DISSERTATION

# TOWARDS MACROMOLECULAR SCAFFOLD ASSISTED CRYSTALLOGRAPHY

Submitted by

Thaddaus Huber

Department of Chemical and Biological Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Christopher Snow

Christopher Ackerson
Charles Henry
Nick Fisk

ABSTRACT

TOWARDS MACROMOLECULAR SCAFFOLD ASISSTED CRYSTALLOGRAPHY

The current, dominant method for structure determination in atomic detail is X-ray crystallography; but, this method requires a brute force search through non-physiological solution conditions looking for the “needle-in-a-haystack” condition in which the target protein crystallizes. Unfortunately, despite exhaustive screening, most proteins of interest do not form crystals. Other proteins are difficult to obtain in sufficient quantities to make the attempt. Finally, even successful crystals reveal a structure adopted under artificial conditions-a single snapshot that dramatically underrepresents the protein mobility.

The motivational insight for this work is the recognition that materials diffract X-rays if they consist of a highly-ordered, repeating lattice, but that the lattice need not be composed *only* of target protein. Instead of growing conventional protein crystals, we will take the unprecedented step of attaching target proteins (guests) to specific sites within pre-existing, crystalline scaffolds for a new technique called ***scaffold assisted crystallography***. This approach circumvents the haphazard nucleation and growth process that underlies conventional crystallography. Instead, we face novel challenges. We must engineer scaffolds that have very large pores (>10 nm), withstand significant solution condition changes, yet still diffract to high resolution. We must also ensure that guest proteins tightly tethered to the crystalline scaffold adopt a coherent structure visible via X-ray diffraction.

Instead of taking on the challenge of *de novo* design of porous protein crystals, we decided to search the protein databank for a suitable scaffold. Algorithms for identifying

highly porous protein crystals are covered (Chapter 1) and a select few representative examples are presented. Constructs for high priority candidates were obtained and crystallization of the targets were attempted. One of the candidates crystallized rapidly and presented a platform for developing methods and identifying roadblocks for second generation scaffolds. Working extensively with a single scaffold, a putative periplasmic protein from *Campylobacter jejuni* (CJ), allowed for robust method development that enabled highly optimized expression and extensive knowledge of its crystallization space.

CJ requires high salt for crystallization. Crystals quickly degrade outside of the growth conditions. Most guest macromolecules will have low solubility in the high salt required to preserve the CJ crystalline lattice. Therefore, methods for chemical crosslinking of CJ crystals were developed to withstand significant solution condition changes, yet still diffract to high resolution. The most ubiquitous crosslinking agent glutaraldehyde effectively stabilized the crystal, but resulted in a dramatic loss of diffraction. Three alternative crosslinkers, formaldehyde, glyoxal, and EDC, were tested for their ability to stabilize CJ crystals. The three alternative crosslinkers all stabilized CJ crystals in challenging conditions (no salt) with little degradation in diffraction quality. The crosslinked crystals were subjected to x-ray diffraction; the resulting electron density demonstrates the first known atomic resolution modifications from formaldehyde, glyoxal, or EDC crosslinks in a protein crystal.

In contrast to the weak, noncovalent interactions that hold together typical protein crystals, guest domains can be attached to the host scaffold using strong interactions. For maximum programmability, affinity tags for the desired assembly can be genetically encoded on the guest and scaffold monomers. We demonstrated that non-covalent,

metal-mediated capture and genetically encoded histidine tags provide a significant level of control. Loading and release of guest molecules were fine-tuned to spatially segregate multiple guest proteins. Similarly, by controlling the diffusion of crosslinking agents we engineered a crystalline shell that still diffracts well.

Scaffold assisted crystallography techniques were demonstrated with small molecule guests in CJ crystals. Guest molecules were installed via a single covalent bond to reduce the conformational freedom and achieve high occupancy structures. We used four different conjugation strategies to attach guest molecules to three different cysteine sites within pre-existing protein crystals. In all but one case, the presence of the adduct was obvious in the electron density.

The above methods led to preliminary attempts of scaffold assisted crystallography with macromolecules. Guest mini-proteins variants were obtained with solvent exposed cysteines. These were covalently attached *in vitro* to CJ with an engineered surface thiol. We attempted to crystallize the resulting CJ-mini-protein conjugates. One of the CJ-mini-protein conjugates crystallized and the structure was determined. While the presence of the guest mini-protein was obvious, the electron density past the attachment point was ambiguous. Still, this result demonstrates feasibility of fusing target proteins to engineered CJ monomers for “chaperoned crystallization”.

For targets that fail to crystallize when pre-installed, we can perform asynchronous crystallization and by attaching the guest mini-protein to a preformed CJ crystal. Techniques for *in crystallo* conjugation and quantification are developed. Finally, present strategies for realizing macromolecular scaffold assisted crystallography are presented.

# ACKNOWLEDGEMENTS

My work encompasses six tremendous years of my life. This dissertation is the result of the many people who have influenced me both personally and professionally.

I would like to thank my advisor, Dr. Christopher Snow, for giving me the opportunity. His creative approach to science and relentless optimism turned me into the scientist I am today. I thrived from his “all carrot, no stick” approach to mentoring that allowed me to work within my unorthodox style. His ideas opened my eyes to a whole new field of science and I will forever be indebted to him for that.

I would like to thank all past and present Snow lab graduate students, specifically Ann Kowalski and Luke Hartje for their support throughout my projects.

I want to especially recognize Dr. Lucas Johnson. As the first graduate students in the lab, he was the perfect counter balance to my shoot from the hip approach and ordered chaos. Without his influence on the lab, we certainly would not be where we are today. He was a tremendous positive influence on me both scientifically and personally.

I had a tremendous team of dedicated undergraduate researchers. I would have never guessed that mentoring undergraduate researchers would have been my favorite part of graduate school. Interacting with you all was truly a pleasure. I would like to recognize Jarad Yost and Karina Appel, who were with me in my early years through all the struggles and uncertainties of the project. Additionally, I would like to recognize Carolyn Keating and Eli McPherson for their dedication and hard work. These undergraduate researchers

spent thousands of hours in lab and, truthfully, I would not have made it half as far as I did without their contributions.

I would like to acknowledge all those who enabled the crystallography aspect of my project. Mark van der Woerd taught me the practical approaches to crystallography and was a treasure trove of knowledge. Jay Nix at the Advanced Light Source Beamline 4.2.2, for all the hours he spent assisting with data collection on the synchrotron. Crystal van der Zanden from the Biochemistry Department at CSU for maintaining the Rigaku. Having such an excellent x-ray diffraction home source truly enabled my project.

Finally, I would like to thank my friends and family for keeping me going through the highs and lows of graduate school. My family was extremely supportive, and I would not have been able to make it through graduate school without them. My dogs, Bee Bee and Bubba, provided endless entertainment. My roommate Maren Suny was a beacon of light through the whole experience, without her I would not be where I am today.

Finally, I would like to inform Jon Keyzer that I did not forget to mail his shirt he left at my house. I will send it to New Zealand after submission of my dissertation.

Thank you everyone.

-Thaddaus Huber

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# 1 OVERCOMING THE STRUGGLES OF TRADITIONAL CRYSTALLOGRAPHY

## 1.1 NEEDLE IN A HAYSTACK

Proteins are large complex molecules responsible for carrying out a multitude of functions in the biological world including catalysis, cellular messaging, transport and storage, binding and recognition, and providing structural supports. Advances in genomic sequencing have led to an explosion in the number of identified protein sequences.[1] However, from sequence alone, it is hard to deduce details about a specific protein. The adage of biochemistry is that the structure of proteins determines function. Thus, high resolution structures are invaluable for understanding the biological role and mechanism of proteins. Unlike sequencing, a revolution in techniques has not been achieved in the field of structural biology. This has resulted in an enormous gap between sequences identified and structures determined (**Figure 1.1**).[2–4]

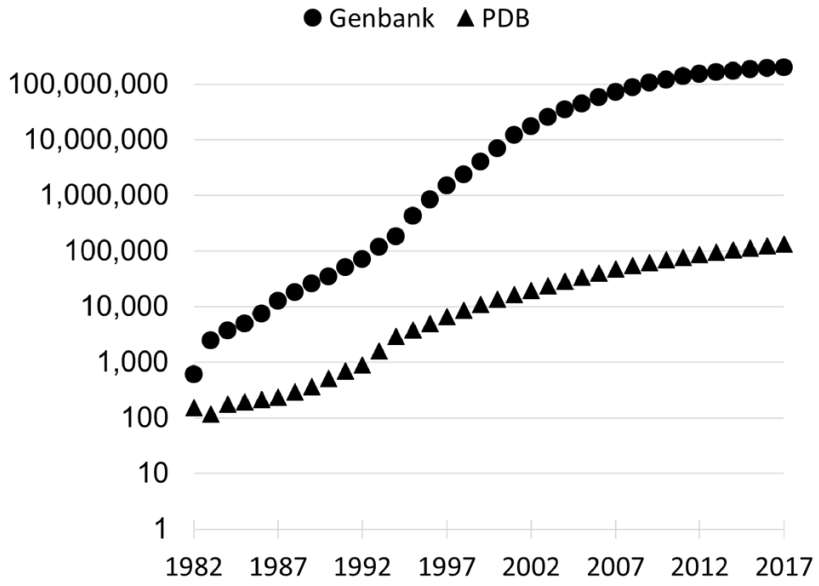



FIGURE 1.1 DEPOSITION IN GENBANK VERSUS PROTEIN DATABANK
With the advent of high-throughput sequencing, the number of genes deposited in Genbank has dramatically outpaced the number of protein structures deposited in the Protein Databank.

The current dominant method for structure determination in atomic detail is X-ray crystallography. This technique requires the target protein to be crystallized in a well-ordered lattice for collection of x-ray diffraction data. To obtain crystals, target proteins are cloned, expressed, and purified in sufficient quantity and purity for crystallization studies. Unfortunately, there is little predictive power in setting up new crystallization trials. Commonly, new targets are subjected to commercial sparse matrix screens to independently test crystallization variables.[5] Overall, the process is laborious and very time consuming with a low success rate.

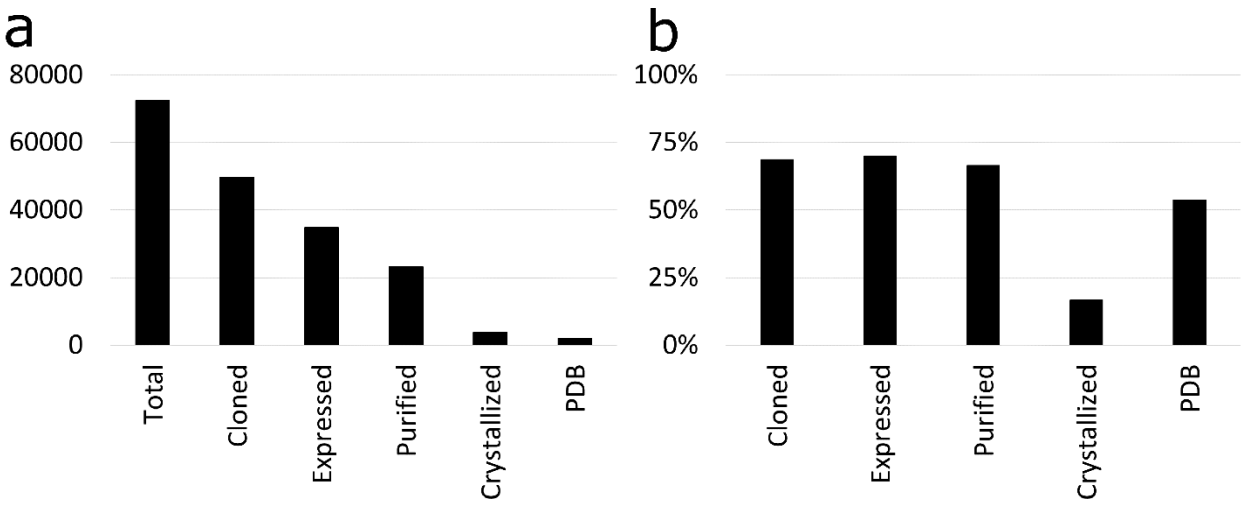



FIGURE 1.2 STATISTICS FROM THE PROTEIN STRUCTURE INITIATIVE
**(a)** Analysis of statistics from Protein Structure Initiative structural genomic centers reveals a 2.9% success rate from target to structure. **(b)** Success rate from previous step of standard structure determination work flow demonstrates significant decrease in successful crystallization trials relative to other steps.

In an attempt to develop high throughput structure determination techniques, the National Institute of Health (NIH) supported the Protein Structure Initiative (PSI), a 10-year grant program with $764 M funding.[6] The PSI demonstrated the difficulty of protein structure determination through a crystallization dependent paradigm (**Figure 1.1.a**). Despite dedicated research centers, trained specialists, and state of the art high throughput robotics; the PSI only resulted in ~2.9% target to protein databank deposition success

rate.[7] Each step of the standard structure determination work flow (cloning, expression, purification, crystallization, x-ray diffraction, and model deposition) from the PSI had attrition due to various factors; but, crystallization was by far the most troublesome with a success rate of 17% from purified protein (**Figure 1.2.b**).

The fundamental problem with crystallization is identifying the “needle in the haystack” condition in which the target protein crystallizes. This is due to the unpredictable nature of protein crystal nucleation and growth. Despite exhaustive screening, most proteins of interest do not form crystals. Even if crystals are formed, the crystals may not be of sufficient diffraction quality for data processing.

The value of atomic resolution structures and problems with x-ray crystallography has led many researchers to investigate new techniques to make protein structure determination more routine. The following sections will review some of the current approaches to salvage proteins that fail to crystallize via conventional means. An additional section covers techniques on the horizon.

## 1.2 SILVER BULLETS, MUTAGENESIS, AND CHEMICAL MODIFICATIONS

To move toward high probability crystallization screening, McPherson and coworkers have worked extensively on identifying crystallization precipitant trends and useful sets of “silver bullet” small molecules that tend to facilitate ordered protein crystals. McPherson identified crystallization precipitants that tend to result in crystals most frequently.[8] This has resulted in the formulation of Tacsimate, a blend of carboxylic acids, that Hampton Research recommends to reduce the screening of salts for crystallization.[9] Tacsimate is primarily composed of malonate — the most successful salt from McPherson’s salt precipitant analysis. From his analysis of crystallization precipitants, McPherson has

suggested new screening techniques.[10] Instead of using sparse matrix screening within the large search space of primary precipitants, researchers should focus on a limited set of high probability fundamental precipitants such as PEG or Tacsimate. The focus of crystallization screens should shift to small molecule additives that may sufficiently form intermolecular non-covalent crosslinks to promote lattice formation or biologically relevant molecules that may stabilize the target proteins conformation.[11] McPherson has investigated the propensity of 81 proteins and viruses to crystallize from a screen of 200 small molecules in tandem with PEG or Tacsimate at a fixed pH. The experiment resulted in crystallizing 65 out of 81 proteins, with 35 out of 65 crystallizing only in the presence of small molecule additives and not in the controls lacking additives. McPhersons screen of small molecules is currently sold through Hampton Research.[12] While this is useful insight, the experiment was primarily composed of proteins that had previously been demonstrated to crystallize. With a set of arbitrary proteins, the results would undoubtedly be less favorable.

Growth of highly ordered protein crystals is often impeded by flexibility of the target protein;[13] this is partly because flexible regions are unlikely to be involved in crystal lattice contacts. Flexibility of the target proteins results in heterogeneity of the target which is at odds with forming well-ordered crystals. Thus, protein sequence is undoubtedly one of the most important variables determining protein crystallization. Even small flexible purification tags appended to the termini of proteins — such as poly-histidine tags (his-tags) — can have profound effect on crystallization and diffraction quality.[14] While purification tags are not intrinsically detrimental to crystallization (there are many structures deposited with his-tags),[15] many researchers err on the side of caution and

remove all tags prior to attempting crystallization.[16] Beyond cleaving tags, some crystallization protocols systematically truncate natively flexible N and C termini. Often heterologous production of eukaryotic proteins in bacterial hosts is aided by systematic truncation of the termini.[17] Truncation of flexible termini has been demonstrated to increase crystallization propensity and to increase resolution.[18,19] Deletion or replacement of large disordered loops can induce crystallization and improve resolution.[20] Large multi-domain proteins are often crystallized as fragments if the target protein proves to be too recalcitrant to crystallization efforts. Occasionally, controlled proteolysis is performed to create “crystallizable” fragments.[21–24] This approach was revealed by the serendipitous result that some proteins crystallized as fragments after naturally degrading by proteolysis in storage.[25–27] Success in altering the sequence to limit flexibility is very much trial and error. Many results are anecdotal and most successes in altering the sequence are a means to an end as opposed to general method development.

Derewenda and coworkers have stated that the entropic cost of burying surface amino acids at a crystallographic contact may be significantly detrimental to its ability to crystallize.[28] As a result, they have developed the surface entropy reduction method (SER) for generating low entropy surface patches that may facilitate crystallization. Specifically, clusters of lysine and glutamic acid residues are targeted due to their high conformational flexibility and preference for being solvent exposed. Derewenda and coworkers have extensively demonstrated this technique using the globular domain of human RhoGDI, an easy to express protein that is recalcitrant to crystallization in the wild-type form.[29–32] The studies primarily focused on truncating flexible residues to alanine, a residue lacking conformational flexibility. Designs with multiple mutations frequently

resulted in new crystal forms, which often contained mutated residues at the crystal contacts.[29] A particularly successful case was the RhoGDI mutant E154A/E155A, which yielded data to 1.25 Å resolution.[30] In 2007, a surface energy reduction prediction server was made available online to computationally predict useful patches for mutagenesis.[33] As of 2011, 160 depositions to the PDB with ~60 novel protein depositions referenced using the SER method.[34] This includes many successful implementations of SER to increase diffraction resolution.[35–37] The majority (~90%) of the structures deposited using SER possess crystal contacts including the mutated surface patches.[34] This technique requires time intensive production of modified constructs and may decrease solubility of the target (as noted by several of SER papers).[38] While fundamentally sound in crystallization theory,[39] SER remains a specialized effort with a modest adoption rate.

Chemical modification techniques have also been explored as a rescue strategy for proteins that fail to crystallize. In contrast to mutagenesis techniques, chemical modifications do not require any additional cloning, expression, or purification steps. Chemical modification approaches only alter the surface amino acids while residues at oligomer interfaces and in the hydrophobic core remain protected. Notably, reductive methylation of lysine residues and N terminal amines has been demonstrated to be effective by structural genomic centers.[40] Reductive methylation is performed reacting the target protein with formaldehyde in the presence of dimethyl amine borane complex (DMAB), a weak reducing agent.[41] The chemical modification results in a dimethylated tertiary amine. There have been many dramatic examples of reductive methylation yielding crystals for some targets that are recalcitrant to crystallization or improving diffraction quality.[42,43] The reasons for improved crystallizability are likely complex.[44]

Methylation results in a reduced isoelectric point, which may favor crystallization depending on the biases of the crystallization screen. Additionally, the hydrophobic properties of dimethylated lysine residues may promote better lattice contacts. This phenomena is demonstrated by the several structures of methylated proteins with ordered dimethylated lysines at interfaces.[45,46] A large-scale study found that reductive methylation achieved a 7% success rate from purified protein to structure for proteins that failed with initial structural attempts.[47] While reductive methylation should be considered as a rescue technique, it is a not a routine crystallization strategy.

## 1.3 CHAPERONE ASSISTED CRYSTALLOGRAPHY

The difficulties of protein crystallography have led researchers to attempt to make structure determination efforts more routine. An approach that has gained traction with significant successes is the use of crystallization chaperones. Notably, multiple structures of G protein couple receptors (GPCRs) — which many thought were intractable crystallization targets — have been obtained via chaperone assisted crystallography.[48] The premise of chaperone assisted crystallization is that the addition of a carrier partner may provide more effective surfaces for crystallization. The chaperones may be antibody fragments, small protein binding domains, or protein fusions. Besides providing additional surface for lattice contracts, protein binding domains may reduce flexibility by stabilizing the target protein in a specific conformation. The crystallization chaperone can also provide initial phasing information for structure determination. Finally, chaperones can increase a target protein's solubility, such as membrane proteins, to enable crystallization trials.

There are many successful examples of using antibody fragments (Fab and scFv) to crystallize membrane proteins and large protein complexes.[49,50] The first structure of a non-rhodopsin G protein-couple receptor (GPCR), human β2 adrenoceptor (β2AR), was crystallized via a chaperone Fab derived from a mouse monoclonal antibody.[51,52] A historical factor limiting use of antibodies as crystallization chaperones was the high cost and low throughput. However, recent approaches using phage display libraries and recombinant bacterial expression have made generating antibody fragments as crystallization chaperones more routine.[53,54] Still, the size of Fabs (~550 amino acids and multimeric) and low-level production in bacterial systems limits applicability as a general crystallization chaperone.

To overcome problems with traditional antibody chaperone assisted crystallization, researchers have attempted to develop smaller proteins with high affinity binding. Single chain camelid “nanobodies” are a popular platform for development.[55,56] Nanobodies are small (~125 amino acids) and composed of only the variable heavy chain ($V_HH$) of conventional antibodies. Despite their limited size, researchers have engineered nanobodies for use in phage display mutagenesis to produce highly specific binding partners to target proteins.[57,58] Further, researchers have demonstrated that $V_HH$ are amenable to incorporation of SeMet, facilitating collection of anomolouos dispersion data.[59] There are several impressive structures aided by nanobody chaperone crystallography including a structure of an intrinsically disordered protein[60] and structures of $\beta_2$AR GPCR trapped in both inactive and active agonist bound conformations.[61,62]

More recently, the Designed Ankyrin Repeat Protein (DARP-in) framework has been explored for use as a binding partner to induce crystallization.[63] DARPins present a very

modular structure due to their composition of alpha helices and evolvable loops.[64,65] Ease of production and robust *in vitro* selection protocols make them an attractive starting point for evolving a crystallization chaperone. However, DARPins are small, highly repetitive, and very hydrophilic, which limits chances for useful surfaces for crystal contacts. Pluckthün and coworkers have tried to overcome this by fusing DARPins to TEM-1 β-lactamase.[66] This strategy induced multiple lattice changes depending on geometry of the fusion. It is an interesting concept because orientation of the fusion can be added as an additional screening variable. Overall, use of DARPins as a crystallization chaperone shows promise; but, it is not as widely adopted as other chaperone assisted techniques.

In the same vein as chaperone assisted crystallography, researchers have investigated use of crystallization inducible fusions/tags. The premise is that the majority of proteins for structural studies are expressed with tags to ease downstream purification.[67,68] Fusion partners, such as maltose binding protein (MBP), have been demonstrated to increase expression yields and target protein solubility.[69] However, removal of these tags can pose problems downstream such as high cost of proteases to cleave the tag, low reaction yields, and decreased solubility of target protein without tag.[16] If a fusion tag could promote crystallization, it would streamline the structure determination process. To limit flexibility of the fusion partner that might inhibit crystallization, crystallization tags are connected to the target protein via a rigid linker.[70] As of 2016, Basic Local Alignment Search Tool (BLAST) detected 102 structures deposited in the PDB as MBP fusions.[71] One of the more interesting applications of the MBP fusion strategy is a synergistic approach by applying surface entropy reduction to MBP (MBP-SER) to create a more crystallizable fusion. Pederson and coworkers demonstrated that MBP-SER produced

diffraction quality crystals for three targets that had never been crystallized before.[72] Beyond expression and purification tags, general crystallization protein fusions have been pursued. T4 lysozyme (T4L) is a well-studied model protein for crystallization. As of October 2017, there are 697 structures of T4L deposited in the PDB.[3] Due to its ubiquity, researchers have explored T4L as a fusion protein to induce crystallization. In parallel with the Fab chaperone assisted approach,[51] an engineered human $\beta_2$-adrenergic GPCR structure replaced with T4 lysozyme ($\beta_2$AR-T4L) was crystallized.[73,74] The resolution was markedly improved from 3.4 Å with the Fab chaperone to 2.4 Å with the T4L fusion. Since this initial structure of $\beta_2$AR-T4L, several structures of additional GPCRs have been determined using this approach.[48]

## 1.4 NANOTECHNOLOGY APPROACHES ON THE HORIZON

A primary goal of nanotechnology is to develop self-assembling materials with atomic accuracy. By analogy, protein crystals are serendipitously occurring highly ordered self-assembling materials. To circumvent crystallization, researchers have proposed organizing proteins on nanostructures for structure determination. Host-guest structure determination of small molecules using metal organic frameworks has had recent high-profile success. Fujita and coworkers developed the “crystalline sponge” method for host-guest crystallographic structure determination.[75] This method relies on adventitious, non-covalent interactions to adsorb and order guest molecules.[76] Recent work by Yaghi and coworkers have proved successful structure determination of various guest molecules covalently attached in a MOF.[77] However, pore diameters of MOFs are generally not large enough capture proteins. Recent large pore MOF crystals (10 nm diameter) from Yaghi

and coworkers have not been suitable for single crystal diffraction.[78] As a result, MOFs will not easily scale to structure determination of larger guests such as proteins.

DNA nanotechnology is a fairly mature method for producing self-assembling materials. The ease of design principles has resulted in truly impressive architectures.[79] Creation of DNA structures capable of specifically capturing proteins for structure determination applications has been an explicit goal of the DNA nanotechnology community since its inception.[80,81] However, no atomic level guest structures have been reported, this is possibly due to the low resolution of designed DNA crystals to date. The closest approaches have been 3-dimensional DNA crystals composed of tensegrity triangles that diffract to ~4 Å[82] and porous hexagonal DNA crystals (9 nm pore diameter) that diffract to ~5.[83–85]

While design of protein nanomaterials remains a difficult undertaking, there have been some notable achievements. Tezcan and coworkers provided multiple precedents for the proposed research by demonstrating metal-driven assembly of a variety of protein building-blocks.[86,87] Saven, Degrado, and coworkers successfully designed a hexagonal protein crystal using a coiled-coil protein building-block.[88] Yeates and Baker used computational protein design to engineer new protein oligomer cages.[89–92] Yeates and Waldo developed split GFP variants for programmed assembly.[93,94] The Turberfield, Sinclair, and Baker groups have engineered 2-D crystalline lattices.[95–97] These technologies share the implicit goal of organizing proteins and determining the structure of the resulting material. However, these studies have not provided methods capable of routinely designing functional atomic resolution materials from arbitrary building blocks. Development of protein nanomaterials for organizing arbitrary guest proteins for structure

determination has not yet been realized. One exception is the 2-D crystalline lattice technology developed by Sinclair, which has been spun out to company named Crysalin Ltd.[98] On their website, they present electron microscopy images of a guest protein structure to ~20 Å.[99] To date, this is the only public information Crysalin has released indicating successful guest structure determination. There are many interesting nanotechnology platforms, however none have yet delivered the elusive goal of routine macromolecular structure determination.

## 1.5 THESIS OVERVIEW

The solvent channels and voids within protein crystals are widely used in classical protein crystallography for diffusion of biological ligands or heavy metals (to solve the phase problem).[100–102] Heavy atom cluster soaks are particularly useful for phasing large proteins or assemblies because they result in the addition of many electrons.[103] Despite lacking specific high-affinity interactions with the protein and despite partial occupancy, it is often possible to localize the position of heavy atom cluster guests.[104] By analogy, this finding suggests that it may be possible to adsorb guest macromolecules at specific sites within host crystals with suitable pores. While some researchers have begun to explore the use of protein crystal pores to template the growth of inorganic structures,[105–107] no one has yet immobilized macromolecules within protein crystals, much less attempted to develop the control necessary to enable structure determination.

The motivational insight for this work is the recognition that materials diffract X-rays if they consist of a highly-ordered, repeating lattice, but that the ***lattice need not be composed only of target protein***. Instead of growing conventional protein crystals, I developed methods to take the unprecedented step of attaching target proteins (guests) to specific

sites within pre-existing crystalline scaffolds. The resulting materials will enable structure determination of arbitrary target proteins via ***scaffold assisted crystallography***. This approach completely circumvents the haphazard nucleation and growth process that underlies conventional crystallography. Instead, we will face novel challenges. Scaffolds must be engineered that have very large pores (>10 nm), withstand significant solution condition changes, yet still diffract to high resolution. Guest proteins must also be tightly tethered to the crystalline scaffold to adopt a coherent structure visible via X-ray diffraction. To take on this sizable challenge, the project was divided into four phases.

## ***Phase 1:***

Scaffold identification and method development, covered in **Chapter 2** and **3**. Specifically, algorithms for identifying protein scaffolds from the protein databank with pore diameters large enough for macromolecule transport are presented in **Chapter 2**. Additionally, initial structure determination method development for a particularly successful scaffold is covered in **Chapter 2**. **Chapter 3** includes detailed information on transforming a porous protein crystal into a robust biomaterial while maintaining high diffraction quality.

## ***Phase 2***:

Non-covalent macromolecular guest binding and release, covered in **Chapter 4**. Methods for loading and unloading guest fluorescent proteins into a porous protein crystal are developed. Additionally, binding and release of guest fluorescent proteins via metal mediated host-guest histidine-tag dimerization is demonstrated. Finally, control of binding and release is tuned to spatially segregate multiple guest proteins.

***Phase 3***:

Scaffold assisted crystallography with small molecule guests, covered in **Chapter 5**. Guest molecules were covalently installed at three different sites within highly porous pre-existing scaffold crystals, using four different conjugation strategies. The conjugates are subjected to x-ray diffraction for structure determination. In all but one case, the presence of the adduct was obvious in the electron density

***Phase 4:***

Scaffold assisted crystallography with macromolecular guests, covered in **Chapter 6**. Preliminary results for covalent installation of multiple mini-proteins are reported. Further, installation is quantified *in* vitro and *in crystallo*. The current state of the resulting structures is discussed. Future designs and experiments are considered.

# 2 SCAFFOLD IDENTIFICATION AND METHOD DEVELOPMENT

## 2.1 MINING POROUS PROTEIN CRYSTAL SCAFFOLDS FROM THE PDB

While there have been notable efforts to design protein crystalline materials,[88,95–97] the Protein Databank has over 130,000 deposited biological structures (October 2017).[3] Instead of taking on the difficult task of *de novo* crystal design, the diverse collection of structures deposited in the protein databank was analyzed for crystalline materials suitable for macromolecular guest capture.

As the Protein Data Bank is quite large, the search was focused on proteins that were solved via x-ray diffraction, found to express in *Escherichia coli*, possessed limited number of unresolved residues, and already possessed high resolution (<4.0 Å). Next, we developed a set of computational tools to assess the likely accessibility of each crystal to guest domains. For each crystal structure, we used custom algorithms to determine a robust lower bound for the probe diameter that can successfully penetrate a central unit cell **(Figure 2.1)**. Specifically, we built a grid representation of the solvent volume for all crystals in the PDB, then constructed a graph representation in which the grid points became nodes, and edges appeared between grid points that permit the transit of a guest domain. With escalating probe radius, we computed the set of grid points that are reachable from the exterior of a 3x3x3 block of unit cells, thereby obtaining the maximum guest diameter for each crystal. Finally, a list was constructed of candidates with a pore diameter >10 nm to represent a class of highly porous protein crystals.

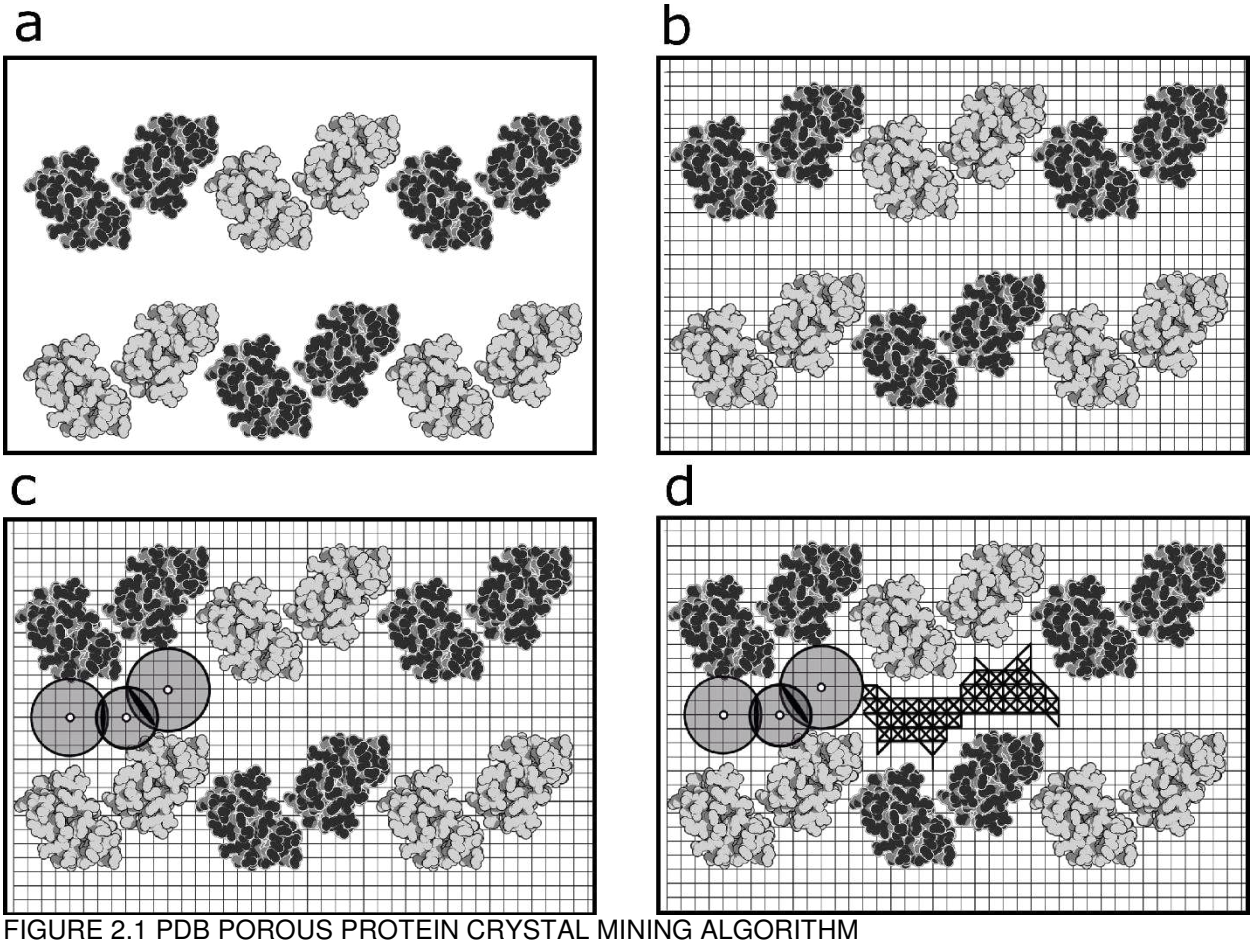


FIGURE 2.1 PDB POROUS PROTEIN CRYSTAL MINING ALGORITHM
**(a)** A 3x3x3 protocrystal is generated for each entry in the protein databank from deposited crystallographic information. **(b)** We superpose a 3x3x3 block of unit cells upon a Cartesian grid. **(c)** For each grid point, we find the minimum distance (radius) to the protein, and for each grid point pair we find a sphere intersection diameter (black ellipses). **(d)** Dijkstra's algorithm is used to determine which grid points in the central unit cell are reachable from the exterior, given a graph in which edges (heavy lines) connect grid point pairs with a sufficiently large intersection diameter.

Even with the restrictions imposed by the algorithm, there are >100 proteins that crystallize with pore diameters larger than 10 nm. With all considerations equal, candidates that could be rapidly obtained and crystallized were fast-tracked to identify challenges that could inform second generation scaffold decisions. One candidate, a putative periplasmic protein from *Campylobacter jejuni* (CJ) (PDB accession code 2fgs), was obtained through the Protein Structure Initiative Material Repository (PSI-MR).[108] CJ quickly became a top scaffold for rapid development due to its ease of crystallization.

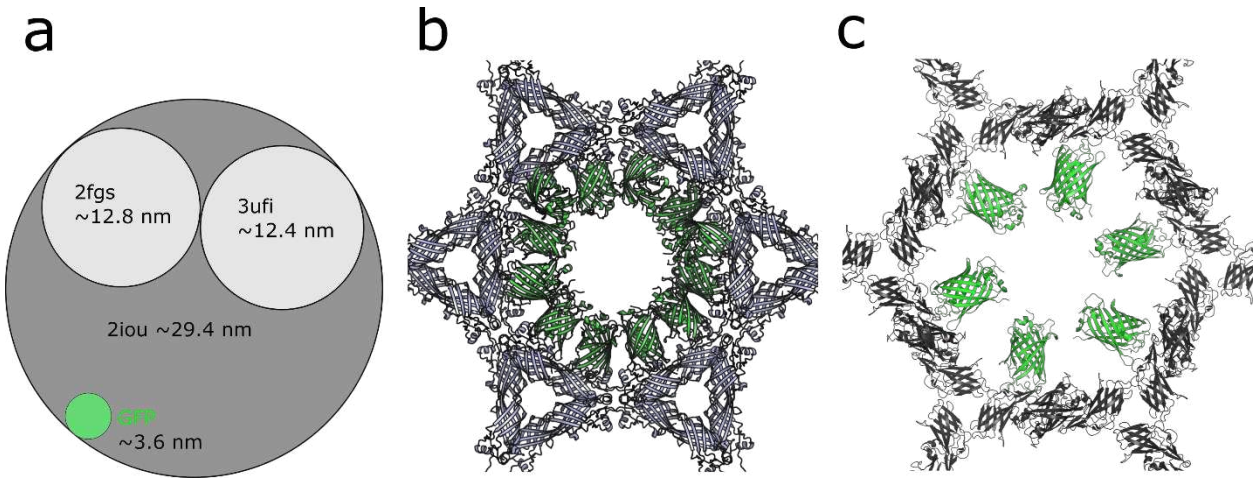



FIGURE 2.2 REPRESENTATIVE POROUS PROTEIN CRYSTALS
**(a)** Approximate minimum diameter in nanometers for 3 porous crystal examples (2fgs, 3ufi, 2iou) and green fluorescent protein (GFP). While 2fgs and 3ufi are both hexagonal space groups with similar diameters, to achieve saturation of binding **(b)** 2fgs P622 symmetty requires 12 copies of GFP while **(c)** 3ufi $P6_3$ symmetry only requires 6 copies of GFP

It is useful to keep other candidates in mind as future roadblocks are encountered. For example, a cellular adhesion subunit protein from *Bacteroides ovatus* was also obtained via the PSI-MRB (PDB accession code 3ufi). This alternate scaffold is very similar to CJ in that it is hexagonal, crystallizes in the presence of high ammonium sulfate concentrations, and has comparably large pores (~12.6 nm). However, the deposited resolution is improved (2.2 Å), and the superior symmetry ($P6_3$) results in half as many scaffold monomers per unit cell, making it more possible to saturate the symmetry guest attachment sites (**Figure 2.2**b & c). One drawback of 3ufi: the reported structure was obtained after proteolysis and reductive methylation. Since obtaining the plasmid, 3ufi has yet to crystallize in our lab. Another standout candidate is 2iou (Figure 2.2a). With a >29 nm pore, this host crystal is the largest pore scaffold identified and may allow us to capture large guest molecules. However, the plasmid was never obtained and the system may be tricky to crystallize due to hetero-8mer formation (A6B2).[109]

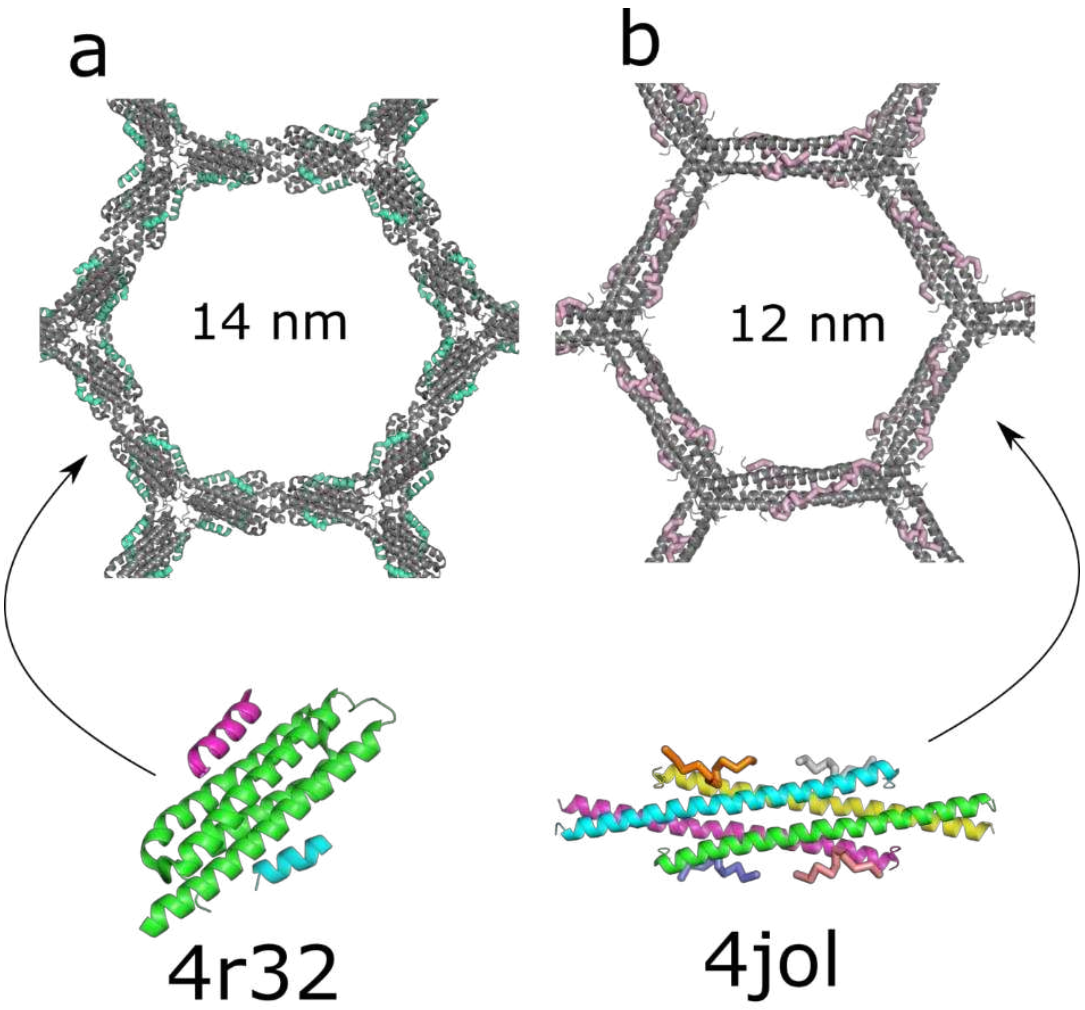



FIGURE 2.3 REPRESENTATIVE HELICAL SCAFFOLD CANDIDATES
**(a)** A view of the pore network from proline-rich tyrosine kinase 2 (Pyk2) complexed with paxillin LD motifs crystallized in a hexagonal space group ($P6_222$) with 14 nm aperture (PDB code 4r32). **(b)** A view of the pore network from AML1-ETO NHR2 domain complexed with HEB fragment crystallized in a trigonal space group ($P3_2$) with 12 nm aperture (PDB code 4jol).

Other scaffolds that may hold future interest are those with high helical content. A structure of proline-rich tyrosine kinase 2 (Pyk2) complexed with paxillin LD motifs crystallizes in a hexagonal space group ($P6_222$) with 14 nm pores (PDB code 4r32) (**Figure 2.3**a).[110] AML1-ETO NHR2 domain complexed with HEB fragment crystallizes in a trigonal space group ($P3_2$) with 12 nm pores (PDB code 4jol) (**Figure 2.3**b).[111] Both 4r32 and 4jol are ~80% helical content and crystallized in complex with a small peptide. One obvious advantage of these two scaffolds is the ability to design guest capture sites

based on robust helical design principles.[112,113] Beyond designed sites, the peptide fragments could be fused to guest proteins for capture. While it has a smaller pore diameter, 4jol may be a more interesting scaffold due to its higher resolution (3.5 Å for 4r32 vs. 2.9 Å for 4jol) and extremely mild crystallization conditions (100 mM TRIS, 20% ethanol, pH 8.5, 293K).

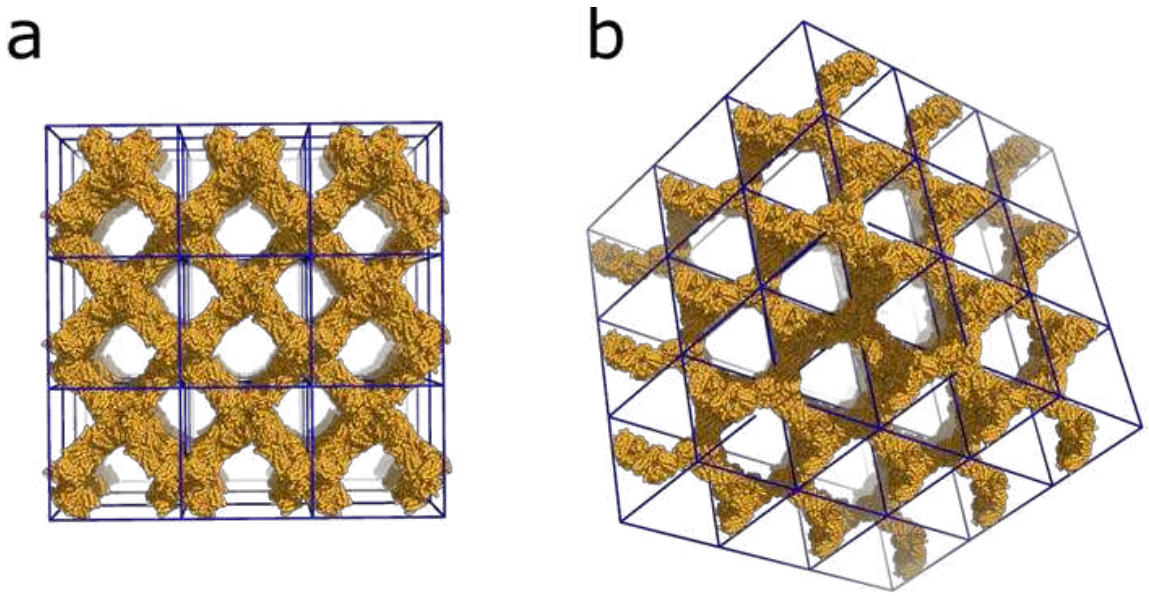



FIGURE 2.4 REPRESENTATIVE CUBIC SCAFFOLD CANDIDATE
A view of the pore network from artocarpin-mannopentose (PDB code 1vbp) crystallized in the cubic space group $P4_132$. **(a)** View from the cubic face. **(b)** View from the cube "corner".

Most identified candidate scaffolds are hexagonal. Hexagonal periodic arrays optimize the amount of scaffolding material to amount of storage capacity ratio (bees figured this out long ago).[114] However, the protein databank does contain some interesting highly porous crystalline materials that are neither hexagonal nor trigonal. A crystalline lattice of the artocarpin-mannopentose (PDB code 1vbp) crystallizes as the cubic space group $P4_132$. Due to its high symmetry, 1vbp has multiple distinct apertures in regular rotations (45º and 90º), the largest of which is a 15-nm aperture (**Figure 2.4**). A crystal composed of antimicrobial peptidase lysostaphin from *Staphylococcus simulans* (PDB code 4lxc)

crystallizes as the tetragonal space group $P4_332$ with 14 nm pores. Due to their high symmetry (24 monomers per unit cell for both 1vbp and 4lxc), these space groups will likely be difficult to engineer for high-occupancy macromolecular guest structure determination. However, exploring these alternative space groups with large pores in multiple axes will result materials with drastically different properties than trigonal and hexagonal space groups. These pore networks may be exploited for interesting macromolecular anisotropic guest diffusion patterns. Additionally, non-hexagonal, non-trigonal scaffolds may present interesting pore networks for biotemplating of inorganic nanostructures.

## 2.2 CJ CRYSTALLIZATION PIPELINE OPTIMIZATION

Besides model protein crystals such as lysozyme, structural biologists have had little motivation to study and improve crystal growth after initial crystal hit and sufficient diffraction quality. In contrast, years of experience with our scaffold protein has led to robust knowledge of growth habits, crystal stability across diverse solutions, and optimized workflow for new variant production.

Expression of CJ protein was frequent due to feedback from downstream experiments, constant generation of new designs, and the need for thousands of crystal plates to explore and troubleshoot CJ as a biomaterial. Preliminary expression trials found that CJ expressed best under cold temperatures (17 $^{o}$C) and lower IPTG (0.2 mM) in BL21-DE3-pLySs and TB media resulting in ~10 mg/L media yield. After this preliminary expression, an initial capture with $Ni^{+2}$-NTA followed by a polishing step size exclusion was performed to obtain initial CJ crystal hits. Chromatograms from size exclusion revealed two peaks corresponding to CJ domain swapped dimer and CJ monomer. Even with high purity

(>95%), the samples resulted in significant aggregation in the required crystallization conditions. In-well aggregation later proved to complicate macromolecular guest loading by “poisoning” the surfaces of the protein crystal thus leading to irregular and unreliable guest loading. From the crystal structure, it is observed that the dimeric form is required to form CJ crystals. Analysis of the size exclusion chromatograms revealed a shoulder corresponding to a slightly larger protein coeluting with the CJ monomer. Later, CJ protein was identified to have a periplasmic signaling peptide via analysis on the Signal P4.1 server.[115] Due to the lack of shoulder on the dimer peak, we inferred that the immature CJ protein (with signal peptide) was not favorable in the dimeric form required for crystallization. While high quality mature protein can be extracted from the periplasm, periplasmic extraction often resulted in very low yields (~2 mg mature CJ protein/L of protein).[116] The requirement of transport to the periplasm for signal peptide cleavage was identified as a bottle neck for obtaining high yield homogenous protein samples. Thus, a construct with the signal peptide deleted was cloned, expressed, and crystallized without any changes to the crystal space group. The truncated CJ increased protein production to ~100 mg/L of culture in C41-DE3, a strain optimized for expression of toxic proteins, with induction at 20 ºC, 0.4 mM IPTG, and 16 hours.

Single step IMAC affinity has resulted in sufficient purity for crystallization, albeit with aggregation due to the extremely high ammonium sulfate (AS) required for crystallization (>3 M AS). Reduction in aggregation can be achieved by ammonium sulfate precipitation prior to crystallization.[117] An addition of 50-75% saturated AS to samples results in high recovery of CJ in the soluble fraction, while less stable proteins salt out. After ammonium sulfate precipitations, the samples are buffer exchanged in an Amicon concentrating spin

column (10 kDa cutoff) and concentrated to desired crystallization concentration (5 mg/mL to 50 mg/mL). If enhanced purity is required, additional purifications have been applied such as size exclusion chromatography or mixed modal chromatography (Nuvia cPrime resin). Optimized CJ cloning, expression, and purification protocols have led to crystallization of over 20 distinct variants.

