (Luo 2016). It can be anticipated these modules, and the defined, programmable nanoarchitectures built from them, will have far reaching applications across the fields of medicine, energy, and technology (Luo 2016).

Towards this broad goal, we are re-purposing a protein domain naturally found across bacteria (protein family: PF00936) as the base-material for rationally constructing protein-based nanostructures (Young 2017). Typically, modules bearing this domain will form pseudo hexagon-shaped building blocks which can then tessellate to a variety of distinct higher-order nanostructures (**Fig. 5.4a**) (Young 2017). Recent publications have begun to demonstrate the value of these building blocks for both *in vivo* and *in vitro* applications, yet finer details concerning the *in situ* molecular arrangements and dynamics necessary for rational design and predictable assembly are currently lacking (Young 2017; Lee 2017; Schmidt 2018). It is our proposal to utilize the high-resolution information afforded by neutron instrumentation to aid in reverse engineering nanostructures built from PF00936-containing proteins and couple this data to computational approaches to further refine our picture of nanoscale structure and behavior.

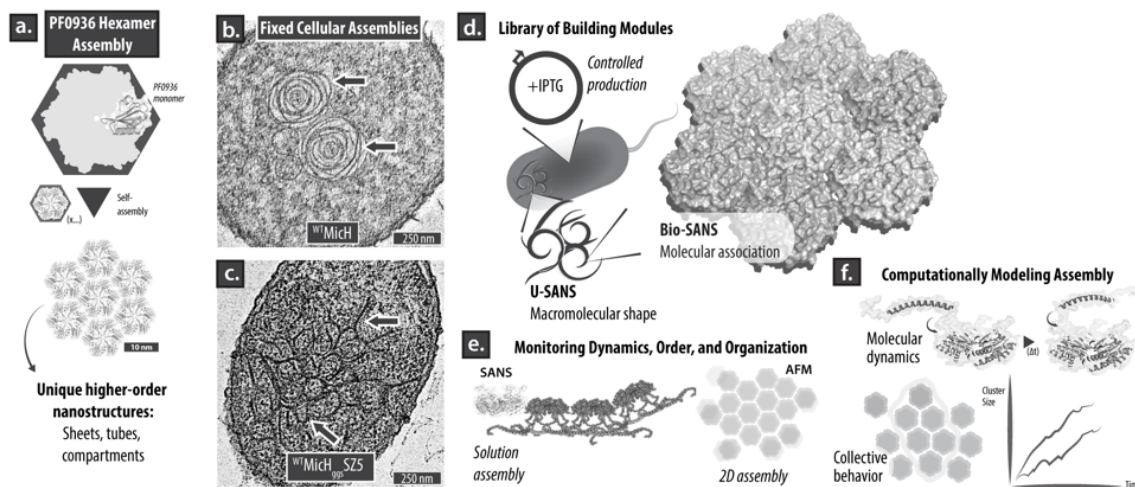


Figure 5.4: Applying neutron scattering and computational simulation in nanoscale self-assembly. A.) The protein-domain building material properties **B-C.)** *In vivo* higher-order architectures produced **D.)** Using U-SANS and BIO-SANS to uncover macromolecular shape and association **E.)** Comparing SANS and AFM dynamics **F.)** Computational avenues to explore dynamics

We currently understand PF00936 nanostructures to arise through a self-assembled process where the shape of an architecture visualized can be influenced by the module selected. It is hypothesized that the differences in primary structure between modules (e.g. substitution of a single amino acid at the protein-protein interface) can alter the landscape of macromolecular association leading to a change in the overall shape of an assembly (Young 2017). This plasticity is exemplified in Fig. 1b-c where two PF00936-containing modules were overexpressed, fixed, and imaged with transmission electron microscopy thin section. Routinely used, electron microscopy has been a powerful tool in observing PF00936 nanostructures, but technical limitations prevent details beyond assembly appearance. SANS instrumentation is one approach uniquely situated to provide the increased resolution and other details of macromolecular assembly in PF00936-built nanostructures that has been eluded by other approaches.

We propose profiling the scattering patterns of nanostructure building modules first *in vitro* then applying this information to test pilot *in vivo* conditions. Modules would be isolated via established methods under standard or deuterated production conditions to generate an *in vitro* library for downstream “plug and play” assessment. Bio-SANS application can be anticipated to reveal the differences in the molecular shape and quaternary associations, while U-SANS would be used to decipher mesoscale shape and structuring (**Fig. 5.4d**). After investigating solutions comprised of a single type of building module, we will create mixtures between modules with the same identity, but selectively deuterated, or entirely different modules to observe modifications to structure and assembly that might occur (**Fig. 5.4e**). Investigating modules with SANS would then move within a cellular system capable of

selectively triggering assembly to begin to describe hierarchical assembly *in situ*. While information is being collected with SANS instrumentation, we will also examine isolated modules using HS-AFM to view assembly confined on a 2D substrate. Experimental information, such as 2D/3D associations and component distributions, would then be applied to benchmark computational simulations under development (**Fig. 5.4f**).