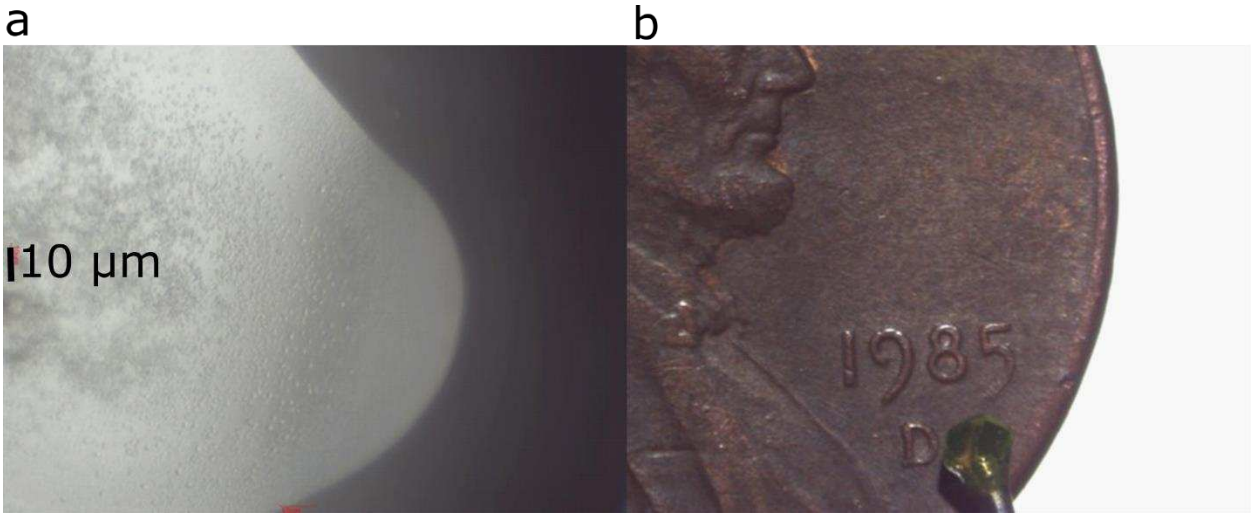



FIGURE 2.5 CONTROL OF CJ CRYSTAL SIZE
CJ crystals can be controlled to grow to **(a)** nanocrystals (<1 um) or **(b)** visible to the eye (>1 mm).

Screening of crystallization conditions has yielded extensive knowledge of CJ crystal growth habits. Both nanocrystals (<1 um) and crystals big enough to be seen by eye (>1 mm) can be generated as a function of precipitant concentration (**Figure 2.5**). The aspect ratio of the crystal (radius/height) can be tuned to yield plates, barrels, or rods, as a function of precipitant pH (**Figure 2.6**). Furthermore, new crystallization hits that change solution properties while maintaining the crystal space group (data not shown) have been identified. These include the use of precipitants such as sodium malonate (and organic acid blends such as Tacsimate) or trimethylamine N-oxide (TMAO).[11,118,119] The new crystallization conditions have led to identification of new cryoprotectants that result in obtaining structures < 2 Å on synchrotron x-ray source.

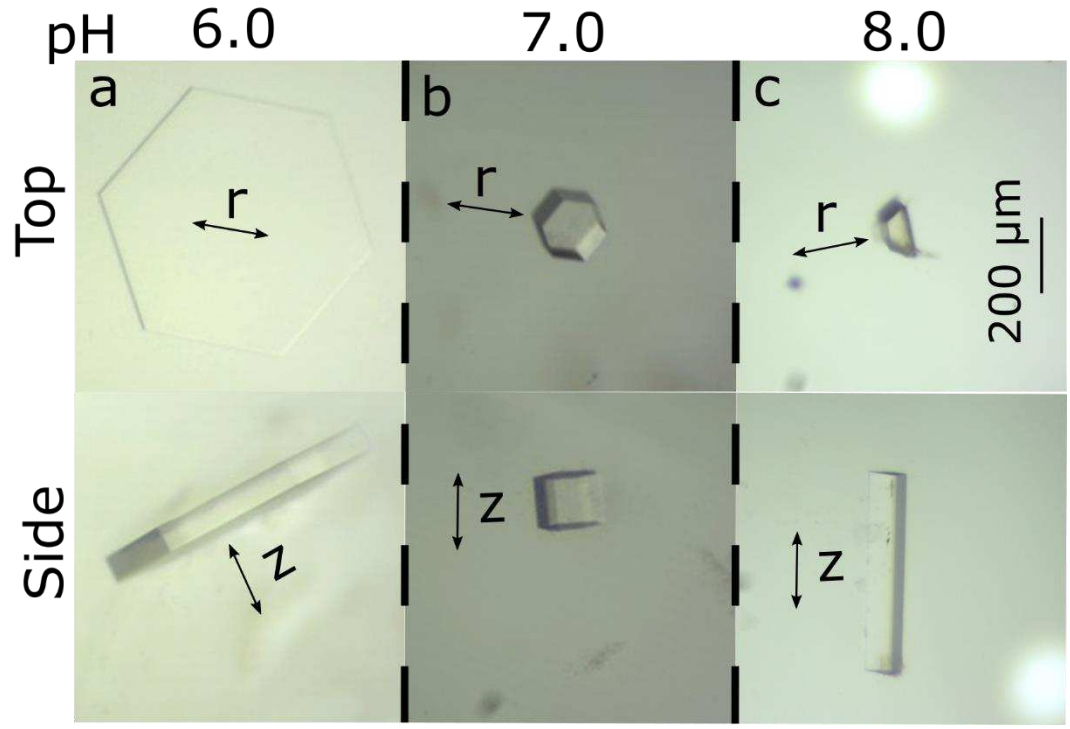



FIGURE 2.6 EXAMPLE CJ CRYSTAL GROWTH HABITS
Growing crystals is usually a haphazard process that is more of an art form than a science. Years of experience and trials has led to a robust knowledge base of CJ crystal growth habits. Morphology is strongly controlled by additives and buffers. Here the effect of pH in Tacsimate (mixture of organic acids) is demonstrated as the aspect ratio changes from **(a)** plates at pH 6.0 (radial growth is faster than axial growth), **(b)** barrels at pH 7.0 (axial and radially growth are balanced), **(c)** rods at pH 8.0 (axial growth is faster than radial growth.

Knowledge of the crystallization space has led to diversification of crystallization techniques. Crystal hits have been obtained using vapor diffusion, dialysis, and batch crystallization techniques. Notably, batch crystallization conditions in a 200 uL PCR tube have been obtained, offering the possibility of scaling up crystal production (**Figure S2.1**). Crystallization using a reusable crosslinked seed has also been demonstrated (**Figure S2.2**).[120] This offers the possibility to reliably grow extremely large crystals. Further, one could imagine cross seeding crystals with orthogonal capture sites to create materials with different capture zones. Cross seeded growth may be useful to encourage nucleation of difficult targets, such as functionalized CJ variants.

## 2.3 CONCLUSIONS

Due to the large number of structures already deposited in the protein databank, there are many interesting existing protein crystal materials. Considering that protein structures are continuously deposited, the algorithms presented here for mining the protein databank should be run routinely to identify potential scaffolds that present easy crystallization, ideal geometries, or interesting pore network topologies. Further, the method development presented here demonstrates that while protein crystals are finicky and fragile, progress in turning protein crystals into biomaterials can be made given sufficient time and materials.

## 2.4 ADDITIONAL FIGURES

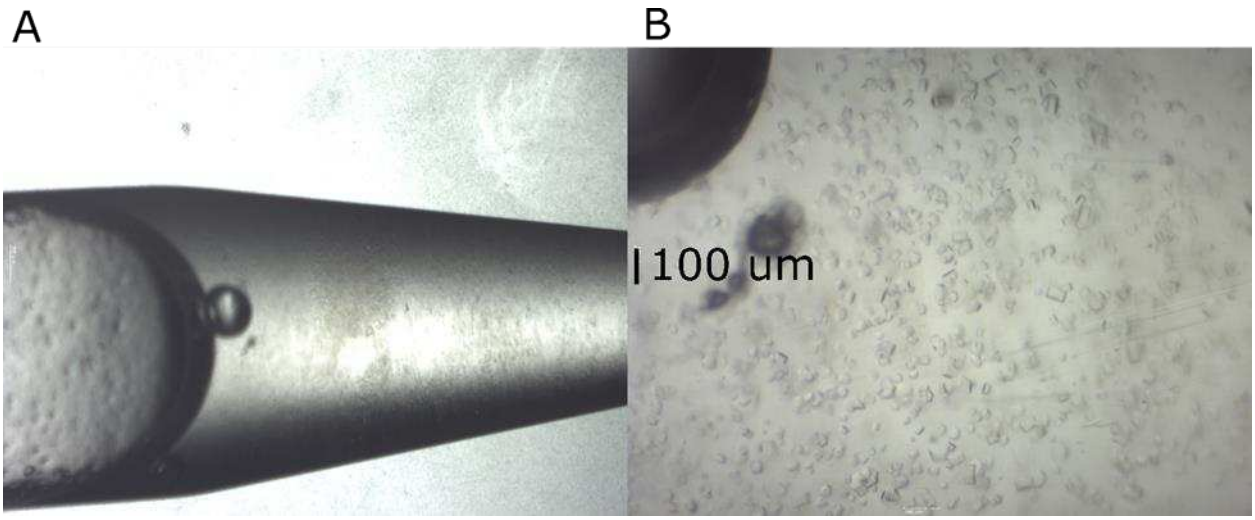



FIGURE S2.1 CJ BATCH CRYSTALLIZATION CJ IN A 200 MICROLITER TUBE

**(a)** By eye, the tube looks hazy. **(b)** Zooming in (5x objective) reveals crystal suspension. Optimization of batch growth will enable scaling up crystal production.

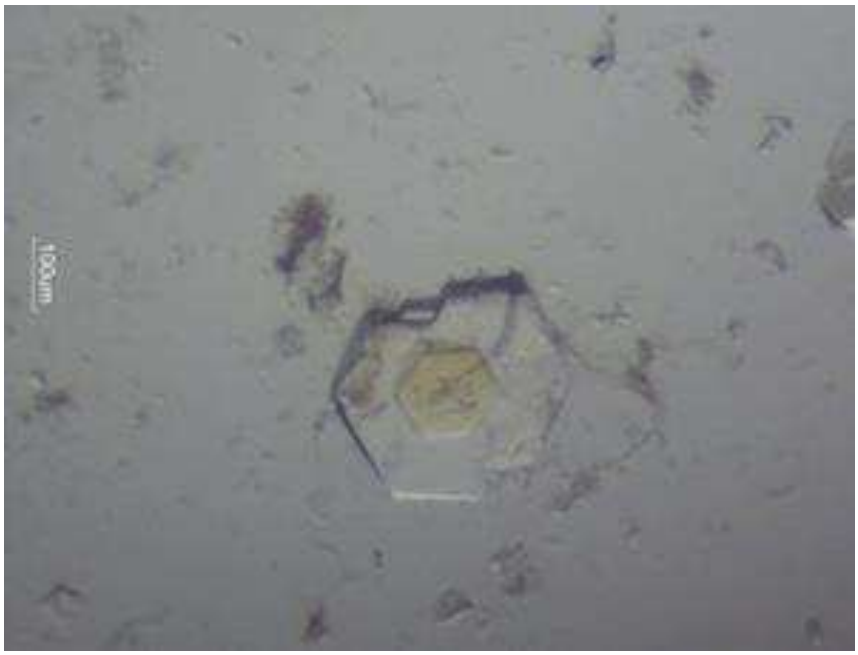



FIGURE S2.2 GROWTH OF CJ CRYSTAL ON REUSABLE CROSSLINKED SEED

A CJ crystal was crosslinked with glutaraldehyde and subsequently placed in a vapor diffusion crystal growth well, slightly undersaturated for nucleation (2.9 M ammonium sulfate, 0.1 M Bis-Tris, pH 6.5). The crosslinked seed (slight yellow) promoted growth of a CJ crystal. This technique may be used to grow larger CJ crystals.

# 3 BEYOND GLUTARALDEHYDE: ALTERNATIVE CROSSLINKERS FOR PROTEIN CRYSTAL STABILIZATION

## 3.1 INTRODUCTION

Most commonly considered solely a means for structure determination, protein crystals have recently been envisioned as nanoporous solid state biomaterials consisting of highly ordered protein assemblies.[121,122] Unfortunately, most protein crystals are incredibly fragile with respect to mechanical force or small changes in crystallization solutions. Weak adventitious intermolecular contacts hold together protein crystals. While there have been attempts to redesign protein crystal interfaces for enhanced stability,[123–125] these are specialized efforts and not routine for general structural biology efforts.

Most commonly, researchers have chemically crosslinked protein crystals in attempts to make a robust material. Glutaraldehyde is the most ubiquitous crosslinker for stabilizing protein crystals;[126] it is inexpensive and highly effective at stabilizing crystals due to its high reactivity. Crosslinking with glutaraldehyde has been used to stabilize cross linked enzyme crystals (CLEC) resulting in robust materials with increased enzyme stability.[127–129] Additionally, glutaraldehyde has been used to stabilize protein crystals for inorganic biotemplating.[106,107,130] In regards to structure determination, there are successful examples where glutaraldehyde retained or improved crystal diffraction quality.[131] Glutaraldehyde crosslinking has also been used in biochemical studies to stabilize crystals for ligand soaks.[100]

The solution properties and reaction mechanism for glutaraldehyde remains controversial. This is because the simplest crosslinking mechanism, Schiff base

formation between the dialdehydes and ε-amino groups of lysine, is labile to hydrolysis, which contrasts with the robustness of glutaraldehyde crosslinked materials (**Figure S3.1**a). An extensive review by Migneault *et al.* details the history of proposed glutaraldehyde solution composition and lists 13 proposed forms of glutaraldehyde depending on pH, concentration, temperature, etc.[132] Depending on various forms of glutaraldehyde in solution, several reaction mechanisms are proposed to be primarily responsible for crosslinking, including aldol condensation or Michael-type addition (Figure S3.1b).

While quick and effective at stabilizing the bulk crystalline form, glutaraldehyde can just as rapidly degrade the protein function or severely degrade diffraction quality. Surprisingly, there is little literature evidence of chemically stabilizing protein crystals with any of the numerous alternatives to glutaraldehyde.[121] In contrasts, practitioners of advanced mass spectrometry have used a variety of crosslinking chemistry to trap protein-protein complexes for analysis.[133,134]

Formaldehyde has extensively been used as a crosslinker in histology and immunology.[135] Given its pervasiveness, it is surprising that to our knowledge there are no literature reports of crosslinking protein crystals with formaldehyde. Formaldehyde crosslinking with proteins proceeds primarily via Schiff base formation with an amine followed by reaction with another nucleophile, resulting in a single carbon methylene bridge between the two molecules (**Figure S3.2**). Formaldehyde crosslinking reactions can result in end products containing a variety of amino acids including lysine, histidine, glutamine, asparagine, cysteine, tryptophan, arginine, tyrosine and free amines from N-terminus amino acids.[136] Under normal formaldehyde crosslinking conditions (<1% v/v),

lysine is most commonly involved in Schiff base formation with formaldehyde. This primarily due to its polar nature (typically found on protein surfaces) and relative nucleophilicity compared to other amino acids. However, the specificity of reactions between the activated immonium group and various amino acids is not apparent.

Glyoxal is the smallest of the dialdehydes with a chain length of only two carbons. One of the primary advantages of using glyoxal over glutaraldehyde is that glyoxal is primarily monomeric at low concentrations (<1 M).[137] Crosslinking with glyoxal appears to proceed through a simple Schiff base reaction mechanism, which would likely be susceptible to hydrolysis (**Figure S3.3**a). However, a weak reducing agent such as dimethyl amine borane complex (DMAB) will reduce Schiff bases to two secondary amines (Figure S3.3b). Beyond Schiff base addition, a second glyoxal can be added to make an imidazolium crosslinking product called a glyoxal lysine dimer (GOLD) (Figure S3.3c).[138,139]

Carbodiimides are zero-length crosslinkers that react primarily between carboxylic acid groups and nucleophiles. The simplest of the reactions between carbodiimide and proteins is amide bond formation between glutamic or aspartic acid and lysine (**Figure S3.4**a).[140] 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) is a water soluble carbodiimide which has seen extensive use in mass spectrometry to capture transient protein-protein interactions.[133] One of the problems with EDC is the unstable intermediate which is susceptible to hydrolysis with concomitant release of non-reactive N-unsubstituted urea.[141] The intermediate can be stabilized by the addition of a catalyst such as N-hydroxysulfosuccinimide (S-NHS) to form an S-NHS ester. The S-NHS-ester is a more stable intermediate that will react with a nucleophile (Figure S3.4b).[142]

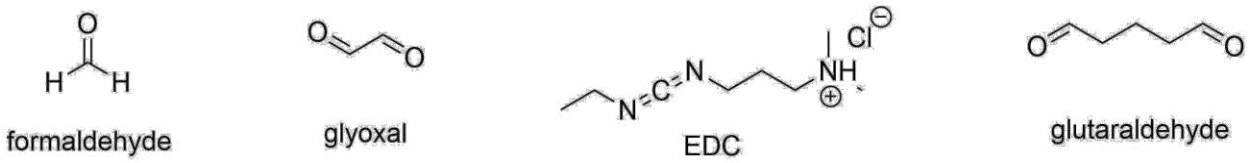



FIGURE 3.1 PANEL OF SELECT CROSSLINKERS

This study compares the crosslinking chemistry of glutaraldehyde, formaldehyde, glyoxal, and EDC for their ability to effectively stabilize a model protein crystal (**Figure 3.1**). A gentle method for crosslinking via vapor diffusion with glutaraldehyde was also tested.[143] Finally, EDC crosslinking with and without addition of S-NHS was performed to determine if addition of catalyst provided added benefit in the context of chemically stabilizing a protein crystal. The current study greatly expands upon the limited literature precedent for protein crystal crosslinking chemistry and provides a strong recommendation for future researchers to look beyond glutaraldehyde.

## 3.2 RESULTS AND DISCUSSION

### 3.2.1 CROSSLINKING MODEL SCAFFOLD SELECTION

A protein crystal composed of CJ, a putative periplasmic protein from *Campylobacter jejuni*, was selected as a model system to demonstrate the efficacy of alternative crosslinkers. CJ is an ambitious target for crosslinking because of its high solvent content (~80%) and large ~13 nm axial pores. CJ crystals require high salt for crystallization and are sensitive to small changes in conditions. They have previously been used to selectively bind and release inorganic nanoparticles or fluorescent proteins.[144,145] The prospect of nanotechnology applications such as scaffold assisted crystallography of macromolecules makes CJ an especially interesting candidate for developing general stabilization techniques. Further, CJ has lysine rich interfaces which will likely make stabilization with chemical crosslinkers effective. To fully realize CJ crystals as a

nanotechnology platform, reliable chemically crosslinking protocols is essential for producing a robust material while maintaining diffraction resolution.

TABLE 3.1 COMPARISON OF DATA QUALITY BETWEEN AS AND TMAO SOLUTIONS

| | Resolution (Å) | Mosaicity (deg.) | Unit Cell Dimensions (a,b,c) (Å) | | |
|---|---|---|---|---|---|
| **AS-6** | 2.90 ± 0.03 | 1.04 ± 0.12 | 177.93 ± 0.39 | 177.93 ± 0.39 | 50.82 ± 0.08 |
| **TMAO-8** | 2.67 ± 0.09 | 0.79 ± 0.01 | 178.98 ± 0.15 | 178.98 ± 0.15 | 50.26 ± 0.01 |

* ± reflects standard deviation from three crystals.

One of the initial concerns of crosslinking CJ was identifying a solution that would not interfere with crosslinking. CJ crystallizes in high ammonium sulfate (AS) (>3.2 M), which would likely interfere with crosslinking. Additionally, the preferred crystallization condition is at pH 6.0, which is significantly below the pKa of lysine residues (~9.5). At pH 6.0, proton transfer between the protonated and unprotonated form (nucleophilic form) of lysine will likely limit reaction kinetics for most of the crosslinkers. Trials in cryoprotecting CJ crystals identified that solutions containing >4 M trimethylamine N-oxide (TMAO) improved resolution, lowered mosaicity, and vitrified more reliably in comparison to cryoprotectants composed of AS (**Table 3.1**).[146] Crystals looped directly from mother liquor at pH 6.0 into 4.2 M TMAO and 100 mM HEPES at pH 8.0 (TMAO-8) did not result in any cracking or noticeable loss in diffraction in comparison to gradual transfer methods. The lack of reactive groups, optimal pH, and improved resolution makes TMAO-8 an ideal solution for crosslinking trials.

### 3.2.2 CROSSLINKING TIME COURSE

TABLE 3.2 DATA QUALITY IN 50% GLYCEROL AFTER EXPOSURE TO CROSSLINKERS

| | | Resolution (Å) | Mosaicity (deg.) | Unit Cell Dimensions (a,b,c) (Å) | | |
|---|---|---|---|---|---|---|
| **Glutaraldehyde** | **30 min** | >6 Å | - | - | - | - |
| | **4 hr** | >6 Å | - | - | - | - |
| | **24 hr** | >6 Å | - | - | - | - |
| **Glutaraldehyde VD** | **30 min** | Dissolved | - | - | - | - |
| | **4 hr** | >6 Å | - | - | - | - |
| | **24 hr** | >6 Å | - | - | - | - |
| **Formaldehyde** | **30 min** | 2.83 ± 0.12 | 0.77 ± 0.03 | 179.96 ± 0.14 | 179.96 ± 0.14 | 49.74 ± 0.04 |
| | **4 hr** | 2.84 ± 0.14 | 0.80 ± 0.06 | 178.98 ± 0.10 | 178.98 ± 0.10 | 49.76 ± 0.03 |
| | **24 hr** | 3.10 ± 0.11 | 0.85 ± 0.03 | 179.44 ± 0.04 | 179.44 ± 0.04 | 49.90 ± 0.02 |
| **Glyoxal** | **30 min** | 3.14 ± 0.08 | 0.78 ± 0.04 | 179.26 ± 0.17 | 179.26 ± 0.17 | 50.58 ± 0.04 |
| | **4 hr** | 3.18 ± 0.03 | 0.85 ± 0.07 | 179.52 ± 0.14 | 179.52 ± 0.14 | 50.40 ± 0.01 |
| | **24 hr** | 3.26 ± 0.05 | 0.82 ± 0.02 | 179.98 ± 0.16 | 179.98 ± 0.16 | 50.22 ± 0.04 |
| **EDC** | **30 min** | Dissolved | - | - | - | - |
| | **4 hr** | 3.02 ± 0.06 | 0.86 ± 0.06 | 179.81 ± 0.14 | 179.81 ± 0.14 | 49.96 ± 0.04 |
| | **24 hr** | 3.32 ± 0.08 | 0.86 ± 0.01 | 180.10 ± 0.11 | 180.10 ± 0.11 | 50.00 ± 0.05 |
| **EDC S-NHS** | **30 min** | 3.27 ± 0.07 | 0.98 ± 0.04 | 177.93 ± 0.15 | 177.93 ± 0.15 | 50.71 ± 0.07 |
| | **4 hr** | 3.30 ± 0.05 | 0.84 ± 0.01 | 177.43 ± 0.18 | 177.43 ± 0.18 | 50.93 ± 0.02 |
| | **24 hr** | 3.36 ± 0.01 | 0.84 ± 0.03 | 177.56 ± 0.12 | 177.56 ± 0.12 | 51.01 ± 0.03 |

* ± reflects standard deviation from three crystals.

To compare the efficacy of the crosslinkers, CJ crystals were crosslinked in TMAO-8 and transferred to a challenging condition containing no salt (50% glycerol) after varying incubation times (30 minutes, 4 hours, 24 hours). The crystals were incubated in the challenging condition for 5 minutes, which proved to be sufficient to completely dissolve non-stable crystals. Stable crystals (no change in morphology) were flash frozen in liquid nitrogen. X-ray diffraction (XRD) resolution estimates using a threshold of ($I/\sigma(I)>1$) were obtained for three different stable crosslinked crystals. Due to the highly anisotropic resolution of CJ crystals, the crystals were regularly aligned with the camera to attempt to prevent fluctuations in resolution due directionality of data collection (**Figure S3.5**). **Table 3.2** presents a summary of the crosslinking time course resolution estimates in the challenging solution.

There was significant loss in diffraction quality from exposure to glutaraldehyde, regardless of applying via direct addition or vapor diffusion. While it effectively stabilizes the crystalline material, exposure to glutaraldehyde significantly degraded diffraction quality of CJ crystals. This is reflected by the fact that the crystals diffract poorly (>6 Å) prior to transfer to the challenging solution (**Figure 3.2**). In contrast, all other crosslinkers tested did not dramatically degrade crystal quality within 24 hours. While there is a potential to find an optimal crosslinking time for glutaraldehyde, it is likely to be highly dependent on crystal size and hitting the crystal form with the "perfect dose". Considering that we already attempted vapor diffusion at a low concentration, it is unlikely that lowering the concentration of glutaraldehyde will prevent diffraction loss. Because we successfully identified several options for crosslinking CJ crystals that both stabilize the crystal and preserve diffraction quality, further analysis of glutaraldehyde crosslinking methods will not be considered.

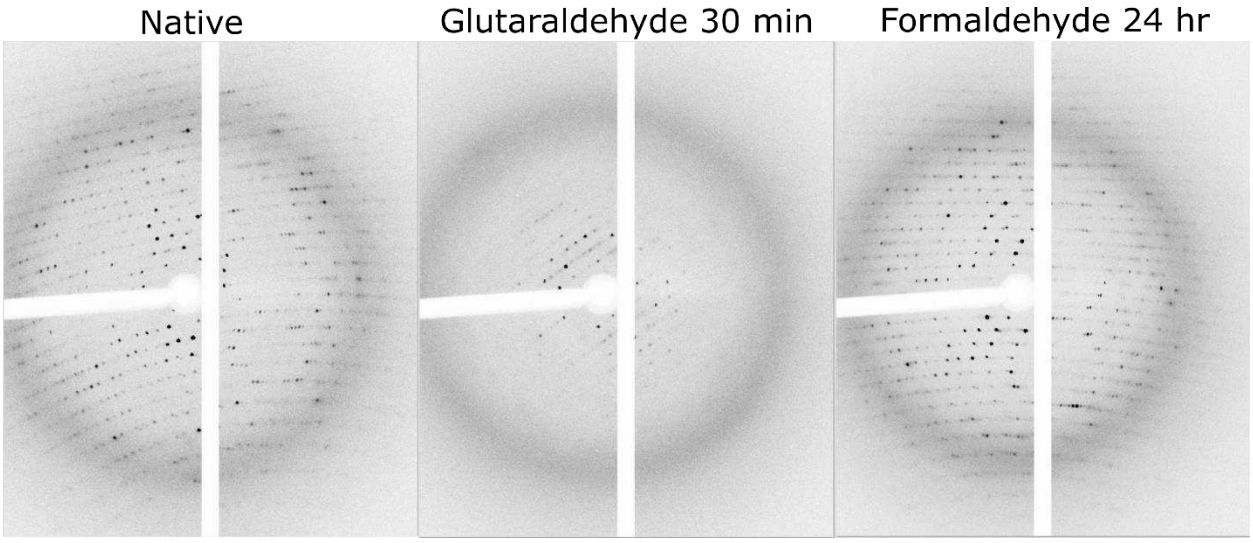



FIGURE 3.2 COMPARISON OF DIFFRACTION PATTERNS BETWEEN GLUTARALDEHYDE AND FORMALDEHYDE CROSSLINKED CRYSTALS
Even without transfer to a challenging solution, glutaraldehyde results in significant diffraction loss within 30 minutes of exposure. In contrast, formaldehyde does not result in detectable loss of diffraction within 24 hours of exposure.

Addition of S-NHS to EDC crosslinking decreased the time needed to achieve a stable crystal relative to EDC crosslinking. Interestingly, the addition of S-NHS did not improve the overall resolution at long time points. At 24 hours EDC crosslinked crystal diffracted to 3.32 Å ± 0.08 in the challenging solution, while EDC+S-NHS diffracted to 3.36 Å ± 0.01. However, there are likely different crosslinking end products between EDC and EDC+S-NHS as indicated by the noticeably different unit cell dimensions, (180.1 Å, 180.1 Å, 50.0 Å) vs. (177.56 Å, 177.56 Å, 51.01 Å) for EDC and EDC+S-NHS respectively. Possibly EDC is achieving the same crosslinks as EDC+S-NHS, albeit at lower reaction yield. Due to the instability of EDC in solution and the labile reaction intermediate, addition of S-NHS for crosslinking CJ is recommended.

We identified conditions, non-catalyzed EDC at 30 minutes and vapor diffusion glutaraldehyde at 30 minutes, under which the crosslinking dose was insufficient to stabilize the crystals, but where a longer exposure sufficiently crosslinked the crystals. We did not capture instances where diffraction substantially improved from additional exposure to crosslinker. Such an effect might be captured with an increase in the frequency of time points. Formaldehyde, glyoxal, and EDC+S-NHS all provided high-quality resolution in the challenging condition with a short incubation time: as low of a dose as 30 minutes. Long exposure with these crosslinkers did not improve diffraction quality in the challenging condition.

While it provided the best resolutions in the challenging conditions, the initial exposure to formaldehyde caused a fraction of the CJ crystals to crack upon addition of the crosslinker. EDC+S-NHS also tended to crack CJ crystals upon addition of the reaction mixture, which was not observed in the EDC reactions. Anecdotally, cracking may be

avoided by using thinner crystals, by decreasing the concentration of crosslinker/catalyst, or by performing more gentle transfers. Despite the cracking, acceptable data was consistently collected from both crosslinkers.

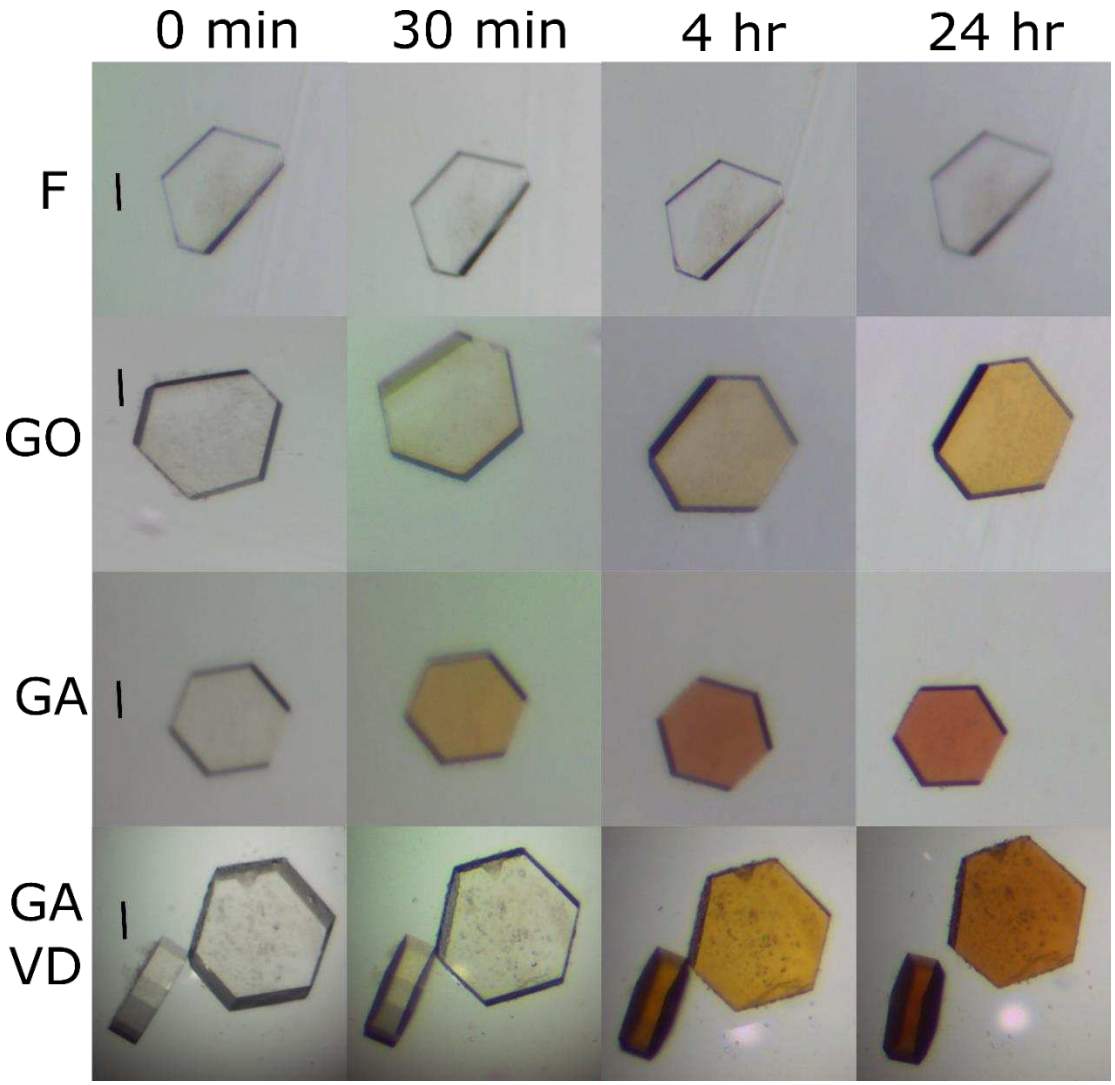



FIGURE 3.3 COLOR CHANGE OVER TIME FROM CHEMICAL CROSSLINKERS

There was a noticeable color difference after prolonged exposure to dialdehydes (**Figure 3.3**). Glyoxal turned CJ crystals a deep gold and glutaraldehyde turned the crystals a deep red over time. Unlike the dialdehydes, crosslinking with formaldehyde did not result in a color change.

Further stability analysis will be required to provide a recommendation for “optimal” crosslinking time and dose. Recommendation of an optimal time will likely be highly subjective depending on what properties are desired. Also, optimal reaction conditions are unlikely to translate between new solution conditions. Considering that many of the crosslinkers tested are very short in length, subtle changes in the conformational states of the interfaces may dramatically alter the crosslinkers ability to stabilize the crystal.

### 3.2.3 OBSERVATION OF CHEMICAL CROSSLINKING MODIFICATIONS

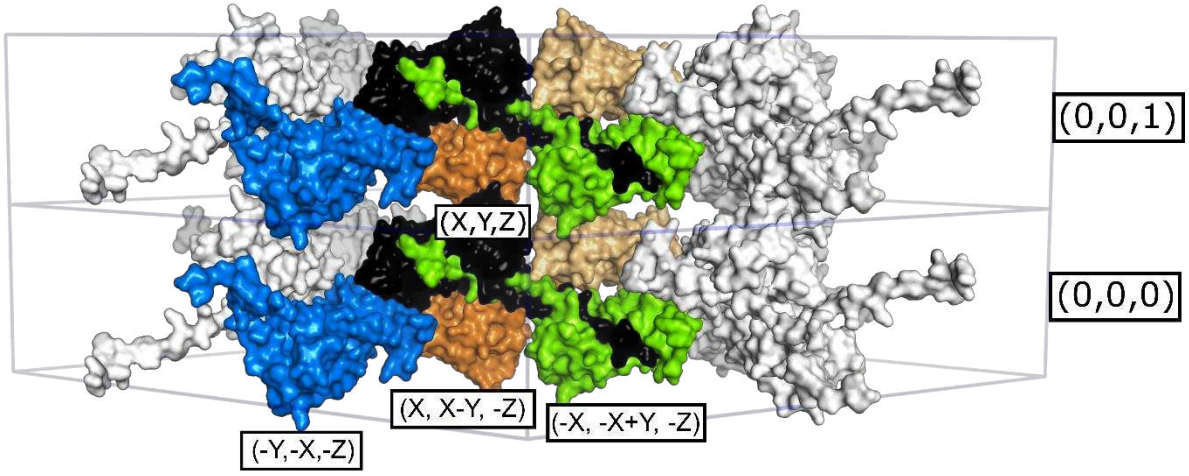



FIGURE 3.4 SYMMETRY RELATION BETWEEN CROSSLINKING SITES
Unique interfaces targeted by crosslinkers are colored and labeled with their corresponding symmetry operator relative to the black monomer (X,Y,Z).

Structure determination was performed on crystals crosslinked with formaldehyde, glyoxal, and EDC+S-NHS. To attempt to obtain high occupancy modifications, CJ crystals were exposed to crosslinkers for 24 hours. While there is an opportunity to attempt to gather time resolved modifications, that will be the subject of future work. Following crosslinking, crystals were directly flash frozen in liquid $N_2$ and XRD was performed. A summary of XRD data collection results is presented in **Table S3.1**. Notably, formaldehyde resulted in the highest resolution crosslinked structure at 2.69 Å, which was on par with the native crystal in TMAO-8 at 2.59 Å resolution. Glyoxal and EDC+S-NHS

crosslinked crystals diffracted to 3.18 Å and 3.15 Å respectively. Crosslinking with glyoxal or EDC+S-NHS resulted in a small degradation of diffraction quality in comparison to the native crystal. However, the crosslinked crystals did not suffer further diffraction loss upon transfer to the challenging solution (Table S3.1 and Table 3.2).

There were multiple sites where formaldehyde and glyoxal targeted the same amino acid pairs for crosslinking. One of these sites, the interface containing K95 and K98, stacks lysine residues across a two-fold symmetry axes perpendicular to the six-fold axis in the native crystal (**Figure 3.4**a). From the significant positive electron density observed at this site, it is obvious that both formaldehyde and glyoxal modify this interface. In the case of the formaldehyde crosslinked crystal, the electron density is slightly more interpretable. Two states of K95 are proposed, with possible methylene bridges between K95 and K95 (X, X-Y, -Z) as well as K95 and K98 (X, X-Y, -Z) (Figure 3.4b). There is also a possible non-productive intramolecular crosslink between K95 and K98. The electron density for the glyoxal crosslinked crystal across this interface is complicated (Figure 3.4c). Modeling across symmetry axes at special symmetry points (marked by black ellipses) is typically difficult, and the potential for multiple states confounds the issue further.

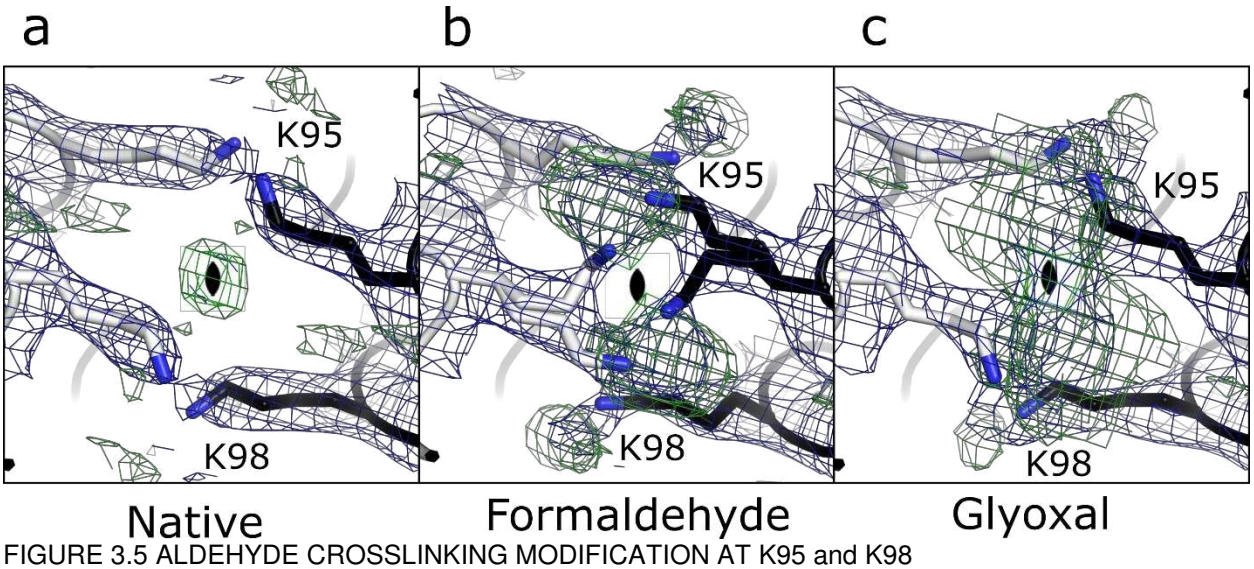



FIGURE 3.5 ALDEHYDE CROSSLINKING MODIFICATION AT K95 and K98
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native, **(b)** formaldehyde, and **(c)** glyoxal modified crystals centered at K95 and K98 two-fold symmetry interface.

There were also sites where formaldehyde and glyoxal targeted a similar amino acid, but resulted in different crosslinking pairs. Both formaldehyde and glyoxal modified K165 (**Figure 3.5**). Formaldehyde resulted in well resolved product between K165 and K97 (X, X-Y, -Z) (Figure 3.5b). Interestingly, this formaldehyde crosslink resolves a highly flexible residue from the native structure (Figure 3.5a) to form an intermolecular crosslink. In contrast, glyoxal resulted in an intramolecular crosslink between K165 and K30 (Figure 3.5c).

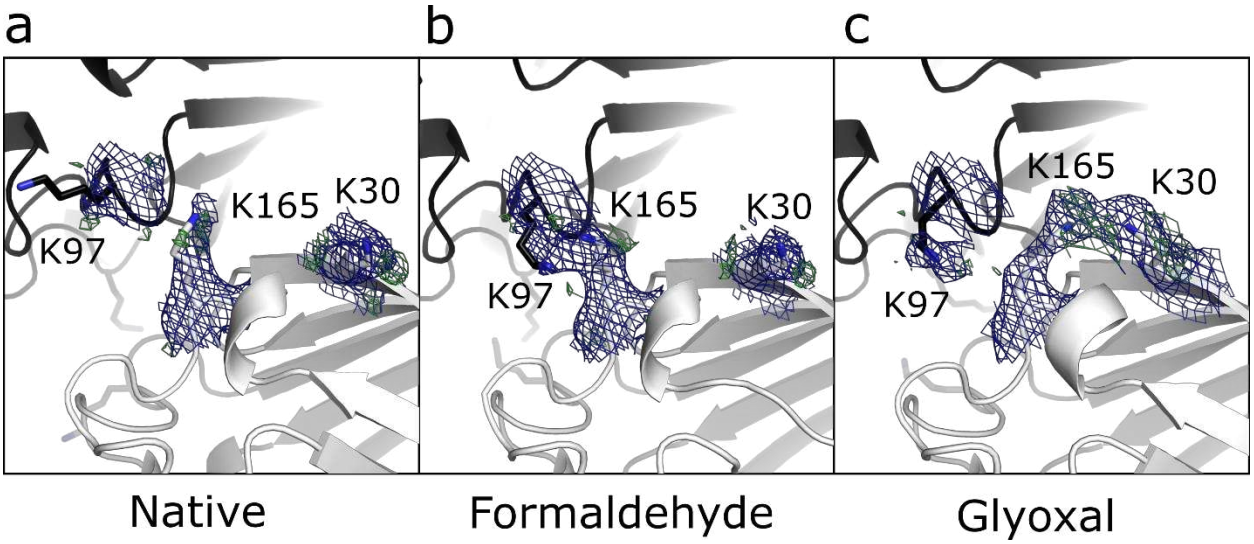



Figure 3.6 ALDEHYDE CROSSLINKING MODIFICATION AT K165
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native, **(b)** formaldehyde, and **(c)** glyoxal modified crystals centered at K158 and K160 two-fold symmetry interface.

Additionally, there were sites where formaldehyde resulted in a crosslink, but glyoxal did not. After formaldehyde crosslinking there was a potential methylene crosslink between K118 and K118 (X, X-Y, -Z) (**Figure S3.6**). Formaldehyde also formed a complex mixture of apparent crosslinks between K158 and K160 and their symmetry partners (X, X-Y, -Z) at a twofold symmetry axis (**Figure S3.7**). There were sites where glyoxal resulted in a crosslink, but formaldehyde did not. Electron density suggested a partial glyoxal crosslink between K108 and K116 (X, X-Y, -Z) (**Figure S3.8**). However, this site is complex as there is additional electron density suggested a crosslink between K118 and K116 (X, X-Y, -Z).

Due to its orthogonal chemistry, EDC+S-NHS presented a new set of crosslinks. In comparison to the aldehydes, there were not as many obvious crosslinking sites. There are many existing lysine-carboxylic acid salt bridges in the native crystal. At a typical CJ crystal resolution (~2.8 Å) and a contouring threshold of 1 σ, the electron density between such pairs is already contiguous. As a result, it is difficult to differentiate between existing

salt bridges and EDC mediated amide bond formation. However, there were two clearly resolved reaction end products. From the electron density, there appears to be amide bond formation between K22 and E189 (-X, -X+Y, -Z) that is not present in any native crystal structure (**Figure S3.9**).

EDC+S-NHS also dramatically modified a loop between residues 143 and 148. Also, contiguous electron density appears between K143 and E137 (-Y, -X, -Z) (**Figure 3.6**b), density that is not present in the native structure (Figure 3.6a). In the native structure, the ε-nitrogen K143 is ~7.4 Å from the $\delta$-carbon E137. However, with prospective modeling, distance between the two respective atoms changes to ~1.5 Å after modification with EDC+S-NHS. This crosslink results in a dramatic shift in the placement of the flexible loop (Figure 3.6c).

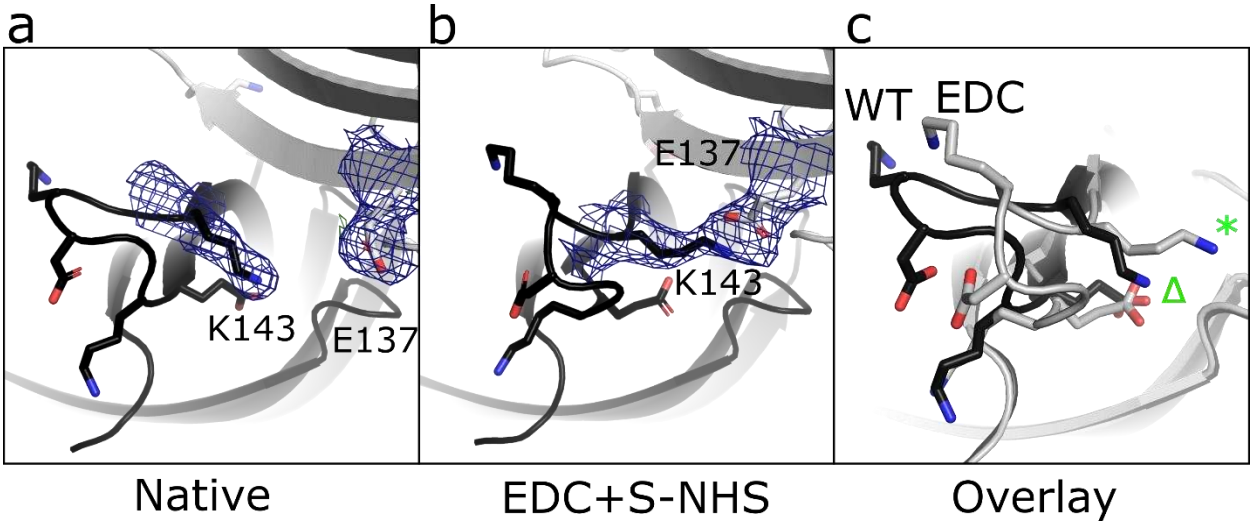



FIGURE 3.7 EDC+S-NHS CROSSLINKING MODIFICATION AT E137 AND K143
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a**) native and **(b)** EDC+S-NHS modified crystals centered at K158 and K160 two-fold symmetry interface. **(c)** The loop drastically moves as a result of chemical modification. The EDC modified K143 (green star) is presented in comparison to the native K143 (green triangle)

### 3.2.4 DISCUSSION

We have successfully demonstrated the use of alternative chemical crosslinkers to stabilize CJ crystals. When analyzing the resolution estimates, it is important to note that the modest resolution of these datasets is in keeping with the intrinsic attributes of the crystal. Specifically, the crystal is highly porous, with a solvent fraction of about 80%. Crystals with a high solvent fraction have a marked tendency to diffract more poorly than crystals with a low solvent fraction.[147–149] Additionally, data was collected on a Rigaku Homelab. Resolution estimates would undoubtedly be higher with a synchrotron light source. With that said, it is remarkable that chemical crosslinking permits high quality diffraction data to be collected in a solution that causes the native crystal to dissolve immediately. In many cases, transferring the crystal back to a high salt solution may reanneal the crystal and restore higher resolution.

Despite the modest resolution, several potential crosslinks sites have been elucidated (**Table S3.2**). While glutaraldehyde end products have been partially resolved in a prior x-ray structure,[150] to our knowledge, these current structures represent the first atomically resolved crosslinks for formaldehyde, glyoxal, and EDC in a protein crystal. From these observed crosslinks, we can make inferences about the covalent connected network topology (**Figure 3.7**). Both formaldehyde and glyoxal yielded high occupancy crosslinks across multiple 2-fold symmetry axes perpendicular to the 6-fold axis. While glyoxal and formaldehyde crosslinking resulted in slightly different end-products, the interfaces with the best resolved linkages were the same. As a result, we propose that the covalent topology for glyoxal and formaldehyde are similar.

One of the interfaces for crosslinking, incorporates crosslinks between K95, K97, K98, and K165, resulting in axial crosslinking perpendicular to the 6-fold axis <u>across unit cells</u> (e.g. (0,0,0) to (0,0,1)) (Figure 3.7a). The other interface for crosslinking, involving K106, K116, K118, K158, and K160 results in crosslinking perpendicular to the 6-fold axis <u>within a single unit cell</u> (Figure 3.7b). In combination with CJ dimer formation through domain swapping, the intra-unit cell crosslink results in tetramer formation. High occupancy combination of axial crosslinks, stabilized tetramer, and domain swapping (Figure 3.7c) is sufficient to stabilize axial threads. Six of these axial threads results in the hexagonal morphology of the honeycomb like lattice (Figure 3.7d).

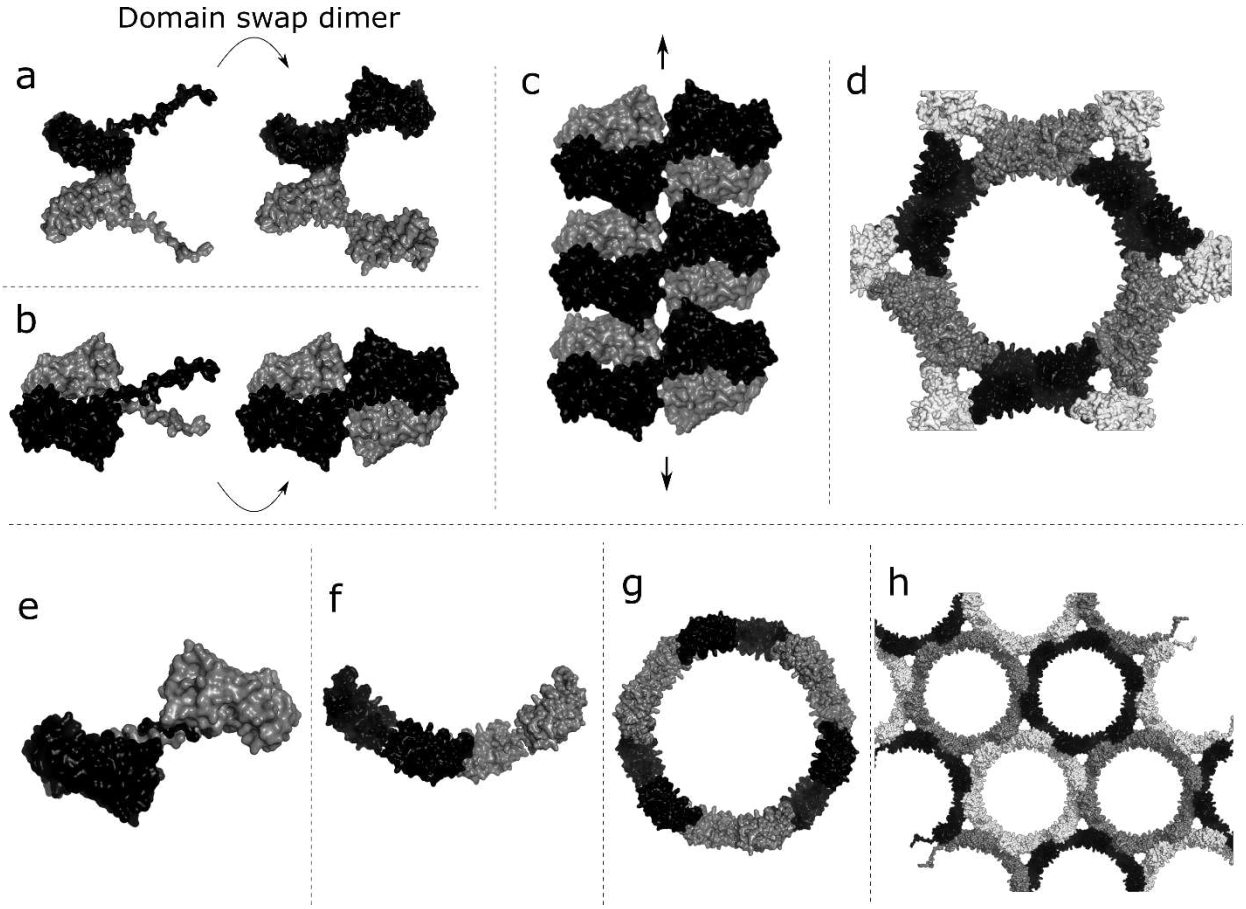



FIGURE 3.8 COVALENT PROTEIN TOPOLOGY FROM OBSERVED CROSSLINKS
Formaldehyde and glyoxal both crosslink axially **(a)** across the unit cell and **(b)** within the unit cell. In combination with the domain swapped dimer, **(c)** axial threads are created. **(d)** Six axial threads compose the honeycomb walls of the crystalline lattice. **(e)** EDC induces amide bond formation between the N and C-termini of the domain swapped dimer. **(f)** EDC crosslinks radially across unit cells. **(g)** The combination of domain swap fusion and radial crosslinks results in covalent ring formation. **(h)** The covalent rings are tiled in the honeycomb lattice.

Visible crosslinks incurred by EDC tratment would be sufficnet to stabilize a quite different topology. EDC fuses residues near the N and C-termini of the domain swapped dimer (K22 to E189). This results in a pseudo circularly permuted domain swapped dimer (Figure 3.6e). Also, an intermolecular crosslink of K143 from the flexible loop to E137 results in fusing of dimers radially across unit cells(Figure 3.6f). With the symmetry at this site, the resulting topology of fused dimers would result in covalent “ring” formation (Figure 3.6g,h).

At this time, it is not known to what extent a fully covalent network is required to withstand dramatic changes in the solution condition. Low occupancy or hard to resolve crosslinks may have dramatic influence on the overall crystal stability. For example, occasional crosslinks across the 6-fold axes for the aldehyde crosslinkers or perpendicular for EDC crosslinking would result in a fully covalent network. This is likely the case, but creative experimentation will be required to resolve these important additional crosslinks.

Currently, between formaldehyde, glyoxal, and EDC there is only a modest difference in resolution in the challenging solution. From these results, it is tempting to select formaldehyde as the optimal crosslinker because it consistently afforded the highest resolution. However, further analysis of crystalline stability and material properties will provide more information on the benefits and pitfalls of each crosslinked material. This analysis should include long term stability analysis across a wide range of challenging conditions including pH, organic solvents, temperature, etc.

The various crosslinkers will result in protein crystals with different surface material properties than the native crystal. Due to hydrolysis of reactive intermediates, EDC putatively offers a “scarless” approach. In contrast, formaldehyde is known to chemically modify many amino acids.[136] Quenching reactive groups will be important for nanotechnology applications and will allow for further modification of crystal properties. Hydroxylamine and carbohydrazide are common reagents for quenching aldehydes and NHS-esters.[151] Both molecules are routinely functionalized for bioconjugation applications and present opportunities for tuning the protein crystal properties. For example, a crosslinked crystal could be quenched with boc-hydrazide to introduce

hydrophobic groups or with hydroxylamine-O-sulfonic acid to introduce groups with a negative charge.

Quenching reactive groups will be important for preserving diffraction. When crosslinking is not quenched, we have observed a phenomena of post crosslinking diffraction loss (data not shown). We hypothesize that upon transfer to solutions different than the crosslinker solution, unquenched crosslinkers may react with newly flexible regions, resulting in creating a highly mosaic crystal. EDC will likely be the easiest to quench due its labile intermediate. Addition of a strong nucleophile, such as hydroxylamine, is likely to be effective at quenching all the crosslinkers (albeit with possible degradation of resolution). Analyzing the efficacy of different quenching methods will be the focus of future work.

## 3.3 CONCLUSIONS

This data goes beyond the literature precedent in protein crystal stabilization by chemical crosslinkers. To this end nanotechnology and structural biology now has more tools available. Optimization of crosslinking protocols is challenging primarily due to the large number of adjustable parameters. Consideration of alternative reagents for crosslinking will allow researchers to explore creative approaches to stabilizing protein crystals and develop new biomaterials. Developing methods on higher resolution crystals may lead to enhanced knowledge of bioconjugation products through new atomic resolution models.

## 3.4 MATERIALS AND METHODS

### 3.4.1 CROSSLINKING GENERAL METHODS

CJ crystals were transferred directly into 4.2 M TMAO, 100 mM HEPES, pH 8.0 and incubated for at least 5 minutes. TMAO solutions at high concentration exhibit the unusual property of dilution dependent pH. Therefore, for accurate solution preparation, a 10 mL solution was reliably prepared containing 4.66 g TMAO-dihydrate, 0.23 g HEPES, 5 mL H2O, 30 uL 37% (wt./vol.) HCl and filter sterilized. Next, 200 uL crosslinking solutions were made to contain TMAO-8 and 200 mM crosslinker. In the case of EDC+S-NHS, 100 mM S-NHS was added to the solution. All crosslinkers were taken from stocks that were opened fresh, aliquoted and immediately stored at -30 C. Approximately 5 crystals per condition were placed into the crosslinking solution. All experiments were conducted in Pyrex borosilicate glass depression well plates, and sealed with VWR glass cover slips and DOW vacuum grease. Each timepoint had a separate well allocated for various time points. After exposure to crosslinking solution for the desired amount of time, crystals were transferred to 50% glycerol in water and incubated for 5 minutes. Crystals were then quickly transferred into liquid nitrogen and cryogenically stored under liquid nitrogen until analyzed with cryogenic XRD.

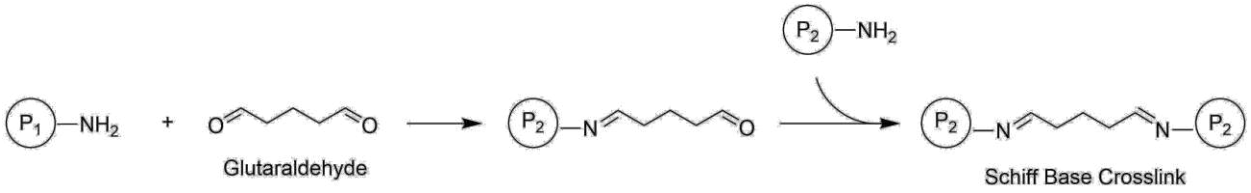
### 3.4.2 XRD TIME COURSE DATA COLLECTION

CJ crystals were prepared using standard protocols. In all cases, crystals were kept in their cryoprotectant solution (50% glycerol in water) under liquid nitrogen in a cryogenic vial prior to XRD. Loops were then serially transferred to a Rigaku HomeLab, exposing the crystal to a liquid nitrogen stream (T = 100 K) to prevent the crystal from thawing. Crystal integrity was quantified via a 10 frame (0.5° per frame, 30 s exposure)

with a fixed detector distance (70 mm, 3.05 A edge) data collection strategy using a microfocus X-ray generator and a Pilatus 200K detector. Data were integrated and scaled using the HKL3000 program suite. Resolution cutoffs were determined by the last resolution shell with I/sigma(I)>1. Full datasets for TMAO-8, formaldehyde, glyoxal, and EDC+S-NHS were index, refined, and scaled using HKL3000. Molecular replacement was performed on TMAO-8 using a high resolution native crystal structure in ammonium sulfate cryoprotectant. The TMAO-8 was used as a molecular replacement model for the crosslinked structures.

## 3.5 ADDITIONAL FIGURES

**A**




**B**

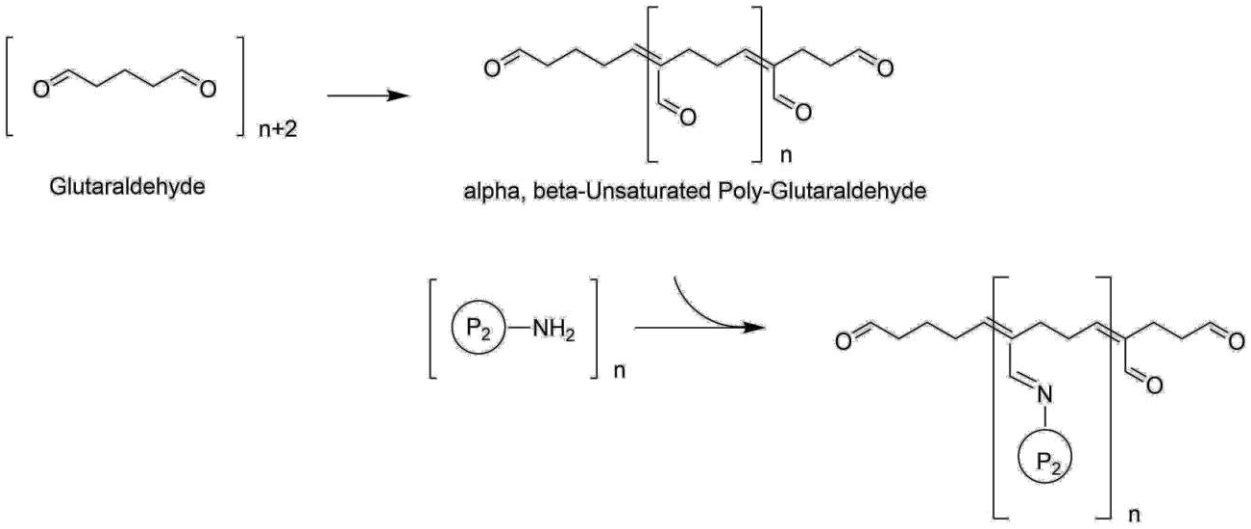



FIGURE S3.1 REACTION SCHEME FOR GLUTARALDEHYDE CROSSLINKING
**(a)** Glutaraldehyde reaction is proposed to proceed primarily via Schiff base reaction between two lysines. However, simple Schiff base formation would not explain the extraordinarily robust glutaraldehyde crosslinked materials. **(b)** Glutaraldehyde can also react with lysine residues via Michael addition to create a more permanent linkage.

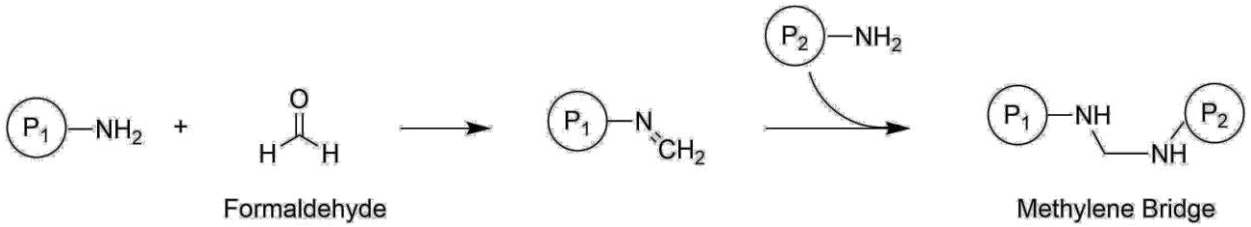
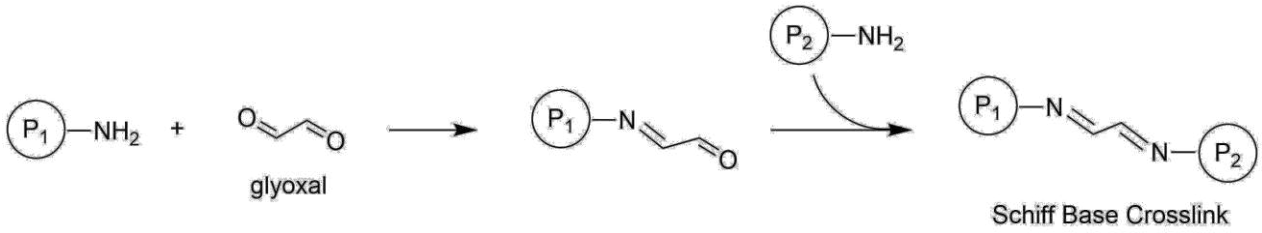
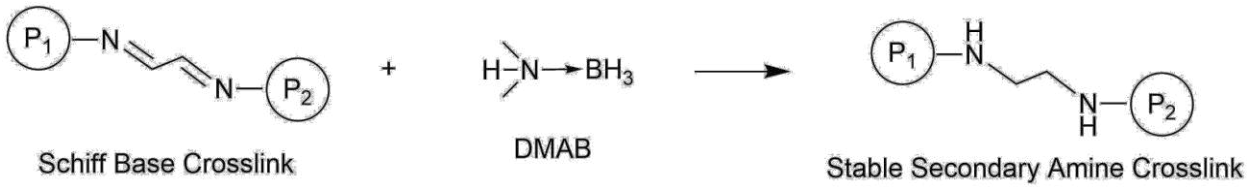



FIGURE S3.2 REACTION SCHEME FOR FORMALDEHYDE CROSSLINKING
Formaldehyde primarily reacts with proteins via Schiff base formation with primary amines followed by addition of a nucleophile.

**A**




**B**




**C**

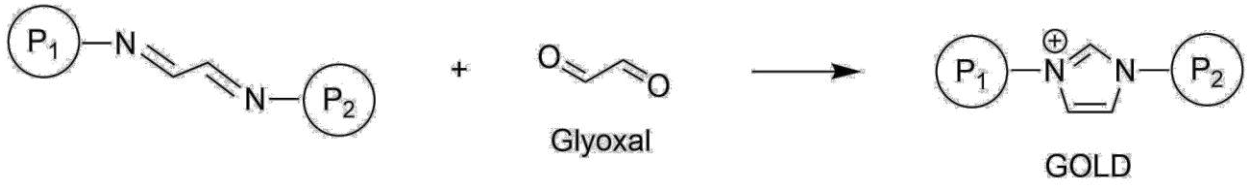
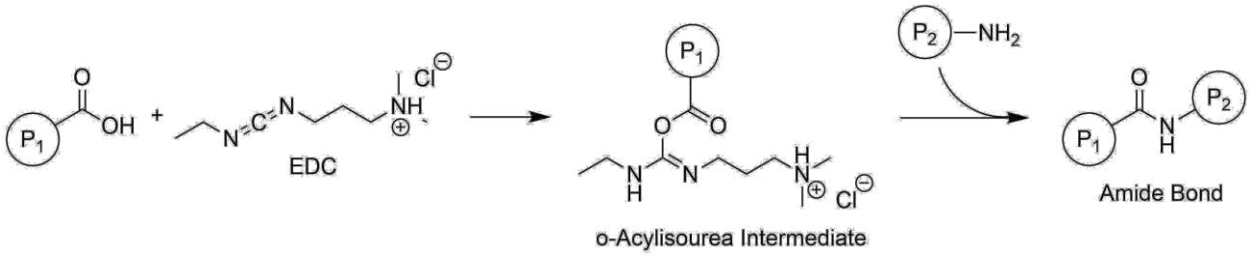



FIGURE S3.3 REACTION SCHEME FOR GLYOXAL CROSSLINKING
**(a)** Glyoxal reaction is proposed to proceed primarily via Schiff base reaction with two lysines. **(b)** The Schiff base product can be reduced with DMAB to generate stable secondary amine linkages. **(c)** A second glyoxal molecule can be added to an existing linkage to form a imidazolium product called a glyoxal lysine dimer (GOLD).

**A**




**B**

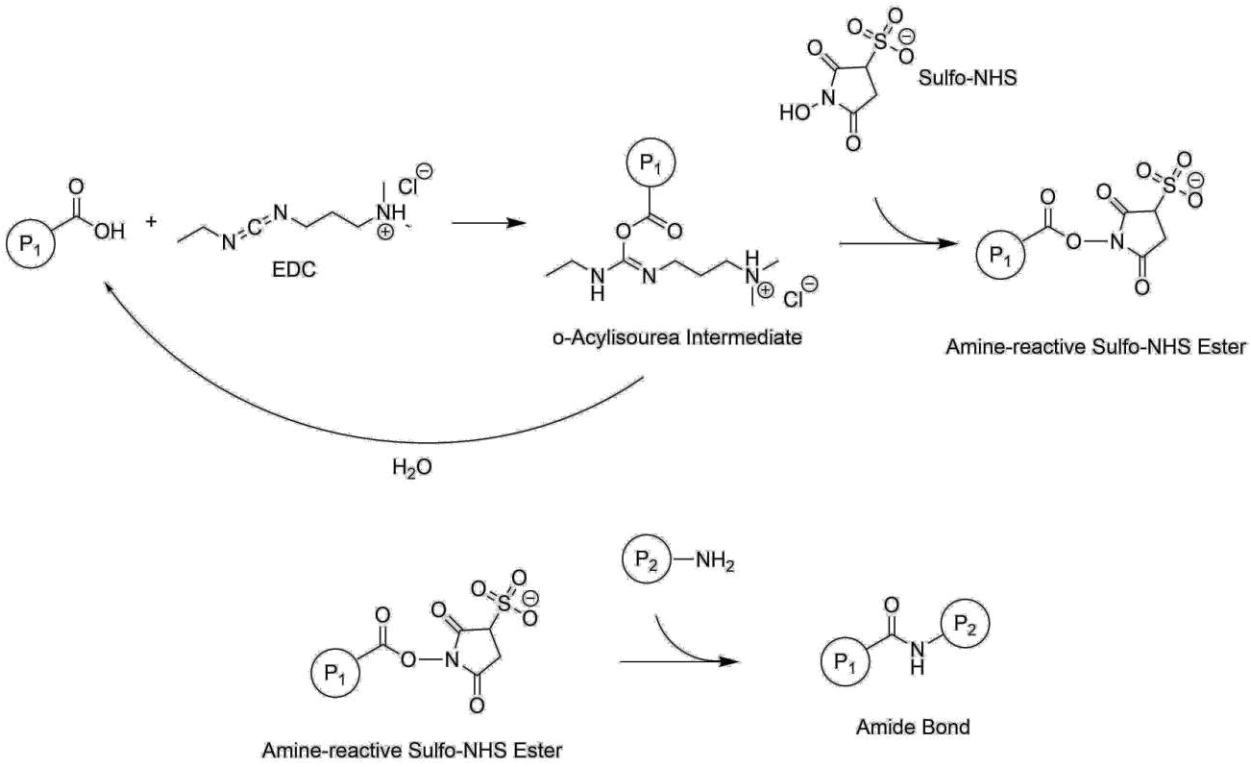



FIGURE S3.4 REACTION SCHEME FOR EDC CROSSLINKING
**(a)** EDC reacts with a carboxylic acid to form an o-Acylisourea active ester intermediate. This active ester intermediate will either hydrolyze or be displaced by an amine to form an amide bond. **(b)** Addition of S-NHS to the EDC reaction can stabilize the labile o-Acylisourea ester intermediate by formation of an NHS-ester intermediate.

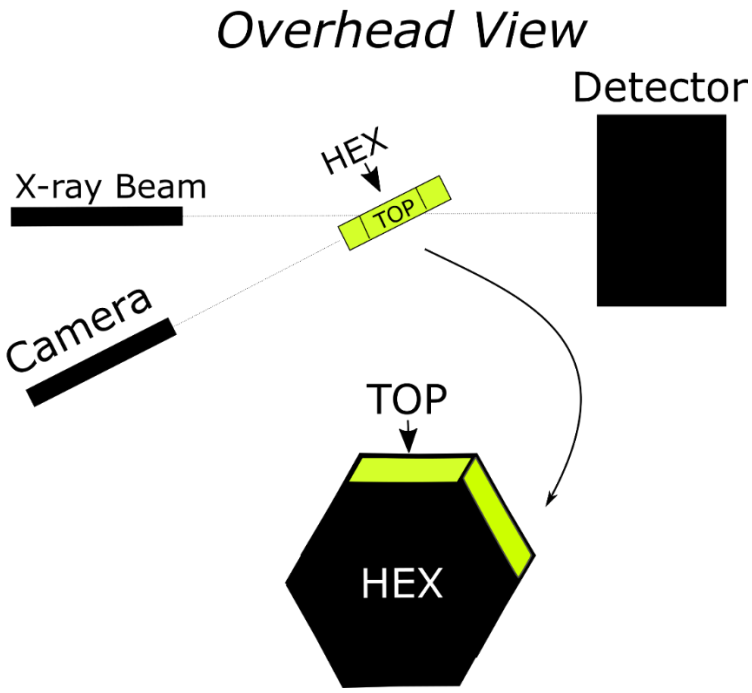



FIGURE S3.5 XRD SETUP FOR RESOLUTION ESTIMATES
Due to the highly anisotropic properties of native CJ crystals, care was taken to consistently align the crystal in the same orientation with respect to the beam for resolution estimates. Specifically, the “side” of the crystal was aligned facing the camera.

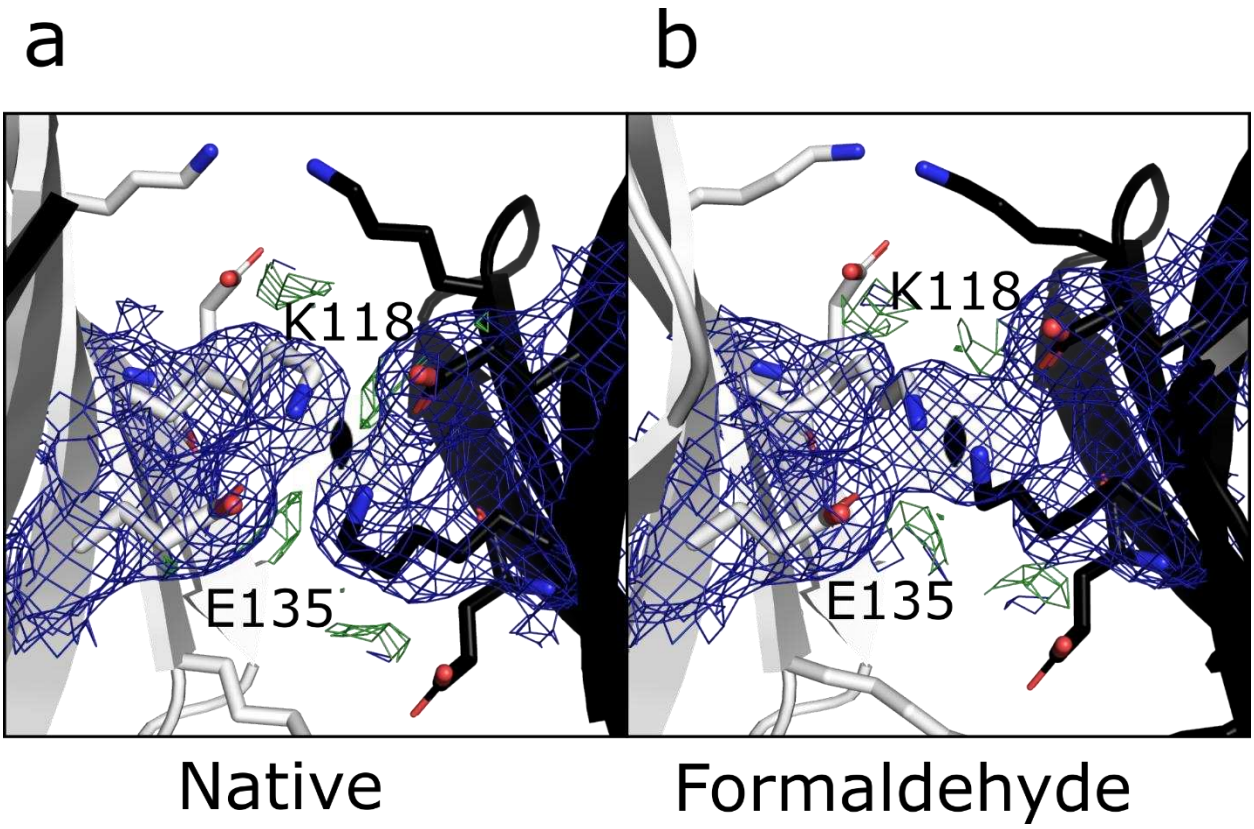



FIGURE S3.6 FORMALDEHYDE CROSSLINKING MODIFICATION AT K118
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native and **(b)** formaldehyde modified crystals centered at K118 2-fold symmetry interface. In the native crystal structure, there is a salt bridge between K118 and E135. After formaldehyde crosslinking there is addition of contiguous electron density between K118:K118 that is not observed in the native crystal structure.

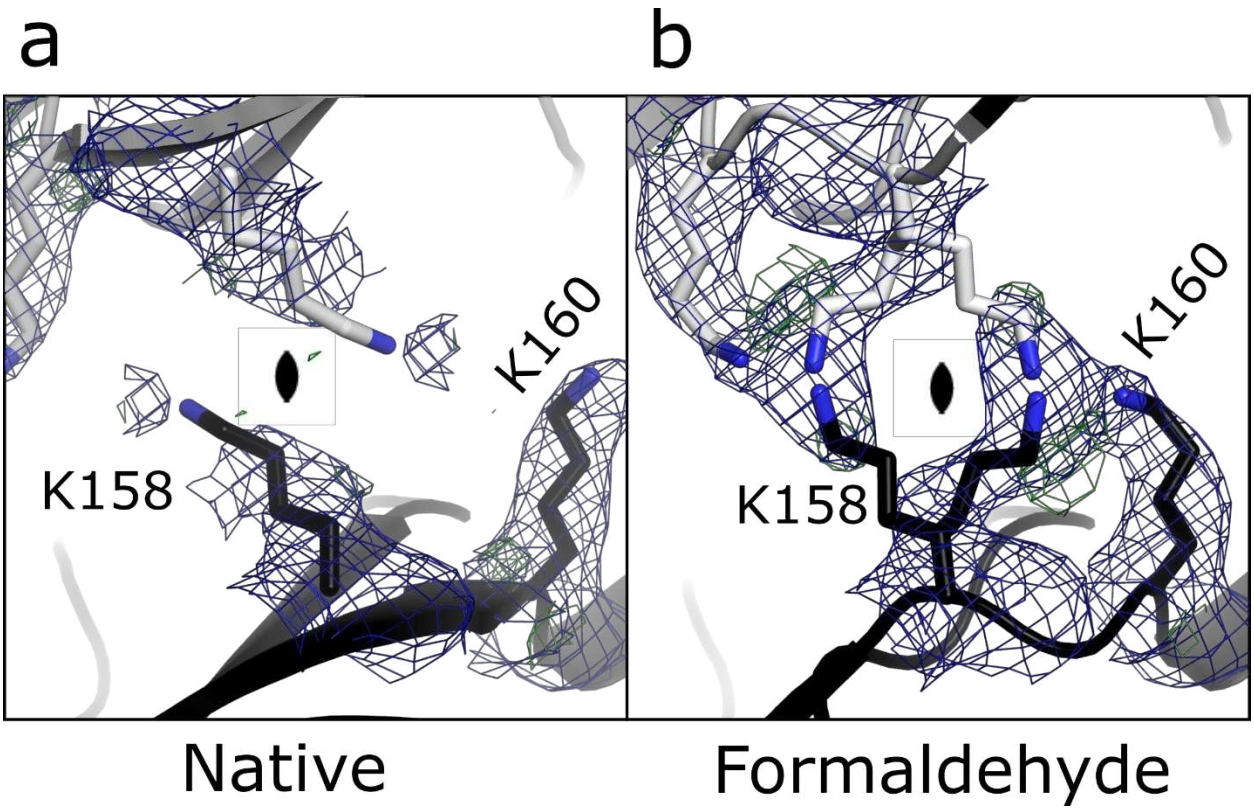



FIGURE S3.7 FORMALDEHYDE CROSSLINKING MODIFICATION AT K158 AND K160
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native and **(b)** formaldehyde modified crystals centered at K158 and K160 two-fold symmetry interface. In the native structure, K158 is not well resolved past the $\delta$-carbon. After crosslinking with formaldehyde, there is addition of electron density between K158 symmetry partners. Additionally, there may be crosslinks between K158 and K160.

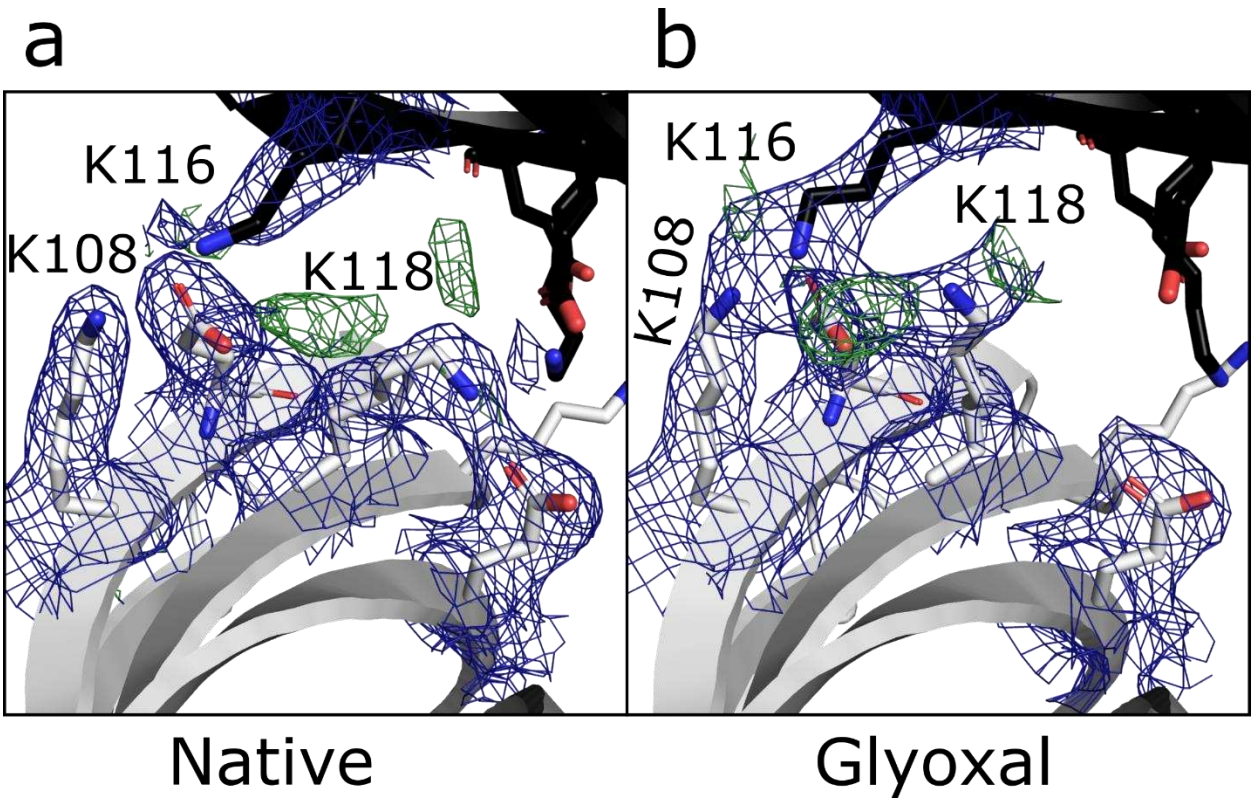



FIGURE S3.8 GLOXAL CROSSLINKING MODIFICATION AT K118
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native and **(b)** glyoxal modified crystals centered at K108, K116, and K118 interface. In the native structure, K118 forms a salt bridge with E135. After crosslinking with glyoxal, there is addition to the electron density between K116 and K118 as well as K108 and K116.

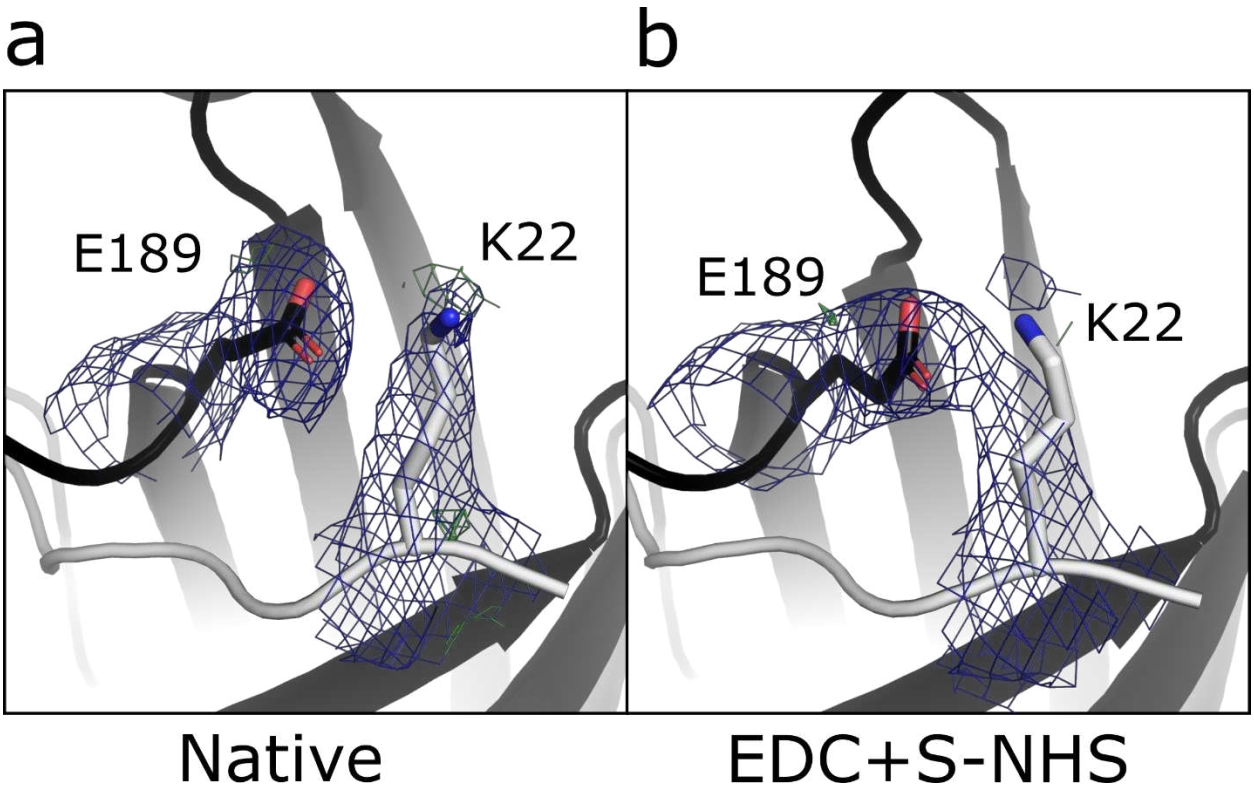



FIGURE S3.9 EDC+S-NHS CROSSLINKING MODIFICATION AT K22 AND E189
2Fo-Fc maps contoured to 1σ and Fo-Fc maps contoured to 3σ for **(a)** native and **(b)** EDC+S-NHS modified crystals centered at K22 and E189 interface. There is a distinct change in the electron density between the native crystal and EDC+S-NHS crystal suggesting amide bond formation. The result is an N and C-terminal fusion of the domain swapped dimer.

## 3.6 ADDITIONAL TABLES

TABLE S3.1 DATA COLLECTION STATISTICS FOR TMAO-8, FORMALDEHYDE, GLYOXAL, AND EDC+S-NHS MODIFIED CRYSTALS

| | **TMAO-8** | **Formaldehyde-24 hr** | **Glyoxal-24 hr** | **EDC+S-NHS-24 hr** |
|---|---|---|---|---|
| **Data Collection** | | | | |
| Light Source | Rigaku Homelab | Rigaku Homelab | Rigaku Homelab | Rigaku Homelab |
| Wavelength (Å) | 1.54187 | 1.54187 | 1.54187 | 1.54187 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | | | |
| a, b, c (Å) | 179.15, 179.15, 50.27 | 179.80, 179.80, 49.56 | 180.08, 180.08, 50.13 | 177.80, 177.80, 50.53 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 18.22-2.55 (2.71-2.59)* | 19.80-2.69 (2.82-2.69)* | 21.49-3.18 (3.4-3.18)* | 20.55-3.15 (3.37-3.15)* |
| Measured reflections | 168252 | 170195 | 142016 | 141982 |
| Unique reflections | 15159 | 13422 | 8459 | 8531 |
| Completeness (%) | 99.3 (100.0) | 99 (95.4) | 99.6 (100) | 99.6 (99.7) |
| Redundancy | 11.1 (11.7) | 12.7 (6.8) | 16.8 (13.8) | 16.6 (12.7) |
| $cc_{1/2}$ | 0.996 (0.756) | 0.996 (0.520) | 0.994 (0.657) | 0.995 (0.622) |
| $I/\sigma(I)$ | 10.7 (1.2) | 11.0 (1.3) | 11.2 (1.0) | 10.6 (1.0) |

* values in parantheses are for high resolution shell

TABLE S3.2 OBSERVED MODIFICATIONS AND RESULTING TOPOLOGY FROM CHEMICAL CROSSLINKED STRUCTURES

| | Crosslinking Pair | | Topology |
|---|---|---|---|
| **Formaldehyde** | K36 | K38 | Intramolecular |
| | K95 | K95 | Axial threads |
| | K97 | K165 | Axial threads |
| | K118 | K118 | Tetramer |
| | K158 | K158 | Tetramer |
| | K158 | K160 | Tetramer |
| **Glyoxal** | K36 | K38 | Intramolecular |
| | K95 | K95 | Axial threads |
| | K95 | K98 | Axial threads |
| | K108 | K116 | Tetramer |
| | K116 | K118 | Tetramer |
| | K165 | K130 | Intramolecular |
| **EDC-S-NHS** | K22 | E189 | Dimer fusion |
| | E137 | K143 | Rings |
| | K73 | D101 | Intramolecular |
| | K118 | E135 | Intramolecular |

*The inferred topologies assume stable domain swapped complex.

# 4 MACROMOLECULAR GUEST LOADING

## 4.1 INTRODUCTION

An alluring goal of nanotechnology is the development of macroscopic materials in which the position of constituent atoms can be readily programmed,[152] and crystalline materials are ideal because X-ray diffraction can elucidate the resulting atomic structure. Building crystals out of biomolecular blocks is a promising route to obtain materials that are economical, biodegradable, and can incorporate a vast known repertoire of functional domains. Notably, proteins and enzymes confined within crystals have often been reported to enjoy enhanced thermal and solvent stability.[153–155] Designed crystals composed of organic molecules, DNA, or proteins have met recent success.[88,156,157] However, crystallizing biomolecules is notoriously challenging and changes to the constituent monomers can easily disrupt the crystallization process.[158,159] No system for easy, modular crystal assembly has been described. Instead, designed crystals have required standalone design efforts by specialist laboratories.[88,157]

We seek to develop comparably precise crystalline materials (nanometer precision) that are nonetheless modular and more easily engineered. Our approach uses highly porous protein crystals (pore diameter > 8 nm) to decouple crystallization from subsequent assembly steps. We first prepare a porous scaffold crystal and afterwards load guest domains within the host crystal.

In contrast to the weak, noncovalent interactions that hold together typical protein crystals, guest domains can be attached to the host crystal using strong interactions. For maximum programmability, affinity tags for the desired assembly can be genetically

encoded on the guest and scaffold monomers. The current manuscript demonstrates that non-covalent, metal-mediated capture and genetically encoded histidine tags provide a significant level of control. Loading and release of guest molecules can be fine-tuned to spatially segregate multiple guest proteins. Similarly, by controlling the diffusion of crosslinking agents and the protein monomer that comprises our scaffold crystals, it is possible to engineer crystalline shells that still diffract.

## 4.2 RESULTS AND DISCUSSION

### 4.2.1 SCAFFOLD PREPARATION AND GUEST LOADING

We began by selecting a candidate protein crystal with pores large enough to allow diffusion of guest proteins (e.g., mNeonGreen[160] with radius ≈3.6 nm) that displayed histidine tags for guest capture. The scaffold crystal is composed of a putative Campylobacter jejuni periplasmic protein (CJ) that crystallizes in a P622 space group with a ≈13 nm pore diameter (Protein Data Bank entry 2fgs) (**Figure 4.1**).[161] High salt (e.g., 3 to 4 M ammonium sulfate) sitting drop crystallization experiments yielded (overnight) hexagonal CJ crystals with a typical maximum span of ≈200–400 μm.

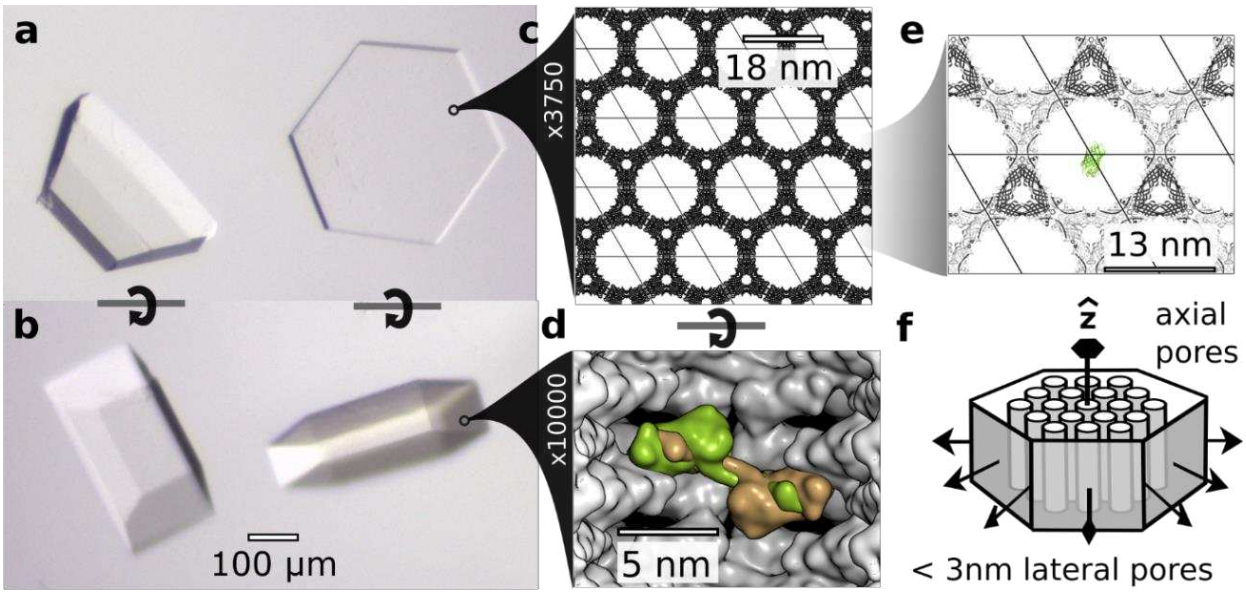



FIGURE 4.1. CJ CRYSTAL TOPOLOGY
Two representative CJ crystal morphologies viewed from **(a)** "top view" and **(b)** "side view" (~90° rotation). **(c)** Th e crystals large axial pores are packed into an 18 nm hexagonal array, with **(e)** each 13 nm pore diameter larger than typical proteins such as mNeonGreen (green) and **(d)** the pore wall composed of CJ domain-swapped dimer (green/tan) with gaps that are too small for guest protein diffusion **(f)** A crystal schematic, not to scale, illustrates our convention for defining crystal orientation, with the hexagonal nanopores parallel with the z-axis ($\hat{z}$).

To serve as a useful scaffold, these crystals must be stable in a wide range of solvent conditions. Optimizing washing and crosslinking of the crystals proved to be technically challenging since we required host crystals that were solvent stable, competent to uptake guest domains, and still capable of yielding usable diffraction data. Ultimately, we found that glyoxal (a chemical conjugation agent not previously used for protein crystal crosslinking) yielded crystals that allowed full guest uptake of mNeonGreen as assessed by confocal microscopy (**Figure 4.2**), while retaining robust diffraction (**Figure S4.1**). mNeonGreen could be passively loaded throughout the crystal within 10 min (Figure 4.2a; for full z-stacks **Figure S4.7**), and passively unloaded within 20 min (Figure 4.2b; for full z-stacks **Figure S4.8**)

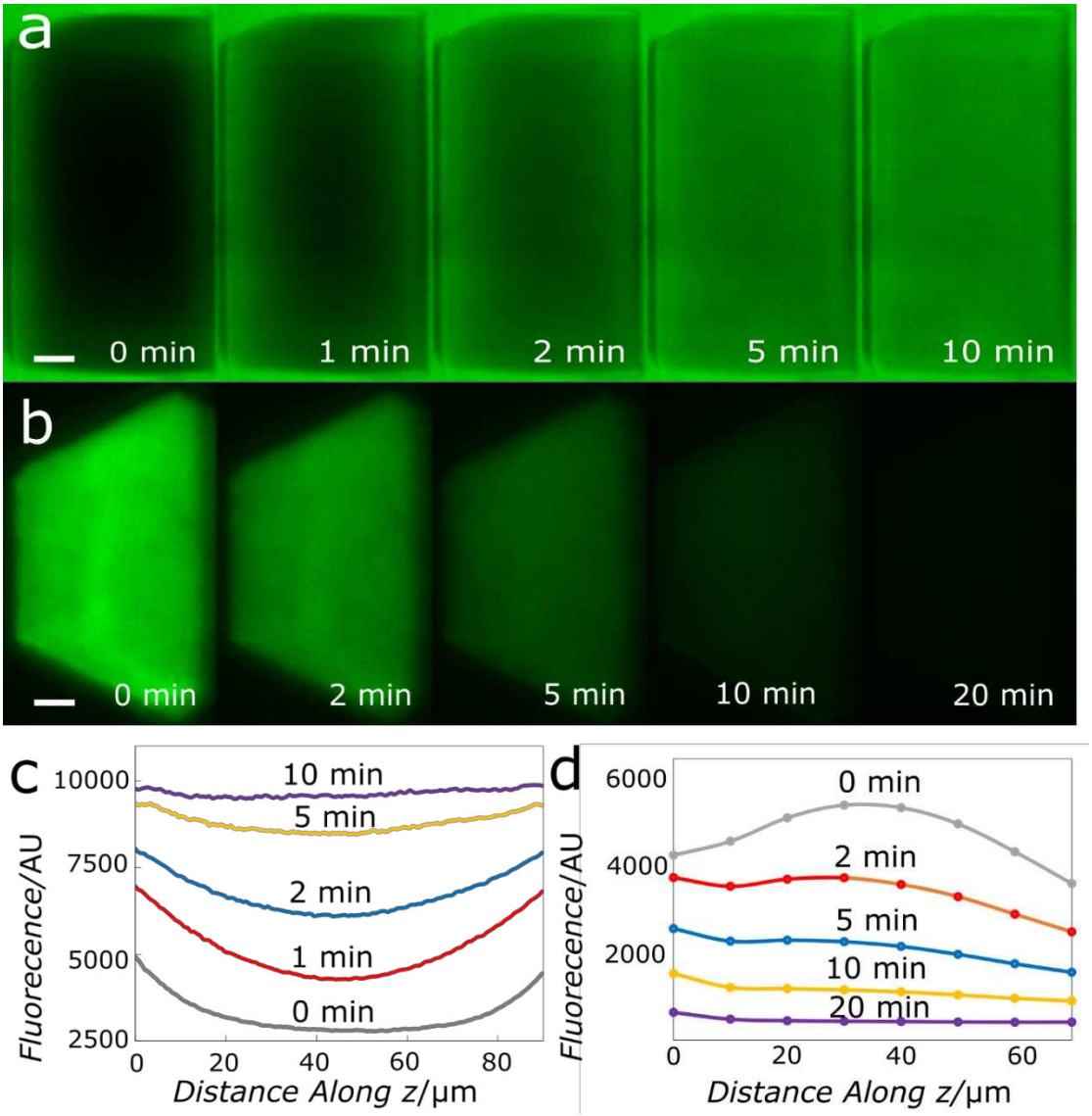



FIGURE 4.2 PASSIVE mNEONGREEN LOADING AND RELEASE
**(a)** Confocal time series (10 min) for mNeonGreen loading throughout crystal. The major crystal pores ($\hat{z}$) are oriented horizontally. **(b)** Subsequent passive mNeonGreen release from an interior CJ crystal plane with major pores ($\hat{z}$) perpendicular to the page. Mean time resolved fluorescence intensity along **(c)** central horizontal stripe of panel (a) or **(d)** z-stack for panel (b) with 10 μm steps. Scale bars are 25 μm.

In the course of extensive crosslinking tests, we found protocols that yield novel protein crystalline containers. In these cases, the crosslinking agent links the outer layers of the crystal, while leaving the inner portion of the crystal unmodified. Upon gradual exchange of the mother liquor, the non-crosslinked interior can dissolve and exit the crystal through the large pores, leaving behind a liquid center. The crystal shell architecture depends on crosslinker kinetics and dose. Crystals with a roughly cylindrical cavity can be generated

with a high dose of very reactive crosslinker, such as glutaraldehyde, for a short time (**Figure 4.3**). With an optimized shell generation protocol, crystalline containers can load guest proteins and maintain diffraction, confirming an intact porous nanostructure (**Table S4.2** and **S4.3** and **Figure S4.9**). To assess the possibility of batch shell synthesis, we proceeded to simultaneously crosslink multiple crystals from a single growth well (**Figure S4.10**). Crystals of similar size and shape yielded shells with consistent crosslinked shell volume. However, smaller crystals were fully crosslinked with the same dose. Optimizing the initial growth of uniform crystals is therefore a key prerequisite for the synthesis of larger batches of crystalline shells.

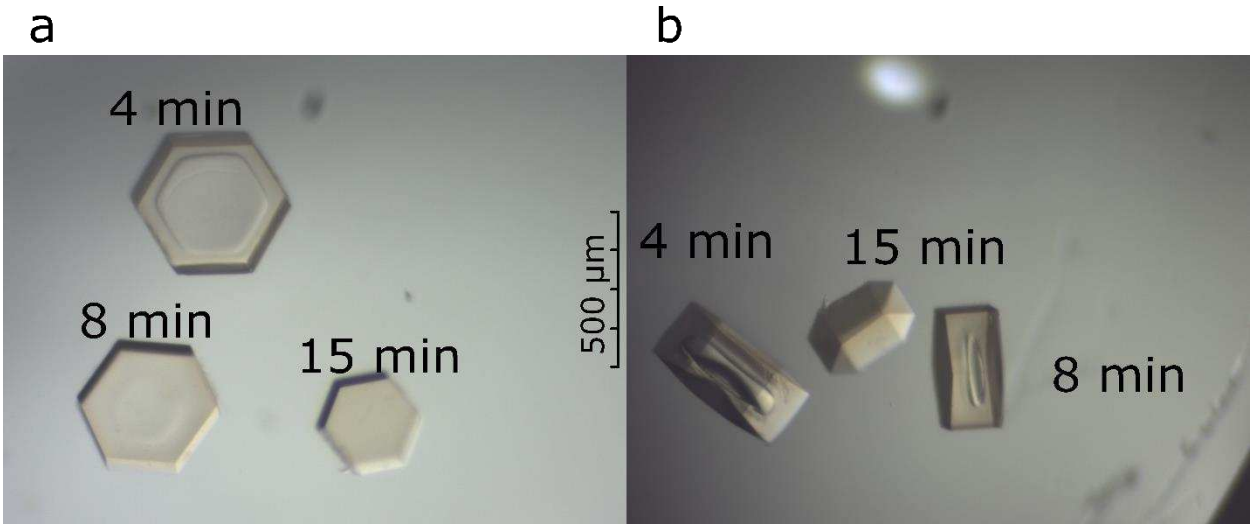



FIGURE 4.3 DOSE DEPENDENT CJ SHELL GENERATION
CJ crystal shells with varying exposure to glutaraldehyde contain cavities of varying size (15 minutes is fully crosslinked). **(a)** Top view of shell generation. **(b)** Side view of shell generation.