Moving from a focus on assembly characteristics, we also propose to investigate the molecular motion of extensions appended to building modules (**Fig. 5.5a**). Ongoing atomistic molecular dynamics depict extensions, designed for downstream molecular scaffolding applications, to differ in motion based on the design simulated (**Fig. 5.5b**). Our hope is to use NSE to compare how manipulating factors (e.g. linker rigidity, extension composition/number) influences nanoscale fluctuations of isolated building modules. Samples to be tested include building modules lacking extensions, building modules anticipated to be saturated with extensions, and mixtures between (**Fig. 5.5c**). Trajectories derived from atomistic molecular dynamics of models with known composition will be used to refine models to include any potential changes that may occur to building modules (e.g. extension degradation) upon isolation (**Fig. 5.5d**).

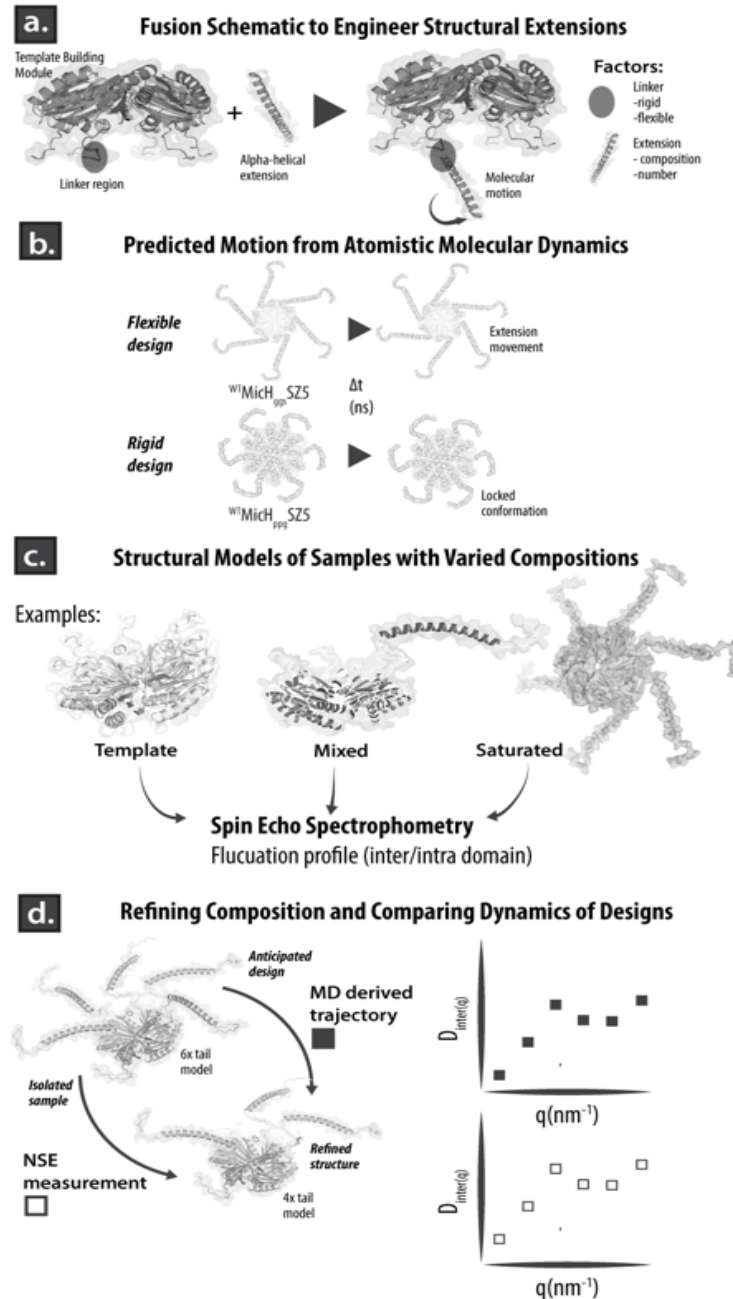


Figure 5.5: Detailing cross-application of neutron spin echo and molecular dynamic simulation to investigate inter-domain dynamics. A.) Fusion schematic in engineering functional extension B.) Using MD to sample potential dynamics C.) Investigating domain motion differences with NSE D.) Comparing MD and NSE trajectory to elucidate extension composition

This proposal seeks to describe modules containing the PF00936 protein domain and their formation of higher-order nanostructures by integrating information gathered by SANS and NSE methods in conjunction with complementary computational approaches. Successful application of this proposal also works to establish a background for exploring additional external parameters such as changes in temperature, pressure, and macromolecular crowders, in addition studying nanoscale assembly within a cellular system. It is our hope that knowledge gained not only fosters our capabilities in predictively fabricating nanostructures built from PF00936 modules, but works to create methodology adaptable to other investigations of biological nanoscale assembly.

Preliminary Work

All samples proposed to be studied have been characterized both *in vitro* and *in vivo* at the PIs home institution or at ORNL including development of the methodology for visualizing samples with HS-AFM as part of previous ORNL CNMS User Proposal. In regard to computational approaches, ORNL CNMS collaborator Miguel Fuentes-Cabrera, also in association with OSU faculty Xiaolin Cheng, has developed the pipeline for atomistic molecular dynamics adaptable in introducing newly designed structural modules, while code has been developed and implemented in CADES for 2D assembly.