### 4.2.2 NONCOVALENT CAPTURE VIA METAL-MEDIATED HIS-TAG DIMERIZATION

Due to the ubiquity of histidine tags (histags), metal-mediated histag dimerization is a particularly convenient binary affinity motif. Evers et al. used Fluorescence Resonance Energy Transfer (FRET) based sensors to quantify histag dimerization in the presence of Zn(II) (with fairly high affinity: $47 \pm 4 \times 10^{-9}$ M and Ni(II) (with 20-fold lower affinity: $0.88 \pm 0.07 \times 10^{-6}$ M).[162] This is in rough accord with other estimates ($10^{-7}$ M) for the equilibrium constant between histag and Ni(II).[163] An advantage of the histag dimerization strategy is the prospect of tuning host–guest affinity by changing the cation identity and concentration.

Fluorescence microscopy demonstrated that histag dimerization between CJ crystal scaffolds and fluorescent proteins, mNeonGreen and mCherry, was an effective capture strategy. Porous protein crystals loaded with mNeonGreen were transferred to a solution containing 10 mM $Zn^{2+}$. In the presence of $Zn^{2+}$, minimal decrease in mNeonGreen signal was observed in the center of the crystal (**Figure 4.4**a). However, when mNeonGreen-$Zn^{2+}$-CJ crystals were transferred to a solution that outcompetes histidine metal coordination (≈100 mM ethylenediaminetetraacetic acid, EDTA), crystals lost their fluorescence in minutes (Figure 4.4b). Strikingly, when EDTA was titrated gradually, mNeonGreen was selectively released from the edges of the crystal (Figure 4.4c), despite the absence of lateral pores of sufficient size to permit mNeonGreen diffusion (Figure 4.1). We attribute this diffusion pattern to rate limiting diffusion of EDTA into the crystal from all sides.

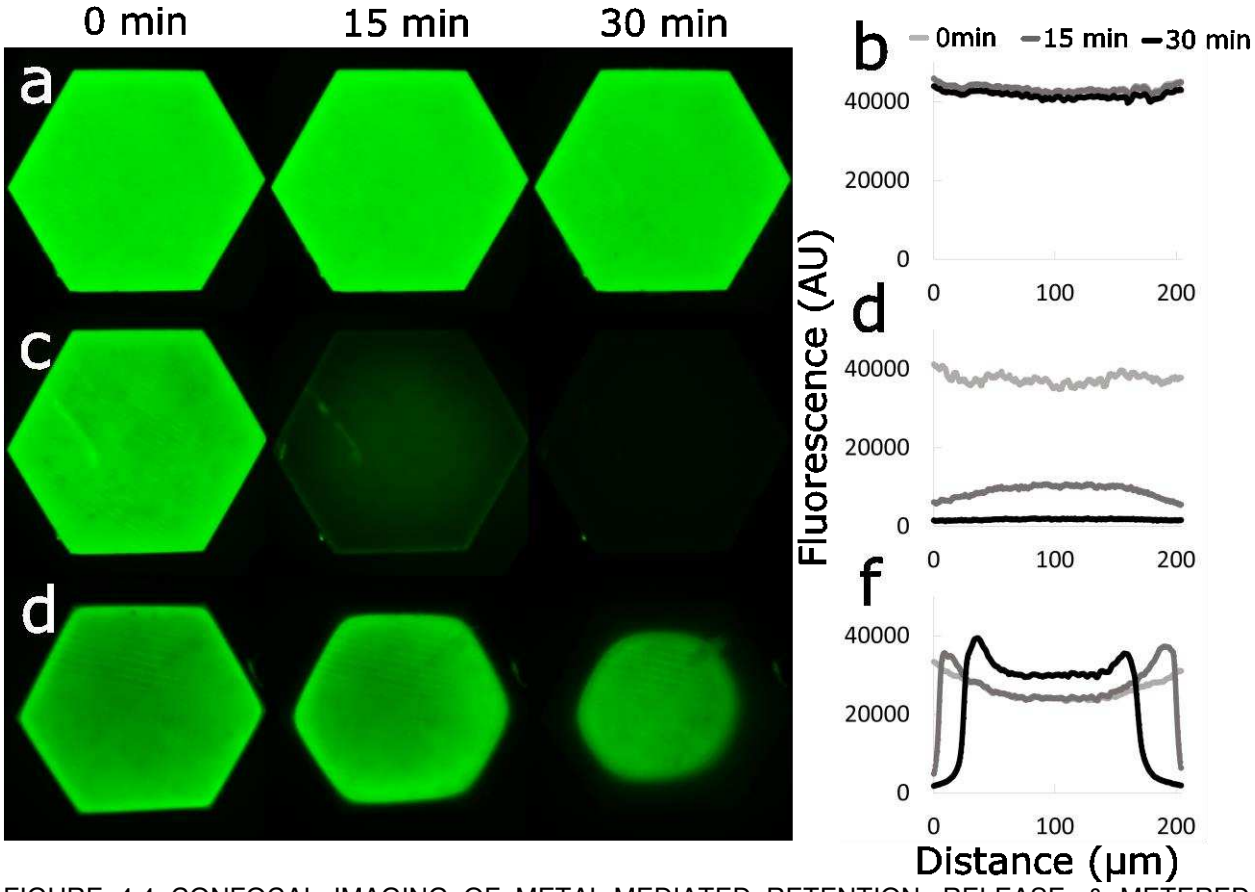



FIGURE 4.4 CONFOCAL IMAGING OF METAL-MEDIATED RETENTION, RELEASE, & METERED RELEASE

**(a)** & **(b)** In the presence of $Zn^{2+}$, mNeonGreen is retained in an interior plane of CJ over 30 minutes but **(c)** & **(d)** can be released within 30 min by adding EDTA (~100 mM). In contrast, **(e)** & **(f)** titration with EDTA until crystal edge fluorescence decreases (~50 μM EDTA), results in a sharp radial pattern. Fluorescence intensity for (b), (d), and (f) was quantified (perpendicular to $\hat{z}$) along a horizontal vector across the cognate crystal interior in (a), (c), and (e).

To demonstrate that metal-mediated histag dimerization contributes to guest protein retention, we designed variants of mCherry and CJ crystals with and without histags (**Figures S4.1–S4.5**). Upon testing all variant combinations, we observed the strongest metal-mediated mCherry retention when both the scaffold and the guest domains carried histags (**Figure S4.11-S4.12**). Notably, mCherry with a histag adsorbed fairly strongly to CJ-Δ6xHIS. We suspect that this is due to high effective concentration (≈42 mM) of histidine residues displayed on the pore of the CJ-Δ6xHIS crystal (**Figure S4.13**).

### 4.2.3 MULTI STAGE GUEST LOADING

We can readily co-load CJ crystals with two different guest proteins (**Figure S4.14**) Multistage loading, in which different guest macromolecules are adsorbed within different crystal sectors, can be achieved via a simple extension of the controlled adsorption described above. Porous crystals were soaked in the first guest (e.g., mNeonGreen) for metal-dependent adsorption. The guest molecule was then selectively released from the crystal edges by titrating with EDTA. Primary guest release was halted by removing EDTA from solution. At this point, a secondary guest (e.g., mCherry) can be supplied (**Figure 4.5**a). Interestingly, strong mCherry adsorption was observed in the crystal interior where mNeonGreen remained bound with $Zn^{2+}$. This could correspond to “active loading” wherein mCherry is adsorbed to the crystal via metal-mediated his–tag interactions in the interior, versus passive loading in the exterior (**Figure S4.15**). The shape of the resulting loading pattern depends on the elapsed loading and release times and the guest concentration supplied to the crystal exterior. Multistage loading was also applied to crystal shells. For example, mNeonGreen was compartmentalized to the cavity and mCherry to the exterior crystalline matrix (Figure 4.5b,c and **Figure S4.16-17**).

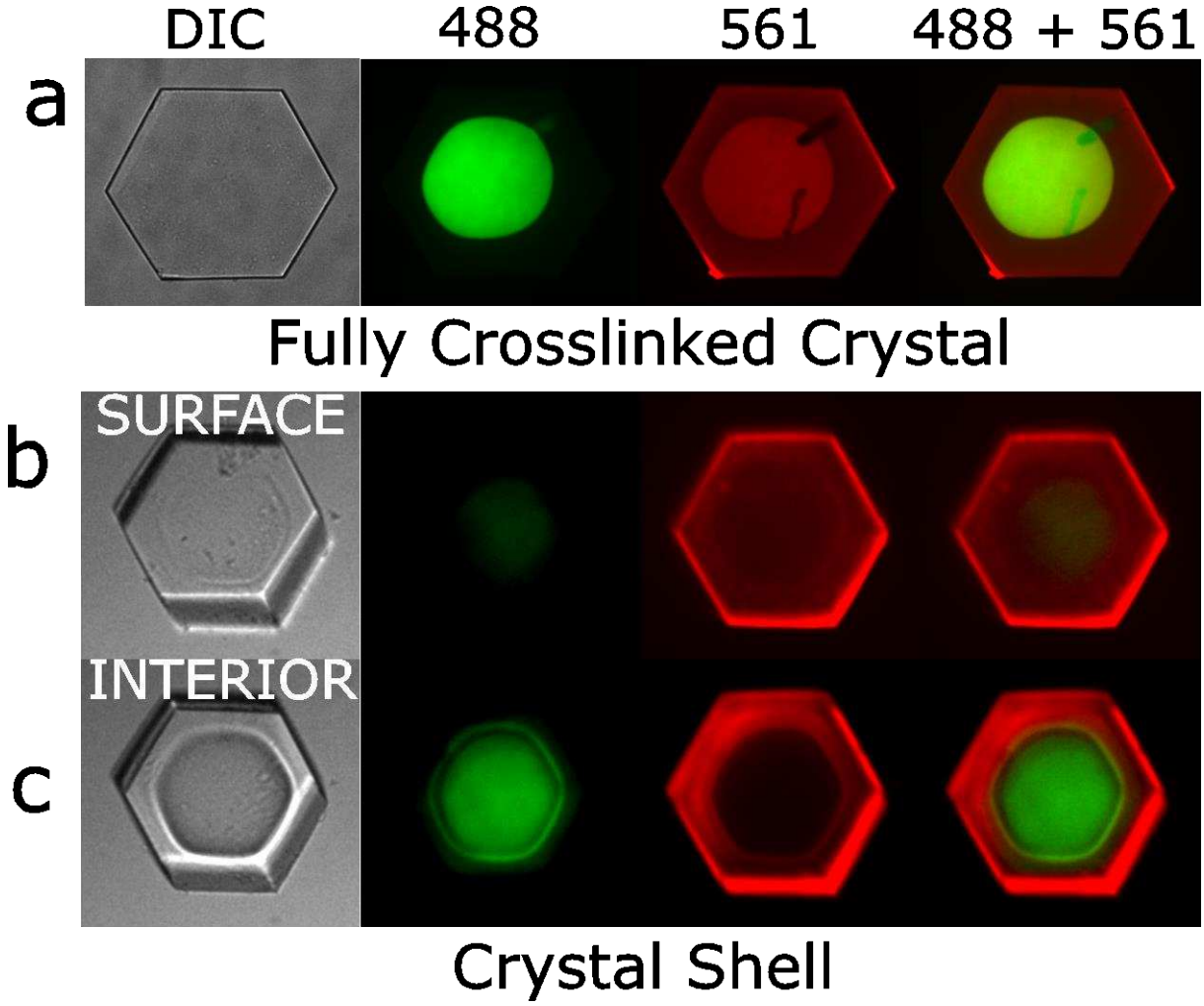



FIGURE 4.5 MULTISTAGE LOADING
After slow mNeonGreen release from the crystal exterior via EDTA titration (Fig, 3e), mCherry was loaded. **(a)** Confocal imaging of an interior plane within the crystal shows spatial segregation of mNeonGreen and mCherry. Multistage loading was also applied to crystal shells. mNeonGreen was loaded throughout a shell crystal, retained with $Zn^{2+}$, and gradually released via EDTA titration such that mNeonGreen was spatially segregated to the shell cavity prior to mCherry loading. DIC, 488 nm, 561 nm, and 488 + 561 nm merged images of confocal planes at the **(b)** surface of the crystal and **(c)** interior of the crystal show spatial segregation of mCherry and mNeonGreen to predominantly occupy the shell exterior and cavity respectively. Major pores ($\hat{z}$) are perpendicular to the page.

### 4.2.4 NONCOVALENT CAPTURE VIA IONIC INTERACTIONS

For downstream applications it may be important to encourage or disrupt ionic interactions between the scaffold and guest proteins. To interrogate electrostatic contributions to guest adsorption we took advantage of the ability to co-load multiple guest proteins. Specifically, mNeonGreen (pI ≈7.1) and mCherry-Δ6xHIS (pI ≈5.5) were co-loaded into a CJ crystal. Loaded crystals were transferred to solutions of varying ionic

strengths to monitor release. The solutions were buffered at pH 7.5 and included glycerol and EDTA to attenuate hydrophobic or metal-mediated interactions between the guest proteins and the host crystals. While guest proteins escaped at similar rates at medium ionic strength, low ionic strength resulted in markedly different release rates (**Figure S4.18**). In principle, we could promote the selective adsorption of guest proteins via chemical modification of the scaffold surface during the crosslinking or quenching steps (e.g., introducing negative charges by quenching with hydroxylamine-O-sulfonic acid).

## 4.3 CONCLUSIONS

Notably, the 13 nm CJ nanopore diameter significantly exceeds the size of the guest fluorescent proteins used herein. We have observed successful uptake of diverse larger macromolecules in preliminary work for a variety of downstream projects. For example, large enzymes such as cytochrome P450 heme domain (≈52 kDa, data not shown) or glucose oxidase dimer (≈160 kDa, data not shown) appear to load into the host crystals without difficulty.

Applications such as enzymatic catalysis and biosensors may benefit from simultaneous control of both the nanoscale (guest capture within precisely defined nanopores) and mesoscale (guest spatial segregation on the tens of micrometer scale). For example, spatial segregation of enzymes could prove helpful by placing oxygen scavenging enzymes in the surface layers and oxygen sensitive enzymes in the core, or sequestering enzyme cascade members that produce toxic or unstable intermediates in the core. Converting protein crystals to integrated biosensors will require methods to couple nanopore opening to analyte detection. Here, EDTA leads to the loss of fluorescence from a very bright protein crystal “pixel.” Crystal shells may provide a similar, albeit highly

amplified biosensor response wherein nanopore opening leads to the metered release of trillions of guest macromolecules.

Ultimately, scaffold crystals may serve as a chassis to organize trillions of guest molecules in 3D. Controlled capture and release of guest proteins within host crystals is a beginning. Fully realizing host–guest protein crystals as nanotechnological materials will require additional engineering. Changes to the scaffold, to the capture tags, and to the solution conditions will lend additional control over the guest domain orientation and function. Guest domains that can be loaded at high capacity and adopt coherent structure should enable the unprecedented opportunity to interrogate guest structure as a function of solution conditions using x-ray diffraction (XRD). To date, XRD tests with mNeonGreen loaded crystals have yielded well-defined scaffold structures. We have not yet observed defined electron density for guest proteins, nor have we saturated host crystals with guest domains (a likely prerequisite for observing guest structures via XRD). Multiple orthogonal attachment sites in the crystal will enable the programmed assembly of increasingly complex materials. The combination of modular assembly via standard genetically encoded tags, with the possibility of resolving structural details via XRD, makes host–guest crystals a privileged platform for nanotechnology.

## 4.4 MATERIALS AND METHODS

### 4.4.1 CJ PROTEIN CRYSTAL PREPARATION

A codon optimized gene encoding a Putative Periplasmic protein (Genebank ID: cj0420, Protein Data Bank code: 2fgs) from Campylobacter jejuni was obtained through Life Technologies and cloned into pSB3 vector at NdeI and XhoI. It was expressed in E. coli BL21 (DE3) pLySs using Studier's autoinduction protocol at 25 °C for 24 hours.[14] The periplasmic contents of the cell were selectively extracted using TSE extraction protocol with both the sucrose and $MgCl_2$ fractions combined.[15] The extracted contents were diluted in HNGI buffer (50 *mM* HEPES, 500 *mM* NaCl, 10% Glycerol, 40 *mM* Imidazole at pH 7.5) and purified via immobilized metal chromatography IMAC. A single chromatography step provided sufficient purity for crystallization. The purified protein was dialyzed into TST Buffer (10 *mM* HEPES, 150 *mM* NaCl, 10% glycerol at pH 7.5) and concentrated to 15 mg $mL^{-1}$. The average CJ yield was 5 mg $L^{-1}$ of culture.

A CJ variant was subcloned to insert a tobacco etch virus (TEV) protease cleavage site between the protein and an N-terminal hexahistidine tag (CJ-Δ6xHIS). This variant was expressed and purified similarly to CJ, but extracted via sonication. After initial purification, the protein was dialyzed into TST Buffer + 40 *mM* Imidazole. TEV cleavage was performed by addition of 1 mg TurboTEV Protease (Eton Biosciences) per 100 mg CJ-Δ6xHIS and 5 mM DTT at room temperature for 4 hours. After digest, the protein was reverse purified on (IMAC), dialyzed into TST, and concentrated to 15 mg $mL^{-1}$.

The purified proteins were characterized with SDS-Page and stained with Invitrogen His-Tag Stain to demonstrate presence or absence of histag (**Figure S4.6**). CJ variants were crystallized overnight by sitting drop vapor diffusion at 20˚C in 80-100% Tadsimate at pH

6.5, defined as: 1.83 *M* malonic acid, 0.25 *M* sodium citrate, 0.12 *M* succinic acid, 0.3 *M* D-L malic acid, 0.4 *M* acetic acid, 0.5 *M* sodium formate, 0.16 *M* sodium tartrate, and 10% glycerol titrated with sodium hydroxide. This is a modified form of Tacsimate™ (Hampton Research) that removes primary amines for crosslinking and adds glycerol as a stabilizing agent). Crystals were generally 50-100 µm in height x 200-500 µm in diameter. Prior to crosslinking, crystals were washed via transfer to washing solution containing 100% Tadsimate at pH 7.5 for 30 minutes. Crystals were then transferred to a crosslinking mixture of 100% Tadsimate at pH 7.5 supplemented with the addition of 1% glyoxal (Acros Organics, 40% Glyoxal, pure solution in water) and 25 *mM* borane dimethylamine complex (DMAB). The crosslinking reaction was allowed to proceed for 2 hours and was quenched by transfer into a solution of 1 *M* hydroxylamine and 100 *mM* DMAB in 150 *mM* NaCl and 100 *mM* citric acid at pH 5.0. After crosslinking and washing, crystals retained smooth, hexagonal morphology and clear color. Coloration was indicative of "overcrosslinking" or "underquenching" and was correlated with guest uptake incompetence.

### 4.4.2 CJ CRYSTAL SHELL GENERATION

CJ crystals for shells were grown and washed using a standard protocol described above for CJ crystal preparation. Crystals were then transferred to a crosslinking mixture of 100% Tadsimate at pH 7.5 supplemented with 0.5% glutaraldehyde (Ted Pella, INC., 25% Glutaradehyde, EM Grade). These crystals were allowed to crosslink for varying amounts of time to produce different shell thickness. Crosslinking was stopped by transfer to a solution of 100% Tadsimate at pH 5.5 supplemented with 0.1 *M* Hydroxylamine and 25 *mM* DMAB for 20 minutes. The crystal interior was gently released by stepwise transfer

to solutions with decreasing Tadsimate concentration until the crystal interior started to dissolve. Abrupt transfers resulted in cracking or popping of the shell. Shells could be reliably generated this way, but due to the reactive nature of glutaraldehyde, crystals could gain a yellow tint and lose the ability to uptake guests. Guest uptake competency was restored by addition of 25 *mM* DMAB to the crosslinking mixture (crystals remained clear colored). However, shell thickness variability was increased when using DMAB.

### 4.4.3 FLUORESCENT PROTEIN PREPARATION

mNeonGreen, mCherry, and mCherry- Δ6xHIS were each subcloned into the pSB3 vector and expressed via IPTG induction. mNeonGreen and mCherry were sonicated and purified via IMAC. Samples were dialyzed into PBS (20 *mM* $Na_2PO_4$ and 150 *mM* NaCl at pH 7.4) and frozen as aliquots of 1 mg $mL^{-1}$ for mNeonGreen and 5 mg $mL^{-1}$ for mCherry. mCherry-Δ6xHIS was prepared by sonication followed by ammonium sulfate precipitation. Ammonium sulfate was added gradually in 0.5 *M* increments from 0 to 4 *M*. At each fraction, precipitated proteins were removed via centrifugation. Each fraction was resuspended in PBS and analyzed for fluorescence vs. total protein content. Enriched fractions were pooled together, polished with size exclusion chromatography, and frozen as 2 mg $mL^{-1}$ aliquots.

### 4.4.4 X-RAY DIFFRACTION AND DATA PROCESSING

CJ crystals and shells were prepared using standard protocols. In all cases, a single crystal or shell was briefly swished through a cryoprotectant solution containing 100% Tadsimate at pH 7.5 prior to flash freezing in a liquid nitrogen stream (T=100 K). Crystal integrity was determined via a 10 frame (0.5°/frame, 60 second exposure) data collection

strategy on a local Rigaku Compact HomeLab with a micro-focus X-ray generator and a Pilatus 200K detector. Data was integrated and scaled using the HKL3000 program suite.

### *4.4.5* FLUORESCENT PROTEIN IMAGING

Fluorescent protein imaging was performed on a Nikon Eclipse Ti spinning-disk confocal microscope with an Andor iXon Ultra 897U EMCCD camera. mNeonGreen images were collected with 488 nm excitation and 525 nm emission wavelengths. mCherry images were collected with 561 nm excitation and 605 nm emission wavelengths. Optical conditions were selected to best represent each fluorescent protein and kept constant through

## 4.5 ADDITIONAL FIGURES

**DNA Sequence CJ**

ATGAAAAAAGTTCTGCTGAGCAGCCTGGTTGCAGTTAGCCTGCTGAGTACCGGTCT
GTTTGCAAAAGAATATACCCTGGATAAAGCCCATACCGATGTTGGCTTTAAAATCAA
ACATCTGCAGattAGCAATGTGAAAGGCAACTTTAAAGATTATAGCGCAGTGATCGA
TTTTGATCCGGCAAGTGCAGAATTCAAAAAACTGGATGTGACCATTAAAATCGCCA
GCGTGAATACCGAAAATCAGACCCGTGATAATCATCTGCAGCAGGATGACTTCTTC
AAAGCCAAAAAATACCCGGATATGACCTTTACCATGAAAAAATACGAGAAAATCGAT
AACGAAAAAGGCAAAATGACCGGCACCCTGACCATTGCCGGTGTTAGCAAAGATA
TTGTTCTGGATGCAGAAATTGGTGGTGTTGCCAAAGGTAAAGATGGCAAAGAAAAA
ATTGGCTTTAGCCTGAACGGCAAAATCAAACGTAGCGATTTCAAATTTGCAACCAG
CACCAGCACCATTACCCTGAGTGATGACATTAATCTGAACATTGAAGTGAAAGCCA
ACGAGAAAGAAGGTGGTAGTCATCACCACCACCATCACTAATAA

**Amino Acid Sequence CJ**

MKKVLLSSLVAVSLLSTGLFA/***8<***/KEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDF
DPASAEFKKLDVTIKIASVNTENQTRDNHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKG
KMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKEKIGFSLNGKIKRSDFKFATSTSTITLSDD
INLNIEVKANEKEGGSHHHHHH**

FIGURE S4.1 DNA AND AMINO ACID SEQUENCE FOR CJ.

Signal peptide cleavage site indicated by */8<*/.

**DNA Sequence CJ-Δ6xHIS**

ATGCATCACCACCACCATCACGAAAATTTGTATTTCCAGGGAAAAGAATATACCCT
GGATAAAGCCCATACCGATGTTGGCTTTAAAATCAAACATCTGCAGattAGCAATGT
GAAAGGCAACTTTAAAGATTATAGCGCAGTGATCGATTTTGATCCGGCAAGTGCAG
AATTCAAAAAACTGGATGTGACCATTAAAATCGCCAGCGTGAATACCGAAAATCAG
ACCCGTGATAATCATCTGCAGCAGGATGACTTCTTCAAAGCCAAAAAATACCCGGA
TATGACCTTTACCATGAAAAAATACGAGAAAATCGATAACGAAAAAGGCAAAATGAC
CGGCACCCTGACCATTGCCGGTGTTAGCAAAGATATTGTTCTGGATGCAGAAATTG
GTGGTGTTGCCAAAGGTAAAGATGGCAAAGAAAAAATTGGCTTTAGCCTGAACGG
CAAAATCAAACGTAGCGATTTCAAATTTGCAACCAGCACCAGCACCATTACCCTGA
GTGATGACATTAATCTGAACATTGAAGTGAAAGCCAACGAGAAAGAATGATGAACC
GCCGATGATTAT

**Amino Acid Sequence CJ-Δ6xHIS**

*MHHHHHHENLYFQ/**8<**/*GKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDFDPASA
EFKKLDVTIKIASVNTENQTRDNHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTG
TLTIAGVSKDIVLDAEIGGVAKGKDGKEKIGFSLNGKIKRSDFKFATSTSTITLSDDINLNI
EVKANEKE**TADDY

FIGURE S4.2 DNA AND AMINO ACID SEQUENCE FOR CJ-Δ6XHIS

TEV cleavage site indicated by */8<*/

**DNA Sequence mNeonGreen**

ATGGTGAGCAAAGGCGAAGAGGATAATATGGCAAGCCTGCCTGCAACCCATGAAC
TGCATATTTTTGGTAGCATTAACGGCGTGGATTTTGATATGGTTGGTCAGGGCACC
GGTAATCCGAATGATGGTTATGAAGAACTGAATCTGAAAAGCACCAAAGGCGATCT
GCAGTTTAGCCCGTGGATTCTGGTTCCGCATATTGGTTATGGTTTTCATCAGTATCT
GCCGTATCCGGATGGTATGAGCCCGTTTCAGGCAGCAATGGTTGATGGTAGCGGT
TATCAGGTTCATCGTACCATGCAGTTTGAAGATGGTGCAAGCCTGACCGTTAATTA
TCGTTATACCTATGAAGGCAGCCACATTAAAGGTGAAGCACAGGTTAAAGGCACCG
GTTTTCCGGCAGATGGTCCGGTTATGACCAATAGTCTGACCGCAGCAGATTGGTGT
CGTAGCAAAAAAACCTATCCGAACGATAAAACCATCATCAGCACCTTCAAATGGTC
ATATACCACCGGCAATGGTAAACGTTATCGTAGCACCGCACGTACCACCTATACCT
TTGCAAAACCGATGGCAGCAAACTATCTGAAAAATCAGCCGATGTATGTGTTCCGT
AAAACCGAACTGAAACACAGCAAAACAGAGCTGAACTTTAAAGAATGGCAGAAAGC
CTTTACCGATGTGATGGGTATGGATGAACTGTATAAACATCATCACCATCACCATTA
A

**Amino Acid Sequence mNeonGreen**

MVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQ
FSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRY
TYEGSHIKGEAQVKGTGFPADGPVMTNSLTAADWCRSKKTYPNDKTIISTFKWSYTTG
NGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSKTELNFKEWQKAFTDVM
GMDELYKHHHHHH*

FIGURE S4.3 DNA AND AMINO ACID SEQUENCE FOR mNEONGREEN

**DNA Sequence mCherry**

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCG
AGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACC
AAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACG
GCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC
CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGT
GGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTG
AAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCA
TGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAG
GGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAG
GTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACA
GTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAA
GGGCCACCATCACCATCATCATTGA

**Amino Acid Sequence mCherry**

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGG
PLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQ
DSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLK
LKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTG
GMDELYKGHHHHHH*

FIGURE S4.4 DNA AND AMINO ACID SEQUENCE FOR mCHERRY

**DNA Sequence mCherry-del6xHIS**

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCG
AGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACC
AAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACG
GCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC
CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGT
GGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTG
AAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCA
TGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAG
GGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAG
GTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACA
GTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAA
GTGA

**Amino Acid Sequence mCherry-Δ6xHIS**

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGG
PLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQ
DSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLK
LKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTG
GMDELYK*

FIGURE S4.5 DNA AND AMINO ACID SEQUENCE FOR mCHERRY- Δ6XHIS

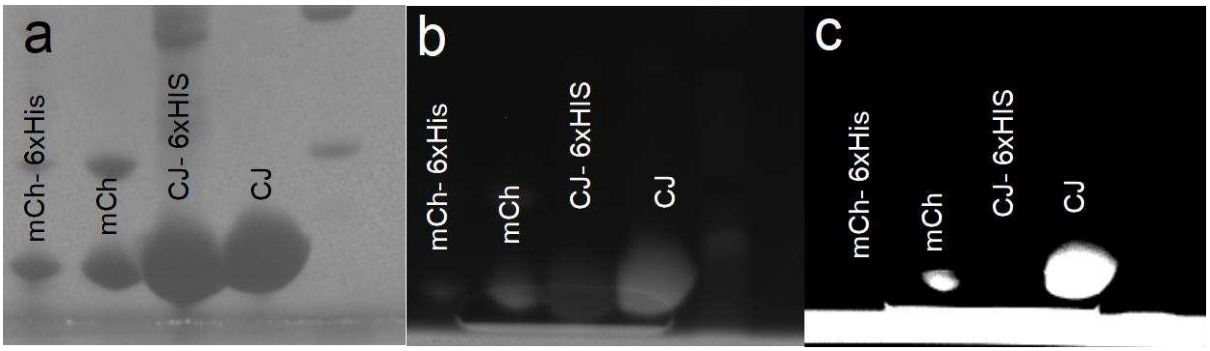



FIGURE S4.6 SDS PAGE GEL WITH HIS-TAG STAIN.
**(a)** Imperial protein stain of SDS PAGE containing mCherry and CJ variants. **(b)** Invision His-Tag stain of SDS PAGE containing mCherry and CJ variants demonstrates increased fluorescence of proteins bearing 6xHIS tags. **(c)** Thresholding of Invision His-Tag stain elucidates presence of his-tag on appropriate samples.

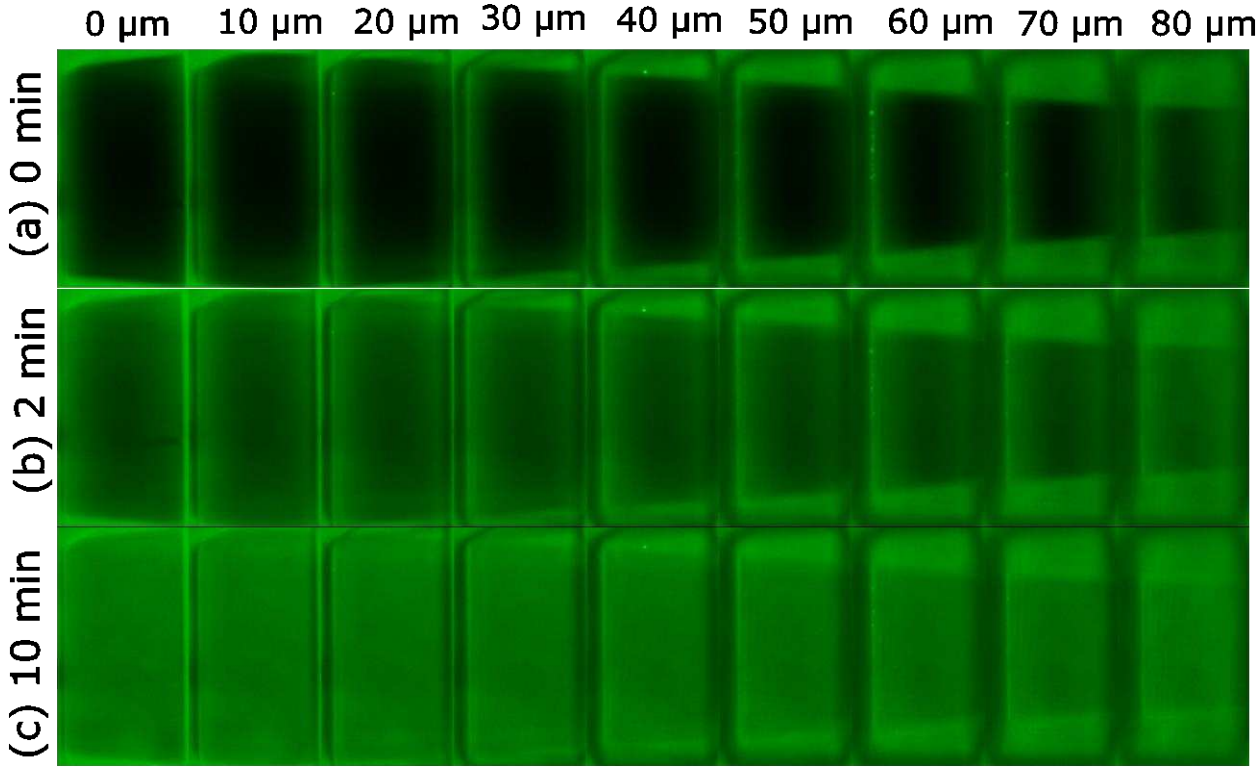



FIGURE S4.7 mNEONGREEN PASSIVE LOADING INTO CJ CRYSTAL CONFOCAL Z-STACK
Change in fluorescence intensity between **(a)** 0 minutes, **(b)** 2 minutes, and **(c)** 10 minutes is indicative of mNeonGreen diffusing into the crystal. The major pores of the crystal ($\hat{z}$) are oriented horizontally for all images.

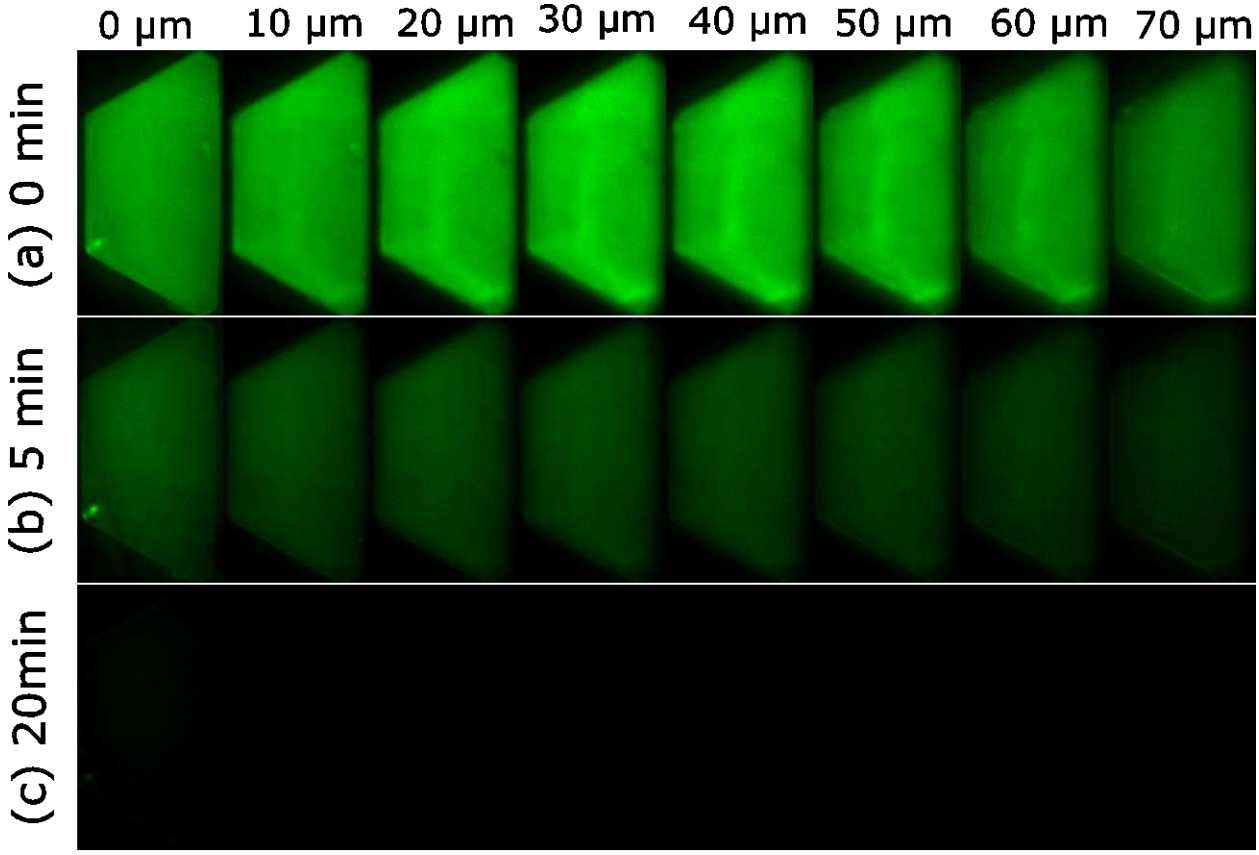



FIGURE S4.8 mNEONGREEN PASSIVE RELEASE FROM CJ CRYSTAL CONFOCAL Z-STACK
Change in fluorescence intensity between **(a)** 0 minutes, **(b)** 5 minutes, and **(c)** 20 minutes is indicative of mNeonGreen diffusing out of the crystal. The major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page for all images.

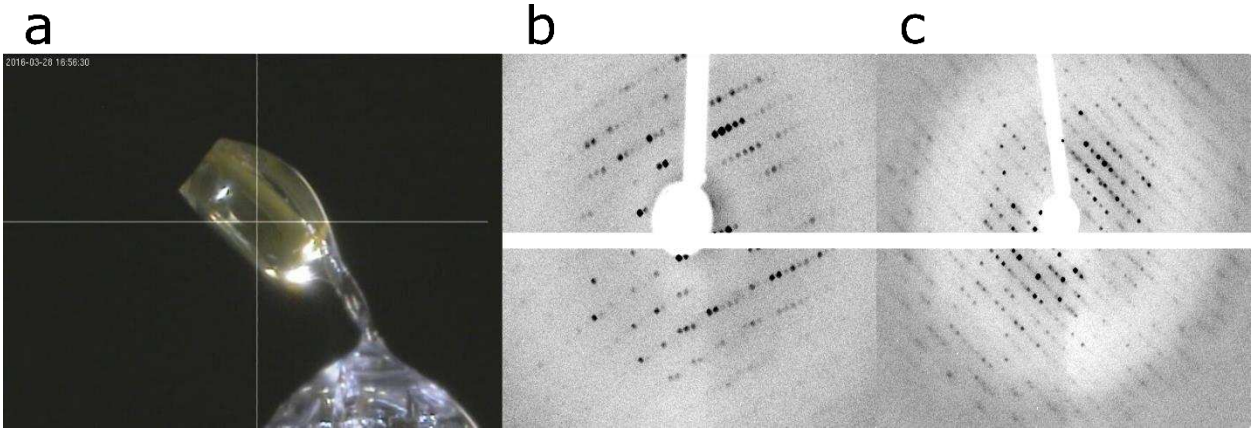
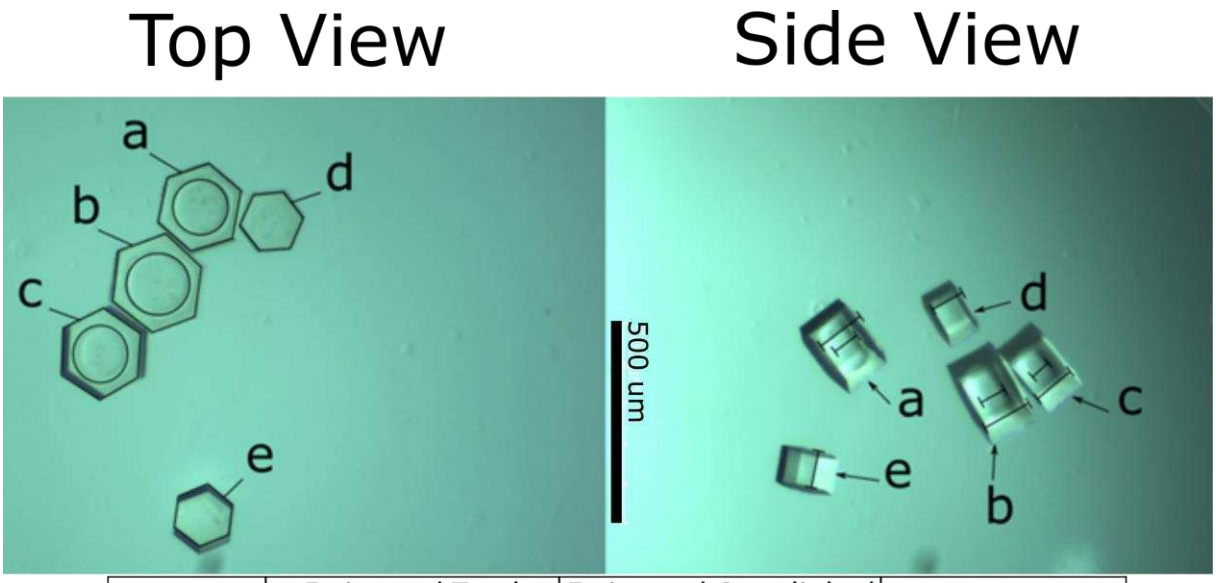


FIGURE S4.9 SAMPLE CJ SHELL DIFFRACTION PATTERNS
**(a)** Image of beam placement for diffraction check on CJ shell treated with glutaraldehyde for 8 minutes. Beam was centered to collect diffraction while shooting directly into the cavity of the shell. **(b)** Sample diffraction image of CJ shell treated with glutaraldehyde for 4 minutes. **(c)** Sample diffraction image of CJ shell treated with glutaraldehyde for 8 minutes.




| | Estimated Total Volume (µm3 x $10^5$) | Estimated Crosslinked Volume (µm3 x $10^5$) | Result |
|---|---|---|---|
| a | 41 | 32 | Shell |
| b | 49 | 38 | Shell |
| c | 39 | 31 | Shell |
| d | 16 | 16 | Fully Crosslinked |
| e | 13 | 13 | Fully Crosslinked |

FIGURE S4.10 BATCH PREPARATION OF CROSSLINKED CRYSTALLINE SHELLS
Batch preparation of crosslinked crystals illustrated the sensitivity of shell generation to crystal-to-crystal variation. CJ crystals were crosslinked on a glass depression plate with 0.5% glutaraldehyde in Tadsimate pH 7.5 for 15 minutes and quenched with direct addition of 1 M hydroxylamine. When exposed to the same dose of crosslinking agent, shells can be reproducibly made amongst crystals of similar size (crystals a, b, and c). However, full crosslinking of the smaller crystals (d and e) occurred. In addition to size, shell generation is expected to be sensitive to crystal growth habit. Under typical conditions for the mass transport of the crosslinking agent it will be very difficult to synthesize complete crystalline shells from crystals with high aspect ratios (plates or rods). Thus, to scale up shell production it will be essential to optimize the growth of crystals with uniform size and favorable aspect ratio (i.e. balanced axial and lateral growth to form barrel shaped crystals). Crosslinked volume was estimated by subtracting the volume of the internal cavity (modeled as a cylinder) from the total volume.

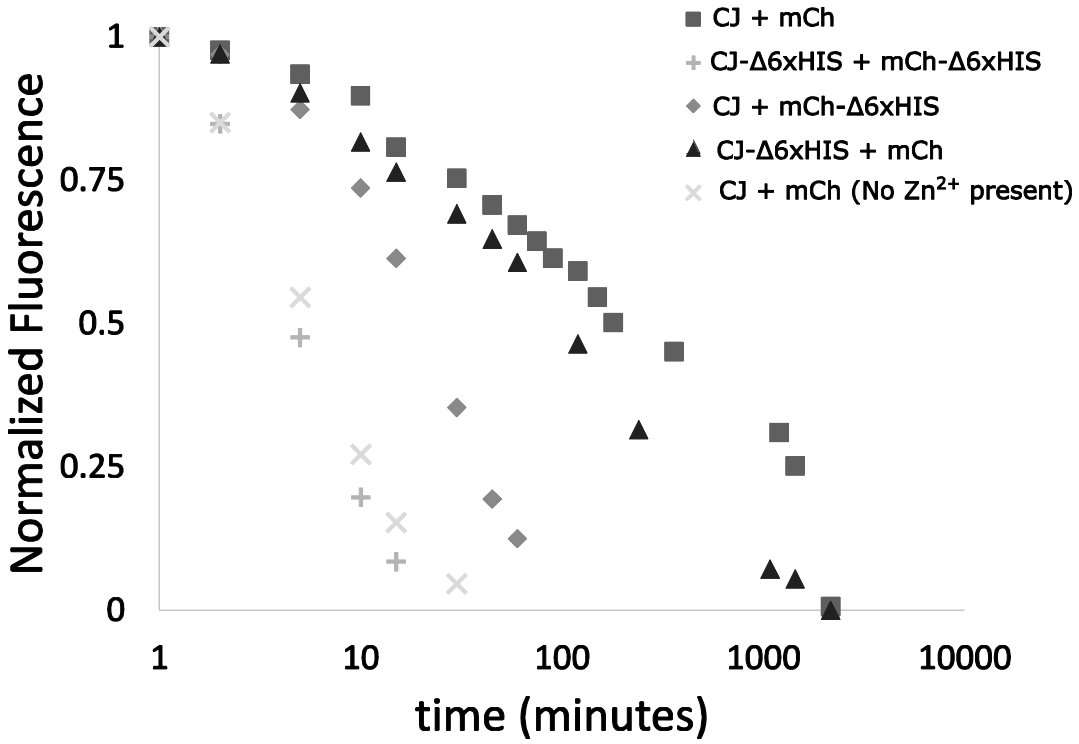



FIGURE S4.11 FLUORESCENCE VS TIME FOR mCHERRY VARIANT RETENTION IN THE PRESENCE OF ZN2+

Each mCherry variant was loaded into the corresponding CJ crystal variant over 10 minutes and transferred to a solution containing Zn2+ (50 mM MES, 300 mM NaCl, 10% glycerol, 10 mM ZnSO4, pH 6.0). The average pixel fluorescence intensity (561 nm excitation) was calculated for the center hexagonal plane. When both CJ and mCherry lack his-tags (CJ-Δ6xHIS + mCherry-Δ6xHIS, + symbols) and in Zn2+, the timescale for mCherry release from the crystal was similar to release in the absence of Zn2+ (x symbols). When the scaffold has a his-tag, but mCherry does not (CJ + mCherry-Δ6xHIS, □□symbols), the release timescale was approximately an order of magnitude longer. This is most likely due to non-specific binding of the scaffold his-tag to mCherry surface groups (i.e. acids or histidine). The alternative control, in which mCherry has a his-tag and the CJ crystal does not not (CJ- Δ6xHIS + mCherry, Δ symbols), is another order of magnitude slower. We attribute the asymmetry to the topology. Scaffold his-tags are present at high concentration (14 mM), but are close enough to dimerize, reducing the likelihood of binding mCherry surface groups present at much lower concentrations. In contrast, a mCherry his-tag in the crystal is surrounded by a high concentration of scaffold surface histidines (3 per CJ monomer, so ~42 mM). When collecting XRD data in the presence of Zn2+, we have observed apparent metal binding to these surface histidines (electron density adducts), indicating high occupancy (Figure S14). Finally, retention is longest in the scenario where both CJ and mCherry have his-tags (CJ + mCherry, □). Future directions will include finding solution conditions that maximizes binding of this scenario versus the other conditions.

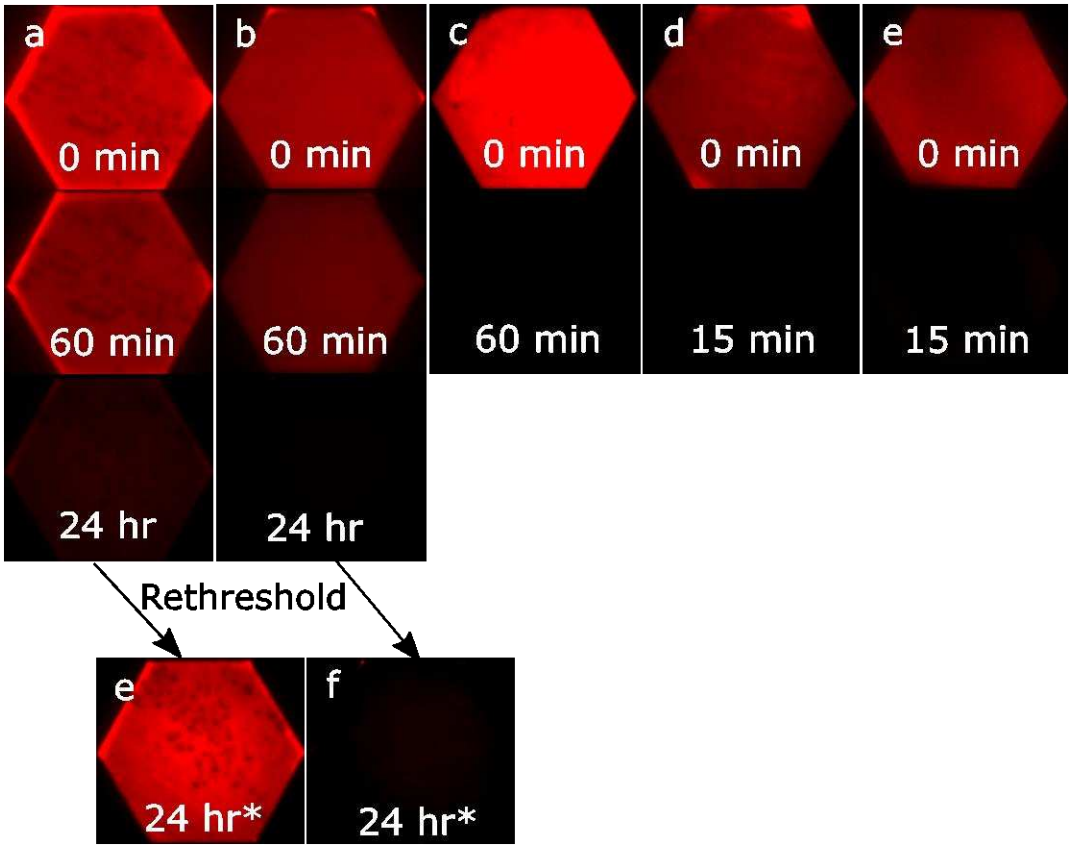



FIGURE S4.12 mCHERRY VARIANT RETENTION IN THE PRESENCE OF ZN2+

Here we demonstrate differential mCherry retention for CJ variants in the presence of Zn2+. Each mCherry variant was loaded into the corresponding CJ crystal variant for 10 minutes and then transferred to a solution containing Zn2+ (50 mM MES, 300 mM NaCl, 10% glycerol, 10 mM ZnSO4, pH 6.0). **(a)** mCherry and CJ, **(b)** mCherry and CJ-Δ6xHIS, **(c)** mCherry-Δ6xHIS and CJ, **(d)** mCherry-Δ6xHIS and CJ-Δ6xHIS, and **(e)** mCherry and CJ in the absence of Zn2+ (passive unloading), are presented at an initial time (0 minutes) and at a time point within 60 minutes. All images were taken with 561 nm excitation at the central hexagonal focal plane. Due to differences in mCherry concentration, crystal path length, host droplet pathlength, and crystal loading, thresholding for the images was selected to best represent the loaded state at 0 minutes. Specifically, the upper threshold was chosen as ~10% higher than maximum pixel value in center of crystal at time 0, while the lower threshold was kept constant through all images. For a more direct visual comparison, threshold values were the same for the (a)/(b) pair as well as (c)/(d). On short time-scales (less than 2 hours) mCherry retention in the presence of Zn2+ was qualitatively similar for host CJ and CJ- Δ6xHIS crystals. After 24 hours, there was a noticeable loss of fluorescence for both CJ and CJ-Δ6xHIS crystals. With a new upper image processing threshold, (lower threshold was fixed, same upper threshold applied to both images), the difference between the long term mCherry retention in the presence of Zn2+ for (e) CJ and (f) CJ- Δ6xHIS was clear. The major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page for all images.

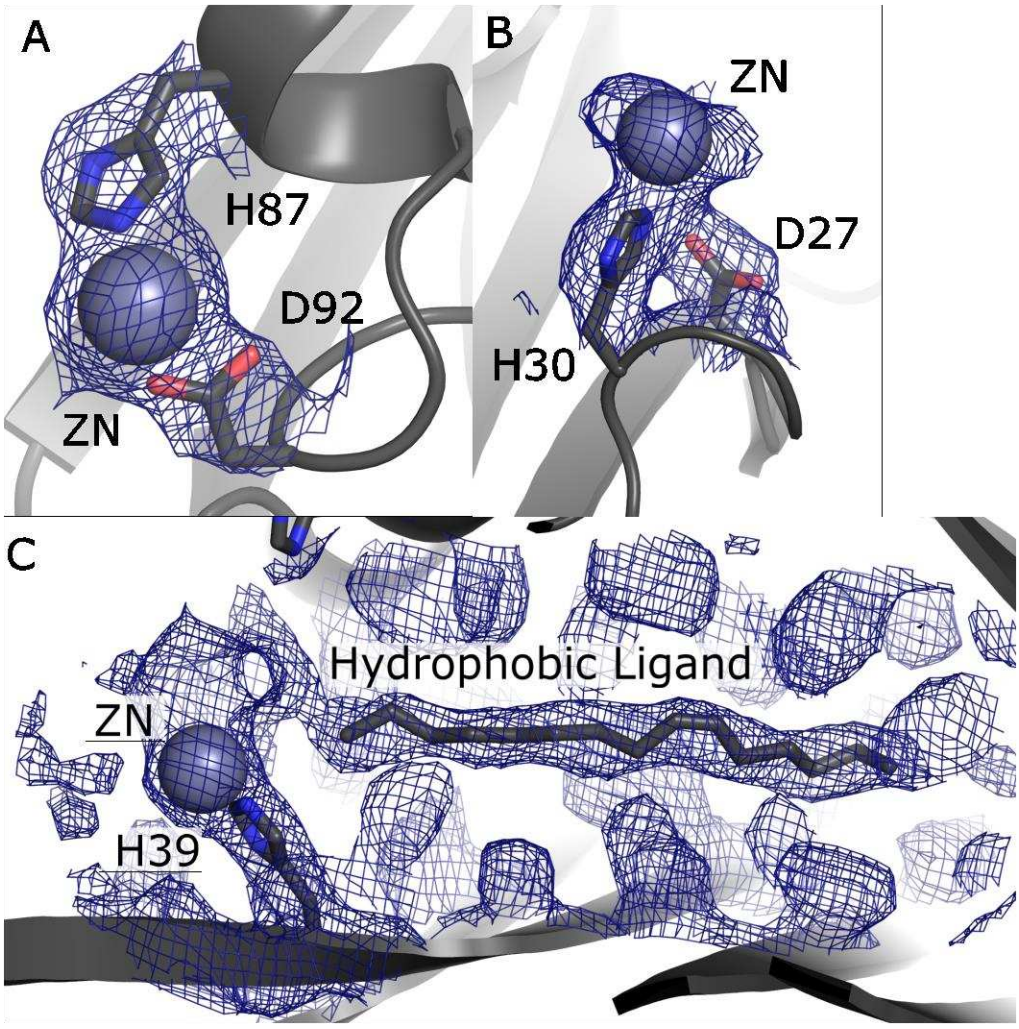



FIGURE S4.13 ELECTRON DENSITY APPARENT FOR SURFACE HISTIDINE ADDUCTS

**(A)** We refined a ZN atom chelated by H87 and D92. Electron density was contoured at 2.0 sigma. **(B)** We refined a ZN atom within the large electron density feature associated with H30. D27 is 4.3 Angstroms from the metal center in this model. The histidine rotamer (and interaction network) is ambiguous; the electron density may reflect a multiply ordered site. Electron density was contoured at 1.0 sigma. Figure prepared in PyMOL (www.pymol.org). **(C)** The third histidine, H39, also appears to be coordinated, but by a large feature in the electron density that is contiguous with density for the hydrophobic ligand carried in the core of the CJ barrel. Since ligand identity is extraneous to the current research, and CJ is being expressed heterologously, we have not identified the ligand. Homology to better known proteins (PDB codes 1Y0G, 3Q34, 2X32) suggests that CJ is likely to be an isoprenoid binding protein. The large electron density feature proximal to H39 might correspond to a pyrophosphate head group. Future efforts (high resolution X-ray diffraction and mass spectroscopy) will help elucidate the identity of the ligand or ligand mixture.

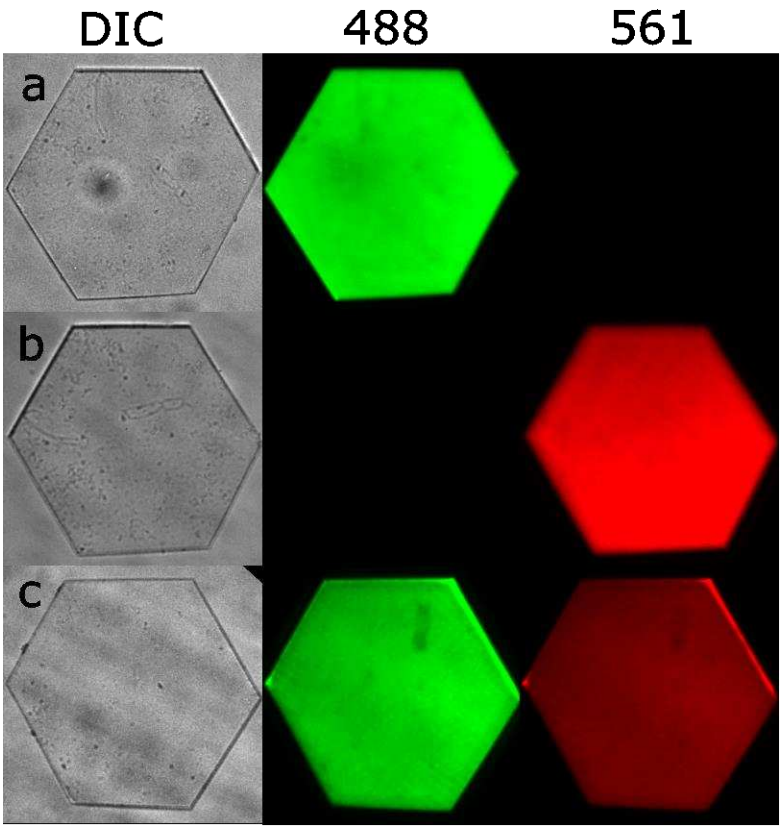



FIGURE S4.14 CO-LOADING mNEONGREEN AND mCHERRY

CJ crystals loaded with mNeonGreen and mCherry were easily distinguished from one another with the excitation at 488 and 561 nm. **(a)** CJ crystal loaded with mNeonGreen only. **(b)** CJ crystal load with mCherry only. **(c)** CJ crystal loaded with mNeonGreen and mCherry simultaneously. The major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page for all images.

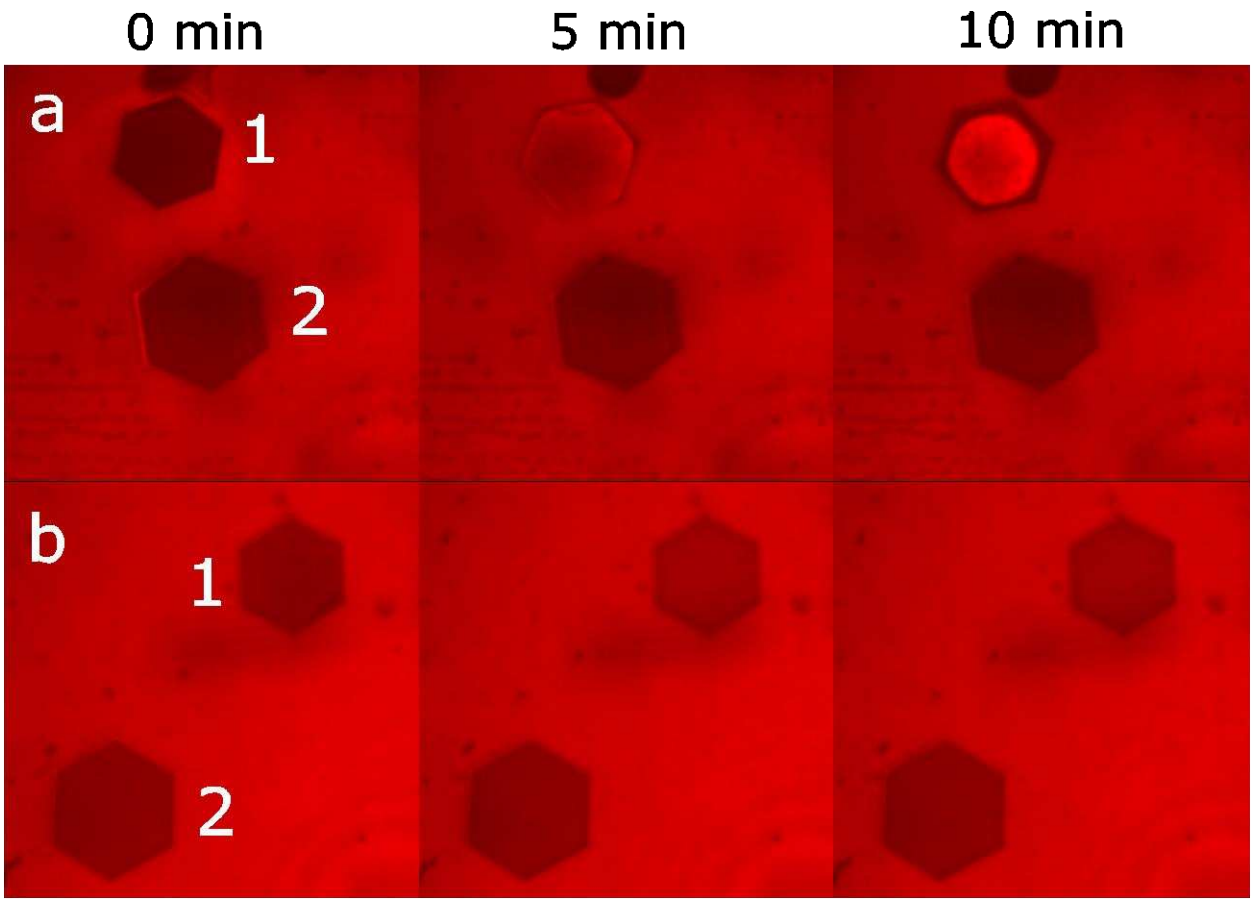



FIGURE S4.15 ACTIVE mCHERRY LOADING IN THE PRESENCE OF ZN2+

In Figure 4a, multi stage loaded crystals demonstrated strong mCherry adsorption in the interior fraction of the crystal (where mNG remained bound with $Zn^{2+}$) relative to the exterior fraction (where titrated EDTA removed $Zn^{2+}$, thus causing mNeonGreen to diffuse out of crystal). This could correspond to "active loading" wherein mCherry adsorbed to the crystal via metal mediated his-tag in the interior, versus passive loading in the exterior. To demonstrate this effect in the absence of mNG, we compared mCherry loading for crystals that were either lacking or precharged with $Zn^{2+}$. (a) A clear mCherry concentrating effect was seen for *(1)* a CJ crystal charged with $Zn^{2+}$ versus *(2)* a CJ crystal prepared without $Zn^{2+}$. Over time (5 min), mCherry was visibly concentrated in crystal 1. After 10 min, mCherry was even more concentrated. We attribute the striking circular diffusion pattern to the slow escape of $Zn^{2+}$ from the scaffold crystal periphery. (b) When loading mCherry-Δ6xHIS, the same concentrating effect was not seen for *(1)* a CJ crystal charged with $Zn^{2+}$ or *(2)* a CJ crystal prepared without $Zn^{2+}$. This data further supports a dominant role for site-specific guest capture via metal-mediated his-tag dimerization. The major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page for all images.

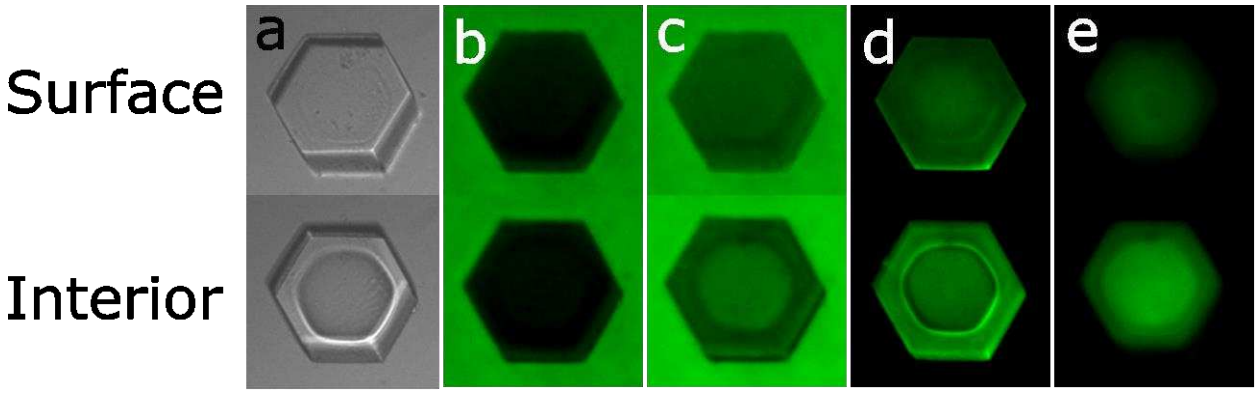



FIGURE S4.16 mNEONGREEN LOADING, RETENTION, AND SELECTIVE RELEASE FROM CJ CRYSTALLINE SHELLS

Confocal images of surface (top) and interior slices (bottom) for a CJ Crystalline Shell. (a) Sample DIC image for surface and interior plane. Other panels show mNeonGreen fluorescence given excitation at 488 nm. Loading mNeonGreen after (b) 0 minutes or (c) 30 minutes. (d) After transfer of the crystal to a solution containing $Zn^{2+}$ (50 *mM* MES, 300 *mM* NaCl, 10% glycerol, 10 *mM* ZnSO4, pH 6.0). (e) mNeonGreen was selectively removed from the exterior layers of the CJ shell by titration with a ~50 *μM* EDTA solution. This image was taken after 5 min of EDTA incubation. For all images, the major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page.

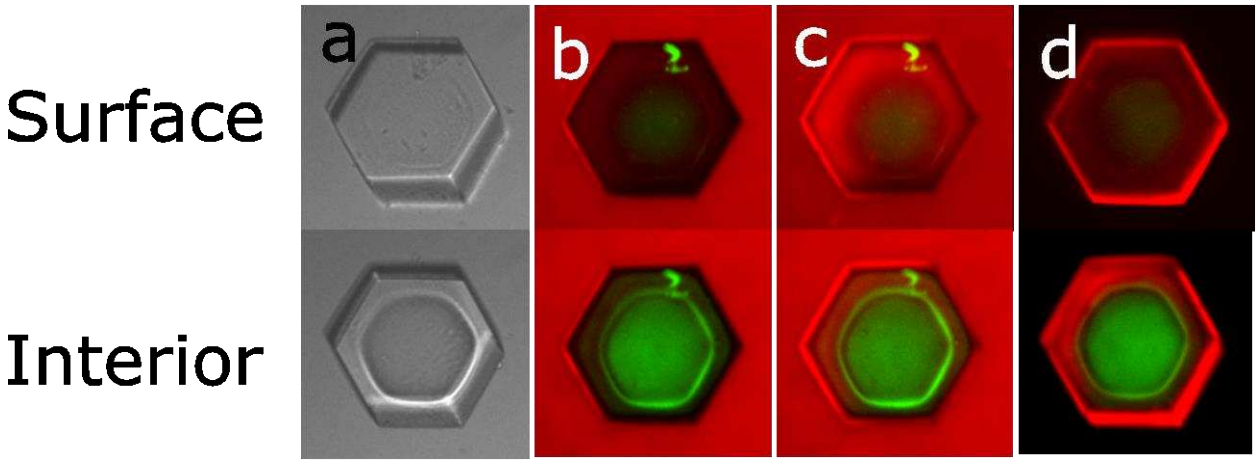



FIGURE S4.17 CJ CRYSTALLINE SHELL SPATIALLY SEGREGATED CO-LOADING OF mNEONGREEN AND mCHERRY

Select confocal images of surface and interior slices of the CJ crystalline shell. (a) Sample DIC image for reference of surface and interior plane. This CJ crystalline shell was loaded with mNeonGreen, transferred to $Zn^{2+}$, and titrated with EDTA to selectively release exterior layers (See SI Figure 17). The crystal was then transferred to a 1 mg/mL mCherry solution. Merged images for 488 and 561 nm excitation illustrate loading of mCherry from (b) 0 minutes and (c) 15 minutes. (d) The resulting co-loaded crystal was transferred to a solution containing $Zn^{2+}$. Using CJ crystalline shells, significant qualitative spatial segregation of guest proteins was achieved. mCherry was predominantly loaded in the exterior, while mNeonGreen mostly remained in the cavity. The major pores of the crystal ($\hat{z}$) are oriented perpendicular to the page for all images.

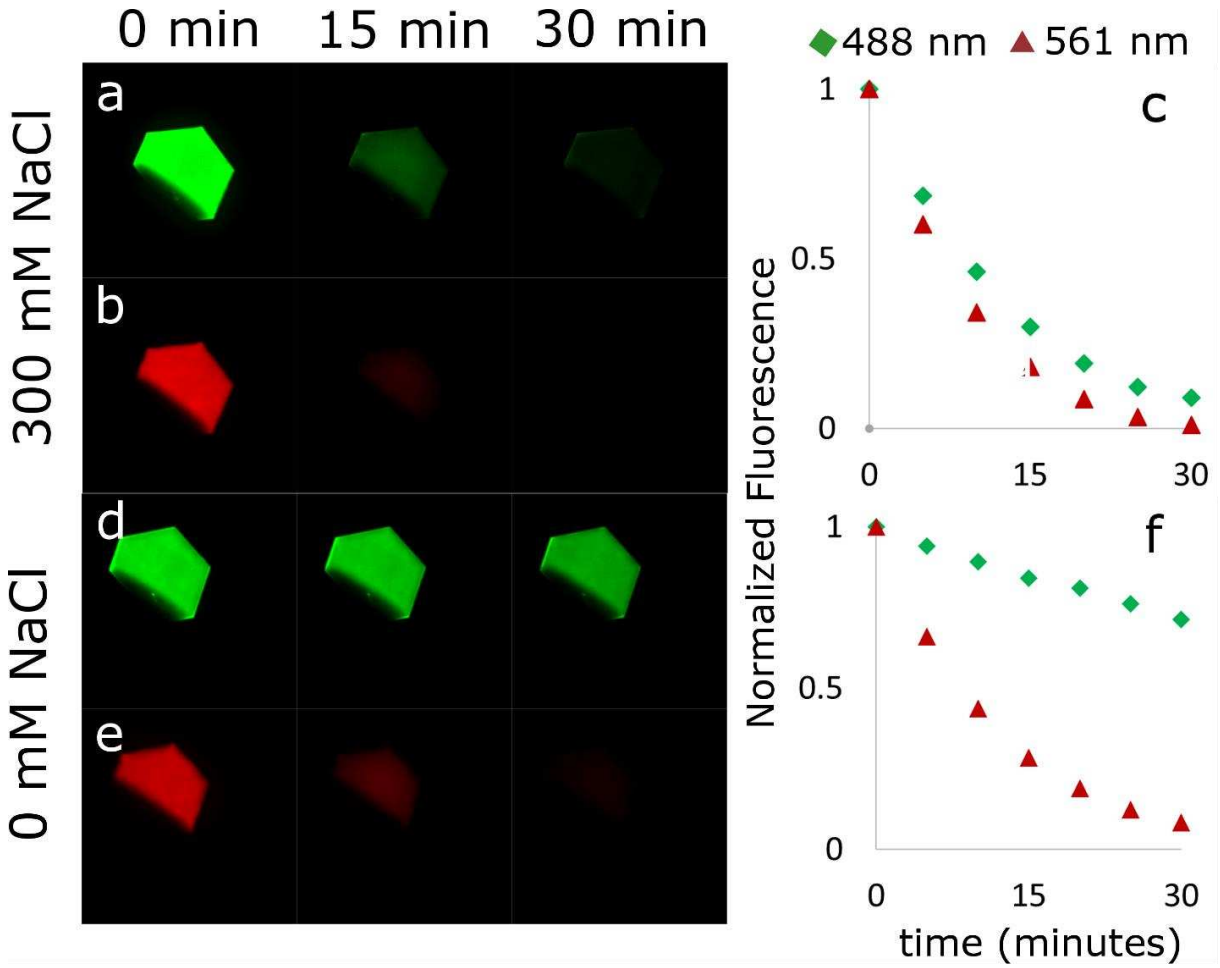



FIGURE S4.18 IONIC BINDING AND RELEASE OF mNEONGREEN AND mCHERRY

Ionic binding effects between host CJ crystals and two guest proteins: mNeonGreen (pI ~7.1) and mCherry-Δ6xHIS (pI ~5.5) were interrogated by co-loading both guest proteins over 10 minutes and unloading in solutions of varying ionic strength. With the goal of removing confounding metal affinity and hydrophobic effects, these experiments were performed in the presence of 1 mM EDTA and 10% glycerol. Similar diffusion release rates for (a) mNeonGreen (488 nm) and (b) mCherry (561 nm) were observed in a solution with moderately high ionic strength (50 mM HEPES, 300 mM NaCl, 1 mM EDTA, 10% Glycerol, pH 7.5). In contrast, release in a low ionic strength solution (50 mM HEPES, 1 mM EDTA, 10% Glycerol, pH 7.5), resulted in a clear separation between (d) mNeonGreen and (e) mCherry. (c & f) To quantify these images, we plot the average fluorescent intensity curves at 488 nm and 561 nm for a crystal interior confocal plane. Again, the varying ionic strength results in a large difference in retention between mNeonGreen and mCherry. The isoelectric point for the two guest proteins was estimated using the ExPASy server[2]. We interpret the differential retention in terms of ionic effects, due to the different net charge expected for the two guests at pH 7.5. We have observed similar apparent ionic effects when performing diffusion experiments for diverse additional guest macromolecules.

## 4.6 ADDITIONAL TABLES

TABLE S4.1 DIFFRACTION ESTIMATE FOR INCUBATED IN mNEONGREEN

| Lower Resolution (Å) | High Resolution (Å) | Average I | Average Error | Average(I/Error) | |
|---|---|---|---|---|---|
| 50 | 9.96 | 78 | 3.4 | 22.9 | |
| 9.96 | 7.92 | 29.1 | 1.9 | 15.3 | |
| 7.92 | 6.92 | 18.7 | 1.8 | 10.4 | |
| 6.92 | 6.29 | 11 | 1.8 | 6.1 | |
| 6.29 | 5.84 | 10.4 | 1.8 | 5.8 | |
| 5.84 | 5.5 | 8.4 | 2 | 4.2 | |
| 5.5 | 5.22 | 7.9 | 2.1 | 3.8 | |
| 5.22 | 4.99 | 13.5 | 2.4 | 5.6 | |
| 4.99 | 4.8 | 9.3 | 2.6 | 3.6 | |
| 4.8 | 4.64 | 11.3 | 2.7 | 4.2 | |
| 4.64 | 4.49 | 12.7 | 2.9 | 4.4 | |
| 4.49 | 4.36 | 9.5 | 3 | 3.2 | |
| 4.36 | 4.25 | 11 | 3 | 3.7 | |
| 4.25 | 4.14 | 9.4 | 3.1 | 3.0 | |
| 4.14 | 4.05 | 7.2 | 3.3 | 2.2 | ← High Resolution Estimate |
| 4.05 | 3.96 | 6.2 | 3.3 | 1.9 | |
| 3.96 | 3.88 | 4.9 | 3.5 | 1.4 | |
| 3.88 | 3.81 | 5.7 | 3.6 | 1.6 | |
| 3.81 | 3.74 | 4.5 | 3.8 | 1.2 | |
| 3.74 | 3.68 | 3.6 | 3.6 | 1.0 | |

Scala output log for 10 frame diffraction check on a CJ crystal incubated in 1 mg $mL^{-1}$ mNeonGreen for 48 hours. Crystal was swished through mother liquor (100% Tadsimate pH 6.5) prior to freezing.

TABLE S4.2 POST-GLUTARALDEHYDE CRYSTAL SHELL DIFFRACTION 4 MINUTES

| Lower Resolution (Å) | High Resolution (Å) | Average I | Average Error | Average(I/Error) | |
|---|---|---|---|---|---|
| 50 | 15.96 | 9.1 | 0.8 | 11.4 | |
| 15.96 | 12.73 | 7.4 | 0.9 | 8.2 | |
| 12.73 | 11.14 | 10.2 | 1.1 | 9.3 | |
| 11.14 | 10.13 | 13.5 | 1.3 | 10.4 | |
| 10.13 | 9.41 | 5.8 | 1.2 | 4.8 | |
| 9.41 | 8.86 | 6.6 | 1.3 | 5.1 | |
| 8.86 | 8.42 | 3.5 | 1.3 | 2.7 | |
| 8.42 | 8.06 | 2.4 | 1.3 | 1.8 | |
| 8.06 | 7.75 | 4.9 | 1.4 | 3.5 | |
| 7.75 | 7.48 | 4.5 | 1.4 | 3.2 | |
| 7.48 | 7.25 | 3.8 | 1.2 | 3.2 | |
| 7.25 | 7.04 | 3 | 1.3 | 2.3 | ← High Resolution Estimate |
| 7.04 | 6.86 | 2 | 1.4 | 1.4 | |
| 6.86 | 6.69 | 1 | 1.4 | 0.7 | |
| 6.69 | 6.54 | 1.3 | 1.3 | 1.0 | |
| 6.54 | 6.4 | 2.1 | 1.5 | 1.4 | |
| 6.4 | 6.27 | 0.8 | 1.5 | 0.5 | |
| 6.27 | 6.15 | 3 | 1.6 | 1.9 | |
| 6.15 | 6.04 | 1.9 | 1.5 | 1.3 | |
| 6.04 | 5.94 | 2.7 | 1.7 | 1.6 | |

Scala output log for 10 frame diffraction check on a CJ crystal shell with 4 minutes of glutaraldehyde exposure (see Supplementary Figure 8). Crystal was swished through mother liquor (100% Tadsimate pH 6.5) prior to freezing. Even with the majority of the crystal volume dissolved, the remaining crystalline exterior diffracted and was indexed correctly as P622.

TABLE S4.3 POST-GLUTARALDEHYDE CRYSTAL SHELL DIFFRACTION 8 MINUTES

| Lower Resolution (Å) | High Resolution (Å) | Average I | Average Error | Average(I/Error) | |
|---|---|---|---|---|---|
| 50 | 10.82 | 39 | 2.1 | 18.6 | |
| 10.82 | 8.6 | 26.1 | 2.4 | 10.9 | |
| 8.6 | 7.52 | 14.9 | 2.3 | 6.5 | |
| 7.52 | 6.84 | 12.3 | 2.4 | 5.1 | |
| 6.84 | 6.35 | 12.4 | 2.9 | 4.3 | |
| 6.35 | 5.97 | 6.7 | 2.8 | 2.4 | |
| 5.97 | 5.67 | 6.3 | 3.1 | 2.0 | |
| 5.67 | 5.43 | 9.8 | 3.4 | 2.9 | |
| 5.43 | 5.22 | 4.1 | 3.4 | 1.2 | |
| 5.22 | 5.04 | 9.5 | 3.8 | 2.5 | |
| 5.04 | 4.88 | 11.1 | 4 | 2.8 | |
| 4.88 | 4.74 | 15.2 | 4 | 3.8 | |
| 4.74 | 4.62 | 13.3 | 4.4 | 3.0 | |
| 4.62 | 4.5 | 19.4 | 4.6 | 4.2 | |
| 4.5 | 4.4 | 15.3 | 4.7 | 3.3 | |
| 4.4 | 4.31 | 15.3 | 4.6 | 3.3 | |
| 4.31 | 4.22 | 13.9 | 4.8 | 2.9 | |
| 4.22 | 4.14 | 11 | 5.1 | 2.2 | ← High Resolution Estimate |
| 4.14 | 4.07 | 8.9 | 5.1 | 1.7 | |
| 4.07 | 4 | 9.5 | 5.2 | 1.8 | |

Scala output log for 10 frame diffraction check on a CJ crystal shell with 8 minutes of glutaraldehyde exposure (see Supplementary Figure 10). Crystal was swished through mother liquor (100% Tadsimate pH 6.5) prior to freezing. Even with a fraction of the crystal volume dissolved, the remaining crystalline exterior diffracted just as well as a fully crosslinked crystal (see Supplementary Table 1 as a comparison).

# 5 SCAFFOLD ASSISTED CRYSTALLOGRAPHY WITH SMALL MOLECULES

## 5.1 INTRODUCTION

Precise position control of functional molecules in 3-dimensions will result in materials with unprecedented performance for diverse applications including biosensing, catalysis, energy conversion, biomedicine, and biotechnology. Researchers have repurposed diverse natural self-assembled architectures including oligomers, fibers[164–166], cages[89,90,167], capsids[168,169], 2-D S-layers[170–172], and protein crystals[173] in pursuit of nanotechnology applications. Protein crystals are highly porous materials and x-ray diffraction (XRD) can elucidate the resulting atomic structure. Thus, we hypothesized that protein crystals could be a favorable platform for scaffold-assisted structure determination.

By soaking small molecules into metal organic frameworks (MOFs), Fujita and coworkers developed the “crystalline sponge” method for host-guest crystallographic structure determination.[75] This method relies on adventitious, non-covalent interactions to adsorb and order guest molecules.[76] We hypothesized that guest molecule installation via a single covalent bond might sufficiently reduce the conformational freedom to provide a feasible alternative approach for scaffold-assisted crystallography. Recent work by Yaghi and coworkers supports this idea, with their successful structure determination of various guest molecules covalently attached in a MOF.[77] Thus, we aimed to engineer unique capture sites for covalent installation of molecules in a protein crystal.

One hypothetical barrier would be a lack of protein crystal plasticity; changes to the constituent monomers could disrupt crystallization or reduce crystal quality.[158,174]

Previous successes in functionalizing protein crystals have relied upon modification of the protein prior to crystallization,[175] which can alter or abrogate crystallization. Even trace labeling protein monomers with fluorophores (<10 mol%) can disrupt crystal nucleation.[176] The approach demonstrated herein decouples crystallization from subsequent modification steps. We first prepare porous scaffold crystal variants that present cysteine proximal to large solvent channels and subsequently install small molecules at these sites. The resulting modified crystalline scaffolds can then be validated using XRD. By performing asynchronous crystallization and covalent small molecule installation, we demonstrate unprecedented control over molecule position in three-dimensional space.

## 5.2 RESULTS AND DISCUSSION

In previous work, we engineered a putative periplasmic polyisoprenoid-binding protein from Campylobacter jejuni (CJ) that crystallized with large (13 nm) solvent channels. We demonstrated that CJ crystals could selectively bind gold nanoparticles or fluorescent proteins.[144,145] Here, solvent exposed residues on the surface of the CJ axial pores were individually mutated to cysteine. These “installation sites” were selected to maximize the inter-site distance between symmetry copies throughout the crystal (**Figure 5.1**). Thus, each installed guest molecule will experience symmetry-equivalent interactions with neighboring amino acids and can independently adopt a preferred conformation.

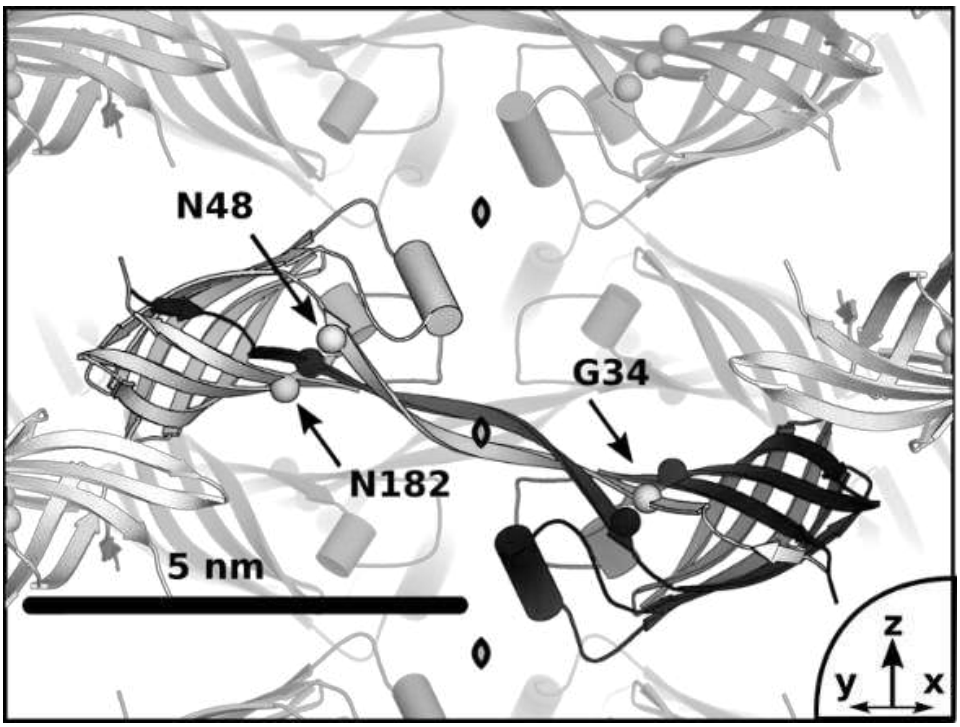



FIGURE 5.1 THIOL MUTATION SITES
Candidate thiol mutation sites (spheres) were selected for maximal inter-site distance and accessibility to the large axial pore (~13 nm). A cleft on the nanopore surface presented three candidate mutation sites, G34C, N48C, N182C. One domain swapped dimer is accentuated (black and white), and 3 proximal 2-fold symmetry axes are marked (curved diamonds).

Chemical conjugation via engineered surface cysteine residues is appealing due to the absence of cysteine residues in wild-type CJ crystals (**Figure S5.1**) and the diverse, established chemistry for thiol conjugation.[151] While cysteine mutations might, in principle, alter or abrogate crystallization, we readily obtained the expected P622 crystals for three variants: G34C, N48C, and N182C (**Figure 5.2b, h, and n** and **Figures S5.2-4** and **Table S5.1**).

Solvent channels within protein crystals are used in classical protein crystallography for diffusion of heavy atoms for protein crystal phasing.[177] To this end, mercury derivatives are useful due the strong propensity of thiols to covalently bind mercury (**Figure S5.5**).[178] To demonstrate engineered thiol accessibility, CJ cysteine mutant crystals were exposed to hydroxymercuribenzoate (MBO), and XRD was subsequently performed. Obvious

electron density features were observed for all three cysteine-variants (Figure 5.2c, i, and o). MBO ligand was fit with 85% occupancy for N48C and 90% occupancy for G34C. The electron density suggested clear directionality for the benzoic acid group for the N48C adduct, but not for G34C. N182C had an extended electron density feature suggesting multiple states and was fit with two 50%-occupancy states (Figure 5.2o).

Free thiols can be oxidized to mixed disulfides via disulfide exchange reactions (**Figure S5.6**).[151] Ellman's reagent, 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB), a compound useful for the quantitative determination of sulfhydryls, was selected as a target chromogenic reagent to demonstrate disulfide exchange on the crystal.[179–181] Addition of DTNB to CJ variants in solution proceeded to near completion as confirmed by absorbance at 412 nm (**Figure S5.7**). When CJ cysteine mutant crystals were exposed to DTNB, the crystals temporarily acquired a faint yellow hue before the reaction product 5-thio-2-nitrobenzoic acid ($TNB^{-2}$) could diffuse out of the crystals. After washing the crystals to remove excess DTNB and $TNB^{-2}$, an intense yellow dye release could be triggered for all three cysteine variant crystals by adding 20 µL of 10 mM 2-mercaptoethanol (BME) (**Figure S5.8**). XRD diffraction elucidated a single conformation for installed 5-mercapto-2-nitro-benzoic acid (MNB) ligand on all three cysteine variants at 100% occupancy (Figures 5.2d, j, and p).

Thiols can also undergo rapid interchange reactions with diselenide compounds[182,183] to form mixed thiol-selenide compounds (**Figure S5.9**). After incubation with selenocystine, XRD revealed modifications on all three cysteine variants indicative of selenocysteine (SEC) installation (Figures 5.2e, k, and q). SEC ligand could be fit to two conformations at 50% occupancy for G34C and a single 90%-occupancy conformation for N48C.

Beyond the heavy selenium atom, the rest of the conjugate was not obvious for both variants. While the S-Se bond formation appears to have a preferred geometry (±90° dihedral), the rest of the conjugate has free rotatable bonds leading to disorder. SEC installation at N182C better resolved the conjugate conformation, with a single state at 100% occupancy (Figure 5.2q).

Haloalkyl derivatives are some of the most widely used conjugates for modifying thiols. The nucleophilic thiolate of the protein reacts with the halogenated alkyl via an SN2 reaction forming a stable thioether linkage (**Figure S5.10**). Due to its fluorogenic properties, monobromobimane (mBBr) ($\lambda_{exc}/\lambda_{em}$=394/490), was selected as the target reagent for demonstrating halo-alkyl installation.[184] mBBr is essentially non-fluorescent until conjugated to a thiol. CJ cysteine-bearing crystals were exposed to mBBr and the fluorescence was monitored via confocal microscopy. Only crystals containing thiol mutations were fluorescent above background (**Figure S5.11**). XRD was performed on the resulting crystals and installation was observed on N48C and N182C. Bimane adducts (MBB) were fit to single conformations for N48C and N182C at 100% and 90% occupancy respectively (Figure 5.2l and r).

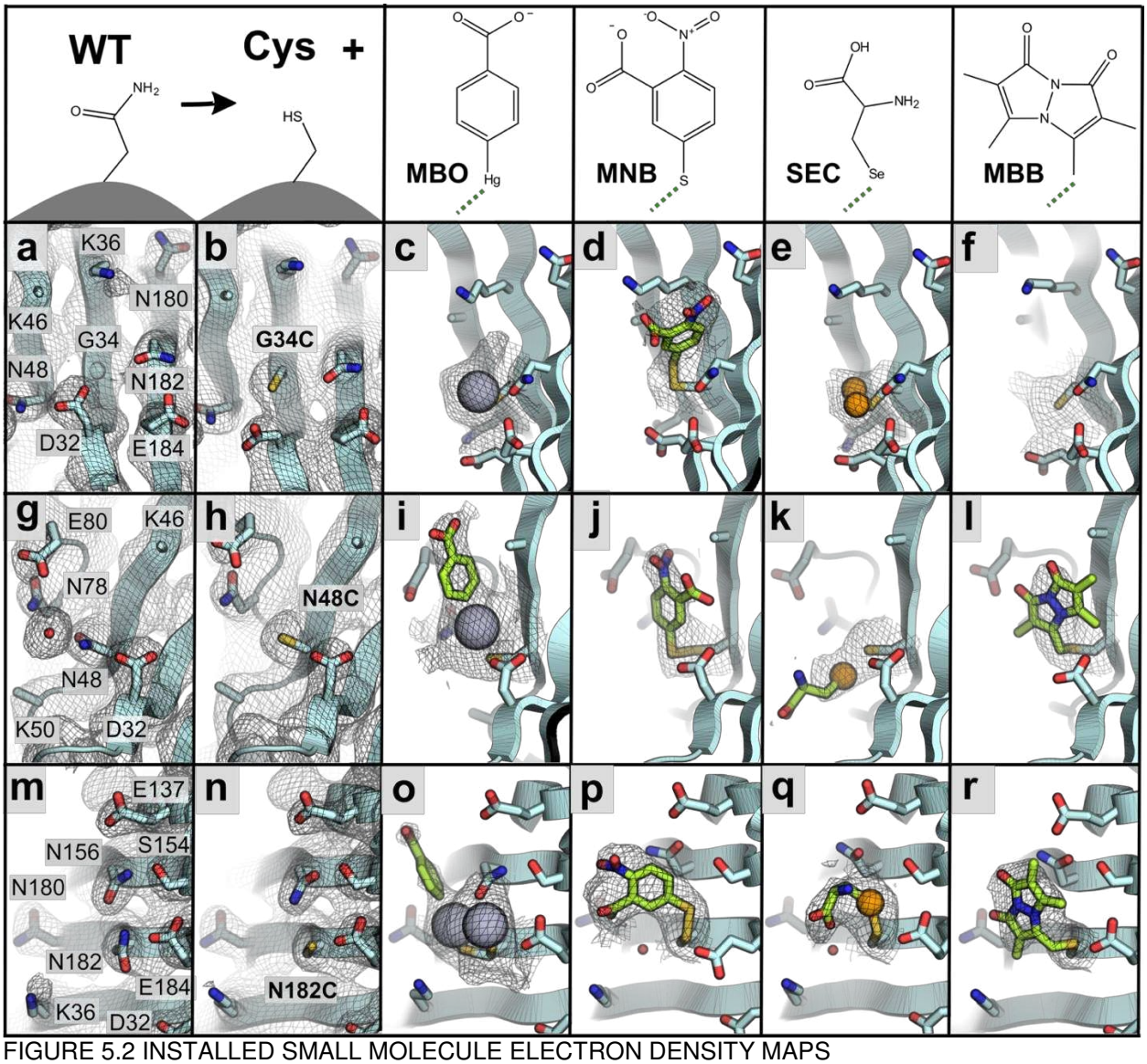



FIGURE 5.2 INSTALLED SMALL MOLECULE ELECTRON DENSITY MAPS
2mFo-DFc maps contoured to 0.8 σ for **(a)** CJ wild-type (G34 alpha carbon marked with a sphere) and **(b)** G34C. After installation at G34C: **(c)** One 90%-occupancy mercury position for 2-hydroxymercuribenzoic acid (MBO) **(d)** One 100%-occupancy conformation for 5-mercapto-2-nitro-benzoic acid (MNB). **(e)** For selenocysteine (SEC), two 50%-occupancy selenium positions **(f)** The electron density was insufficient to place the bimane adduct (MNB); **(g)** CJ wild-type centered at N48 and **(h)** N48C. After installation at N48C: **(i)** One 85%-occupancy conformation for MBO. **(j)** One 100%-occupancy conformation for MNB. **(k)** One 90%-occupancy conformation for SEC was modeled, though part of the flexible adduct was not resolved. **(l)** One 100%-occupancy conformation for MBB; **(m)** CJ wild-type centered at N182 and **(n)** N182C. After installation at N182C: **(o)** Two 50%-occupancy conformations for MBO. **(p)** One 100%-occupancy conformation for MNB. **(q)** One fully resolved 100%-occupancy conformation for SEC. **(r)** One 90%-occupancy conformation for MBB.

We have confirmed multiple types of post-crystallization conjugation chemistry at multiple sites on a protein. The XRD results are summarized in **Table 5.1** (for full x-ray statistics see **Table S5.1-S5.5** and **Figure S5.14**). Reaction yields sufficient for high occupancy

installation were achieved despite the high salt environment required for preserving protein crystal integrity. To validate the standard $2mF_o$-$DF_c$ maps (Figure 5.2) we compared them (**Figures S5.15**, **S5.16**, and **S5.17**) to discovery maps,[185] FEM maps,[186] Polder omit maps,[187] and classical simulated annealing omit maps.[188] Real space correlation coefficient values from the $2mF_o$-$DF_c$ maps were calculated for each installed guest molecule using Phenix.[189]

Of the attempted structure determination targets, only one cysteine:adduct pair (mBBr installation at G34C) has yet to yield interpretable electron density for the guest molecule. Notably, G34C crystals became highly fluorescent when incubated with mBBr. We therefore speculate that installation succeeded, but the resulting adduct was too mobile to be clearly resolved in the electron density map. It is not uncommon for surface sidechain disorder to lead to ambiguous or absent electron density contours (e.g. Lys46 and Lys50 in Figure 5.2).

Subtle conformational changes on the scaffold surface were observed upon small molecule installation. The least subtle example was N156. The apparent native hydrogen bond network (Figure 5.2m) was disrupted upon mutating N182C. For the N182C model, we placed N156 to accept a hydrogen bond from S154 (Figure 5.2n). N156 moved again when small molecules were installed at the N182C site, to avoid steric clashes (Figures 5.2p, q, and r) or to make favorable interactions with the adduct (Figure 5.2o).

TABLE 5.1 SUMMARY OF SMALL MOLECULE GUEST INSTALLATION RESULTS AND DEPOSITED STRUCTURES

| | | MBO | MNB | SEC | MBB |
|---|---|---|---|---|---|
| G34C | Number of States | 1 | 1 | 2 | - |
| | Occupancy | 90% | 100% | 50%/50% | |
| | Real Space cc | 0.89 | 0.75 | 0.93/0.92 | - |
| | Notes | Hg atom only | Fully resolved | Se atom only | Small peak unresolved |
| | PDB Code | 5W2K | 5W2R | 5W2V | - |
| N48C | Number of States | 1 | 1 | 1 | 1 |
| | Occupancy | 85% | 100% | 90% | 100% |
| | Real Space cc | 0.99 | 0.85 | 0.81 | 0.90 |
| | Notes | Fully Resolved | Fully resolved | Se and alpha carbon only | Fully Resolved |
| | PDB Code | 5W31 | 5W2Z | 5W32 | 5W30 |
| N182C | Number of States | 2 | 1 | 1 | 1 |
| | Occupancy | 50%/50% | 100% | 100% | 90% |
| | Real Space cc | 0.98/0.86 | 0.90 | 0.95 | 0.87 |
| | Notes | Multiple states, complex | Fully resolved | Fully resolved | Fully Resolved |
| | PDB Code | 5W3B | 5W3A | 5W3C | 5W39 |

*Real space cc calculated from 2mFo-DFc map using Phenix.

As is typical for protein crystals, many surface sidechains were too disordered to be easily placed in electron density (15 out of 180 sidechains were truncated or unresolved). *A priori*, it was not clear if small molecule adducts would be likely to adopt ordered conformations, especially considering the absence of any design or selection pressure. Indeed, adducts were quite flexible, leading to high B-factors (computed for all adduct atoms). However, apart from MBO and SEC adducts at G34C, B-factors were within the range (Z score < 3) observed for other non-truncated sidechains (**Figure S5.19-S22** and **Table S5.7**). The high real space correlation coefficients for the modeled guest molecules (**Table 5.1**) and our omit map analysis (**Figures S5.15**, **S5.16**, and **S5.17**) suggested that the molecules were correctly modeled despite high B-factors. In theory, we could use anomalous diffraction to gather more information on the position of selenium and mercury atoms. However, anomalous scattering would be more useful if the atoms in question are not directly attached to the cysteine and are not already evident due to dramatic increases

in the electron density. Additionally, our long-term goal is to determine how and when guest molecules can adopt coherent structure without the aid of anomalous scattering sites.

This study supports our original supposition that limiting conformational flexibility is pivotal to resolving guest molecules via scaffold-assisted crystallography. We only observed guest molecules at the intended covalent installation sites. Additionally, the most readily resolved guest small molecule was MNB. It seems likely that the observed coherent MNB conformations were adopted due to the preferred geometry of the disulfide dihedral (±90°) and the rigidity of the subsequent planar ring structure. In contrast, molecules with multiple rotatable bonds such as SEC more often yielded poorly resolved structures past the initial attachment point. We are currently investigating the use of chemical crosslinking to stabilize the host crystals, thereby enabling diffraction under widely varying solution conditions and cryoprotectants. Varying the solution conditions may ultimately resolve multiple coherent guest conformations.

While the modest resolution of CJ crystals (>2.4 Å) is not ideal for high-resolution structure determination of installed small molecules, a major long-term advantage of developing the CJ crystal platform is the promise of scalability to large guest molecules. The techniques developed herein could be adapted to protein crystals with higher resolution, which might result in more detailed adduct structures. However, increased crystal resolution will not necessarily improve adduct detail. Ueno, studying myoglobin crystal adducts (installed prior to crystallization) found little interpretable density despite the superior resolution of myoglobin crystals (~1.5 Å ).[175]

More likely, the current work suggests that variation of installation sites, optimization of neighboring amino acids, and perhaps provision of strong secondary anchoring interactions, will be key to fully realizing guest molecule structure determination. In this study, installation sites were purposely selected to be highly exposed and proximal to the 13 nm axial solvent channels. Despite a lack of designed secondary interactions, we were able to model adducts at high occupancy and high B-factor (Table S5.1-S5.5). Clear patterns emerged, in that one site (N182C) led to more coherent adduct structures than another site (G34C). Future small molecule adduct structure determination may be improved if installation sites are partially buried in a surface pocket. Less accessible installation sites might reduce adduct flexibility, though perhaps conjugation efficiency could suffer. Additionally, engineering the environment near the installation sites may increase success in determining coherent structures. For example, mutagenesis of neighboring amino acids to hydrophobic side chains might promote favorable interactions with some guest molecules. Ultimately, the plasticity of the protein crystal makes this system an evolvable platform for scaffold-assisted crystallography.

## 5.3 CONCLUSIONS

We have demonstrated that several established thiol conjugation strategies are suitable for installing small molecules upon engineered cysteines in a pre-existing three-dimensional protein crystal. This strategy enables diverse nanotechnological applications. The visibility of the resulting small molecule conjugates is promising for advancing techniques in scaffold-assisted crystallographic structure determination. The conjugation strategies demonstrated here could be adapted to conjugate small molecules of unknown structure to alternate protein crystal scaffolds with superior resolution and

engineered local environment to promote favorable guest-scaffold contacts. Alternately, in contrast to the MOFs currently used for guest structure determination,[75,77] the 13-nm pores of CJ crystals used here are large enough to accommodate macromolecules such as proteins, inorganic nanoparticles, and DNA. The methods developed herein lay the groundwork for site-specific installation of macromolecules and structure determination of the resulting co-crystals.

## 5.4 MATERIALS AND METHODS

### 5.4.1 CJ PROTEIN CRYSTAL PREPARATION

A codon optimized gene encoding a putative periplasmic protein (Genebank ID: cj0420) from *Campylobacter jejuni* was obtained from Life Technologies and cloned into pSB3 vector at NdeI and XhoI. For cytosolic expression, the gene was truncated to remove the signaling peptide. Thiol variants were generated via single primer mutagenesis with Q5 polymerase (New England Biolabs) and sequenced verified. All variants were expressed in *E. coli* C41 (DE3) (Lucigen) grown in Terrific Broth and induced with 0.4 mM IPTG at 25 °C for 16 hr. The cells were harvested and sonicated into a lysis buffer (50 mM HEPES, 500 mM NaCl, 10% glycerol, 25 mM imidazole, pH 7.4). The lysate was clarified and purified via $Ni^{2+}$-NTA chromotagraphy (Thermo Fisher Scientific HisPur™ Ni-NTA). A single chromatography step provided sufficient purity for crystallization. The purified protein was dialyzed into a storage buffer (10 mM HEPES, 500 mM $(NH_4)_2SO_4$, 10% glycerol at pH 7.4), aliquoted, and stored at -20 $^oC$. The final concentration was ~20 mg/mL with an average CJ yield of >100 mg per 1 L culture. CJ variants were crystallized overnight by sitting drop vapor diffusion at 20 °C in >3.0 M $(NH_4)_2SO4$, 0.1 M Bis-Tris pH 6.0.