Choice of Instrument

Bio-SANS beam time is requested as a means to characterize nanoscale shape and structuring, while U-SANS data is anticipated to complement this information by describing mesoscale assembly. HS-AFM is requested as a comparative instrument for describing 2D

dynamics and assembly. We also request NSE instrumentation and allocation of high-performance computing resources to compare any unique molecular motions that may accompany certain designs.

Experiment Plan

E. coli—a routinely used host for studying PF00936 proteins—will be the chassis organism for controlled protein production. For the SANS measurements, we will initially measure a contrast series (0, 20, 40, 60, 80, 100% D₂O) to determine the contrast match point of the deuterated protein. We will then study how the different components in the assembly interact by performing measurements at the contrast match point of the deuterated protein (as determined above) and also at the match point for protonated protein (42% D₂O). We envision that these measurements will require 4 days beam time. NSE measurements will be applied over Q-ranges in which SANS reveals locally ordered features. Detected motions will be compared between samples and MD models. We apply for 12 days time for NSE related experiments.

Safety Considerations

We will follow all established safety measures necessary for protein production within *E. coli*.

Example public-orientated communication piece—

Just like biomolecules elegantly coming together to produce machines and structures at the nanoscale, it will take scientists working closely together across disciplines to form the nanoscale systems of tomorrow. Still, fully realizing the potential of mastery at the nanoscale requires cooperation amongst not just the scientists, but of non-scientists alike. Funding in the science world often relies on a substantial contribution from publicly-driven sources, so keeping people inspired from science, not disconnected and afraid of it, plays an important role moving forward. Communicating cutting-edge science with engaging and accessible language can help in this regard

<https://prl.natsci.msu.edu/news-events/news/perspectives-on-building-nanofactories-for-energy-and-medical-uses/>

(Describes Chapter 2 material—Co-authors: Igor Houwat and Danny Ducat)

LITERATURE CITED

LITERATURE CITED

- Gahlmann, A. & Moerner, W. E. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat. Rev. Microbiol.* 12, 9–22 (2014).
- Gray, R. D. M. *et al.* VirusMapper: open-source nanoscale mapping of viral architecture through super-resolution microscopy. *Nature Publishing Group* 6, 29132 (2016).
- Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Current Opinion in Biotechnology* 22, 485–491 (2011).
- Lassila, J. K., Bernstein, S. L., Kinney, J. N., Axen, S. D., & Kerfeld, C. A. Assembly of Robust Bacterial Microcompartment Shells Using Building Blocks from an Organelle of Unknown Function. *Journal of Molecular Biology*, 426(11), 2217–2228. (2014).
- Lee, M. J., Mantell, J., Hodgson, L., Alibhai, D., Fletcher, J. M., Brown, I. R., *et al.* Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nature Chemical Biology*, 181, 5967. (2017).
- Luo, Q., Hou, C., Bai, Y., Wang, R., & Liu, J. Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chemical Reviews*, 116(22), 13571–13632.(2016).
- MacCready, J. S. *et al.* Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* 7, 850 (2018).
- Mahalik, J. P., Brown, K. A., Cheng, X., & Fuentes-Cabrera, M. Theoretical Study of the Initial Stages of Self-Assembly of a Carboxysome's Facet. *ACS Nano*. (2016).
- Papapostolou, D. & Howorka, S. Engineering and exploiting protein assemblies in synthetic biology. *Mol. BioSyst.* 5, 723–11 (2009).
- Schmidt-Dannert, S., Zhang, G., Johnston, T., Quin, M. B., & Schmidt-Dannert, C. Building a toolbox of protein scaffolds for future immobilization of biocatalysts. *Applied Microbiology and Biotechnology* (2018).
- Schrad, J. R., Young, E. J., Abrahão, J. S., Cortines, J. R. & Parent, K. N. Microscopic Characterization of the Brazilian Giant Samba Virus. *Viruses* 9, (2017).
- Sutter, M., Greber, B., Aussignargues, C., & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science*, 356(6344), 1293–1297. (2017).
- Sutter, M., Faulkner, M., Aussignargues, C., Paasch, B. C., Barrett, S., Kerfeld, C. A., & Liu, L.-N. Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Letters*, 16(3), 1590–1595. (2016).

- Turkowsky, B., Virant, D. & Endesfelder, U. From single molecules to life: microscopy at the nanoscale. *Anal Bioanal Chem* 408, 6885–6911 (2016).
- Weiss, T. L., Young, E. J. & Ducat, D. C. A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. *Metab. Eng.* 44, 236–245 (2017).
- Wheeldon, I. *et al.* Substrate channelling as an approach to cascade reactions. *Nat Chem* 8, 299–309 (2016).
- Young, E. J., Burton, R., Mahalik, J. P., Sumpter, B. G., Fuentes-Cabrera, M., Kerfeld, C. A., & Ducat, D. C. Engineering the Bacterial Microcompartment Domain for Molecular Scaffolding Applications. *Frontiers in Microbiology*, 8, 527–9. (2017).