Prior to installation, crystals were washed via transfer to the installation solution (3.4 M $(NH_4)_2SO4$, 100 mM HEPES, pH 7.5) for 15 min to equilibrate the crystals and remove excess free protein. Crystals were then transferred to 20 µL of the installation solution with 500 µM of the molecule to be conjugated and incubated for 2 hours.

### 5.4.2 X-RAY DIFFRACTION AND DATA PROCESSING

In all cases, individual crystals were briefly swished through a cryoprotectant solution containing 3.2 M $(NH_4)_2SO4$ and either 10% glycerol or 10% ethylene glycol at pH 7.5 and flash frozen in liquid nitrogen. X-ray diffraction data was collected on beamline 4.2.2 at the Advanced Light Source (ALS) or on a local Rigaku Compact HomeLab with a microfocus X-ray generator and a Pilatus 200K detector. The collected data was processed with XDS.[190] The wild-type structure was determined by molecular replacement (MR) with the Campylobacter jejuni putative periplasmic protein (PDB entry 2fgs) as a search model. Model refinement was performed in COOT using sigma weighted ($2mF_o$-$DF_c$) and ($mF_o$-$DF_c$) electron density maps and REFMAC5 from the CCP4 suite.[191–193] The resulting wild-type model was used as the starting MR model for G34C, N48C, and N182C with the same refinement scheme. Each cysteine variant model was then used as a MR search model for their corresponding small molecule adducts. The molecular refinement workflow is summarized in **Figure S5.12**. For each thiol structure with an installed small molecule, a scheme of discovery map generation, ligand building, refinement, and omit map generation was implemented to reduce model bias. The model building scheme is summarized in **Figure S5.13**.

## 5.5 ADDITIONAL FIGURES

MKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRN

HLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKEI

GFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

FIGURE S5.1 AMINO ACID SEQUENCE FOR CJ WITHOUT CYSTEINE MUTATION
The leader peptide on the native putative periplasmic polyisoprenoid-binding protein from *Campylobacter jejuni* was deleted to ease expression and purification.

***G34C***

MKEYTLDKAHTDV**C**FKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

KIGFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

FIGURE S5.2 AMINO ACID SEQUENCE FOR CJ-G34C.

***N48C***

MKEYTLDKAHTDVGFKIKHLQISNVKG**C**FKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

KIGFSLNGKIKRSDFKFATSTSTITLSDDINLNIEVKANEKEGGSHHHHHH

FIGURE S5.3 AMINO ACID SEQUENCE FOR CJ-N48C.

MKEYTLDKAHTDVGFKIKHLQISNVKGNFKDYSAVIDFDPASAEFKKLDVTIKIASVNTENQTRD

NHLQQDDFFKAKKYPDMTFTMKKYEKIDNEKGKMTGTLTIAGVSKDIVLDAEIGGVAKGKDGKE

***N182C***

KIGFSLNGKIKRSDFKFATSTSTITLSDDINL**C**IEVKANEKEGGSHHHHHH

FIGURE S5.4 AMINO ACID SEQUENCE FOR CJ-N182C.

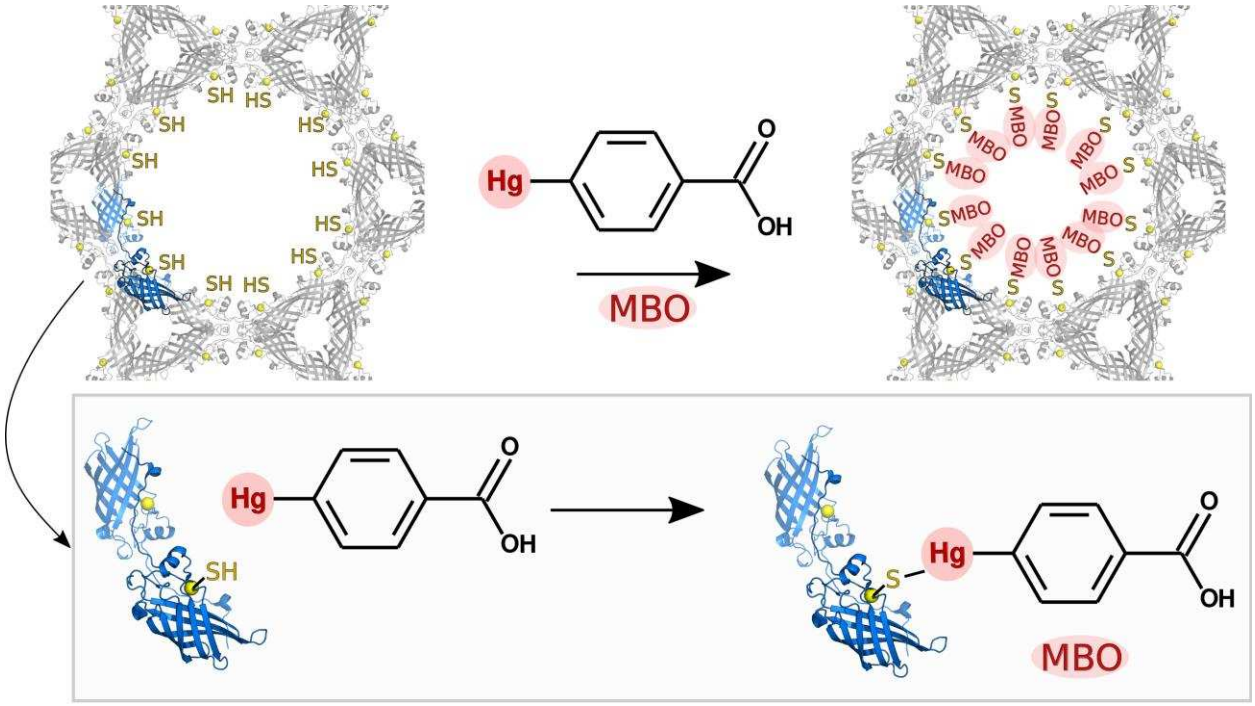



FIGURE S5.5 REACTION SCHEME FOR MERCURIBENZOIC ACID

Due to the strong propensity of thiols to covalently bind mercury, mercuribenzoic acid (MBO) was used to demonstrate the accessibility of the engineered cysteines to the solvent channels. The resulting heavy atom adducts were easy to observe via XRD.

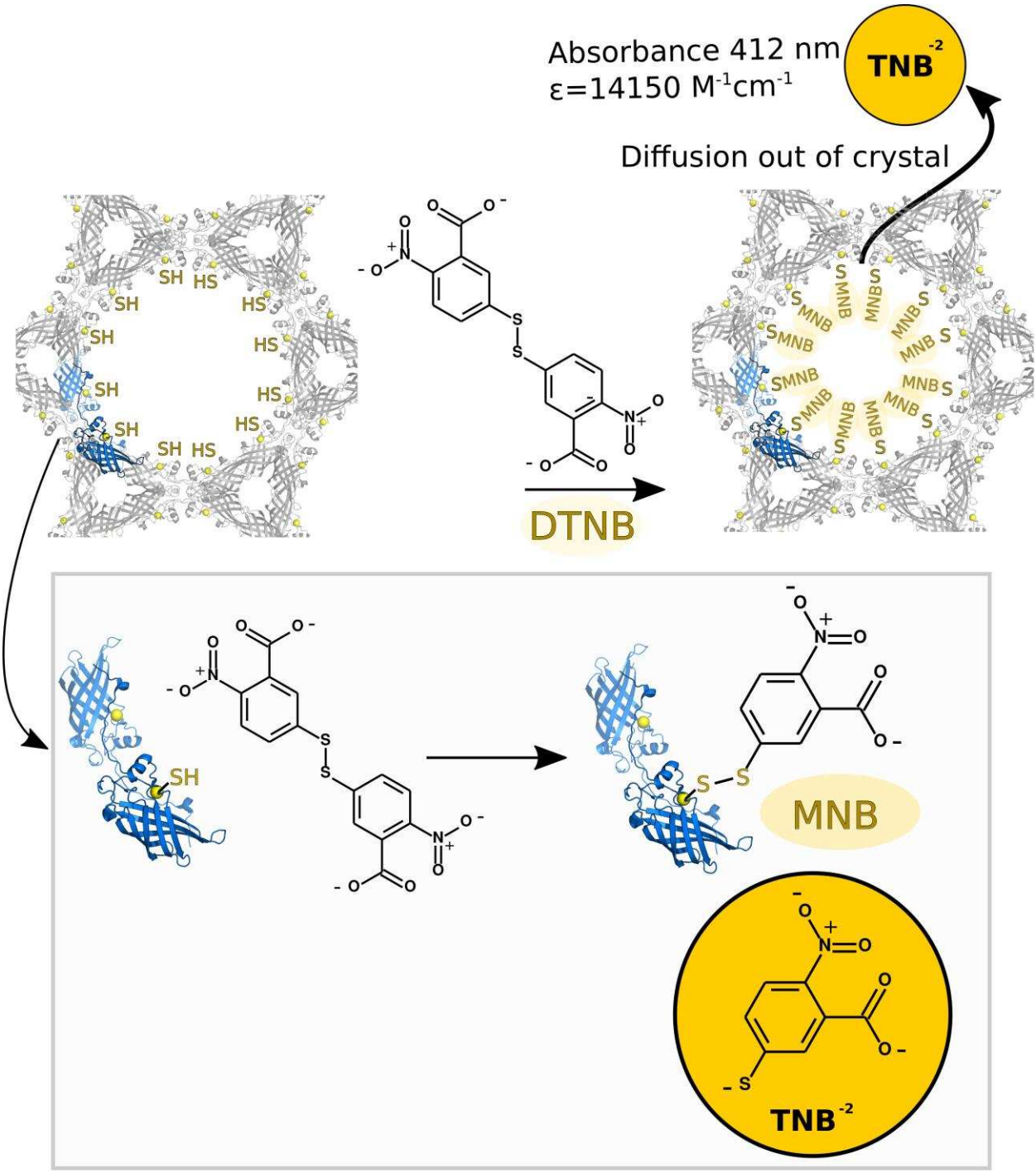



FIGURE S5.6 REACTION SCHEME FOR ELLMANS REAGENT

Ellman's reagent (5,5'-dithio-bis-[2-nitrobenzoic acid]) reacts with thiols via an $SN_2$ reaction that forms a mixed disulfide product with the addition of 5-mercapto-2-nitro-benzoic acid (MNB) to reduced thiols. The reaction can be monitored by measuring the release of 2-nitro-5-thiobenzoate anion ($TNB^{-2}$). which absorbs strongly at 412 nm due to the resonance stabilized 4-thiolpyridone tautomer. These properties made Ellman's reagent an attractive choice for demonstrating disulfide exchange in CJ cysteine mutant crystals.

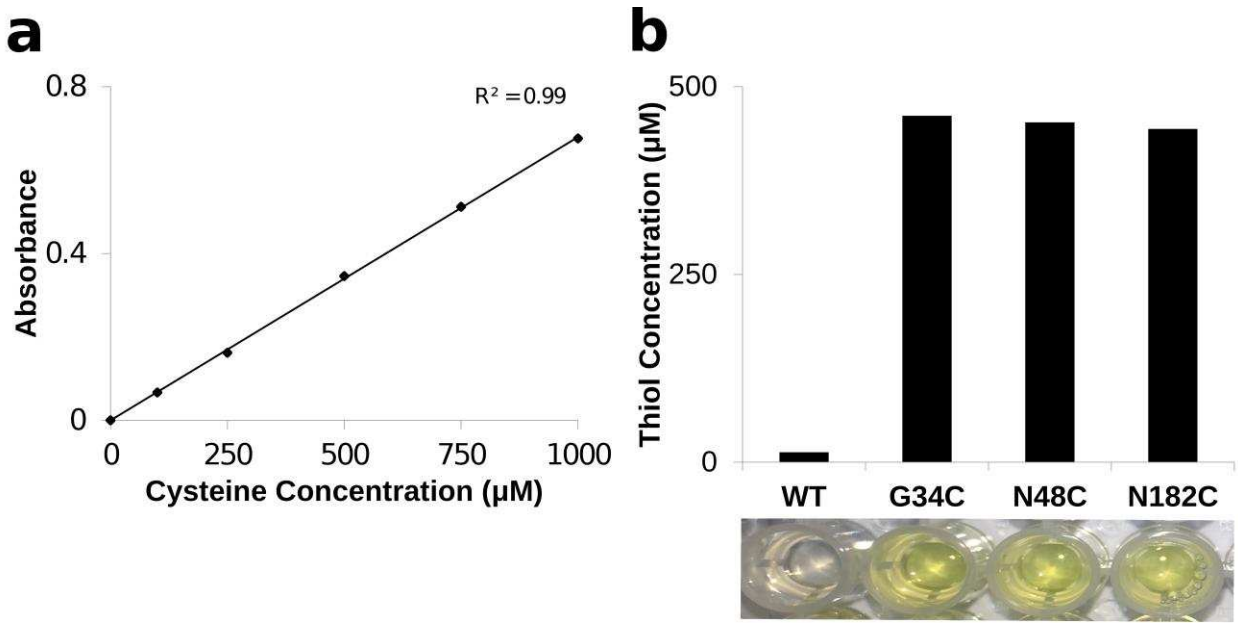



FIGURE S5.7 IN VITRO ELLMANS REAGENT ASSAY

Thiol concentration in solution can be measured by addition of Ellman's reagent and measuring absorbance 412 nm. **(a)** An *in vitro* Ellman's Reagent standard curve was prepared for reduced L-cysteine from 0-1000 µM. **(b)** Purified CJ-variants were diluted to ~10 mg/mL (~500 µM) and Ellman's reagent was added to confirm the presence and accessibility of thiols in solution. Only CJ-variants with engineered cysteines produced a signal at 412 nm.

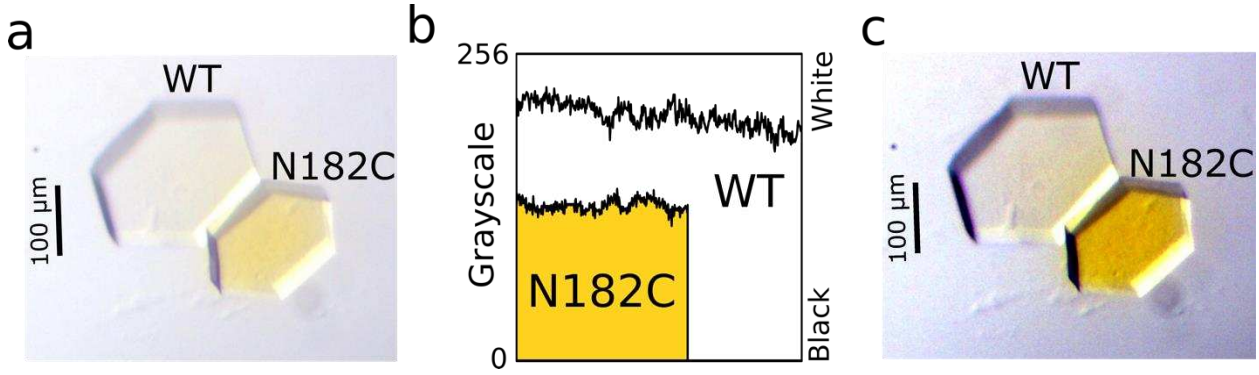



FIGURE S5.8 IN CRYSTALLO MNB CLEAVAGE AND SUBUSEQUENT TNB RELEASE

This figure is best observed in the color images available online. Wild-type and CJ cysteine mutant crystals were exposed to Ellman's reagent and extensively washed to remove unreacted Ellman's reagent. **(a)** The raw image above was obtained later, 30 seconds after adding 10 mM 2-mercaptoethanol (BME). Addition of BME produced an intense yellow signal only on CJ cysteine mutant crystals, suggesting release of covalently bound 5-mercapto-2-nitro-benzoic acid (MNB). **(b)** To quantify the difference, we use ImageJ to extract grayscale values across diagonal vectors that span each crystal (avoiding the edge effects). **(c)** The pathlength of these crystals is small, which results in a somewhat subtle signal. Here, a contrast-enhanced image assists with visualizing the difference between the crystals and better represents what is observed by eye. The contrast between the cysteine-bearing crystal and the WT crystal would likely be even larger if the two were not adjacent, since the TNB dye molecule released by the N182C crystal can diffuse into the WT crystal.

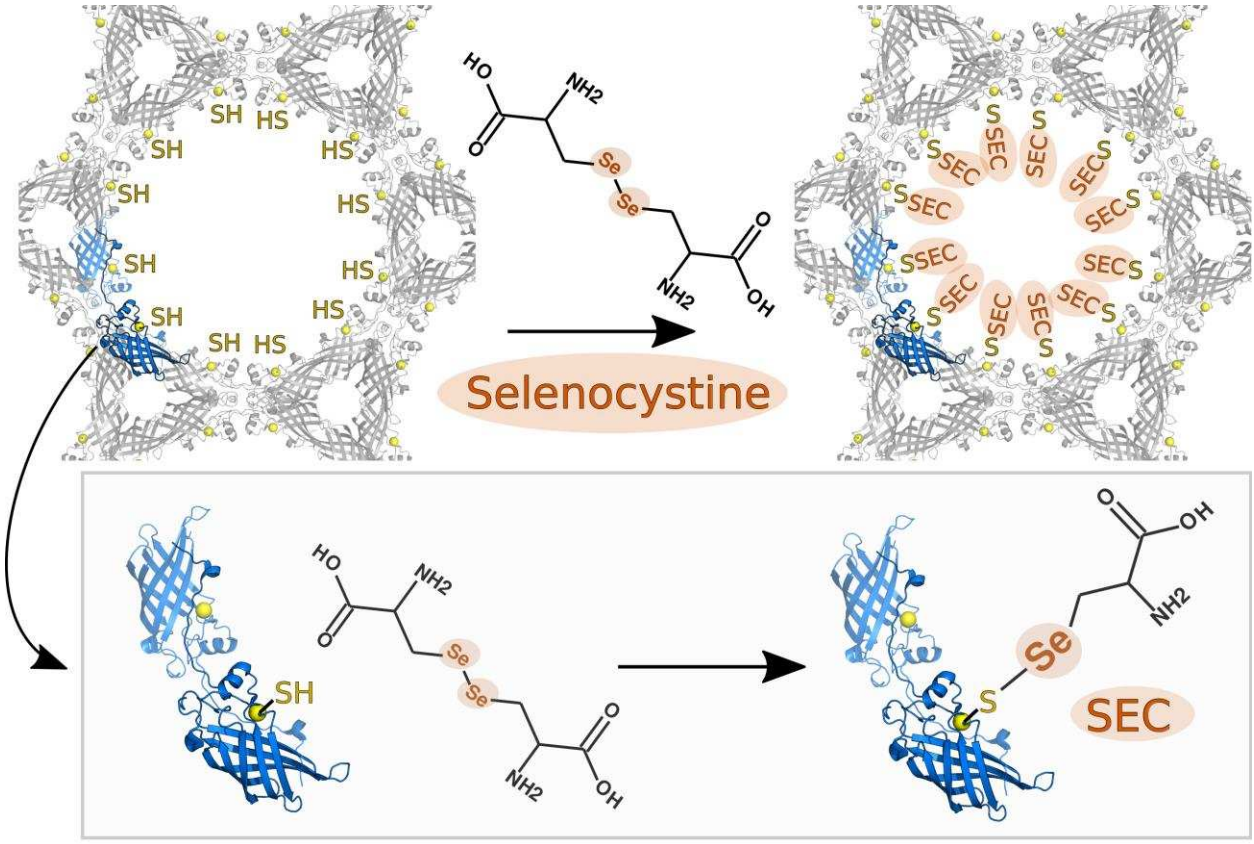



FIGURE S5.9 REACTION SCHEMATIC FOR SELENOCYSTINE

In a reaction analogous to disulfide exchange, thiols can form mixed oxidized products with diselenide compounds. The reaction of thiols with diselenides has the benefit of adding a heavy atom at the attachment point. Selenium addition was readily visible in XRD. However, the remainder of the adduct was too flexible to be clearly resolved in 2 out of 3 cases.

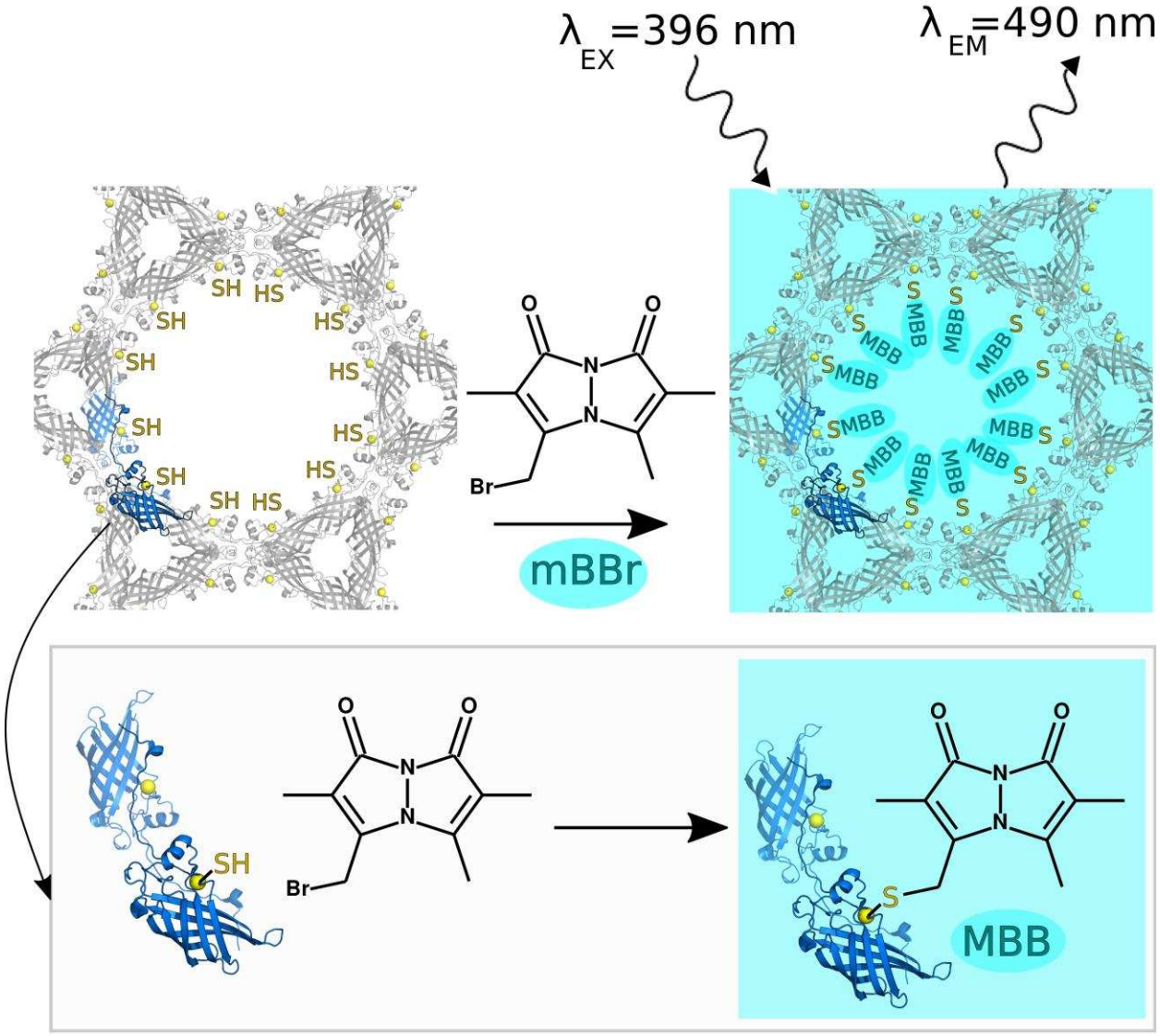



FIGURE S5.10 REACTION SHEMATIC FOR MONOBROMOBIMANE

Haloalkyl groups react with thiols via an SN2 reaction resulting in a stable thioether linkage. We used monobromobimane (mBBr) to demonstrate the CJ cysteine crystals ability to react with haloalkyl derivatives. mBBr is essentially non-fluorescent until forming the resulting bimane (MBB) conjugate.

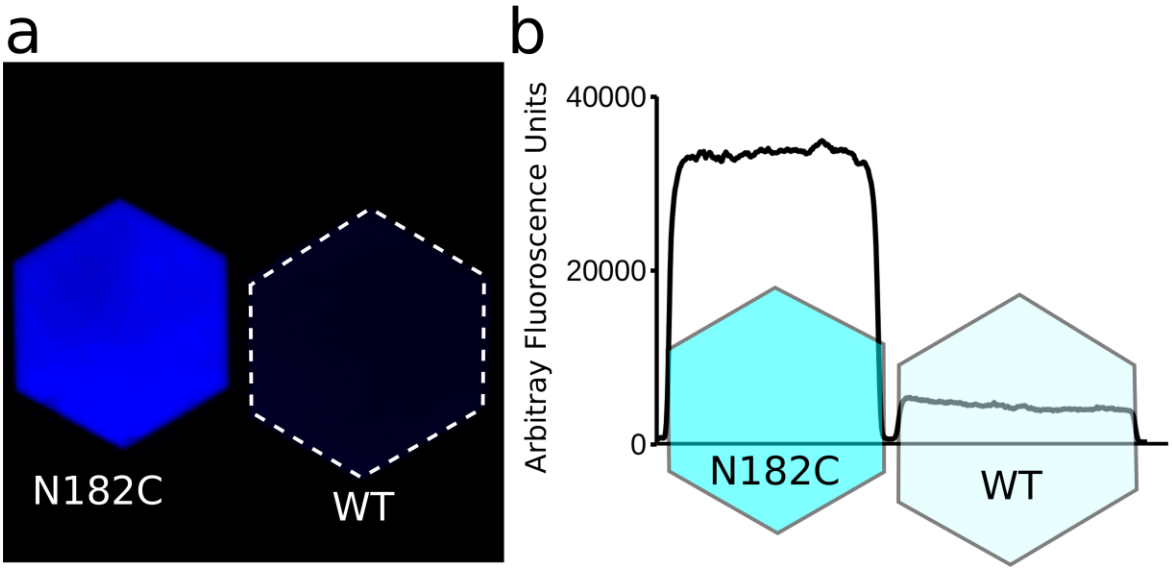



FIGURE S5.11 CONFOCAL MICROSCOPY FOR MONOBROMOBIMANE INSTALLATION

We incubated a CJ without thiol (WT) crystal and N182C crystal in a monobromobimane (mBBr) solution for 30 minutes. The crystals were then washed in 4.2 M trimethylamine-N-oxide (TMAO) pH 7.4 to remove excess reactive species. **(a)** Subsequent confocal microscopy images ($\lambda_{EX}$=405 nm) of N182C (left) and WT (right) after exposure to mBBr. N182C was fluorescent, indicative of monobromobimane installation. **(b)** Raw pixel values for N182C (left) and WT (right) show that mBBr was selectively installed on the engineered cysteine. The WT crystal had minimal background fluorescence relative to the N182C with mBBr installed.

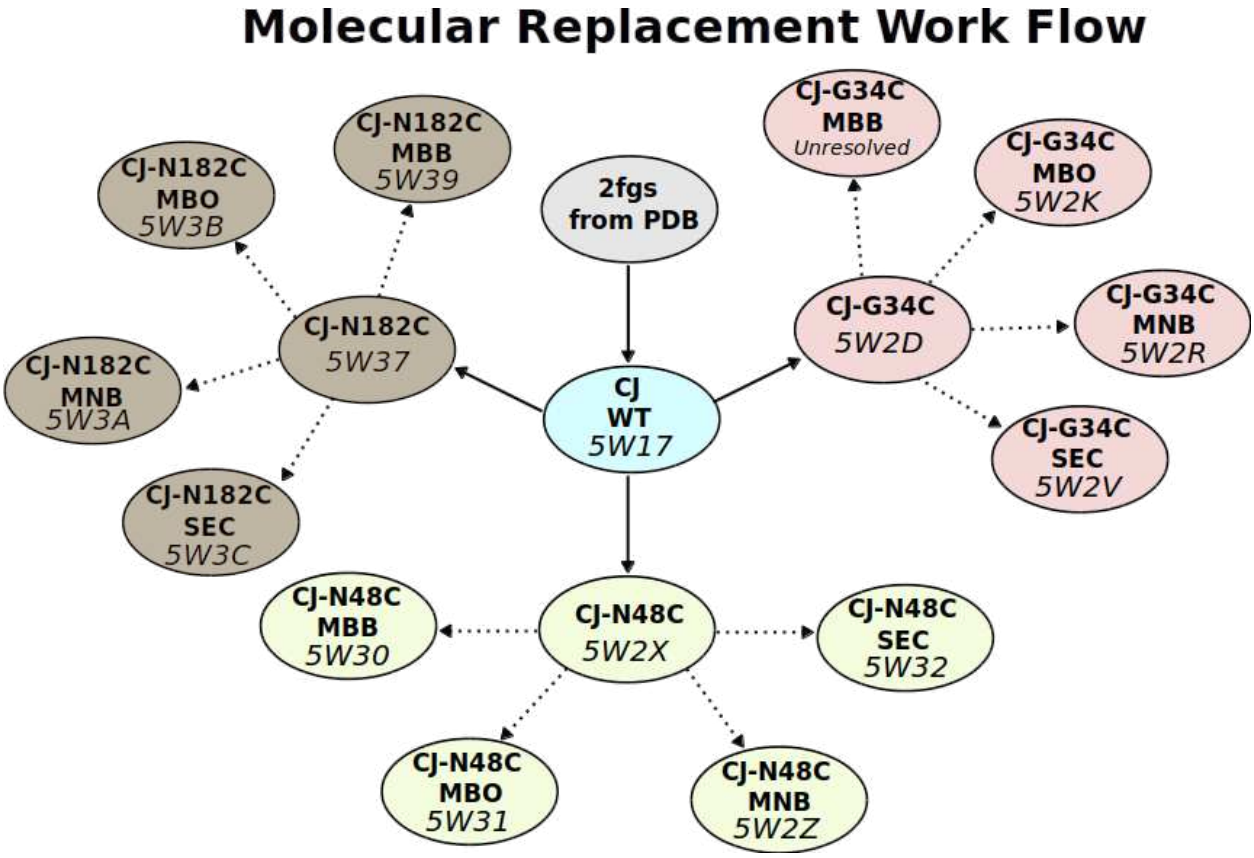



FIGURE S5.12 MOLECULAR REPLACMENT TREE

A structure for *Campylobacter jejuni* YceI periplasmic protein at 2.9 Å resolution was deposited by authors at the New York Structure Genomics team (PDB code 2fgs). We obtained an updated model for this protein at improved resolution (2.58 Å). Improved resolution allowed for further refinement of side chains, modeling of ordered water, and placement of a ligand in the hydrophobic core of the protein. The identity of the hydrophobic ligand remains unknown, but a saturated C18 ligand was modeled as a placeholder (**Figure S18**). This improved model (CJ WT) served as the molecular replacement model for the reduced CJ thiol mutant crystals: G34C, N48C, and N182C. The new thiol mutant models further served as models for the resulting CJ thiol crystals conjugated with small molecules. Non-trivial changes from the input molecular replacement model were only made if there was a strong reason such as improved side chain resolution, disrupted hydrogen bond network, modeling new features in the electron density, etc.

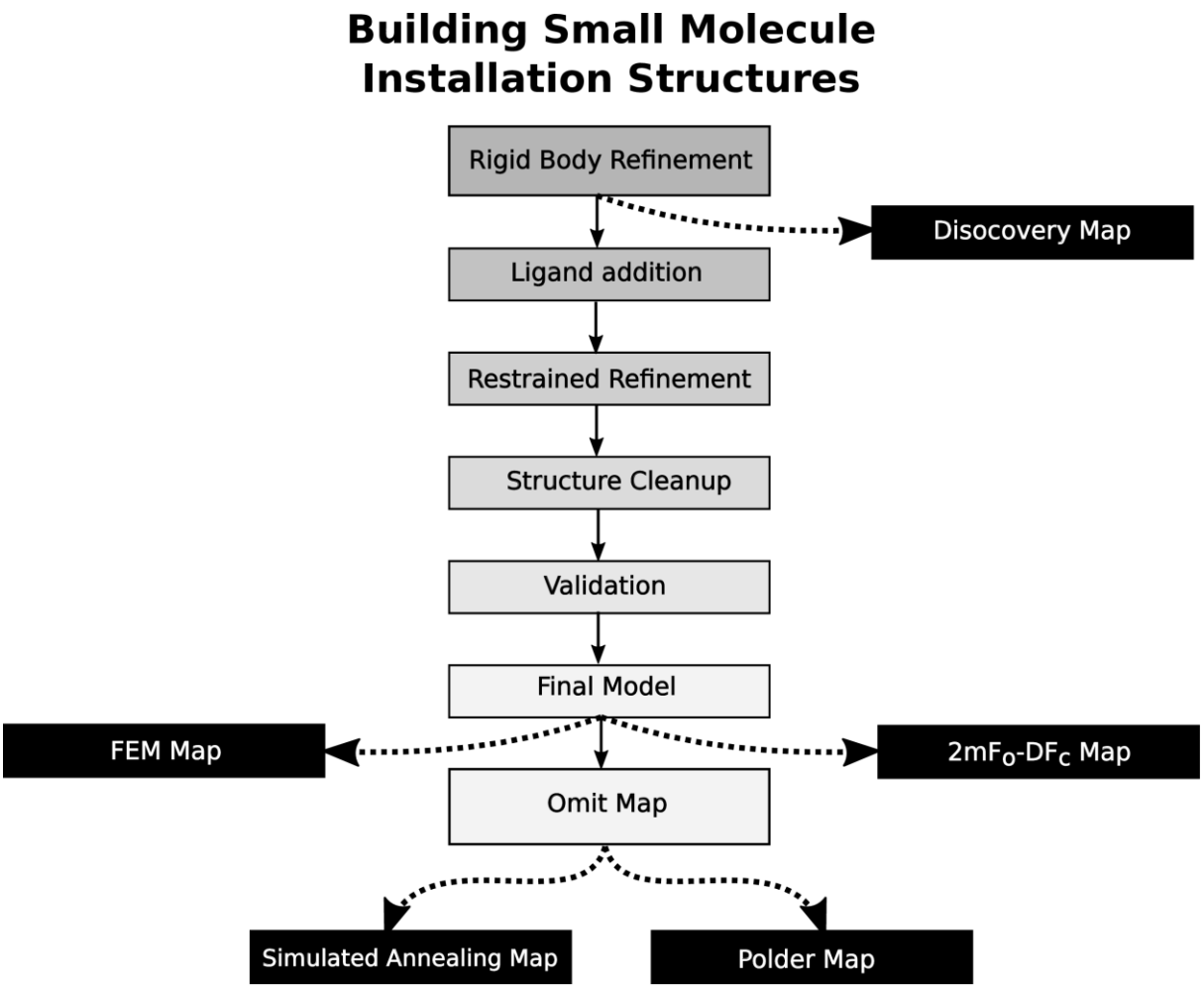



FIGURE S5.13 SMALL MOLECULE MODEL BUILDING WORK FLOW

To prevent bias, all structures with installed small molecules were generated with the same work flow. The corresponding reduced thiol structure (without an adduct) was used as a molecular replacement model. An initial "discovery map" was generated by performing rigid body refinement and identifying large map peaks suggestive of small molecule installation. Next, the probable small molecule is added to the observed electron density and a conformation was selected based on the local environment and preferred geometries. Restrained refinement was performed on the resulting structure and the occupancy of the ligand was adjusted to minimize difference map peaks. The remaining structure was refined to improve interactions and fit observed electron density. The structure was then validated on the PDB server and the final model was prepared. From the final model, we generated $2mF_o$-$DF_c$ maps for Figure 2 in the main text. We additionally generated FEM maps[186] to assess adduct placement. To further guard against bias, we generated Polder omit maps using phenix.polder.[187] The Polder maps also exclude bulk solvent from the immediate neighborhood of the adduct. Finally, we used phenix.composite_omit_map to generate conventional simulated annealing[188] omit maps with harmonic restraints. These maps are compared for each installation site in **Figure S15** (G34C)**, S16** (N48C)**, and S17** (N182C).

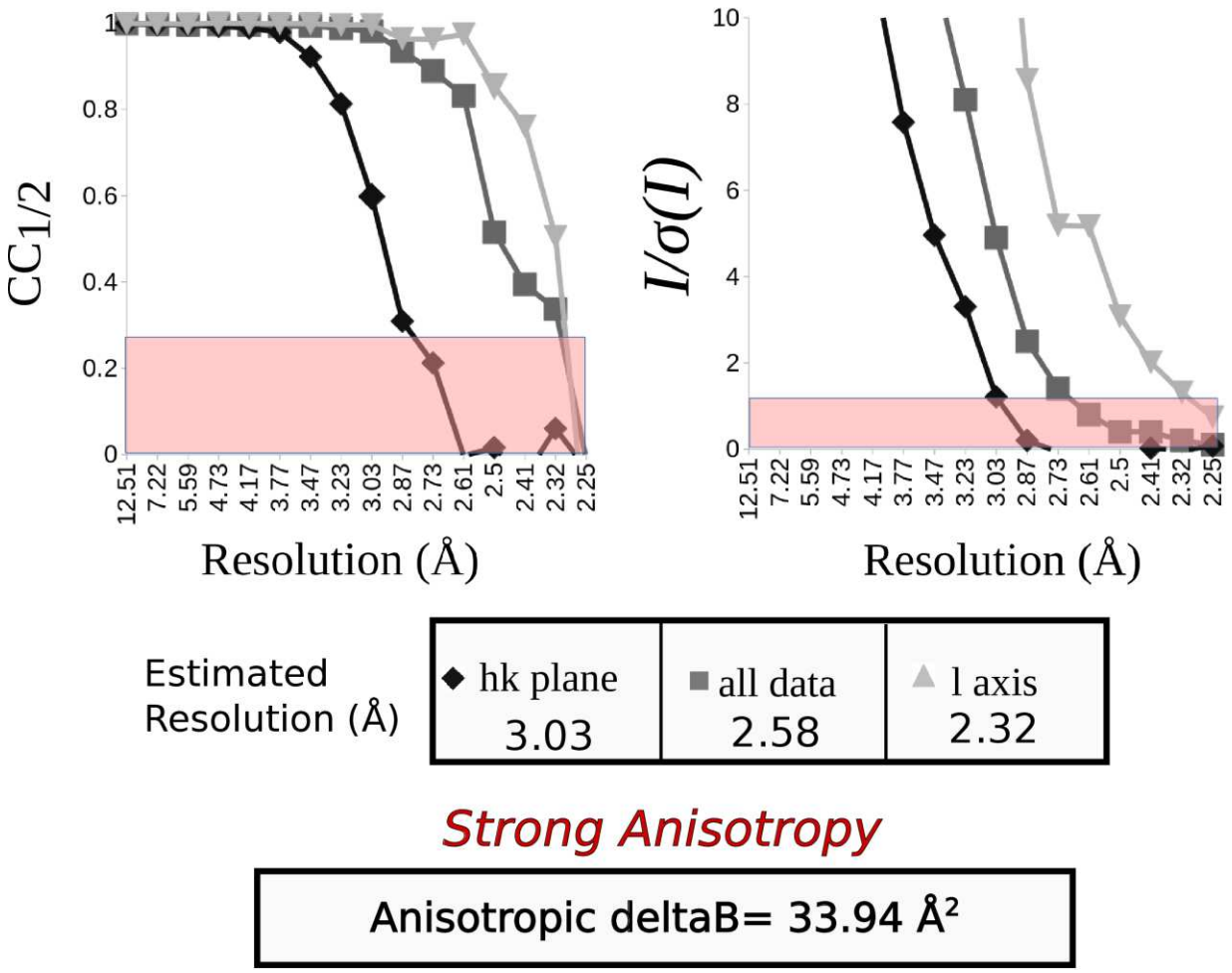



FIGURE S5.14 ANISTROPIC DATA ANALYSIS

Resolution estimates were determined by identifying the highest resolution shell that satisfied the following criteria: I/σ(I) > 1, completeness > 90%, and $cc_{1/2}$ > 0.3. The resulting data refinement statistics revealed $R_{merge}$ values in the high-resolution shell that were mathematically nonsensical (i.e. >100%). Most likely, this can be attributed to the high anisotropy of the datasets. The data scaling program *aimless* provides anisotropic data refinement statistics.[194] For the P622 symmetry of the CJ crystals, the data were separated for analysis of the hk plane and l axis. Data for CJ WT were scaled using aimless and the decoupling revealed strong anisotropy. Anisotropy is not surprising, given the unusually large solvent nanopores (13 nm diameter) that run through these crystals parallel to the c axis.

Applying the same rules of thumb for resolution estimates, the cutoff for the hk plane and l axis are 3.03 Å and 2.32 Å respectively. For this dataset, the $R_{merge}$ for the full dataset would be an acceptable 66.7% if we were to apply the hk plane derived 3.03 Å high-resolution cutoff for the full dataset. However, such a conservative approach would leave valid and useful data unused. At 3.03 Å, the l-axis data is of very high quality: $cc_{1/2}$>0.996 and I/σ(I)=16.17. Furthermore, Diedrichs and Karplus have demonstrated that there is no harm to the model R values when weaker higher resolution data is preserved, and that $cc_{1/2}$ is a better indicator for the resulting model quality than $R_{merge}$.[195,196] Thus, we decided to keep the high-resolution data, rather than maintain traditional $R_{merge}$ values. The resolution cutoff for all but two datasets were determined by the I/σ(I) > 1 rule of thumb (the others, N48C-SEC and N182C-MBO were determined by completeness >90%) and we maintained very high $cc_{1/2}$ for all high-resolution shells.

Structure refinement programs such as *Refmac*, handle anisotropy in the data to some degree. We were able to obtain acceptable refinement statistics with anisotropic corrections from *Refmac*. However, structure refinement may get "stuck" with highly anistropic data, resulting in high R values and poorly resolved side chains. Because most data refinement programs use circular resolution shells, a crystallographer might be tempted to discard higher resolution data to ensure suitable refinement statistics. Recently Strong and coworkers developed techniques for ellipsoidal scaling, which can keep strong anisotropic data, while truncating weak data in other directions at similar resolutions.[197] If future anisotropic data from CJ structures results in poorly resolved guest molecules or high R factors, we will consider ellipsoidal scaling as an alternative approach.

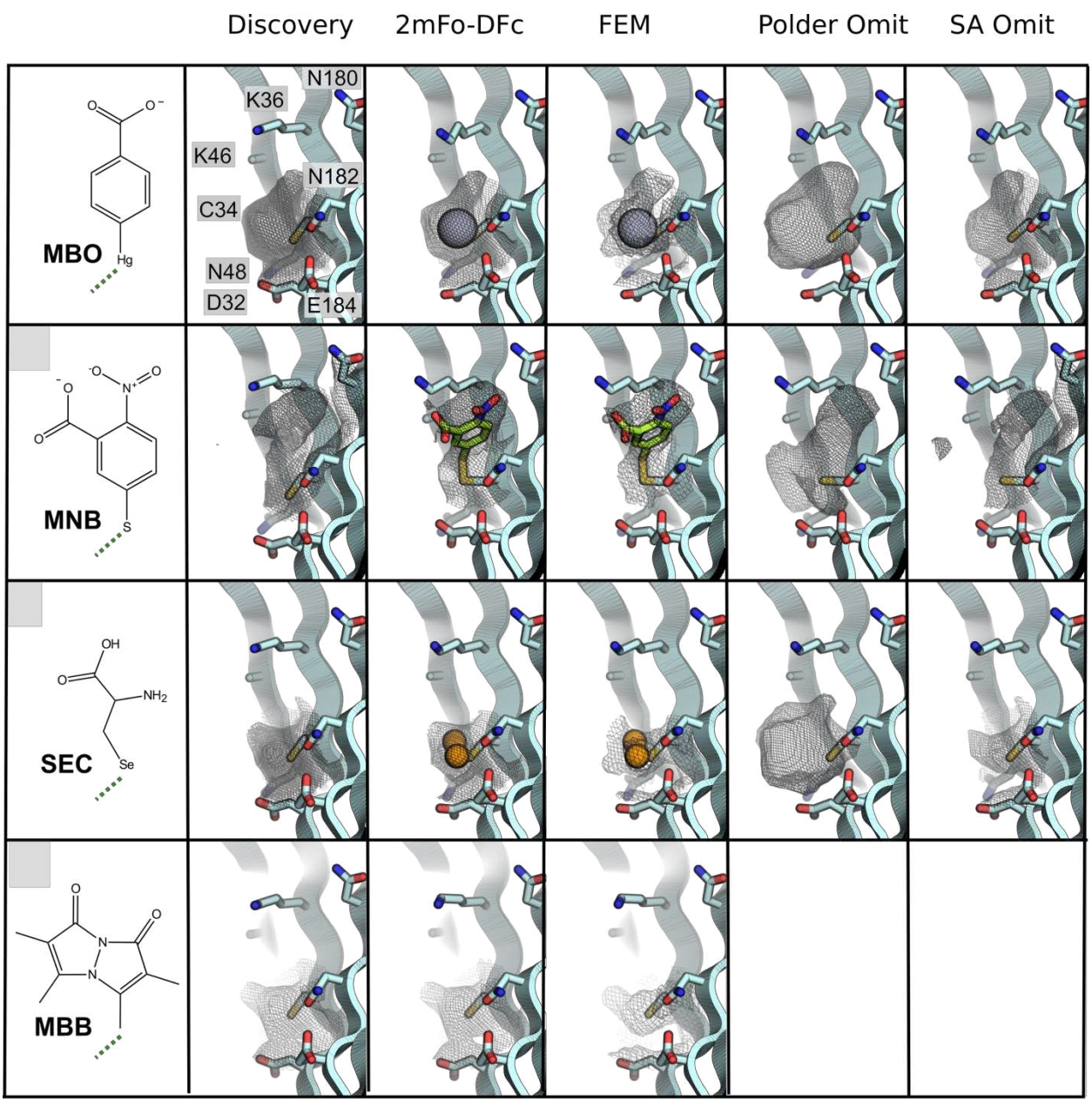



FIGURE S5.15 ADDITIONAL MAP ANALYSIS FOR SMALL MOLECULES INSTALLED AT G34C

Side by side comparison of maps at the G34C installation site, including the Discovery map (contoured at 0.8 σ), the standard $2mF_o - DF_c$ map (contoured at 0.8 σ), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 σ) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same.

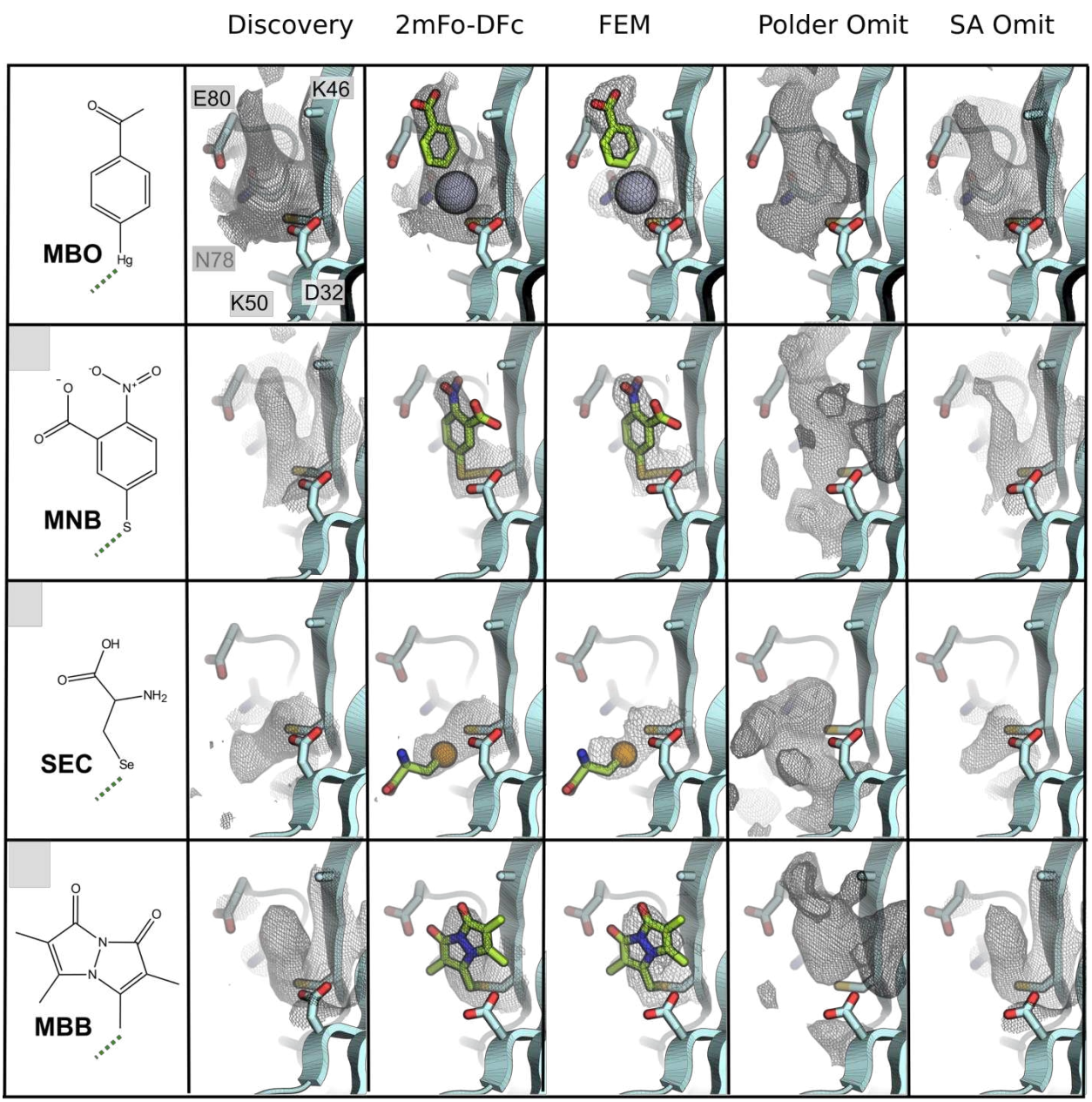



FIGURE S 5.16 ADDITIONAL MAP ANALYSIS FOR SMALL MOLECULES INSTALLED AT N48C

Side by side comparison of maps at the N48C installation site, including the Discovery map (contoured at 0.8 $\sigma$), the standard $2mF_o - DF_c$ map (contoured at 0.8 $\sigma$), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 $\sigma$) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 $\sigma$). In all cases, the illustrated protein coordinates are the same.

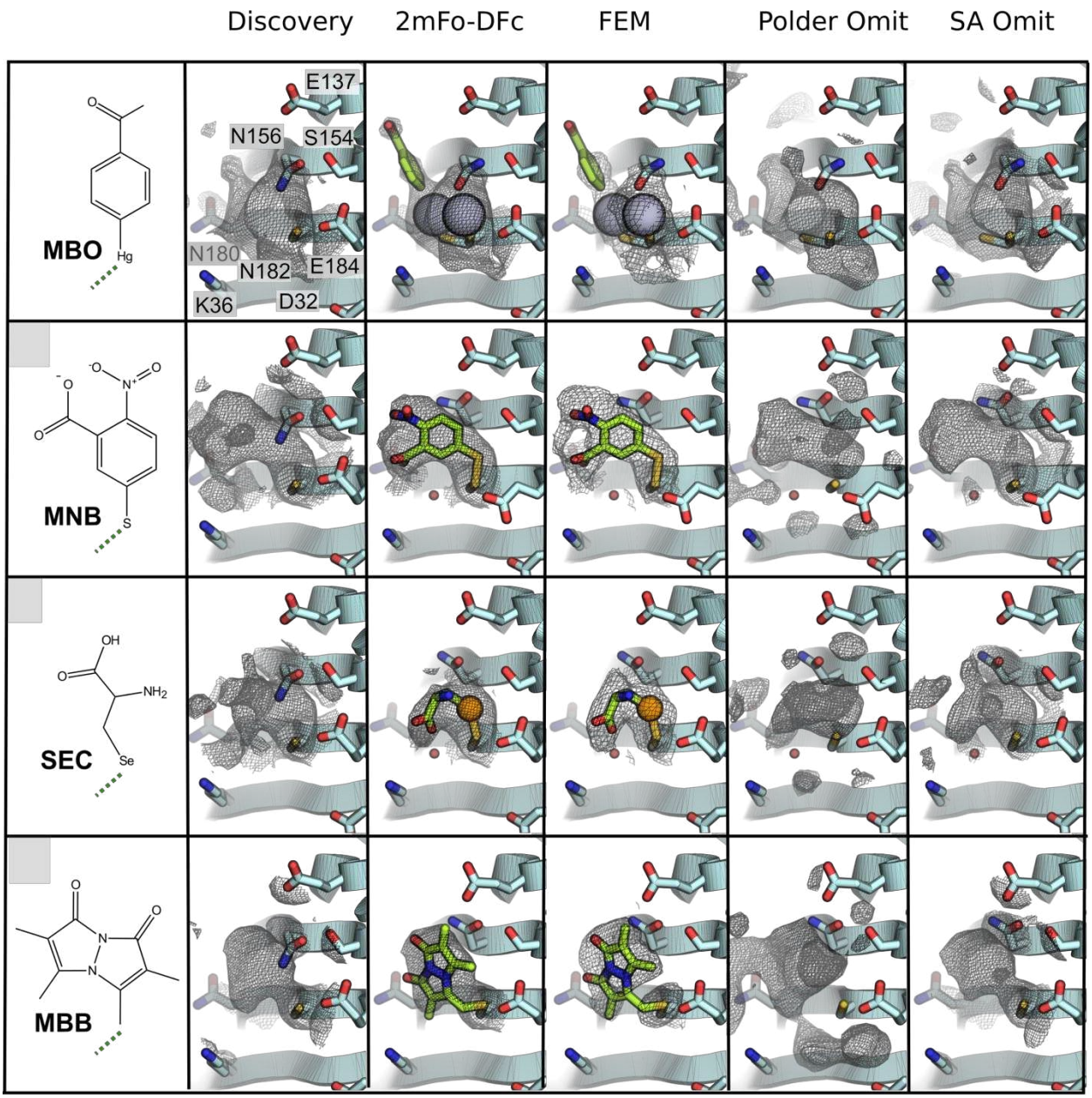



FIGURE S5.17 ADDITIONAL MAP ANALYSIS FOR SMALL MOLECULES INSTALLED AT N182C

Side by side comparison of maps at the N182C installation site, including the Discovery map (contoured at 0.8 σ), the standard $2mF_o - DF_c$ map (contoured at 0.8 σ), the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically the Polder omit map (contoured at 3.0 σ) and a simulated annealing omit map with harmonic restraints (contoured at 0.8 σ). In all cases, the illustrated protein coordinates are the same.

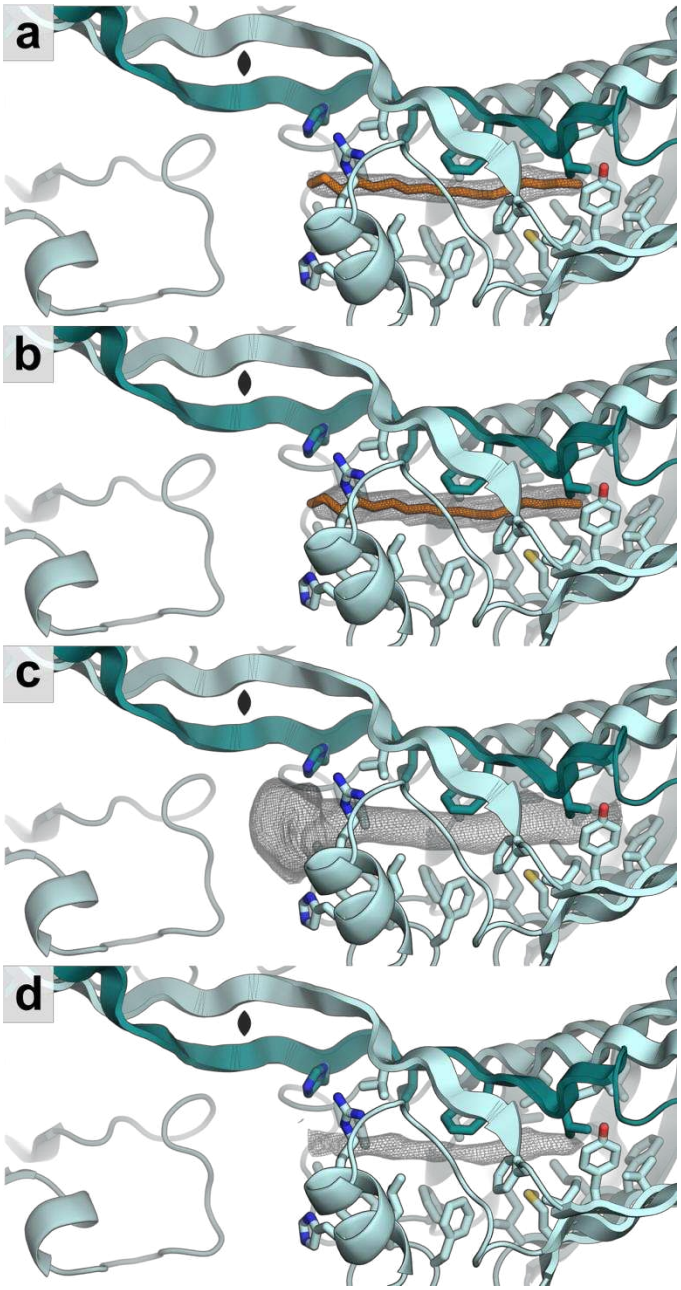



FIGURE S5.18 ADDITIONAL MAP ANALYSIS FOR INTERNAL LIGAND

Comparison of WT maps at the internal ligand site, including **(a)** the standard $2mF_o - DF_c$ map (contoured at 0.8 $\sigma$), **(b)** the FEM map (contoured at 1.0), and two different maps generated with all adduct atoms omitted, specifically **(c)** the Polder omit map (contoured at 3.0 $\sigma$) and **(d)** a simulated annealing omit map with harmonic restraints (contoured at 0.8 $\sigma$). In all cases, the illustrated protein coordinates are the same. In this instance, the $2mF_o - DF_c$ map lacks interpretable density near the binding site opening. The Polder map provides a better illustration of the likely placement for a negatively charged ligand head group in the vicinity of His39, Arg84, and His87. A 2-fold symmetry axis relating the halves of the domain swapped dimer (light blue and dark blue chains) is indicated with the curved diamond shape (black).

| Set of Surface Residues for Bfactor Comparision | | | |
|---|---|---|---|
| MET21 | VAL55 | THR103 | ASN156 |
| LYS22 | ASP57 | THR105 | LYS158 |
| GLU23 | ASP59 | LYS107 | LYS160 |
| THR25 | SER62 | LYS108 | SER162 |
| ASP27 | GLU64 | LYS116 | ASP163 |
| LYS28 | LYS66 | THR120 | LYS165 |
| ALA29 | LYS67 | THR122 | THR168 |
| HIS30 | ASP69 | THR124 | THR172 |
| ASP32 | THR71 | SER129 | ASP178 |
| GLY34 | LYS73 | LYS130 | ASN180 |
| LYS36 | ASN78 | ASP131 | ASN182 |
| LYS38 | GLU80 | ASP135 | GLU184 |
| GLN41 | ASN81 | GLU137 | GLU186 |
| SER43 | GLN82 | ALA142 | ASN188 |
| ASN44 | THR83 | LYS143 | GLU189 |
| LYS46 | ASN86 | LYS145 | LYS190 |
| ASN48 | GLN90 | ASP146 | GLU191 |
| LYS50 | ASP91 | LYS148 | |
| ASP51 | ASP92 | LYS150 | |
| SER53 | LYS97 | SER154 | |

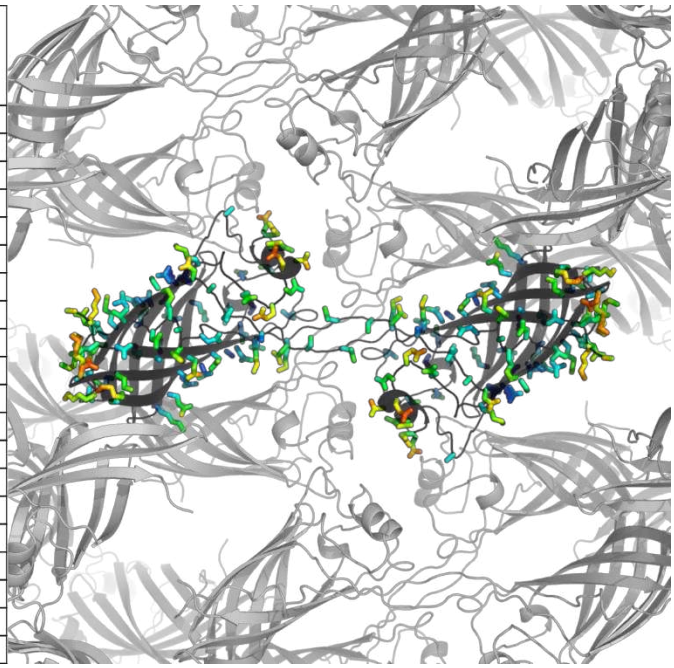
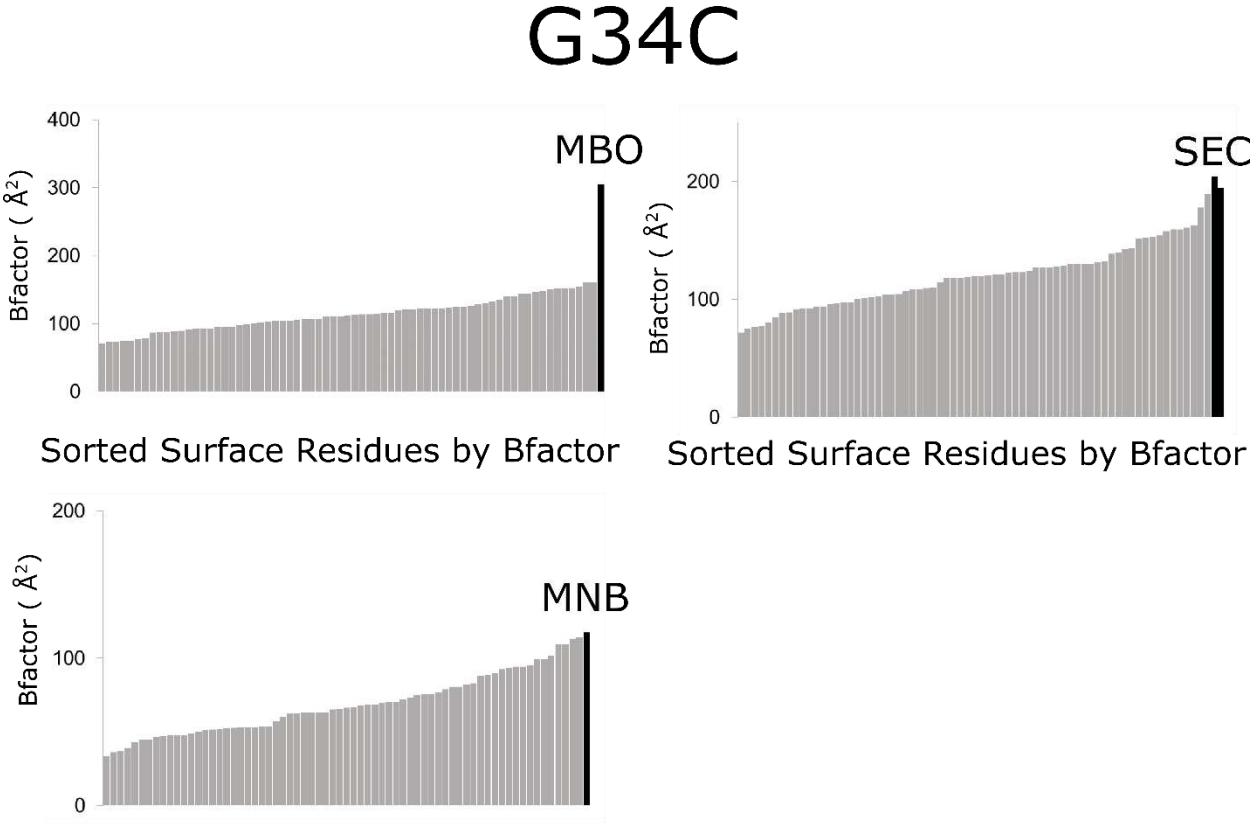


FIGURE S5.19 PANEL OF SOLVENT EXPOSED SURFACE RESIDUES

A set of solvent exposed surface residues was manually identified for B-factor comparison to installed guest molecules. A representative image of the selected residues is displayed with B-factor weighted coloring (blue coloring represents relative low B-factor and red coloring represents relative high B-factor).




FIGURES 5.20 B-FACTOR DISTRIBUTION FOR SMALL MOLECULE INSTALLATION AT G34C

B-factor distribution for a set of solvent exposed surface amino acids sorted low to high (light grey bars) and installed guest molecule B-factor (black bar) for G34C installations. B-factors for each residue was determined as an average for all atoms past the beta carbon (if present). B-factor for the ligand was determined as an average for all atoms present in the model.

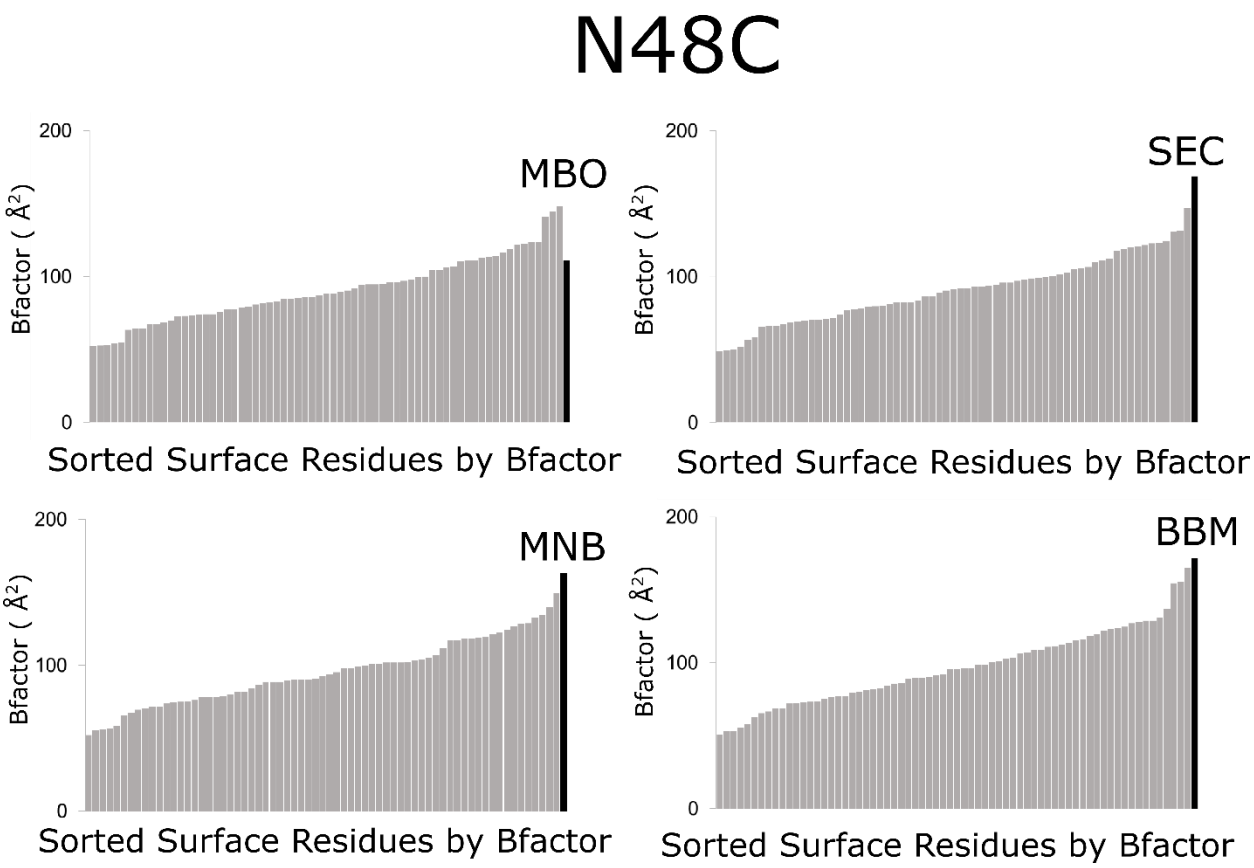



FIGURE S5.21 B-FACTOR DISTRIBUTION FOR SMALL MOLECULE INSTALLATION AT N48C

B-factor distribution for a set of solvent exposed surface amino acids sorted low to high (light grey bars) and installed guest molecule B-factor (black bar) for N48C installations. B-factors for each residue was determined as an average for all atoms past the beta carbon (if present). B-factor for the ligand was determined as an average for all atoms present in the model.

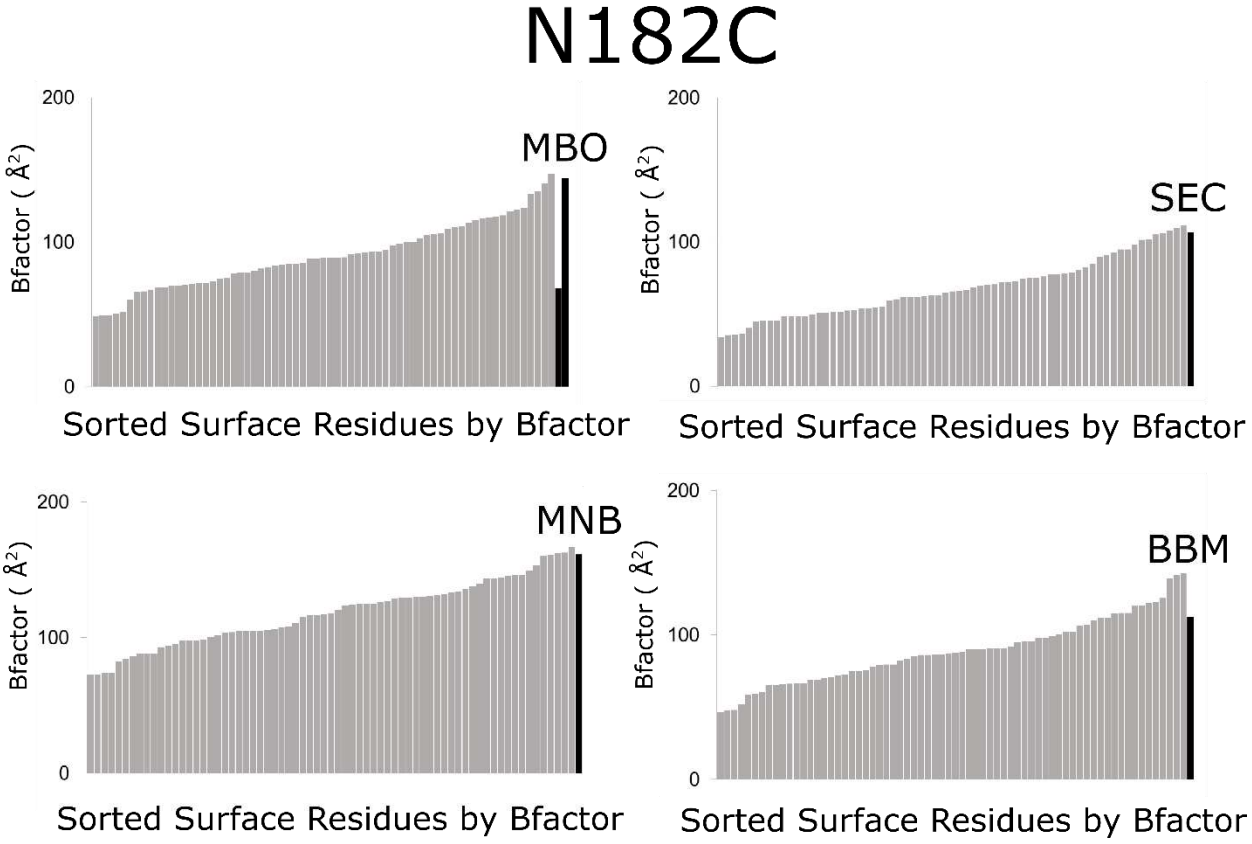



FIGURE S5.22 B-FACTOR DISTRIBUTION FOR SMALL MOLECULE INSTALLATION AT N182C

B-factor distribution for a set of solvent exposed surface amino acids sorted low to high (light grey bars) and installed guest molecule B-factor (black bar) for N182C installations. B-factors for each residue was determined as an average for all atoms past the beta carbon (if present). B-factor for the ligand was determined as an average for all atoms present in the model.

## 5.6 ADDITIONAL TABLES

TABLE S5.1 CRYSTALLOGRAPHY DATA COLLECTION AND REFINEMENT STATISTICS FOR CJ WITHOUT THIOL (CJ-WT), G34C, N48C, AND N182C.

| | **CJ-WT** | **G34C** | **N48C** | **N182C** |
|---|---|---|---|---|
| **Data Collection** | | | | |
| Light Source | Synchrotron | Synchrotron | Synchrotron | Synchrotron |
| Wavelength (Å) | 1.0 | 1.0 | 1.0 | 1.0 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | | | |
| a, b, c (Å) | 178.81, 178.81, 50.34 | 178.88, 178.88, 50.72 | 179.15, 179.15, 50.77 | 178.43, 178.43, 50.58 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 38.71-2.58 (2.72-2.58)* | 38.73-2.70 (2.85-2.70)* | 38.79-2.73 (2.88-2.73)* | 38.63-2.52 (2.66-2.52)* |
| Mesaured reflections | 313158 | 280434 | 278863 | 336425 |
| Unique reflections | 15452 | 13638 | 13248 | 16542 |
| Completeness (%) | 99.9 (99.6) | 99.9 (100.0) | 99.8 (98.9) | 99.9 (99.5) |
| Redundancy | 20.3 (19.9) | 20.6 (21.2) | 21.0 (21.8) | 20.3 (17.4) |
| $cc_{1/2}$ | 0.997 (0.897) | 0.999 (0.761) | 0.999 (0.788) | 0.999 (0.784) |
| $I/\sigma(I)$ | 10.2 (1.0) | 13.2 (1.0) | 13.5 (1.0) | 22.8 (1.0) |
| **Refinement** | | | | |
| *Rwork/Rfree* | 0.2212/0.2537 | 0.2299/0.2713 | 0.2100/0.2393 | 0.2149/0.2578 |
| No. Atoms | | | | |
| Protein | 1321 | 1323 | 1319 | 1319 |
| LFA | 17 | 17 | 17 | 17 |
| SO4 | 10 | 10 | 10 | 10 |
| Water | 21 | 18 | 17 | 21 |
| B-Factors | | | | |
| Protein | 69.69 | 94.25 | 83.69 | 80.47 |
| LFA | 69.11 | 88.35 | 81.92 | 75.38 |
| SO4 | 75.29 | 101.22 | 89.43 | 92.45 |
| Water | 57.53 | 79.65 | 68.6 | 72.66 |
| RMSD | | | | |
| Bond lengths (Å) | 0.013 | 0.013 | 0.014 | 0.014 |
| Bond angles (°) | 1.856 | 1.826 | 1.918 | 1.935 |
| Ramachandran plot (%) | | | | |
| Favored region | 0.947 | 0.941 | 0.935 | 0.935 |
| Allowed region | 0.053 | 0.059 | 0.065 | 0.065 |
| Outlier region | 0 | 0 | 0 | 0 |

* values in parantheses are for high resolution shell

TABLE S5.2 CRYSTALLOGRAPHY DATA COLLECTION AND REFINEMENT STATISTICS FOR INSTALLATION OF MERCURIBENZOIC ACID (MBO) ON G34C, N48C, AND N182C.

| | G34C-MBO | N48C-MBO | N182C-MBO |
|---|---|---|---|
| **Data Collection** | | | |
| Light Source | Synchrotron | Synchrotron | Synchrotron |
| Wavelength (Å) | 1.0 | 1.0 | 1.0 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | | |
| a, b, c (Å) | 178.81, 178.81, 50.64 | 178.79, 178.79, 50.81 | 179.07, 179.07, 50.65 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 38.71-2.78 (2.93-2.78)* | 38.71-2.56 (2.70-2.56)* | 38.77-2.70 (2.85-2.70)* |
| Mesaured reflections | 244820 | 330151 | 245372 |
| Unique reflections | 12465 | 15939 | 13431 |
| Completeness (%) | 99.7 (98.2) | 99.9 (99.9) | 98.8 (92.2) |
| Redundancy | 19.6 (13.9) | 20.7 (19.0) | 18.3 (7.4) |
| $cc_{1/2}$ | 0.999 (0.758) | 0.999 (0.857) | 0.999 (0.875) |
| *I/σ(I)* | 11.4 (1.0) | 14.4 (1.0) | 20.7 (1.4) |
| **Refinement** | | | |
| *Rwork/Rfree* | 0.2214/0.2614 | 0.2133/0.2412 | 0.2047/0.2281 |
| No. Atoms | | | |
| Protein | 1323 | 1315 | 1316 |
| LFA | 17 | 17 | 17 |
| SO4 | 10 | 10 | 10 |
| Water | 18 | 17 | 21 |
| MBO | 1 | 10 | 11 |
| B-Factors | | | |
| Protein | 94.44 | 76.12 | 73.56 |
| LFA | 93.26 | 75.64 | 66.64 |
| SO4 | 106.09 | 78.6 | 79.72 |
| Water | 80.18 | 66.85 | 62.05 |
| MBO | 305.21 | 111.11 | 75.33/144.31 |
| RMSD | | | |
| Bond lengths (Å) | 0.013 | 0.014 | 0.062 |
| Bond angles (°) | 1.843 | 1.94 | 2.013 |
| Ramachandran plot (%) | | | |
| Favored region | 0.917 | 0.935 | 0.976 |
| Allowed region | 0.083 | 0.065 | 0.024 |
| Outlier region | 0 | 0 | 0 |

* values in parantheses are for high resolution shell

TABLE S5.3 CRYSTALLOGRAPHY DATA COLLECTION AND REFINEMENT STATISTICS FOR INSTALLATION OF 5-MERCAPTO-2-NITRO-BENZOIC ACID (MNB) ON G34C, N48C, AND N182C.

| | G34C-MNB | N48C-MNB | N182C-MNB |
|---|---|---|---|
| **Data Collection** | | | |
| Light Source | Synchrotron | Synchrotron | Rigaku Homelab |
| Wavelength (Å) | 1.0 | 1.0 | 1.54187 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | | |
| a, b, c (Å) | 178.34, 178.34, 50.32 | 179.86, 179.86, 50.62 | 180.33, 180.33, 50.57 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 38.61-2.90 (3.06-2.90)* | 38.94-2.80 (2.95-2.80)* | 48.11-2.76 (2.91-2.76)* |
| Mesaured reflections | 219589 | 255445 | 222910 |
| Unique reflections | 10917 | 12372 | 12968 |
| Completeness (%) | 99.7 (99.1) | 99.9 (99.8) | 100.0 (100.0) |
| Redundancy | 20.1 (20.0) | 20.6 (20.9) | 17.2 (9.4) |
| $cc_{1/2}$ | 0.998 (0.909) | 0.998 (0.688) | 0.994 (0.390) |
| *I/σ(I)* | 9.4 (1.0) | 11.2 (1.0) | 9.0 (1.0) |
| **Refinement** | | | |
| *Rwork/Rfree* | 0.2158/0.2732 | 0.2032/0.2358 | 0.2371/0.2746 |
| No. Atoms | | | |
| Protein | 1323 | 1315 | 1315 |
| LFA | 17 | 17 | 17 |
| SO4 | 10 | 10 | 10 |
| Water | 17 | 21 | 22 |
| MNB | 13 | 13 | 13 |
| B-Factors | | | |
| Protein | 101.1 | 79.34 | 53.78 |
| LFA | 94.44 | 78.64 | 52.35 |
| SO4 | 106.39 | 85.1 | 57.6 |
| Water | 84.9 | 73.96 | 40.14 |
| MNB | 161.51 | 163.36 | 106.77 |
| RMSD | | | |
| Bond lengths (Å) | 0.012 | 0.013 | 0.013 |
| Bond angles (°) | 1.688 | 1.884 | 1.77 |
| Ramachandran plot (%) | | | |
| Favored region | 0.917 | 0.947 | 0.947 |
| Allowed region | 0.083 | 0.053 | 0.047 |
| Outlier region | 0 | 0 | 0.006 |

* values in parantheses are for high resolution shell

TABLE S5.4 CRYSTALLOGRAPHY DATA COLLECTION AND REFINEMENT STATISTICS FOR INSTALLATION OF SELENOCYSTEINE (SEC) ON G34C, N48C, AND N182C.

| | G34C-SEC | N48C-SEC | N182C-SEC |
|---|---|---|---|
| **Data Collection** | | | |
| Light Source | Synchrotron | Synchrotron | Rigaku Homelab |
| Wavelength (Å) | 1.0 | 1.0 | 1.54187 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | | |
| a, b, c (Å) | 178.02, 178.02, 50.71 | 178.62, 178.62, 50.81 | 180.34, 180.34, 50.59 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 38.54-2.90 (3.06-2.90)* | 38.67-2.63 (2.77-2.63)* | 48.13-2.87 (3.03-2.87)* |
| Mesaured reflections | 203030 | 287615 | 192807 |
| Unique reflections | 10946 | 14539 | 11574 |
| Completeness (%) | 99.7 (98.3) | 98.8 (92.5) | 100.0 (100.0) |
| Redundancy | 18.5 (12.5) | 19.8 (12.3) | 16.7 (7.5) |
| $cc_{1/2}$ | 0.999 (0.783) | 0.999 (0.888) | 0.994 (0.412) |
| *I/σ(I)* | 10.7 (1.0) | 25.5 (1.6) | 9.4 (1.0) |
| **Refinement** | | | |
| *Rwork/Rfree* | 0.2127/0.2608 | 0.2140/0.2360 | 0.2334/0.2738 |
| No. Atoms | | | |
| Protein | 1323 | 1315 | 1319 |
| LFA | 17 | 17 | 17 |
| SO4 | 10 | 10 | 10 |
| Water | 16 | 16 | 19 |
| SEC | 2 | 7 | 7 |
| B-Factors | | | |
| Protein | 100.35 | 75.55 | 54.61 |
| LFA | 86.03 | 80.95 | 49.6 |
| SO4 | 103.86 | 83.5 | 56.38 |
| Water | 80.47 | 65.49 | 33.78 |
| SEC | 203.74/194.71 | 168.95 | 117.46 |
| RMSD | | | |
| Bond lengths (Å) | 0.012 | 0.015 | 0.012 |
| Bond angles (°) | 1.693 | 2.051 | 1.816 |
| Ramachandran plot (%) | | | |
| Favored region | 0.935 | 0.935 | 0.953 |
| Allowed region | 0.065 | 0.065 | 0.047 |
| Outlier region | 0 | 0 | 0 |

* values in parantheses are for high resolution shell

TABLE S5.5 CRYSTALLOGRAPHY DATA COLLECTION AND REFINEMENT STATISTICS FOR INSTALLATION OF MONOBROMOBIMANE (MBB) ON N48C AND N182C.

| | N48C-MBB | N182C-MBB |
|---|---|---|
| **Data Collection** | | |
| Light Source | Synchrotron | Synchrotron |
| Wavelength (Å) | 1.0 | 1.0 |
| Spacegroup | *P 6 2 2* | *P 6 2 2* |
| Cell dimensions | | |
| a, b, c (Å) | 180.25, 180.25, 50.97 | 178.24, 178.24, 50.60 |
| α, β, γ (°) | 90.00, 90.00, 120.00 | 90.00, 90.00, 120.00 |
| Resolution (Å) | 39.03-2.75 (2.90-2.75)* | 38.59-2.48 (2.61-2.48)* |
| Mesaured reflections | 253774 | 339182 |
| Unique reflections | 13196 | 17262 |
| Completeness (%) | 100.0 (100.0) | 99.6 (97.7) |
| Redundancy | 19.2 (17.9) | 19.6 (14.3) |
| $cc_{1/2}$ | 0.998 (0.678) | 0.999 (0.772) |
| *I/σ(I)* | 11.4 (1.0) | 22.4 (1.0) |
| **Refinement** | | |
| *Rwork/Rfree* | 0.2114/0.2457 | 0.2175/0.2448 |
| No. Atoms | | |
| Protein | 1319 | 1319 |
| LFA | 17 | 17 |
| SO4 | 10 | 10 |
| Water | 18 | 23 |
| MBB | 14 | 14 |
| B-Factors | | |
| Protein | 80.3 | 72.36 |
| LFA | 75.05 | 71.72 |
| SO4 | 86.87 | 65.39 |
| Water | 63.13 | 65.72 |
| MBB | 171.48 | 112.63 |
| RMSD | | |
| Bond lengths (Å) | 0.014 | 0.015 |
| Bond angles (°) | 1.898 | 2.013 |
| Ramachandran plot (%) | | |
| Favored region | 0.953 | 0.941 |
| Allowed region | 0.047 | 0.059 |
| Outlier region | 0 | 0 |

* values in parantheses are for high resolution shell

TABLE S5.6 PYMOL ELECTRON DENSITY CONTOURING

| Protein | Adduct | e/Å$^3$ equivalent to COOT rmsd 0.8 | e/Å$^3$ equivalent to COOT rmsd 1.0 | e/Å$^3$ equivalent to COOT rmsd 3.0 for Polder map | e/Å$^3$ equivalent to COOT rmsd 0.8 for Discovery map | e/Å$^3$ equivalent to COOT rmsd 0.8 for restrained simulated annealing omit map |
|---|---|---|---|---|---|---|
| WT | | 0.0872 | 0.1089 | NA | NA | NA |
| N48C | | 0.0762 | 0.0952 | NA | NA | NA |
| N48C | MBO | 0.0824 | 0.1031 | 0.1683 | 0.0727 | 0.0776 |
| N48C | MNB | 0.0777 | 0.971 | 0.1583 | 0.0699 | 0.0769 |
| N48C | SEC | 0.0833 | 0.1042 | 0.1526 | 0.0685 | 0.0801 |
| N48C | MBB | 0.0785 | 0.098 | 0.1592 | 0.0863 | 0.0749 |
| G34C | | 0.0721 | 0.09 | NA | NA | NA |
| G34C | MBO | 0.0713 | 0.0891 | 0.1168 | 0.0793 | 0.0653 |
| G34C | MNB | 0.066 | 0.0825 | 0.1618 | 0.0776 | 0.0616 |
| G34C | SEC | 0.066 | 0.0825 | 0.1145 | 0.0824 | 0.0603 |
| G34C | MBB | 0.0743 | 0.0928 | NA | NA | NA |
| N182C | | 0.0784 | 0.098 | NA | NA | NA |
| N182C | MBO | 0.0846 | 0.1058 | 0.1583 | 0.0837 | 0.0817 |
| N182C | MNB | 0.1041 | 0.13 | 0.2013 | 0.1076 | 0.0989 |
| N182C | SEC | 0.1009 | 0.1262 | 0.189 | 0.1071 | 0.0979 |
| N182C | MBB | 0.0855 | 0.1068 | 0.1532 | 0.0854 | 0.0833 |

To most faithfully represent COOT contours in PyMOL images, we turned off the PyMOL automatic map normalization ("set normalize_ccp4_maps, off") and instead directly set contour values to match the intended COOT contour level. This Table tracks the e/Å$^3$ values reported by COOT, when contours were set to specific rmsd values. The first two columns are for the primary .mtz file. The Polder map column corresponds to the values found for Polder omit maps ($mF_o$-$DF_c$), contoured to 3 rmsd in COOT. The penultimate column has the values found for the rigid body discovery maps ($2mF_o$-$DF_c$), contoured to 0.8 rmsd in COOT. The final column has the values found for the simulated annealing omit maps ($2mF_o$-$DF_c$), contoured to 0.8 rmsd in COOT. Unlike the other maps, the FEM approach attempts to eliminate the need to choose (arbitrary) map contouring cutoffs. In PyMOL, FEM maps were always contoured with a cutoff of 1.0.

TABLE S5.7 B-FACTOR Z-SCORE FOR SMALL MOLECULE INSTALLATION STRUCTURES

## Ligand Bfactor Z-score

| | G34C | N48C | N182C |
|---|---|---|---|
| **MBO** | 8.13 | 0.9 | -0.97/2.25 |
| **MNB** | 1.81 | 2.69 | 1.83 |
| **SEC** | 3.3/2.95 | 2.86 | 2.29 |
| **BBM** | | 2.73 | 1.04 |

Z-scores were calculated for guest molecule B-factor relative to the set of solvent exposed residue B-factor distribution.

# 6 SCAFFOLD ASSISTED CRYSTALLOGRAPHY WITH MACROMOLECULES

## 6.1 INTRODUCTION

High throughput sequencing has revolutionized the biosciences and expanded the known space of the protein universe.[1] While the evolutionary history of proteins can be deduced from sequence information and grouped into families based on homology, the three-dimensional structure of proteins is key to understanding protein function. Unfortunately, there has not been a similar revolution to enable high throughput three-dimensional structure determination. As a result, the number of macromolecular three-dimensional structures has dramatically lagged the number of sequences identified.[2,3] The predominant method for protein structure determination is x-ray diffraction (XRD). This methodology is laborious and relies on finding the “needle in the haystack condition under which a protein crystallizes”. Some proteins are too difficult to obtain in sufficient quantity and purity to make the attempt, while others fail to crystallize or yield diffraction quality crystals even after exhaustive condition screening.

The motivational insight for this research is the recognition that materials diffract X-rays if they consist of a highly-ordered, repeating lattice, but that the lattice need not be composed only of target protein. Instead of growing conventional protein crystals of the target protein, we will attempt to determine the structure of mini-proteins covalently conjugated to a protein (scaffold) in a predetermined lattice. The selected scaffold, CJ, has large 13 nm axial pores and the engineered attachment sites are designed to avoid clashes with symmetry related neighbors.

Due to the robust crystallization of the scaffold protein, we will attempt to perform conjugation prior to crystallization in a technique similar to the chaperone assisted crystallization method.[198] Unlike chaperone assisted crystallography, guest are covalently attached to the scaffold protein via disulfide bonds, the lattice is predetermined with specific placement of the guest protein in three-dimensional space, and the crystallization condition is known. If guest attachment abrogates crystallization, the accessibility of the engineered thiols to the large solvent channels of the existing crystals allows for attempts at asynchronous crystallization and conjugation. In the long term, asynchronous guest attachment may present a more general structure determination method for target proteins that fail to crystallize via conventional means.

## 6.2 RESULTS AND DISCUSSION

In previous work we engineered solvent exposed thiols proximal to the CJ axial pores (13 nm diameter) and subsequently installed small molecules at these sites. The resulting modified crystalline scaffolds were validated using XRD. These “installation sites” were selected to maximize the inter-site distance between symmetry copies throughout the crystal. Thus, each installed guest molecule will experience symmetry-equivalent interactions with neighboring amino acids and can independently adopt a preferred conformation. One of these variants, CJ-182C, crystallizes rapidly and reliably afforded interpretable electron density of small molecule conjugates (Chapter 5). The thiol at CJ-182C is ~41 Å to its nearest symmetry partner allowing, for macromolecule installation. CJ-182C was selected as the preferred site to develop methods for scaffold assisted crystallography of macromolecules.

Limiting conformational flexibility of guest molecules is important for interpretability of the resulting electron density (Chapter 5). To limit flexibility at the initial attachment point, we decided to attach guest proteins to the scaffold via disulfide bond formation. Disulfide bonds prefer ±90° dihedral angles with thermodynamic penalties for deviation from ideal geometry.[199] Ellman's reagent, 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB), readily undergoes disulfide exchange with a free sulfhydryl to form a mixed disulfide with the release of 5-thio-2-nitrobenzoic acid ($TNB^{-2}$). Previously, we attached DTNB to a pre-existing CJ-182C crystal and used XRD diffraction to elucidate a single conformation for installed 5-mercapto-2-nitro-benzoic acid (MNB) ligand at CJ-182C. A thiol activated with MNB ($R_1S$-MNB) can undergo a second interchange by the addition of another sulfhydryl-containing target molecule ($HS-R_2$) to yield a disulfide crosslink ($R_1-S-S-R_2$). The low pKa of the $TNB^{-2}$ thiol (pKa ~4.5) drives the equilibrium towards the desired disulfide crosslink with the release of $TNB^{-2}$.[200] Further, the desired disulfide crosslink can be monitored by release of $TNB^{-2}$ anion, which absorbs strongly at 412 nm (e=14,150 $M^{-1}cm^{-1}$ at pH 8.0).[201] We decided to use DTNB to generate an "activated" disulfide at CJ-182C (182C-MNB) to promote disulfide formation with a guest molecule thiol. To prepare 182C-MNB, CJ-182C was reacted *in vitro* with DTNB in excess. The resulting 182C-MNB conjugate was subsequently dialyzed, aliquoted, and frozen prior to secondary conjugation.

### 6.2.1 ENGINEERING A GUEST MINIPROTEIN FOR COVALENT CAPTURE

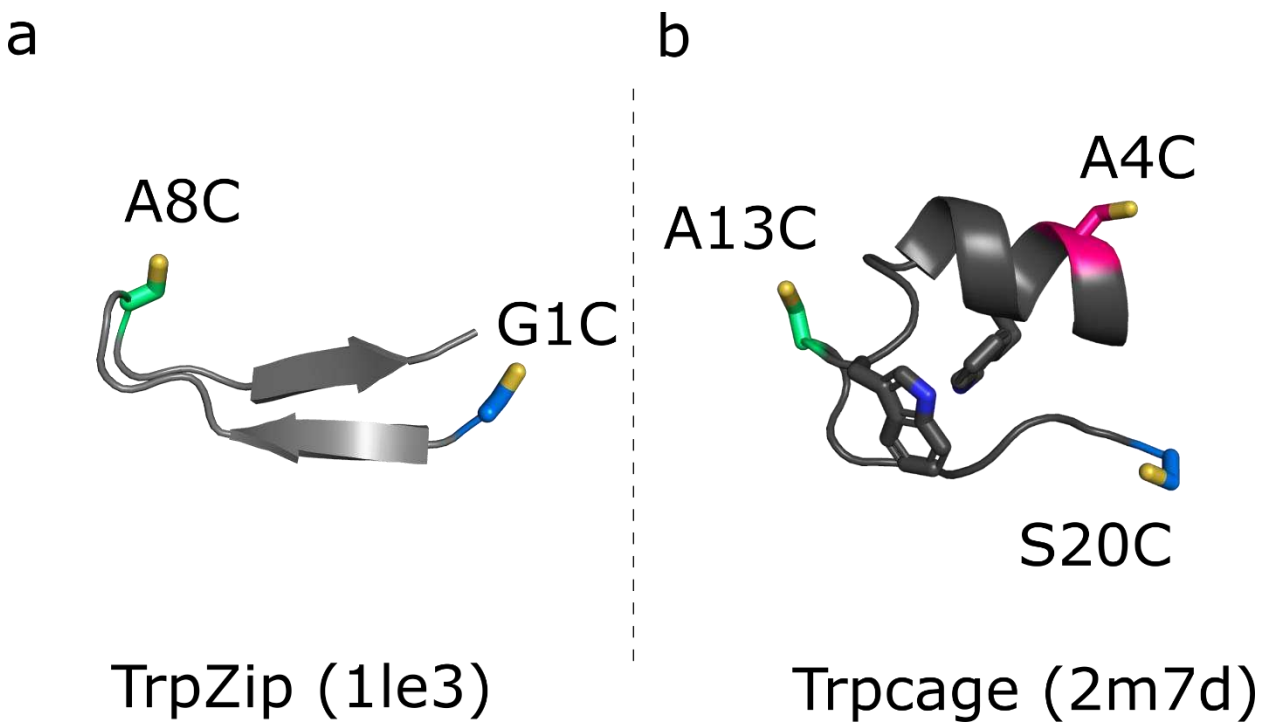



FIGURE 6.1 DESIGN OF ENGINEERED THIOLS ON MINI-PROTEINS
(a) Tryptophan zipper 2 (TrpZip) (PDB accession code 1le3) is a model 13-residue monomeric beta hairpin. Locations for single thiol mutations were selected at the N-terminus (G1C) or in the beta-turn (A8C). (b) TrpCage T2C16b P12W (TrpCage) (PDB accession code 2m7d) is an exceptionally stable model mini-protein. Single thiol sites were selected on the helix (A4C), the loop (A13C), or on the C-terminus (S20C). Between the model systems of TrpZip and Trpcage, we will have the opportunity to test a panel of protein design motifs.

To demonstrate scaffold assisted crystallography we selected TrpCage T2C16b P12W (TC), [202–205] a 20-residue thermostable mini-protein, and TrpZip 2 (TZ),[206] a 13-residue stable monomeric beta hairpin, as model guest mini-proteins. Both TC (2m7d) and TZ (1le3) have structures deposited in the PDB solved by NMR. We used structure guided design from existing NMR models to propose engineered thiols (**Figure 6.1** and **Table S6.1**). Rigid body clash analysis indicates that these designs can form disulfide bonds at 182C without clashing with the neighboring CJ backbone. The designs were obtained via Genscript at >95% purity.

### 6.2.2 MINIPROTEIN IN VITRO CONJUGATION AND CRYSTALLIZATION

Preliminarily, this study focused on two of the obtained peptides:TZ-A8C and TC-A4C. Each of the peptides were resuspended to 1 mM in conjugation buffer (20 mM $Na_2HPO_4$, 300 mM NaCl, 10% glycerol, pH 7.5) and reduced with 5 mM TCEP. The peptide was subsequently dialyzed with a 750 Da G2 Cassette (Fisher Scientific) to remove excess reducing agent. *In vitro* conjugation was performed with slight excess of reduced peptide to 182C-MNB (1.5 mol to 1 mol). Normalization of reaction yield was determined by addition of excess TCEP to N182C-MNB. A reaction yield of 100% and 97% was achieved after two hours for TC-A4C and TZ-A8C respectively (**Figure 6.2**). Conjugated products, 182C-TC-A4C and 182C-A8C were subsequently set up for crystallization.

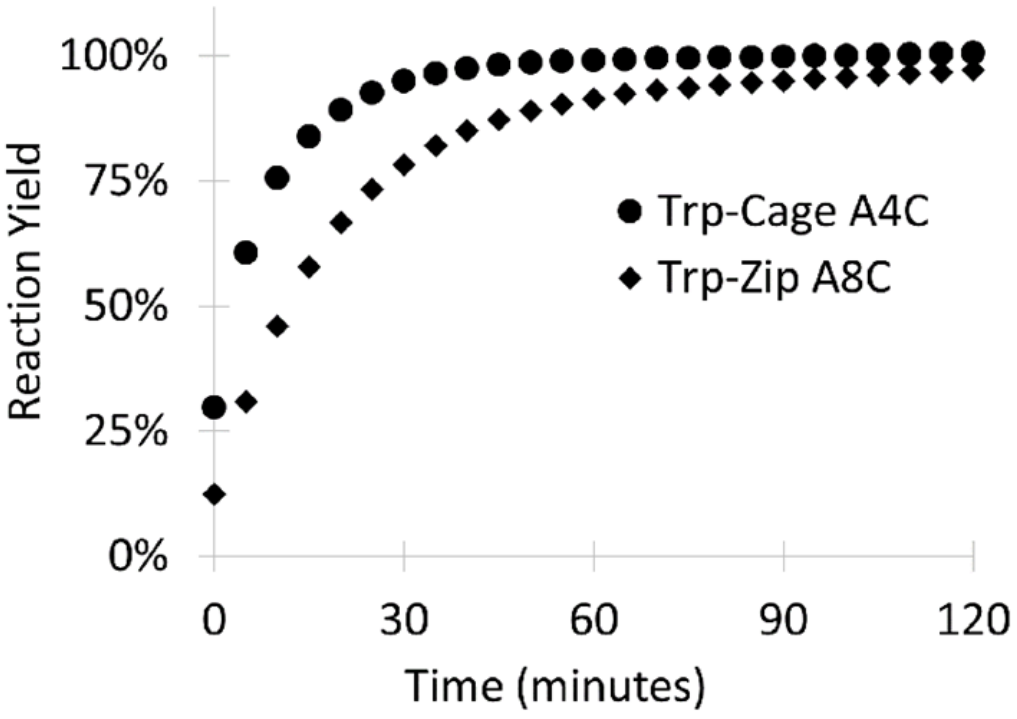



FIGURE 6.2 IN VITRO CONJUGATION OF MINI-PROTEINS TO 182C
The disulfide crosslink between 182C and engineered mini-proteins can be monitored by release of $TNB^{-2}$ anion which absorbs strongly at 412 nm. Absorbance at 412 nm was measured every 5 minutes to record time course data for conjugation between 182C and the target guest mini-protein. A reaction yield of 100% and 97% was achieved after 2 hours for TC-A4C and TZ-A8C respectively.

FIGURE 6.3 CJ-182C-TZ-A8C CRYSTAL HIT

Crystals of CJ-182C-TZ-A8C were obtained within a week with hexagonal morphology in a similar condition to the native CJ-182C condition (3.6 M ammonium sulfate, 100 mM Bis-TRIS, pH 6.0) (**Figure 6.3**). XRD was performed in two different cryoprotectant solutions: 3.2 M ammonium sulfate, 10% ethylene glycol, 100 M Bis-Tris, pH 6.0 (AS-EG-6) and 4.2 M trimethylamine N-oxide pH, 100 mM HEPES, pH 8.0 (TMAO-8). From these two datasets, it is clear that the intended disulfide bond formation is present (**Figure 6.4**a). However, interpretability of the electron density beyond the initial attachment point is difficult. Presumably, this is due to high flexibility of TZ-A8C when installed at 182C (Figure 6b &c). The resulting conjugate for 182C-TC-A4C has yet to crystallize.

If flexibility remains a problem, advanced data processing methods may be pursued. The PanDDA method has been used to create “ground-state” maps for discerning low occupancy and noisy electron density maps.[207] Considering the vast number of CJ native crystal structures we have collected, this may be a promising strategy. Collecting room-

temperature data and analyzing the diffuse scattering patterns may also help determine conformational ensembles.[208]

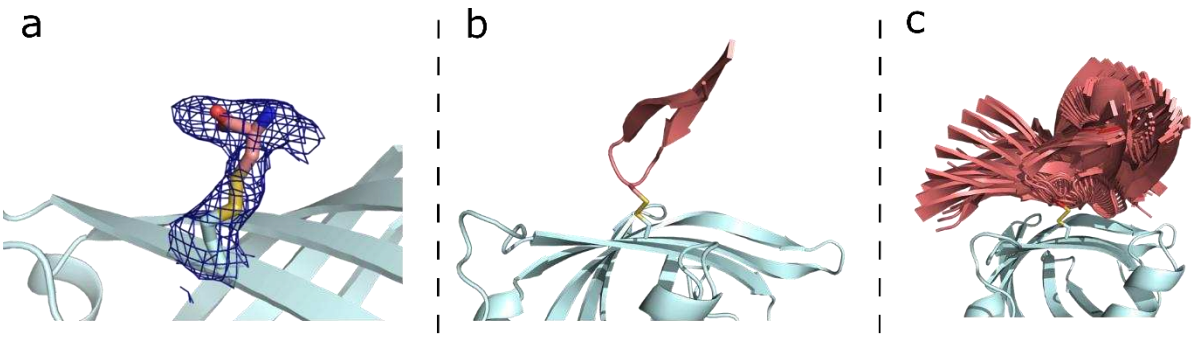



FIGURE 6.4 RESULTING ELECTRON DENSITY FOR 182C-TZ-A8C REVEALS FLEXIBLE ATTACHMENT
**(a)** 182C-TZ-A8C crystal in TMAO-8 reveals obvious disulfide modification in 2Fo-Fc map contoured at 1σ. Beyond this attachment point, the electron density is ambiguous. **(b)** An example conformation of a single TZ-A8C attached at 182C. **(c)** The set of possible TZ-A8C conformations at 182C from rigid body clash analysis.

### 6.2.3 IN CRYSTALLO CONJUGATION QUANTIFICATION

Although MNB had previously been installed at 182C on a reduced pre-existing crystal, crystallization of the pre-conjugated 182C-MNB had not yet been demonstrated. Despite covalent installation of the small molecule with functional groups, 182C-MNB crystallized under similar conditions (3.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris, pH 6.0) and with similar morphology to the native CJ-182C crystal. Addition of 10 mM 2-mercaptoethanol (BME) to N182C-MNB results in a strong yellow color indicative of $TNB^{-2}$ release (**Figure 6.5**).

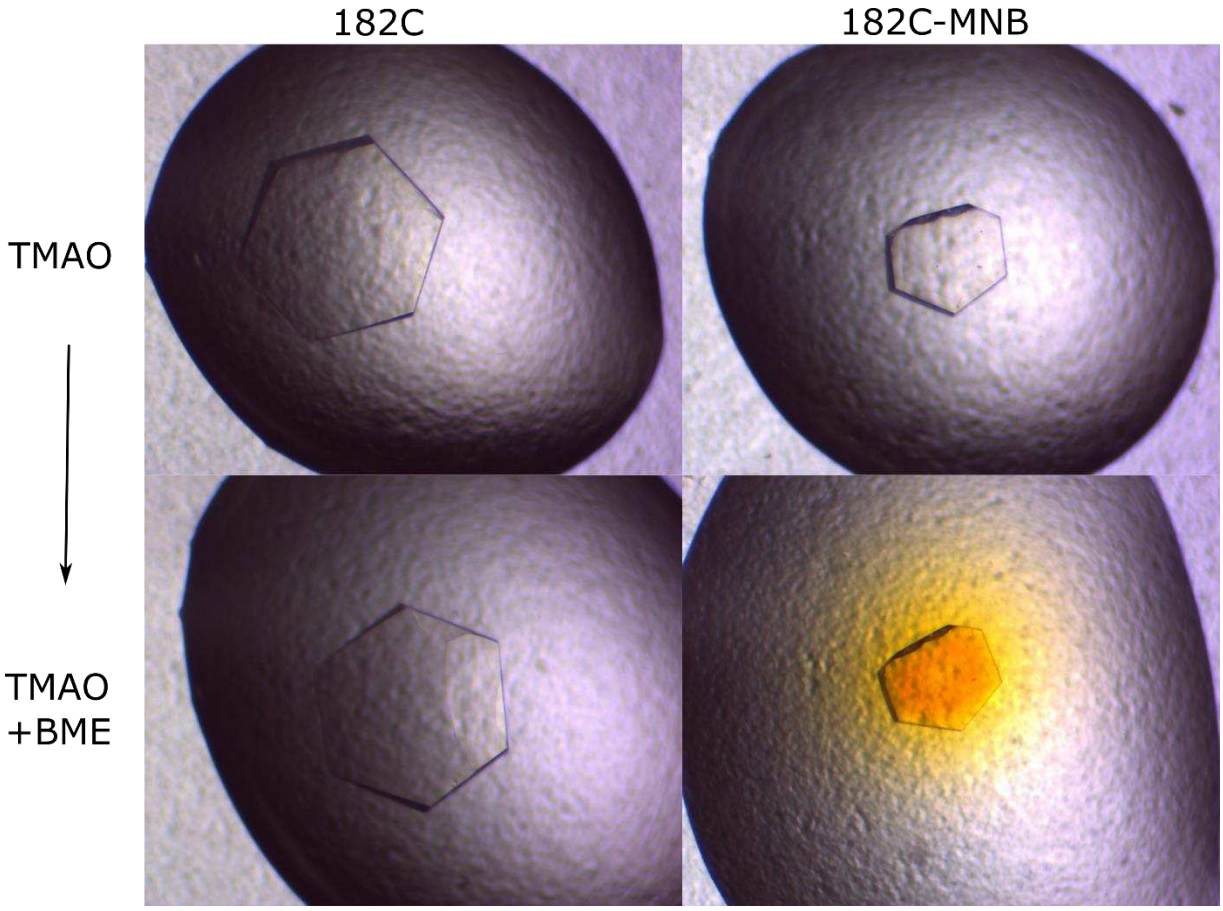



FIGURE 6.5 ADDITION OF BME TO 182C-MNB CRYSTALS
Addition of BME to 182C-MNB results in a strong yellow signal not observed in the crystal lacking MNB.

While the crystal contains a reasonable local concentration for quantification of thiol in the solid-state volume (~14 mM), the volume of the average crystal is typically small (~2 nL) in comparison to a typical solution working volume (>1 µL). While the signal could be amplified by adding more crystals to the solution, the volume of crystal varies significantly from crystal to crystal. Method development was needed to quantify scaffold thiols and conjugation yields *in crystallo*. The absorbance at 412 nm upon reduction of 182C-MNB is equivalent to the amount of crystalline thiol, with a single $TNB^{-2}$ molecule released per thiol. The crystals can be reduced and an end-point measurement can be performed after $TNB^{-2}$ equilibrates into the bulk solution. Reaction volumes should be kept minimal (1 µL)

to maximize the signal and the absorbance can be measured in a microdrop format in a Take3 microplate. Excess MNB-activated guest molecule (e.g. TZ-A8C-MNB) can be added to the reduced crystals and quantified, resulting in a single $TNB^{-2}$ molecule released per guest molecule conjugated to the scaffold thiols. Time-course data of absorbance at 412 nm can be collected by sandwiching the microdrop between two glass slides sealed with grease. Once the reaction yield plateaus, crystals can be collected and subjected to XRD.

While it might be tempting to simply add the reduced mini-protein to the 182C-MNB activated scaffold, the multi-step activated guest technique has advantages. First, this technique allows for reliable quantification that will enable solution screening to optimize reaction yield. Second, MNB activated guest molecules will afford easier to interpret electron density maps. In the case of MNB-activated scaffold, as the reaction yield progresses there is both loss in MNB occupancy and addition in guest molecule occupancy. Due to the potential for mixed occupancy products, modeling low reaction yield products will be extremely difficult. In contrast, with the reduced-scaffold activated guest technique, only additions of the guest molecule will be observed in the electron density. Third, this technique bypasses the laborious and imprecise measurement of crystal dimensions to determine crystal molarity in the drop. Finally, the reaction for this technique can be driven by excess MNB activated guest molecule.

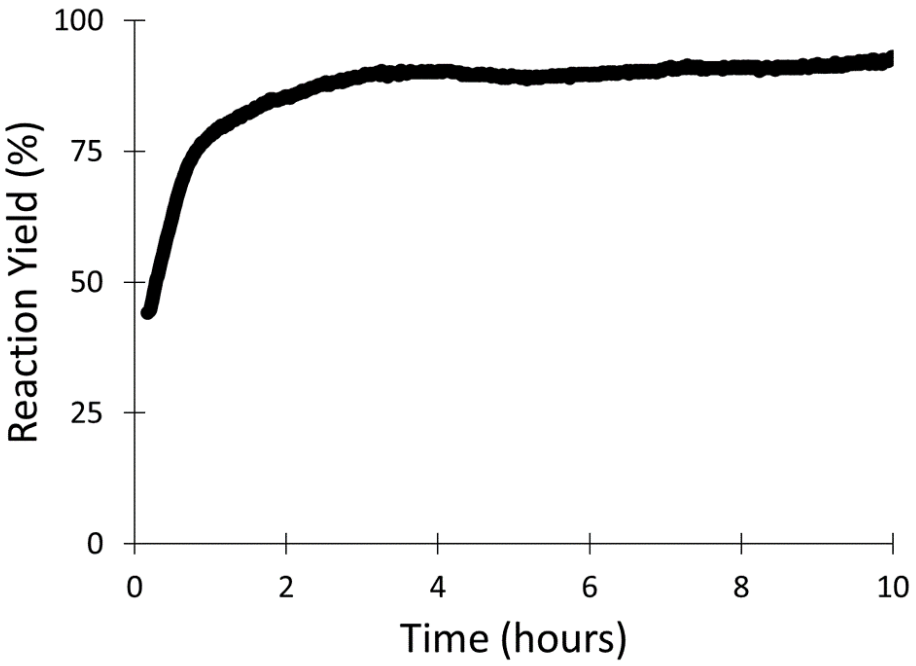



FIGURE 6.6 IN CRYSTALLO DTNB CONJUGATION TIME COURSE DATA
Five 182C-MNB crystals of approximate equal size (~100 µm radius x ~40 µm height) were reduced with 10 mM BME in a 1 µL volume. After the reaction equilibrated, the absorbance at 412 nm was recorded. The crystal was thoroughly washed and transferred to a solution containing 1 mM DTNB. Absorbance at 412 nm was recorded over time and normalized by the maximum signal from 182-MNB BME cleavage for determination of reaction yield.

As proof of concept, time-course data for DTNB reacting with preformed CJ-182C crystals was quantified (**Figure 6.6**). First, multiple 182C-MNB crystals were transferred to TMAO-8 with 10 mM BME in a 1 uL drop. After equilibration (~20 minutes), the crystals were removed and the absorbance was measured. Next, crystals were thoroughly washed to remove excess reducing reagent and $TNB^{-2}$. The reduced 182C crystals were then subsequently exposed to 1 mM DTNB. Absorbance at 412 nm was recorded over time. The reaction yield approached ~90% within four hours, which is significantly slower than that in solution. However, the reaction was performed on a non-crosslinked crystal in TMAO-8, which is a highly viscous solution. A complementary *in vitro* assay for 182C in solution can't be performed due to the inherent low solubility of the CJ in TMAO-8.

Undoubtedly, the *in crystallo* conjugation reaction kinetics would be faster in milder solution condition. Transfer to milder conditions will likely to be necessary for *in crystallo* conjugation guest proteins. At present, reactions in low salt solutions will require crosslinking the crystal. While we have developed robust methods for stabilizing native crystals, crosslinking will invariably complicate the method. Because most crosslinkers react with nucleophiles, the scaffold thiols will have to be protected. Additionally, the crosslinkers result in modified scaffolds that can result in complicated unresolved electron density maps. Solution screening with a poorly defined scaffold will likely complicate interpretation of guest molecule electron density. Future work optimizing the crystal lattice interface via protein engineering may reduce the constraints on solution conditions and enable more routine macromolecule scaffold assisted crystallography.

Nonetheless, the quantification of DTNB reaction yield in the crystal is the first of its kind and confirms observation of high occupancy conjugates from previous small molecule scaffold assisted crystal structures (Chapter 5). This technique should be transferrable to quantifying installation of guest mini-proteins activated with MNB. To further pursue this technique, methods for obtaining MNB activated mini-proteins will need to be established. The size difference between Ellman's reagent and TZ and TC are too close to be separated by standard dialysis or desalting methods. The preferred route forward would be to obtain peptide synthesized with an MNB activated thiol.

Initially it was presumed that working with miniproteins would be the path of least resistance for demonstrating scaffold assisted crystallography of a macromolecule. However, solving structures of mini-proteins such as TC and TZ may prove difficult due to lack of large surface areas to interact with the scaffold. A slightly larger guest protein,

such as B1 IgG binding domain of Protein G (56 residues), may be easier to work with in the future. Constructs for Protein G can be rapidly cloned and expressed with high yields in our lab at a lower cost than purchasing peptides. Also, anomalous residues such as selenomethionine can be introduced into Protein G to aid with structure determination attempts. Finally, Protein G offers more design space and surface contacts to interact with the scaffold.

One of the proposed shortcomings of the initial structure determination attempts with 182C-TZ-A8C was the amount of flexibility. *A priori* it was unclear if multiple conformations of the guest molecule was going to be inherently detrimental. For example, given a conformational search space, the target protein may anneal into a low energy minima state. Constraining the guest might reduce reaction yielding leading to a tradeoff with reaction yield. From preliminary structure guided design, multiple solvent accessible sites on the helix of Protein G are promising for constrained attachment. Using i+4 rules for designing on the surface of the helix, A24, K28, and Q32 were identified as residues for replacement with cysteine. Preliminary rigid body clash analysis demonstrates that these three sites have a wide range of conformational flexibility. For example, A24C can form a disulfide in many orientations without clashing. In contrast, K28C was unable to form a disulfide without resulting in a clash. The best design, Q32C, results in a minimum number of states in a single orientation (**Figure 6.7**). Cloning of these targets has recently begun (**Table S6.2**).

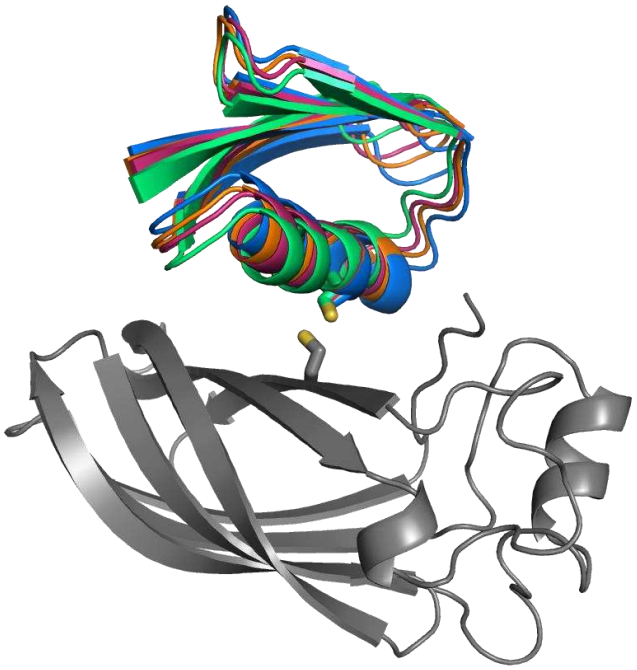


FIGURE 6.7 RIGID BODY CLASH ANALYSIS OF PROTEIN G-Q32C AT 182C
Rigid body clash analysis reveals Protein G Q35C has low conformational freedom when installed at 182C.

## 6.3 CONCLUSIONS

While the results of scaffold-assisted crystallography are still preliminary, there are promising leads. Surprisingly, conjugation of TZ-A8C to 182C did not disrupt crystallization, which enabled the first structures of a mini-protein conjugated to a porous protein scaffold. We suspect this pre-installation technique will be hit or miss with future targets. However, efforts are underway to create an engineered variant of CJ with superior resolution that crystallizes under milder conditions. A highly optimized CJ variant will likely enable more successes with pre-installation followed by crystallization.

Preliminary results from *in crystallo* conjugation experiments indicate that quantification of reaction yields in crystals is possible. Comparing measured *in crystallo* reaction yields

with occupancy in the electron density will be a useful next step. Finally, obtaining more guest macromolecules with engineered thiols will enable a shotgun approach. We suspect that acquisition of the first macromolecular scaffold assisted crystal structures are only a matter of time. From this first structure, we will be able to develop additional design principles. As design principles are developed, this strategy may ultimately become a routine option for high-throughput structure determination.

## 6.4 ADDITIONAL TABLES

TABLE S6.1 GENSCRIPT MINI-PROTEIN SYNTHESIS SEQUENCES

| Peptide | Peptide Sequence | Design Site Topology |
|---|---|---|
| Trpcage-A4C | DAYCQWLAD{D-ALA}GWAS{D-ALA}RPPPS | Helical |
| Trpcage-A13C | DAYAQWLAD{D-ALA}GWCS{D-ALA}RPPPS | Loop |
| Trpcage-S20C | DAYAQWLAD{D-ALA}GWAS{D-ALA}RPPPC | C-terminus |
| TrpZip-G1C | CEWTWDDATKTWTWTE | N-Terminus |
| TrpZip-A8C | GEWTWDDCTKTWTWTE | Loop |

TABLE S6.2 SEQUENCES FOR DESIGNED PROTEIN G VARIANTS

| | |
|---|---|
| **PG** | MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEHHHHHH |
| **PG A24C** | MTYKLILNGKTLKGETTTEAVDA**C**TAEKVFKQYANDNGVDGEWTYDDATKTFTVTEHHHHHH |
| **PG K28C** | MTYKLILNGKTLKGETTTEAVDAATAE**C**VFKQYANDNGVDGEWTYDDATKTFTVTEHHHHHH |
| **PG Q32C** | MTYKLILNGKTLKGETTTEAVDAATAEKVFK**C**YANDNGVDGEWTYDDATKTFTVTEHHHHHH |

# 7 CONCLUSIONS AND FUTURE WORK

The work herein takes steps towards a paradigm shift in elucidating dynamic protein structures. Successful implementation of scaffold assisted crystallography will provide a new opportunity to apply x-ray diffraction to study the structure of macromolecules that do not crystallize by themselves. The resulting structures can benefit society by accelerating biotechnology progress and therapeutic drug development. The research will provide a new tool for makers in many disciplines because the engineered scaffolds constitute an unparalleled nanotechnology platform with applications beyond structure determination. Proteins, oligonucleotides, and inorganic nanoparticles can be precisely positioned within such crystals for diverse research topics, including catalysis, separations, drug delivery, and biosensing. The scaffold assisted crystallography approach could complement conventional crystallization and methods that are on the horizon such as cryo-electron microscopy[209–211] and X-ray free electron laser (XFEL) methods.[212–215] As a result, there are still many fruitful threads to pursue for dedicated researchers to pursue. Recommendations for future work are presented in the following sections.

In ***Chapter 2***, algorithms were presented for identification of highly porous protein crystals. Many researchers perform structure determination with little care to crystal properties beyond the extent to which the target molecule is clearly resolved. Because new structures are continuously deposited, algorithms for mining the protein databank should be performed regularly to identify useful crystal forms. Alternative scaffolds such as those with high helical propensity (e.g. 4r32 and 4jol) and superior symmetry to CJ

(e.g. 3ufi) should be investigated by a future researcher. The lessons learned from method development for CJ should aid future researchers in rapidly developing new highly porous crystalline scaffolds.

Additionally, in ***Chapter 2***, method development was demonstrated for a well behaved highly porous protein crystal. Further fine tuning of CJ production would still be rewarding, especially regarding scaling up will . Performing expression in alternative systems such as *Pichia pastoris* may provide enhanced titres of >1 g/L.[216] Bulk crystal production could be obtained by further optimization and scaling up of initial batch crystallization experiments. Optimization of the scaffold through surface entropy reduction techniques, deletion of flexible regions, and optimization of crystallographic interfaces may result in reduced B-factors, improved resolution, and higher tolerance to solution changes. A domain swap fusion may aid in enhanced crystallizability and reduced symmetry. A highly optimized scaffold could have enhanced crystallizability of guests conjugated prior to crystallization.

In ***Chapter 3***, methods were developed for chemically crosslinking CJ crystals. While glutaraldehyde was effective at stabilizing the solid crystalline state, it resulted in a substantial reduction in diffraction resolution. Formaldehyde, glyoxal, and EDC, were demonstrated to be effective at both stabilizing the crystal and preserving resolution. Furthermore, high occupancy end products could be resolved via X-ray diffraction (XRD). Increased resolution from a synchrotron may assist in elucidation of end-products. Additional techniques such as mass spectrometry may assist with revealing modifications. Crosslinking CJ crystals has been determined to be highly sensitive to all solution components, not just the active reagent. Robust analysis of crosslinking

modifications as a function of pH may promote desired crosslinked structures or higher resolution products. Demonstration of alternative crosslinkers on better known model protein crystals such as hen egg white lysozyme (HEWL), sperm whale myoglobin, xylanase, or glucose isomerase may promote work by demonstrating generalizability. The higher resolution provided by these alternate crystals may allow for more interpretable crosslinks. Demonstrating a panel of more diverse crosslinking, such as demonstrating bis NHS-ester protein crystal crosslinking or imidoesters with varying length, may allow for crosslinks at new sites that could result in enhanced stability. Similarly, adding crosslinking extenders, such as dihydrazides, may promote new crosslinks. In the same vein, introducing cleavable extenders such as cystamine may aid in generating crosslinks that could easily be degraded. Designing orthogonal site-specific crosslinks with thiols or non-natural amino acids may allow more control and designability. Quenching and functionalization may be a productive route for altering bulk material properties. Overall, this is a very fruitful tree for development that could create highly citable papers that will aid many researchers.

In ***Chapter 4***, controlled loading, binding, and release of guest fluorescent proteins were developed. This was an important step in demonstrating the ability to control macromolecules with a porous protein scaffold. More sophisticated controlled binding and release of macromolecules could lead to a plethora of applications. Specifically, enzymes, such as horseradish peroxidase, could be loaded into the crystal to demonstrate enzymes in a porous protein crystal. Enzymes could be coloaded into the crystal for enhanced reaction kinetics by channeling.[217] The split GFP system could be utilized on the porous scaffold to create solid state biosensors.[218] Sophisticated

biosensors could be created by taking advantage of our spatial segregation techniques and novel crystalline shell materials.

In ***Chapter 5***, scaffold assisted crystallography was demonstrated with small molecules. Specifically, small molecules were conjugated to engineered thiols on a preformed crystal. While attachment was obvious, this work was limited by crystal resolution and flexibility of the target molecules. This work could be extended by demonstrating the effect of cryoprotectant solution on conjugate conformational changes. Engineering low entropy surface patches or hydrophobic residues may result in lower flexibility conjugates. Extending work to functional molecules, such as transition metal conjugates, may be useful for creating biohybrid materials. While this was an important proof of principle for demonstrating scaffold assisted crystallography with CJ, there is no reason for future small molecule work to use a protein crystal with such high solvent. Engineering a higher resolution crystal, such as hen egg white lysozyme, for small molecule assisted crystallography may afford better resolved conjugates.

In ***Chapter 6***, preliminary studies for scaffold assisted crystallography of macromolecules were performed. Future efforts will focus on conjugating a diverse panel of mini-proteins to multiple installation sites and attempting to crystallize the resulting conjugates. This will elucidate the effects of flexibility and the ordering of guest molecules on the surface. In lieu of obtaining ordered guests via pre-installed crystallization, pursuing *in crystallo* conjugation with highly constrained adducts, such as Protein G Q32C, is a top priority. Designs that fail to conjugate with a simple disulfide *in crystallo* could be rescued by conjugation with bifunctional linker molecules such as dibromobimane or dibromomaleimide.[219,220] For low occupancy and flexible structures, the PanDDa method

will be pursued. If flexibility of the guest macromolecule remains problematic, dual anchor attachment strategies may be pursued to decrease conformational flexibility. Additionally, crosslinking the conjugate may capture and resolve transient states. The list of existing ideas for this thread is quite long, providing many possibilities for the next generation of graduate students.

BIBLIOGRAPHY


1. Levitt, M. Nature of the protein universe. *Proc. Natl. Acad. Sci.* **106,** 11079–11084 (2009).
2. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **37,** D26–D31 (2009).
3. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
4. Khafizov, K., Madrid-Aliste, C., Almo, S. C. & Fiser, A. Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. *Proc. Natl. Acad. Sci.* **111,** 3733–3738 (2014).
5. Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24,** 409–411 (1991).
6. Grabowski, M., Niedzialkowska, E., Zimmerman, M. D. & Minor, W. The Impact of Structural Genomics: the First Quindecennial. *J. Struct. Funct. Genomics* **17,** 1–16 (2016).
7. Protein Structure Initiative Structural Biology Knowledgebase. Metrics Describing Progress of the Protein Structure Initiative. Available at: http://sbkb.org/metrics/milestonestables.html.
8. Mcpherson, A. A comparison of salts for the crystallization of macromolecules. *Protein Sci. Publ. Protein Soc.* **10,** 418–422 (2001).
9. Hampton Research. Available at: http://hamptonresearch.com/product_detail.aspx?sid=74&pid=225.

10. Larson, S. B., Day, J. S., Cudney, R. & McPherson, A. A novel strategy for the crystallization of proteins: X-ray diffraction validation. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 310–318 (2007).
11. McPherson, A. & Cudney, B. Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *J. Struct. Biol.* **156,** 387–406 (2006).
12. Hampton Research. Available at: https://hamptonresearch.com/product_detail.aspx?cid=30&sid=179&pid=562.
13. Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J. Struct. Biol.* **142,** 88–97 (2003).
14. Bucher, M. H., Evdokimov, A. G. & Waugh, D. S. Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 392–397 (2002).
15. Carson, M., Johnson, D. H., McDonald, H., Brouillette, C. & DeLucas, L. J. His-tag impact on structure. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 295–301 (2007).
16. Waugh, D. S. An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expr. Purif.* **80,** 283–293 (2011).
17. Gräslund, S. *et al.* The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr. Purif.* **58,** 210–221 (2008).
18. Yeh, J. I. *et al.* High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. *J. Mol. Biol.* **262,** 186–201 (1996).

19. Dale, G. E., Kostrewa, D., Gsell, B., Stieger, M. & D'Arcy, A. Crystal engineering: deletion mutagenesis of the 24 kDa fragment of the DNA gyrase B subunit from Staphylococcus aureus. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1626–1629 (1999).
20. Schwartz, T. U., Walczak, R. & Blobel, G. Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor β subunit. *Protein Sci. Publ. Protein Soc.* **13,** 2814–2818 (2004).
21. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4|[thinsp]||[Aring]| resolution. *Nature* **395,** 347 (1998).
22. Taneja, B., Patel, A., Slesarev, A. & Mondragón, A. Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases. *EMBO J.* **25,** 398–408 (2006).
23. Gheyi, T., Rodgers, L., Romero, R., Sauder, J. M. & Burley, S. K. Mass Spectrometry Guided In Situ Proteolysis to Obtain Crystals for X-ray Structure Determination. *J. Am. Soc. Mass Spectrom.* **21,** 1795–1801 (2010).
24. Dong, A., Xu, X. & Edwards, A. M. In situ proteolysis for protein crystallization and structure determination. *Nat. Methods* **4,** 1019–1021 (2007).
25. Holden, H. M. *et al.* Molecular Structure of a Proteolytic Fragment of TLP20. *Acta Crystallogr. D Biol. Crystallogr.* **52,** 1153–1160 (1996).
26. Mandel, C. R., Gebauer, D., Zhang, H. & Tong, L. A serendipitous discovery that in situ proteolysis is essential for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **62,** 1041–1045 (2006).

27. Orth, P., Jekow, P., Alonso, J. C. & Hinrichs, W. Proteolytic cleavage of Gram-positive β recombinase is required for crystallization. *Protein Eng. Des. Sel.* **12,** 371–373 (1999).

28. Derewenda, Z. S. & Vekilov, P. G. Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 116–124 (2006).

29. Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 679–688 (2001).

30. Mateja, A. *et al.* The impact of Glu→Ala and Glu→Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1983–1991 (2002).

31. Czepas, J. *et al.* The impact of Lys-->Arg surface mutations on the crystallization of the globular domain of RhoGDI. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 275–280 (2004).

32. Cooper, D. R. *et al.* Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 636–645 (2007).

33. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci. Publ. Protein Soc.* **16,** 1569–1576 (2007).

34. Derewenda, Z. S. It's all in the crystals…. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 243–248 (2011).

35. Bauman, J. D. *et al.* Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acids Res.* **36,** 5083–5092 (2008).
36. Munshi, S., Hall, D. L., Kornienko, M., Darke, P. L. & Kuo, L. C. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 1725–1730 (2003).
37. Jin, L. *et al.* Mutation of surface residues to promote crystallization of activated factor XI as a complex with benzamidine: an essential step for the iterative structure-based design of factor XI inhibitors. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 1418–1425 (2005).
38. Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **12,** 529–535 (2004).
39. Price, W. N. *et al.* Understanding the physical properties controlling protein crystallization based on analysis of large-scale experimental data. *Nat. Biotechnol.* **27,** 51–57 (2009).
40. Tan, K. *et al.* Salvage of Failed Protein Targets by Reductive Alkylation. *Methods Mol. Biol. Clifton NJ* **1140,** 189–200 (2014).
41. Rayment, I. [12] Reductive alkylation of lysine residues to alter crystallization properties of proteins. in *Methods in Enzymology* **276,** 171–179 (Academic Press, 1997).
42. Rayment, I. *et al.* Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261,** 50–58 (1993).
43. Kobayashi, M., Kubota, M. & Matsuura, Y. Crystallization and improvement of crystal quality for X-ray diffraction of maltooligosyl trehalose synthase by reductive

methylation of lysine residues. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 931–933 (1999).

44. Walter, T. S. *et al.* Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **14,** 1617–1622 (2006).
45. Fan, Y. & Joachimiak, A. Enhanced crystal packing due to solvent reorganization through reductive methylation of lysine residues in oxidoreductase from Streptococcus pneumoniae. *J. Struct. Funct. Genomics* **11,** 101–111 (2010).
46. Sledz, P. *et al.* New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. *Protein Sci. Publ. Protein Soc.* **19,** 1395–1404 (2010).
47. Kim, Y. *et al.* Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat. Methods* **5,** 853–854 (2008).
48. Ghosh, E., Kumari, P., Jaiman, D. & Shukla, A. K. Methodological advances: the unsung heroes of the GPCR structural revolution. *Nat. Rev. Mol. Cell Biol.* **16,** 69–81 (2015).
49. Kovari, L. C., Momany, C. & Rossmann, M. G. The use of antibody fragments for crystallization and structure determinations. *Structure* **3,** 1291–1293 (1995).
50. Hunte, C. & Michel, H. Crystallisation of membrane proteins mediated by antibody fragments. *Curr. Opin. Struct. Biol.* **12,** 503–508 (2002).
51. Rasmussen, S. G. F. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–387 (2007).
52. Day, P. W. *et al.* A monoclonal antibody for G protein-coupled receptor crystallography. *Nat. Methods* **4,** 927–929 (2007).

53. Bradbury, A. R. M. & Marks, J. D. Antibodies from phage antibody libraries. *J. Immunol. Methods* **290,** 29–49 (2004).

54. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23,** 1105–1116 (2005).

55. Muyldermans, S. Single domain camel antibodies: current status. *Rev. Mol. Biotechnol.* **74,** 277–302 (2001).

56. Harmsen, M. M. & De Haard, H. J. Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* **77,** 13–22 (2007).

57. Barthelemy, P. A. *et al.* Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. *J. Biol. Chem.* **283,** 3639–3654 (2008).

58. Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat. Protoc.* **9,** 674–693 (2014).

59. Tereshko, V. *et al.* Toward chaperone-assisted crystallography: Protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Sci. Publ. Protein Soc.* **17,** 1175–1187 (2008).

60. Loris, R. *et al.* Crystal structure of the intrinsically flexible addiction antidote MazE. *J. Biol. Chem.* **278,** 28252–28257 (2003).

61. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β(2) adrenoceptor. *Nature* **469,** 175–180 (2011).

62. Steyaert, J. & K Kobilka, B. Nanobody stabilization of G protein coupled receptor conformational states. *Curr. Opin. Struct. Biol.* **21,** 567–572 (2011).

63. Chaperone-Assisted Crystallography with DARPins. *Structure* **16,** 1443–1453 (2008).
64. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **332,** 489–503 (2003).
65. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22,** 575–582 (2004).
66. al, B. A., et. DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. - PubMed - NCBI. Available at: https://secure.colostate.edu/pubmed/,DanaInfo=www.ncbi.nlm.nih.gov,SSL+26975886. (Accessed: 29th September 2017)
67. Kim, Y. *et al.* Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* **5,** 111–118 (2004).
68. Waugh, D. S. Making the most of affinity tags. *Trends Biotechnol.* **23,** 316–320 (2005).
69. Kapust, R. B. & Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674 (1999).
70. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci. Publ. Protein Soc.* **12,** 1313–1322 (2003).
71. Waugh, D. S. Crystal structures of MBP fusion proteins. *Protein Sci. Publ. Protein Soc.* **25,** 559–571 (2016).

72. Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci. Publ. Protein Soc.* **19,** 901–913 (2010).
73. Cherezov, V. *et al.* High Resolution Crystal Structure of an Engineered Human β2-Adrenergic G protein-Coupled Receptor. *Science* **318,** 1258–1265 (2007).
74. Rosenbaum, D. M. *et al.* GPCR Engineering Yields High-Resolution Structural Insights into β2-Adrenergic Receptor Function. *Science* **318,** 1266–1273 (2007).
75. Inokuma, Y. *et al.* X-ray analysis on the nanogram to microgram scale using porous complexes. *Nature* **495,** 461–466 (2013).
76. Ledford, H. Controversial molecular-analysis tool tries for a comeback. *Nature* (2015). doi:10.1038/nature.2015.17702
77. Lee, S., Kapustin, E. A. & Yaghi, O. M. Coordinative alignment of molecules in chiral metal-organic frameworks. *Science* **353,** 808–811 (2016).
78. Deng, H. *et al.* Large-Pore Apertures in a Series of Metal-Organic Frameworks. *Science* **336,** 1018–1023 (2012).
79. Seeman, N. C. Nanomaterials Based on DNA. *Annu. Rev. Biochem.* **79,** 65–87 (2010).
80. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
81. Seeman, N. C. Structural DNA Nanotechnology: Growing Along with NanoLetters. *Nano Lett.* **10,** 1971–1978 (2010).
82. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461,** 74–77 (2009).

83. Paukstelis, P. J., Nowakowski, J., Birktoft, J. J. & Seeman, N. C. Crystal structure of a continuous three-dimensional DNA lattice. *Chem. Biol.* **11,** 1119–1126 (2004).

84. Paukstelis, P. J. Three-dimensional DNA crystals as molecular sieves. *J. Am. Chem. Soc.* **128,** 6794–6795 (2006).

85. Geng, C. & Paukstelis, P. J. DNA Crystals as Vehicles for Biocatalysis. *J. Am. Chem. Soc.* **136,** 7817–7820 (2014).

86. Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **43,** 661–672 (2010).

87. Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4,** 375–382 (2012).

88. Lanci, C. J. *et al.* Computational design of a protein crystal. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1112595109

89. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).

90. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **336,** 1129–1129 (2012).

91. Lai, Y.-T., King, N. P. & Yeates, T. O. Principles for designing ordered protein assemblies. *Trends Cell Biol.* **22,** 653–661 (2012).

92. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 (2014).

93. Nguyen, H. B., Hung, L. W., Yeates, T. O., Terwilliger, T. C. & Waldo, G. S. Split green fluorescent protein as a modular binding partner for protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 2513–2523 (2013).

94. Leibly, D. J. *et al.* A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **23,** 1754–1768 (2015).

95. Malo, J. *et al.* Engineering a 2D Protein–DNA Crystal. *Angew. Chem. Int. Ed.* **44,** 3057–3061 (2005).

96. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6,** 558–562 (2011).

97. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348,** 1365–1368 (2015).

98. About Us. *Crysalin*

99. Elucidation of protein structure. *Crysalin*

100. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 72–79 (2007).

101. Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F. & Carpenter, E. P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr. Sect. Struct. Biol.* **72,** 303–318 (2016).

102. Mizutani, R. *et al.* Spatiotemporal development of soaked protein crystal. *Sci. Rep.* **4,** srep05731 (2014).

103. Knäblein, J. *et al.* Ta6Br122+, a tool for phase determination of large biological assemblies by X-ray crystallography11Edited by K.Nagai. *J. Mol. Biol.* **270,** 1–7 (1997).

104. Dahms, S. O. *et al.* Localization and orientation of heavy-atom cluster compounds in protein crystals using molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 284–297 (2013).

105. Takeda, Y., Kondow, T. & Mafuné, F. Self-assembly of gold nanoparticles in protein crystal. *Chem. Phys. Lett.* **504,** 175–179 (2011).

106. Wei, H. *et al.* Time-dependent, protein-directed growth of gold nanoparticles within a single crystal of lysozyme. *Nat Nano* **6,** 93–97 (2011).

107. Muskens, O. L., England, M. W., Danos, L., Li, M. & Mann, S. Plasmonic Response of Ag- and Au-Infiltrated Cross-Linked Lysozyme Crystals. *Adv. Funct. Mater.* n/a–n/a (2012). doi:10.1002/adfm.201201718

108. Cormier, C. Y. *et al.* Protein Structure Initiative Material Repository: an open shared public resource of structural genomics plasmids for the biological community. *Nucleic Acids Res.* **38,** D743–D749 (2010).

109. Miller, J. L. *et al.* Selective Ligand Recognition by a Diversity-Generating Retroelement Variable Protein. *PLoS Biol.* **6,** (2008).

110. Vanarotti, M. S., Miller, D. J., Guibao, C. D., Nourse, A. & Zheng, J. J. Structural and mechanistic insights into the interaction between Pyk2 and Paxillin LD motifs. *J. Mol. Biol.* **426,** 3985–4001 (2014).

111. Sun, X.-J. *et al.* A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature* **500,** 93–97 (2013).

112. Fletcher, J. M. *et al.* A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth. Biol.* **1,** 240–250 (2012).

113. Woolfson, D. N. Coiled-Coil Design: Updated and Upgraded. in *Fibrous Proteins: Structures and Mechanisms* 35–61 (Springer, Cham, 2017). doi:10.1007/978-3-319-49674-0_2

114. Hales, T. C. The Honeycomb Conjecture. *Discrete Comput. Geom.* **25,** 1–22 (2001).

115. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2,** 953–971 (2007).

116. Quan, S., Hiniker, A., Collet, J.-F. & Bardwell, J. C. A. Isolation of bacteria envelope proteins. *Methods Mol. Biol. Clifton NJ* **966,** 359–366 (2013).

117. Wingfield, P. T. Protein Precipitation Using Ammonium Sulfate. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **APPENDIX 3,** Appendix-3F (2001).

118. Mcpherson, A. A comparison of salts for the crystallization of macromolecules. *Protein Sci. Publ. Protein Soc.* **10,** 418–422 (2001).

119. Marshall, H., Venkat, M., Hti Lar Seng, N. S., Cahn, J. & Juers, D. H. The use of trimethylamine N-oxide as a primary precipitating agent and related methylamine osmolytes as cryoprotective agents for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 69–81 (2012).

120. Koizumi, H. *et al.* Crystallization Technique for Strain-free Protein Crystals Using Cross-linked Seed Crystals. *Cryst. Growth Des.* **16,** 6089–6094 (2016).

121. Margolin, A. L. & Navia, M. A. Protein Crystals as Novel Catalytic Materials. *Angew. Chem. Int. Ed.* **40,** 2204–2222 (2001).

122. Abe, S. & Ueno, T. Design of protein crystals in the development of solid biomaterials. *RSC Adv.* **5,** 21366–21375 (2015).

123. Salgado, E. N. *et al.* Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci.* **107,** 1827–1832 (2010).

124. Heinz, D. W. & Matthews, B. W. Rapid crystallization of T4 lysozyme by intermolecular disulfide cross-linking. *Protein Eng.* **7,** 301–307 (1994).

125. Quistgaard, E. M. A disulfide polymerized protein crystal. *Chem. Commun. Camb. Engl.* **50,** 14995–14997 (2014).

126. Quiocho, F. A. & Richards, F. M. INTERMOLECULAR CROSS LINKING OF A PROTEIN IN THE CRYSTALLINE STATE: CARBOXYPEPTIDASE-A*. *Proc. Natl. Acad. Sci. U. S. A.* **52,** 833–839 (1964).

127. St. Clair, N. L. & Navia, M. A. Cross-linked enzyme crystals as robust biocatalysts. *J. Am. Chem. Soc.* **114,** 7314–7316 (1992).

128. Jegan Roy, J. & Emilia Abraham, T. Strategies in Making Cross-Linked Enzyme Crystals. *Chem. Rev.* **104,** 3705–3722 (2004).

129. Roy, J. J. & Abraham, T. E. Continuous biotransformation of pyrogallol to purpurogallin using cross-linked enzyme crystals of laccase as catalyst in a packed-bed reactor. *J. Chem. Technol. Biotechnol.* **81,** 1836–1839 (2006).

130. England, M. W. *et al.* Fabrication of polypyrrole nano-arrays in lysozyme single crystals. *Nanoscale* **4,** 6710–6713 (2012).

131. Reményi, A., Pohl, E., Schöler, H. R. & Wilmanns, M. Crystallization of redox-insensitive Oct1 POU domain with different DNA-response elements. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1634–1638 (2001).

132. Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques* **37,** 790–796, 798–802 (2004).

133. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9455–9460 (2014).

134. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **25,** 663–682 (2006).

135. Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes. *J. Biol. Chem.* **290,** 26404–26411 (2015).

136. Metz, B. *et al.* Identification of Formaldehyde-induced Modifications in Proteins REACTIONS WITH MODEL PEPTIDES. *J. Biol. Chem.* **279,** 6235–6243 (2004).

137. Whipple, E. B. Structure of glyoxal in water. *J. Am. Chem. Soc.* **92,** 7183–7186 (1970).

138. Odani, H. *et al.* Imidazolium crosslinks derived from reaction of lysine with glyoxal and methylglyoxal are increased in serum proteins of uremic patients: evidence for increased oxidative stress in uremia. *FEBS Lett.* **427,** 381–385 (1998).

139. Chellan, P. & Nagaraj, R. H. Protein Crosslinking by the Maillard Reaction: Dicarbonyl-Derived Imidazolium Crosslinks in Aging and Diabetes. *Arch. Biochem. Biophys.* **368,** 98–104 (1999).

140. Williams, A. & Ibrahim, I. T. Carbodiimide chemistry: recent advances. *Chem. Rev.* **81,** 589–636 (1981).

141. Gilles, M. A., Hudson, A. Q. & Borders, C. L. Stability of water-soluble carbodiimides in aqueous solution. *Anal. Biochem.* **184,** 244–248 (1990).

142. Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392,** 305–312 (2008).

143. Lusty, C. J. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J. Appl. Crystallogr.* **32,** 106–112 (1999).

144. Huber, T. R., Hartje, L. F., McPherson, E. C., Kowalski, A. E. & Snow, C. D. Programmed Assembly of Host–Guest Protein Crystals. *Small* 10.1002/smll.201602703 (2016). doi:10.1002/smll.201602703

145. Kowalski, A. E. *et al.* Gold nanoparticle capture within protein crystal scaffolds. *Nanoscale* **8,** 12693–12696 (2016).

146. Marshall, H., Venkat, M., Seng, N. S. H. L., Cahn, J. & Juers, D. H. The use of trimethylamine N-oxide as a primary precipitating agent and related methylamine osmolytes as cryoprotective agents for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 69–81 (2012).

147. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci. Publ. Protein Soc.* **12,** 1865–1871 (2003).

148. Andersson, K. M. & Hovmöller, S. The protein content in crystals and packing coefficients in different space groups. *Acta Crystallogr. D Biol. Crystallogr.* **56,** 789–790 (2000).

149. Chruszcz, M. *et al.* Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? *Protein Sci. Publ. Protein Soc.* **17,** 623–632 (2008).

150. Wine, Y., Cohen-Hadar, N., Freeman, A. & Frolow, F. Elucidation of the mechanism and end products of glutaraldehyde crosslinking reaction by X-ray structure analysis. *Biotechnol. Bioeng.* **98,** 711–718 (2007).

151. Hermanson, G. T. *Bioconjugate Techniques*. (Academic Press, 2013).

152. Feynman, R. P. There's Plenty of Room at the Bottom. *Eng. Sci.* 22–36 (1960).

153. St. Clair, N. L. & Navia, M. A. Cross-linked enzyme crystals as robust biocatalysts. *J Am Chem Soc* **114,** 7314–7316 (1992).

154. Zelinski, T. & Waldmann, H. Cross-Linked Enzyme Crystals (CLECs): Efficient and Stable Biocatalysts for Preparative Organic Chemistry. *Angew. Chem. Int. Ed. Engl.* **36,** 722–724 (1997).

155. Govardhan, C. P. Crosslinking of enzymes for improved stability and performance. *Curr. Opin. Biotechnol.* **10,** 331–335 (1999).

156. Beaudoin, D., Maris, T. & Wuest, J. D. Constructing monocrystalline covalent organic networks by polymerization. *Nat. Chem.* **5,** 830–834 (2013).

157. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461,** 74–77 (2009).

158. D'Arcy, Stihle, M., Kostrewa, D. & Dale, G. Crystal engineering: a case study using the 24 kDa fragment of the DNA gyrase B subunit from Escherichia coli. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1623–1625 (1999).

159. Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J. Struct. Biol.* **142,** 88–97 (2003).

160. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10,** 407–409 (2013).

161. Kowalski, A. E. *et al.* Gold nanoparticle capture within protein crystal scaffolds. *accepted*

162. Evers, T. H., Appelhof, M. A. M., Meijer, E. W. & Merkx, M. His-tags as Zn(II) binding motifs in a protein-based fluorescent sensor. *Protein Eng. Des. Sel.* **21,** 529–536 (2008).

163. Peng, L. Controlled Nucleation of His-tag Protein Assemblies Using Rigid Symmetric Multivalent Nitrilotriacetic acid Chelating Ligands. (Purdue University, 2012).

164. Pandya, M. J. *et al.* Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry (Mosc.)* **39,** 8728–8734 (2000).

165. Potekhin, S. A. *et al.* De novo design of fibrils made of short alpha-helical coiled coil peptides. *Chem. Biol.* **8,** 1025–1032 (2001).

166. Ogihara, N. L. *et al.* Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 1404–1409 (2001).

167. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2217–2221 (2001).

168. Douglas, T. & Young, M. Viruses: Making Friends with Old Foes. *Science* **312,** 873–875 (2006).

169. Lee, S.-Y., Lim, J.-S. & Harris, M. T. Synthesis and application of virus-based hybrid nanomaterials. *Biotechnol. Bioeng.* **109,** 16–30 (2012).

170. Moll, D. *et al.* S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl. Acad. Sci.* **99,** 14646–14651 (2002).

171. Sleytr, U. B., Egelseer, E. M., Ilk, N., Pum, D. & Schuster, B. S-Layers as a basic building block in a molecular construction kit. *FEBS J.* **274,** 323–334 (2007).

172. Baneyx, F. & Matthaei, J. F. Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotechnol.* **28,** 39–45 (2014).

173. Wei, H. *et al.* Time-dependent, protein-directed growth of gold nanoparticles within a single crystal of lysozyme. *Nat. Nanotechnol.* **6,** 93–97 (2011).

174. McElroy, H. E., Sisson, G. W., Schoettlin, W. E., Aust, R. M. & Villafranca, J. E. Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. *J. Cryst. Growth* **122,** 265–272 (1992).

175. Koshiyama, T. *et al.* Modification of porous protein crystals in development of biohybrid materials. *Bioconjug. Chem.* **21,** 264–269 (2010).

176. Forsythe, E., Achari, A. & Pusey, M. L. Trace fluorescent labeling for high-throughput crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 339–346 (2006).

177. Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F. & Carpenter, E. P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr. Sect. Struct. Biol.* **72,** 303–318 (2016).

178. Sun, D. P., Alber, T., Bell, J. A., Weaver, L. H. & Matthews, B. W. Use of site-directed mutagenesis to obtain isomorphous heavy-atom derivatives for protein crystallography: cysteine-containing mutants of phage T4 lysozyme. *Protein Eng.* **1,** 115–123 (1987).

179. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82,** 70–77 (1959).

180. Riddles, P. W., Blakeley, R. L. & Zerner, B. Reassessment of Ellman's reagent. *Methods Enzymol.* **91,** 49–60 (1983).

181. Winther, J. R. & Thorpe, C. Quantification of Thiols and Disulfides. *Biochim. Biophys. Acta* **1840,** (2014).

182. Beld, J., Woycechowsky, K. J. & Hilvert, D. Diselenides as universal oxidative folding catalysts of diverse proteins. *J. Biotechnol.* **150,** 481–489 (2010).

183. Steinmann, D., Nauser, T. & Koppenol, W. H. Selenium and Sulfur in Exchange Reactions: A Comparative Study. *J. Org. Chem.* **75,** 6696–6699 (2010).

184. Sardi, F. *et al.* Determination of acidity and nucleophilicity in thiols by reaction with monobromobimane and fluorescence detection. *Anal. Biochem.* **435,** 74–82 (2013).

185. Tronrud, D. [phenixbb] simulated annealing omit map. (2014).

186. Afonine, P. V. *et al.* FEM: feature-enhanced map. *Acta Crystallogr. D Biol. Crystallogr.* **71,** 646–666 (2015).

187. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. Struct. Biol.* **73,** 148–157 (2017).

188. Hodel, A., Kim, S.-H. & Brünger, A. T. Model bias in macromolecular crystal structures. *Acta Crystallogr. A* **48,** 851–858 (1992).

189. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).

190. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).

191. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).

192. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).

193. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2184–2195 (2004).

194. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1204–1214 (2013).

195. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).

196. Diederichs, K. & Karplus, P. A. Better models by discarding data? *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1215–1222 (2013).

197. Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci.* **103,** 8060–8065 (2006).

198. Tereshko, V. *et al.* Toward chaperone-assisted crystallography: Protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Sci.* **17,** 1175–1187 (2008).

199. Aida, M. & Nagata, C. An ab initio MO study on the disulfide bond: properties concerning the characteristic S-S dihedral angle. *Theor. Chim. Acta* **70,** 73–80 (1986).

200. Winther, J. R. & Thorpe, C. Quantification of Thiols and Disulfides. *Biochim. Biophys. Acta* **1840,** (2014).

201. Riddles, P. W., Blakeley, R. L. & Zerner, B. Reassessment of Ellman's reagent. *Methods Enzymol.* **91,** 49–60 (1983).

202. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9,** 425–430 (2002).

203. Barua, B. *et al.* The Trp-cage: optimizing the stability of a globular miniprotein. *Protein Eng. Des. Sel.* **21,** 171–185 (2008).

204. Williams, D. V., Barua, B. & Andersen, N. H. Hyperstable miniproteins: additive effects of D- and L-Ala mutations. *Org. Biomol. Chem.* **6,** 4287–4289 (2008).

205. Byrne, A., Kier, B. L., Williams, D. V., Scian, M. & Andersen, N. H. Circular permutation of the Trp-cage: fold rescue upon addition of a hydrophobic staple. *RSC Adv.* **3,** 19824–19829 (2013).

206. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. Tryptophan zippers: stable, monomeric beta -hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5578–5583 (2001).

207. Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8,** ncomms15123 (2017).

208. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16247–16252 (2011).

209. Zhou, Z. H. Atomic resolution cryo electron microscopy of macromolecular complexes. *Adv. Protein Chem. Struct. Biol.* **82,** 1–35 (2011).

210. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).

211. Campbell, M. G. *et al.* Near-atomic resolution reconstructions using a mid-range electron microscope operated at 200kV. *J. Struct. Biol.* (2014). doi:10.1016/j.jsb.2014.09.008

212. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470,** 73–77 (2011).

213. Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337,** 362–364 (2012).

214. Redecke, L. *et al.* Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science* **339,** 227–230 (2013).

215. Kupitz, C. *et al.* Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513,** 261–265 (2014).

216. Daly, R. & Hearn, M. T. W. Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* **18,** 119–138 (2005).

217. Castellana, M. *et al.* Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* **32,** 1011–1018 (2014).

218. Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **3,** 845–854 (2006).

219. Kim, J. S. & Raines, R. T. Dibromobimane as a fluorescent crosslinking reagent. *Anal. Biochem.* **225,** 174–176 (1995).

220. Smith, M. E. B. *et al.* Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **132,** 1960–1965 (2010).